



ἘΠΈΚΕΙΝΑ

International Journal of Ontology
History and Critics

STEFANA GARELLO, FRANCESCA SUNSERI, NICCOLÒ MONTI, JIM
SCHRUB E RICCARDO CANGIALOSI

Introduction

EPEKEINA, vol. 16, nn. 1-2 (2023), pp. 1-10
Philosophy of Technology

ISSN: 2281-3209

DOI: 10.7408/epkn.

Published on-line by:

CRF – CENTRO INTERNAZIONALE PER LA RICERCA FILOSOFICA
PALERMO (ITALY)

www.ricercafilosofica.it/epekeina



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs 3.0 Unported License.

Introduction

*Stefana Garelo, Francesca Sunseri, Niccolò Monti, Jim Schrub e
Riccardo Cangialosi*

Since when creativity became a problem for computers?

In all its multifaceted forms, creativity has long been (and often still is) regarded as a species-specific human capacity. AI has tried to challenge that; and whether we think of it in terms of symbolic AI or of connectionism, both of these lineages have equally contributed to a new idea of creative behavior.

The history of symbolic AI, however, saw possibly the first instance in which creativity was addressed within this field as a proper issue. John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon – the four scientists who inaugurated the Dartmouth Summer Research Project of 1955 – had decided to include creativity among the central “aspects of the artificial intelligence problem”. Although rudimental and incomplete, by their own admission, the epistemological relevance of this moment can hardly be overstated. According to them, creativity had something to do with randomness and its role in reasoning: “A fairly attractive and yet clearly incomplete conjecture is that the difference between creative thinking and unimaginative competent thinking lies in the injection of some randomness” (McCarthy *et al.* 1955, 2). Could a computer replicate this creative thinking, they wondered, if programmed in the right manner? What *type* of computer – meaning what hardware – was then needed so that a creative behavior emerged? The proposal did not reach a solution, but, nonetheless, some common assumptions were still being challenged: creativity was no longer a prerogative of human agents, and researchers started speculating that even machines were capable of it – or, at least, that, one day, they would have been.

But McCarthy and his colleagues only constitute a small fraction of the larger discourse around AI and creativity. In order to realize this, one only needs to consider the importance held today by connectionism, the paradigm introduced by the first artificial neural network architecture, the perceptron (Rosenblatt 1958). Its approach to creativity, chiefly focused on pattern recognition and prediction—on *statistics*, rather than *semantics*—went on to foster the rise of machine learning starting from the 1980s, becoming the dominant approach to AI. So, if

we follow the history of AI and track its discontinuous development, we may notice that the same problem, now, crowds the field.

The year 2023 has been particularly filled with novelties and worries, starting with the enthusiasm surrounding OpenAI's ChatGPT, probably the AI model that has most stimulated the scientific and popular discourse on the social effects of AI. By the end of the year, Google released Gemini, a multi-modal model that, they claim in their promotional blog post, will “[have] the potential to create opportunities [, to] bring new waves of innovation, economic progress and drive knowledge, learning, creativity and productivity on a scale we haven't seen before”.¹ One may then be tempted to ask: why did that blog post include creativity between learning and productivity? What does creativity mean in that context, and, most importantly, *whose* creativity are we referring to? Is it the humans', or the machines'? The presence of that term, by itself, would seem unimportant, or akin in value to a catchphrase, just another term to better commercialize a product, Google's AI model. One might easily decide to gloss over its use. But we should be wary of how creativity is thrown around by companies that wish to describe the potentialities of their products: in fact, creativity is never a neutral noun, as it responds to scientific, technical, social, and commercial stakes; therefore, it should be taken not at face value, but always through a critical lens.

In connection to this, 2023 saw an increase of opinions which have voiced the possible risks posed by AI technologies, whether contingent or even existential—and, on the other hand, the contrary view that most of these opinions and universal worries are actually too generic, and often convenient, as they fail to account for any labor implications of AI. The risk would then be to overshadow the exploitation of low wage workers that are employed to optimize the outputs of artificial models, as well as ignoring the sometimes illicit acquisition and management of data that power these technologies. As the current historical background shows, many attempts have been made to regulate this evolving scenario and the actors that populate it, which is why the coming to fruition of the EU AI Act might be regarded as a step for-

1. «Introducing Gemini: our largest and most capable AI model » [en ligne], Google, 2023, URL: <https://blog.google/technology/ai/google-gemini-ai/,consultéle10décembre2023>.

ward, not as a bridle to progress (or, rather, to profitability), but as an opportunity to assess what the undesired consequences of AI driven economies and lives might be. Ultimately, the industrial, the political, and the legal dimensions all impact the way in which the discourse on AI creativity is reshaping how we think of and live by these technical and digital objects.

After all, apart from the market and business interests, aside from the political and legal struggles, creativity as applied to computation and computer-generated content has managed to establish a whole domain of research: what several scholars now refer to as *computational creativity* (CC)—whereas others prefer speaking of *artificial*, or even *generative*. Spanning many lines of research, this number of *Epekeina* tries to deal with at least some of the central aspects of CC, which in the work curated by Tony Veale and F. Amílcar Cardoso (2019) has defined as follows: “Computational creativity is an emerging field of research within AI that focuses on the capacity of machines to both generate and evaluate new outputs that, if produced by a human, would be considered creative”. Certainly, between this definition by Veale and the foundational examples that we find in the Dartmouth school and the invention of the perceptron, there is more. CC has been structured by many contributions and perspectives, not least those that were advanced by the likes of Douglas Hofstadter, Margaret Boden, or Marcus du Sautoy. CC has never ceased posing profound philosophical questions regarding the relation between computation and intelligence, computation and creativity, or intelligence and creativity, a triad of dualisms that revolve on a central matter: whether an artificial agent can be ever considered or become creative.

Such questions were, for instance, explored in the 1990s in Hofstadter’s *Fluid Concepts and Creative Analogies*, where the author evidenced the difficulty to tell novel ideas apart from “clever plagiarism” (Hofstadter 1996, 480-1), or to adequately find a way to credit either humans or machines in their respective contributions to a given product—acknowledging the added difficulty related to the lack of access to the process that led to the output under evaluation. This actually leads us back to another aspect of the socio-political impact of computer-generated content. Indeed, one of the main arguments resorted to in the legitimation of AI public and private investment is that the development of AI tools could help laypeople create visual

or written art: anyone would then have access to the same creative tools, the argument often goes, framing AI as a democratizing technology, which broadens the possibility for each and everyone to tap into and develop their own creative skills. But this view, as others have noted, is accompanied by the threat of seeing the replacement of artists' work with AI-supported productions, perceived as either bland or as the result of data misappropriation, to the detriment of working artists. A threat that appears all too real – and foreshadowed by Hofstadter's "clever plagiarism" –, as is reported by a class action taken on by the Joseph Saveri Law Firm earlier this year (Saveri *et al.* 2023). The class action targets Stable Diffusion and Midjourney, whose generative models were trained using LAION, a dataset which contains vast amounts of copyrighted material, the contention being that the material is exploited without consent nor any financial retribution for those it was taken from. Similar recent events, such as the WGA and SAG-AFTRA strikes that lasted throughout the central months of 2023, brought public attention to the challenges that AI poses to the notion of authorship, to how we deal with copyrighted material, to the extent to which we deem "creative" the use of generative technologies.

Given all of this, this issue was conceived as a critical and analytical response to this domain, as well as to the complex scenario in which it is immersed. The overall aim is to engage, adopting a cross-disciplinary approach—philosophical, sociological, linguistic, semiotic, etc.—the issue of computational creativity, and all its consequences for our general idea of creativity. This has resulted in three directions in which the articles gathered here may be articulated.

First, the papers by Cecile Malaspina, Edan O'Riordan and Michele Pavan develop a critical approach by dealing directly with some of the key elements of CC, mostly remaining within the bounds of its technical field. M. Pavan's article interrogates the construction of the concept of creativity as applied to computation. In particular, he develops a critical analysis of the definition of creativity that was proposed by Boden, as "the ability to generate ideas or artifacts that are new, surprising, and valuable" (Boden 1990), a definition that has proven cardinal for the field. Pavan proceeds to oppose similar subjectivist construals of creativity to objectivist ones, so as to criticize the former, insofar as they reinforce a view that risks becoming independent from the technical objects it is supposed to characterize. Pavan, then,

proposes an objectivist definition of creativity, based on the notion of *impossible results*, which he draws from the works of mathematician Alan Turing. Meanwhile, E. O’Riordan investigates two different but mutually implicated issues: first, the metaphysical realism that is supposed by the claims of creativity in the field of CC. Second, how the normative aspect of creativity is often (conveniently) eluded ; a normative aspect which O’Riordan builds on the hegelian-marxian concept of humans as species beings. And, finally, the article by Cecilia Malaspina, which is the Italian translation of her article on “Pure information: on infinity and human nature in the technical object”, originally published in *Culture, Theory and Critique* in 2019, tackles the concept of *information*, quintessential to the definition of what is considered creative in CC, and even before that in cybernetics. Without shying away from some of the paradoxes that are central to the relation of humans with technical objects, Malaspina’s article inquires how the latter are capable of giving shape—through the information they carry—to the transindividual, collective, existence of the former.

Secondly, three more articles reconsider and deal with some specific assumptions about CC and creativity in general. Riccardo Tarantino’s article questions the methodological and epistemological implications of refusing *de jure* the possibility of CC. Through a comparison of the computational models of living and non-living realities, he develops a criteriological approach to creativity, conceived as the control experience for an adequate modelization of agents in a computational environment, specifically focusing on the NetLogo software. Then, Luca Cardone’s article handles the fundamental question of meaning and the relation between meaning and creation. He interrogates the processes by which meaning is produced out of its reception, introducing a dialogue and a proximity between Marc Richir’s phenomenological approach to the schizophrenic creative process as disincarnated, and the generative process of Stable Diffusion. In accordance to this second direction of inquiry, we have also included Jim Schrub’s article, which uses Simondonian philosophy to explore the difference between the notion of invention and Abraham Moles’ notion of “*créativité induite*”, with the purpose of questioning the anthropomorphic interpretative framework applied to computer generativity, a framework that tends to elude the normativity of human creativity that makes it specific.

Third, CC is explored with the purpose of dealing with an issue which relates the scope of AI with problems otherwise external to its field of research. Cerutti and Laterza questioned the theological roots of the creation–generation dualism that we often use either to talk about CC, or to name specific digital objects (i.e., a generative AI such as Chat-GPT), which they prompted and whose output they analyze. They propose the term “computational creative generativity”, in order to insist that generative AIs do not properly create, although we may perceive their outputs as creative. Then, Sanfilippo’s article, written during the months of the strike led by SAG-AFTRA, tackles the issue of (co-)authoriality in the age of AI, reprising the *topos* of the author’s death, and considering the current struggle for the authors’ recognition of their own work. The function and role of authorhood is transforming, Sanfilippo observes, because of computer-generated texts which call for new—and thus far incomplete—legislation. The third article in this group, by Vaccaro, developed an analysis of novelty and what novelty means for computational systems, in particular examining the case of AlphaZero, whose architecture is based on principles such as programming-free learning, a general and nonspecific algorithm, training through concrete observation, and the program’s own activity in the training process. Drawing from postkantians—namely, from Simondon’s *Imagination and Invention*, Merleau-Ponty’s works on aesthetics and Nishida’s work on the productive nature being—Vaccaro proposes to approach CC as another mode for *being productive and open*. Such productiveness and its existence in complex digital objects implies navigating the tensions between object and subject, or between the semantic and the technological. Finally, Gnoffo and Pirrone displace us a little bit and propose a reflection on creativity oriented towards the analysis of transmedial objects and the operations of transcreation they suppose for their production, an operation that navigates aesthetic, technical and socio-economic sets of constraints. This operation of transcreation appears like a fruitful track to explore and compare with, for instance, the interpolation process that enables creative compositions of images (in Stable Diffusion for instance).

Finally, the issue includes two reviews. Claudia Navarra’s review of Gerd Gigerenzer’s book, *Perché l’intelligenza umana batte ancora gli algoritmi* [*How to Stay Smart In a Smart World: Why Human Intelligence Still Beats Algorithms*] (2023), critically examines prevailing

narratives surrounding AI. Gigerenzer, a cognitive psychologist and director of the Max Planck Institute for Human Development in Berlin, aims to counter both catastrophic and messianic views of AI, challenging the notion of AI as a *Superintelligence* capable of surpassing human capabilities across various domains. The central theme is the debate over whether AI, with its increasing computational power, will eventually outperform humans in every aspect. Gigerenzer questions this narrative, asserting that human intelligence, evolved to navigate uncertainty, remains superior in situations characterized by instability and unpredictability. Examining specific cases, such as AI in dating platforms and healthcare systems, Gigerenzer reveals how algorithms designed by profit-driven companies may manipulate human behavior and compromise essential aspects of life. The review underscores the ethical implications of AI, pointing out that the negative consequences in areas like love and health are ultimately driven by corporate interests prioritizing data exploitation over problem-solving. The author critiques the misconception of equating computational power with intelligence, highlighting the absence of *common sense* in AI. The book is described as a perfect synthesis of accessibility and scientific rigor, offering readers essential strategies to navigate life consciously alongside AI.

The review by Angelo Bellofiore of Marcus du Sautoy's book, *The Creativity Code: How Ai is Learning to Write, Paint and Think*, whose Italian translation was published as *Il codice della creatività. Il mistero del pensiero umano al tempo dell'intelligenza artificiale* (2019). Du Sautoy's essay provides an in-depth analysis of themes related to AI and CC. Bellofiore especially highlights the concerns that had emerged in 2022 regarding the developments of artificial intelligence, with particular attention to the implications for writing and art. Du Sautoy, following a tripartition introduced by Boden, argues that AI excels in explorative and combinative creativity, but transformative creativity, which requires a complete departure from predefined rules, still poses a challenge for machines. The review examines du Sautoy's reflections on AI generated artistic production, ranging from several case studies, including visual outputs generated by algorithms, such as the *Mandelbrot set*. Despite the many advancements in AI, du Sautoy maintains that human creativity remains too complex for AI to fully replicate: as

the review remarks as well, machines have not yet fully grasped the *secret* of human creativity.

In conclusion, the scale at which AI is now deployed has changed and is changing our practices, be they the creative practices of artists whose work can be growingly assisted by AI tools (in design jobs with the use of Adobe Photoshop,² in the visual effect industry,³ or in animation with Procreate Dreams,⁴ or the full generation of content based on a prompt provided to the AI by its user (as in Midjourney, DALL·2, or Stable Diffusion).

These changes bring their own lines of questioning. One of them is society-centered, as we are witnessing a growing public debate over what can be done to manage and regulate this avalanche of AI developments, actively – and collectively – dealing with the environmental and human costs of these devices. The other line of questioning, also present throughout this issue, is more theoretical: human creativity, as a problem-solving operation (Shannon 1952; Moles 1956), or instead as the property of a novel, surprising, and valuable output (Boden 1990), undergoes new ways of being technicized – of being formed by techniques – to an unprecedented scale and entity. These technological changes are altering the socio-technical milieu of humans, and by extension the new associated milieu where computational and human creativity coexist (Simondon 2005; 2012), and how peaceful that coexistence is. When it comes to the labor market and the issue of intellectual property, as shown with the SAG-AFTRA’s proposal document, or the Saveri Law Firm class action, peaceful is not the most adequate way to describe the situation. When it comes to human individuation and to the development of imagination, the new distribution of creative processes between AI and users will also transform the socio-material production of the symbolic and, with it, the structure of our social fabric – as well as our own imagination, which, if not

2. «Retouche photo optimisée par l’IA avec Adobe Photoshop» [en ligne], URL: <https://www.adobe.com/fr/products/photoshop/ai.html>, consulté le 10 décembre 2023.

3. «Issue 11 - AI and VFX» [en ligne], before afters, 2023, URL: <https://beforeandafters.com/product/issue-11/>, consulté le 10 décembre 2023.

4. Ian Dean Published, «Procreate Dreams: everything you need to know» [en ligne], Creative Bloq, 2023, URL: <https://www.creativebloq.com/features/procreate-dreams-everything-you-need-to-know>, consulté le 10 décembre 2023.

proletarianized altogether, is at least organologically transformed (Stiegler 2006; 2015).

The intricate interplay of human and computational creativity shapes our social and technological landscape, giving rise to complex challenges that demand subtle and refined solutions. As we navigate the dynamic interaction between artificial intelligence and human creativity, the blending of creative approaches profoundly influences how we depict and interpret the world. In this transformative exploration, a central question emerges, expressed throughout the articles included here: Does our imagination undergo a streamlining process, simplifying its essence, or does it embark on a fundamental structural transformation? This inquiry transcends the realm of technology, delving into the complexity of our shared human experience.

References

Boden M. A. 1990, *The Creative Mind, Myths and Mechanisms*, Routledge, London.

Douglas R. and Hofstadter D. R. 1996, *Fluid Concepts and Creative Analogies: Computer Models Of The Fundamental Mechanisms Of Thought*, 1st edition, Basic Books, New York.

McCarthy J., Minsky M. L., Rochester N., and Shannon C. E. 2006, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence (August 31, 1955)*, "AI Magazine", 27(4), p. 12. <https://doi.org/10.1609/aimag.v27i4.1904>

Moles A. A. 1956, *La création scientifique*, René Kister, Genève.

Pichai S. and Hassabis D. 2023, *Introducing Gemini: our largest and most capable AI model*, "Google", <https://blog.google/technology/ai/google-gemini-ai/>.

Rosenblatt F. 1958, *The perceptron: A probabilistic model for information storage and organization in the brain*, "Psychological Review", 65(6), pp. 386-408. <https://psycnet.apa.org/doi/10.1037/>

Stefana Garelo, Francesca Sunseri, Niccolò Monti, Jim Schrub e Riccardo Cangialosi

Saveri J. R., Zirpoli C., Young, C. K. L., et al. 2023, *Individual and Representative Plaintiffs, v. Stability AI Ltd., a UK corporation; Stability AI, Inc., a Delaware corporation; DeviantArt, Inc., a Delaware corporation; Midjourney, Inc., a Delaware corporation; Runway AI, Inc., a Delaware corporation*, <https://storage.courtlistener.com/recap/gov.uscourts.cand.407208/gov.uscourts.cand.407208.129.0.pdf>

Shannon C. 1952, *Creative Thinking* [online], Bell Labs. <http://www1.ece.neu.edu/~naderi/Claude%20Shannon.html>

Simondon G. 2005, *L'Invention dans les techniques, cours et conférences*, Seuil, Paris.

Simondon G. 2012, *Du mode d'existence des objets techniques*, Flammarion, Paris.

Stiegler B. 2010, *What makes life worth living: on pharmacology*, Polity., Cambridge (UK)

Stiegler B. 2015, *Automatic Society. The future of work*, Polity, Cambridge (UK).

Veale T., and Amílcar C. F. (eds.) 2019, *Computational Creativity: The Philosophy and Engineering of Autonomously Creative Systems*, Springer, London.

Stefana Garelo, Francesca Sunseri, Niccolò Monti, Jim Schrub e Riccardo Cangialosi

Università degli studi di Palermo,

Università degli studi di Torino,

Paris Nanterre

stefana.garelo@unipa.it, francesca.sunseri@unipa.it, niccolo.monti@unito.it,
jim.schrub@gmail.com, riccardo.cangialosi01@unipa.it



ἘΠΕΚΕΙΝΑ

International Journal of Ontology
History and Critics

CLAUDIA NAVARRA

Gerd Gigerenzer, *Perché l'intelligenza umana batte ancora gli algoritmi*, Raffaello Cortina, Milano 2023

EPEKEINA, vol. 16, nn. 1-2 (2023), pp. 1-7
Book Reviews

ISSN: 2281-3209

DOI: 10.7408/epkn.

Published on-line by:

CRF – CENTRO INTERNAZIONALE PER LA RICERCA FILOSOFICA
PALERMO (ITALY)

www.ricercafilosofica.it/epekeina



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs 3.0 Unported License.

Gerd Gigerenzer, *Perché l'intelligenza umana batte ancora gli algoritmi*, Raffaello Cortina, Milano 2023

Claudia Navarra

Gerd Gigerenzer, psicologo cognitivo e direttore del Max Planck Institute for Human Development di Berlino, nel suo ultimo libro intitolato *Perché l'intelligenza umana batte ancora gli algoritmi*, edito nel febbraio 2023 da Raffaello Cortina Editore, si pone come obiettivo quello di salvare il lettore dalla narrazione da un lato catastrofica e dall'altro "messianica" riferita all'Intelligenza Artificiale.

Quella contro cui Gigerenzer concentra la sua attenzione è l'idea di una IA come *Superintelligenza* capace di sostituire l'essere umano in ogni campo, dagli scacchi al Go, dalla guida alla finanza, dall'educazione all'amore. Continuamente lo spettro di una IA che prende coscienza e riduce in schiavitù l'umanità assilla fior di studiosi – facendo la fortuna di altrettanti registi -, mentre altri accademici con la stessa intensità e CEOs di grandi aziende e istituti finanziari auspicano, invece, un mondo in cui ogni attività umana sarà svolta da una Intelligenza non-umana.

L'IA batte gli umani, è questo l'argomento dominante:

«l'IA ha sconfitto i migliori umani negli scacchi e nel Go.

La capacità di elaborazione raddoppia ogni paio d'anni.

Quindi, le macchine ben presto faranno ogni cosa meglio degli umani»
(Gigerenzer 2023, 2).

Ma è davvero così? Siamo, in quanto specie, meno intelligenti delle macchine che abbiamo costruito? Esiste un effettivo pericolo di estinzione per la specie umana? Sarà una Superintelligenza a governarci tutti? L'erronea idea generale sembra essere quella per cui sostituire le persone con un software possa rendere il mondo un luogo migliore a patto, però, di rinunciare a mantenere il controllo sulla propria vita e ad avere una fiducia incondizionata nell'algoritmo. È la fede nella complessità e nell'opacità a garantire la fiducia nell'IA, cioè la credenza in quello che Gigerenzer chiama «dilemma accuratezza-trasparenza: più l'algoritmo è accurato, meno sarà trasparente» (2023, 140). L'uso, negli Stati Uniti, di algoritmi nelle aule di tribunale che aiutino il giudice a

decidere, l'utilizzo di IA da parte della polizia che predicono il luogo di future scene del crimine o quali persone hanno più possibilità di commettere o essere vittima di reati (Sappiamo chi sei!) e ancora le lunghe ed esaustive spiegazioni sul consenso al trattamento dei dati personali rendono il mondo un luogo più sicuro.

L'IA garantisce trasparenza e previene le situazioni di pericolo. Secondo Gigerenzer le cose stanno diversamente. Anzi, lungi dall'essere onnisciente e super intelligente, l'IA non ha speranze di raggiungere le prestazioni umane, almeno quando ci si muove nel mondo vero che è un mondo dominato dall'incertezza la cui causa sono proprio gli esseri umani.

L'analisi pungente, e a tratti provocatoria, dello psicologo tedesco tocca quelli che potremmo chiamare i "cavalli di battaglia" dell'IA: l'amore, la privacy, la salute, il denaro, la giustizia, senza dimenticare uno dei settori di cui ultimamente ci si è interessati di più: la guida automatica. In ognuno di questi settori l'IA ha avuto una evoluzione straordinaria e ha messo in circolazione, più o meno virtuosamente, milioni di dollari promettendo altissime prestazioni e garantendo la capacità di prendere la decisione migliore possibile, nonché di riuscire a prevedere eventi, soprattutto di natura economica, che possano permettere alle aziende di aumentare i loro profitti con sicurezza.

Gigerenzer, in poco meno di 400 pagine e 11 capitoli divisi in due parti, ci mostra come non esista nessuna garanzia e nessuna prova che, in effetti, l'IA, con i suoi algoritmi complessi, riesca a fare quello che promette né a battere l'essere umano quando si tratta di giocare sul serio, quando cioè le situazioni non sono stabili ma cariche di incertezza. L'algoritmo, insiste Gigerenzer, funziona nelle condizioni in cui è possibile applicare quello che egli definisce come il *Principio del mondo stabile*: «gli algoritmi complessi funzionano al meglio in situazioni ben definite, stabili, in cui sono disponibili grandi quantità di dati. L'intelligenza umana, invece, si è evoluta in modo da gestire l'incertezza a prescindere dalla quantità di dati disponibili» (2023, 22). Se è vero che l'IA batte gli umani in ambienti stabili e quando ha a disposizione una enorme quantità di dati (i *big data*), come quando gioca a scacchi, questo non significa che possa farlo in ogni situazione. Lo scopo di Gigerenzer con questo libro è quello di fornire al lettore strategie e metodi per mantenere il controllo della propria vita e, contemporaneamente, quello di comprendere le potenzialità delle nuove

tecnologie. Dunque, nessuna visione catastrofica o “messianica”, ma un’analisi chiara di ciò che l’IA può fare e dei vantaggi, così come dei pericoli, che ne può comportare l’utilizzo.

L’appello di Gigerenzer è di *Restare Intelligenti*. Significa comprendere vantaggi e rischi delle nuove tecnologie e mantenere il controllo. Le nuove tecnologie hanno, e avranno, un impatto profondissimo sugli esseri umani, poiché cambiano il modo in cui sono percepiti il bene e il male. La colpa di questo cambiamento non è dell’algoritmo in sé, ma è delle aziende che progettano l’algoritmo. L’autore definisce le aziende come delle macchine per predire e modificare il comportamento degli esseri umani: propongono soluzioni tecnologiche a problemi umani utilizzando un modello di business “paga con i tuoi dati”; alle aziende non interessa risolvere il problema, interessa il tempo e l’attenzione degli utenti. Infatti, spesso le aziende possono perseguire degli obiettivi che risultano in contrasto con quelli che i loro software dovrebbero riuscire a fare.

Pare opportuno presentare alcuni casi analizzati da Gigerenzer soprattutto nella prima parte del libro.

Il primo degli esempi di questo utilizzo dell’IA riguarda i siti d’incontri e cerca di rispondere ad una domanda sottesa al primo capitolo su come possa un algoritmo garantire di trovare l’anima gemella. Una delle più importanti aziende di questo settore dichiara¹: ogni 11 minuti un single si innamora! È sicuramente uno slogan interessante e funziona visto che sono milioni gli iscritti al sito. Ma a ben guardare – come fa Gigerenzer – se gli iscritti fossero soltanto un milione significherebbe che, in un anno, troverebbe l’amore soltanto il 5% di loro. In più, se sul sito fossero iscritti, come di fatto è, più di un milione di utenti ci vorrebbero più di dieci anni per trovare l’amore. Nonostante tutto però, i milioni di utenti restano e sperano e continuano a compilare questionari nella speranza che qualcun altro compili il questionario meglio compatibile. Gli algoritmi che utilizzano questi siti si fondano su tre principi cui vengono assegnati dei valori numerici. I tre principi base dell’algoritmo dell’amore sono: la somiglianza, la complementarità e l’importanza che insieme concorrono alla creazione

1. Il sito d’incontri cui si riferisce Gigerenzer è Parship, attivo con milioni di utenti e disponibile in 13 Paesi.

di un profilo utente; eppure, è raro che trovino l'anima gemella. Il motivo di questo fallimento risiede, secondo l'autore, nel fatto che un profilo non è una persona, infatti, manca l'apporto degli elementi fisici, chimici ed emotivi che sono fondamentali nelle relazioni umane, ma non solo. L'ambiente in cui operano gli algoritmi di questo tipo è colmo di incertezza che è, inoltre, alimentata anche dal fatto che i dati inseriti, molto spesso, non corrispondano alla realtà: reddito più alto e peso minore sono tra le "bugie" più comuni.

Il consiglio dello psicologo, oltre a calcolare in maniera più precisa le percentuali di successo dichiarate dalle aziende, è quello di andare a prendere un caffè e vedere come va. Un esempio ancora più chiaro di come le aziende progettino gli algoritmi al fine di manipolare il comportamento umano è fornito nel secondo capitolo. In questa sede Gigerenzer mostra come l'adozione da parte di aziende sanitarie statunitensi di un sistema basato su algoritmi per la creazione di cartelle sanitarie digitali che permettesse consultazioni, analisi, registrazioni e archiviazioni in modo veloce e funzionale, si rivelò un completo disastro. Furono previsti, in seguito all'adozione di un tale software già positivamente sperimentato in Gran Bretagna, circa 81 miliardi di dollari risparmiati all'anno. La realtà si rivelò ben diversa: «le fatturazioni degli ospedali aumentarono con costi che passarono da 2000 a 2800 miliardi di dollari» (2023, 31). L'IA avrebbe dovuto fungere da una sorta di velocissimo database interattivo, avendo a che fare con un mondo stabile di dati che venivano registrati e integrati. Il software, invece, era progettato non per snellire il sistema sanitario, ma per aumentare le richieste di interventi medici sui pazienti. Gigerenzer afferma infatti, che erano le stesse aziende sanitarie che si occupavano degli esami o delle operazioni o della produzione di dispositivi medici a progettare i software che, nell'ottica del profitto e della lealtà aziendale, consigliavano al medico inutili ulteriori esami o di cambiare, aggravandola, una diagnosi.

Due degli aspetti più importanti per gli esseri umani, l'amore e la salute, affidati all'Intelligenza Artificiale hanno subito e subiscono danni incalcolabili in nome del profitto. È, quello dello psicologo, un forte ammonimento al capitalismo. Dall'analisi che prende forma nel libro, appare chiaro come, anche se i dati raccolti nelle indagini dello studioso sembrano essere disastrosi, gli esseri umani continuano a fidarsi delle promesse delle aziende che sfruttano gli algoritmi. Ciò

avviene perché spesso le aziende si servono dei bias cognitivi, degli errori insiti nel modo di pensare delle persone. Uno dei più comuni bias, che confonde la predizione con l'adattamento, è la "Fallacia del cecchino texano": un cecchino spara alla parete di un fienile da una distanza considerevole; il 90% dei colpi sono raggruppati all'interno di un cerchio; come ha fatto a centrare così tante volte il bersaglio? Il cecchino texano ha prima sparato e poi disegnato il cerchio. Quando un'azienda tenta di vendere risultati sbalorditivi, consiglia al lettore Gigerenzer, è sempre importante chiedersi se «la performance esaltante di un algoritmo sia stata ottenuta attraverso una predizione o un mero adattamento» (2023, 35).

Il quarto capitolo è dedicato all'analisi dell'affidabilità delle auto a guida autonoma, tema delle più attuali indagini sulle effettive capacità dell'Intelligenza Artificiale. In questo tipo di tecnologia vengono investiti milioni di dollari e tollerati incidenti spaventosi. Le auto di nuova generazione hanno una autonomia di livello 5, il massimo possibile, il che significa che possono guidare in qualsiasi situazione e senza l'intervento dell'essere umano. Un nuovo principio viene introdotto da Gigerenzer, quello dell'adattamento dell'IA: «per migliorare la prestazione dell'IA, bisogna rendere l'ambiente fisico più stabile e il comportamento umano più prevedibile» (2023, 57). Questo significa che non è più la tecnologia a doversi adattare al comportamento umano, ma è l'essere umano a doversi adattare alla tecnologia in modo da evitare il più possibile le situazioni di incertezza. Inoltre, nonostante il migliore contributo possibile di una IA psicologica, che potrebbe istruire l'algoritmo delle auto a guida autonoma con le simulazioni delle percezioni, dei giudizi e dei processi decisionali umani, sarebbe impossibile per una AI imparare a guidare come un essere umano. Una delle regole, non scritte, della guida è: dietro un pallone c'è sempre un bambino. Un pilota umano saprebbe quando è il caso di rallentare, di frenare o di fare qualche altra azione sulla base di questa regola. Un pilota digitale frena, anche se dietro il pallone non c'è alcun bambino, e provoca un incidente a catena. Il vero pilota digitale sono le reti neurali profonde, che però commettono errori e scambiano, a 69 km all'ora, una donna in bicicletta prima per un oggetto sconosciuto, poi per un veicolo e infine identificano una bicicletta, a 4 secondi dall'impatto mortale. Secondo lo psicologo tedesco il problema delle reti neurali non consiste nel non avere abbastanza immagini di biciclette, ma nel

non possederne il concetto. La “Fallacia del carrarmato russo” è il tipo di errore che commettono le reti neurali e riguarda il fatto che spesso queste si focalizzano su elementi secondari che sono presenti sia nell’addestramento sia nella serie di test che vengono fatti. L’Intelligenza Artificiale con le reti neurali cerca pattern basati su certe caratteristiche; invece, l’intelligenza umana è in grado di costruirsi una rappresentazione del mondo. Ancora più importante è il dilemma morale che le auto a guida autonoma impongono: tre anziani stanno attraversando la strada col semaforo rosso mentre sta per raggiungerli un’auto a guida autonoma con tre passeggeri a bordo; chi ucciderà l’IA, i tre anziani o i tre passeggeri, schiantandosi su un muro per evitare i pedoni? Secondo Gigerenzer questo dilemma non interessa nemmeno agli ingegneri, proprio perché risolverlo significherebbe aver progettato una IA capace di prevedere ogni possibile conseguenza e di essere perfettamente autonoma, cosa ancora molto al di là dal divenire realtà.

È particolarmente chiara l’analisi del divario tra l’intelligenza umana e quella artificiale che si va tracciando nelle pagine del libro. Ci sono profonde differenze tra l’intelligenza umana e quella artificiale. La prima e fondamentale differenza è che l’intelligenza umana si è sviluppata in condizioni di incertezza, mentre quella artificiale in condizioni di rischio. Questa differenza di condizioni ha comportato lo sviluppo di attività differenti. Gli esseri umani hanno sviluppato capacità che tutte insieme Gigerenzer chiama “buonsenso”: il pensiero causale, la psicologia intuitiva, la fisica intuitiva e la socialità intuitiva. Il “buonsenso” è la «conoscenza condivisa sulle persone e sul mondo, resa possibile dal cervello biologico, e richiede un’esperienza limitata» (2023, 82). Le reti neurali hanno sviluppato il calcolo veloce, la scoperta di associazioni nei big data, il ritrovamento di pattern nelle immagini o nelle informazioni sonore.

L’errore di fondo, sostiene l’autore, è stato quello di confondere la capacità di calcolo con l’intelligenza; un maggior numero di informazioni con la conoscenza; la memoria con l’intuizione. All’Intelligenza Artificiale manca il buonsenso e questo si riscontra nel linguaggio, che spesso viene stravolto dall’IA che non trova corrispondenze adeguate all’interpretazione, come anche nel riconoscimento facciale oppure di un certo scenario.

Gigerenzer fornisce al lettore altri strumenti per mantenere il controllo sulla propria vita e restare intelligenti, tra cui l’euristica della

recenza: «in situazioni di incertezza, fatela semplice e non scommettete sul passato» (2023, 105); oppure suggerisce di non fidarsi dei grandi numeri, proprio perché i numeri *non* parlano da soli è sempre necessario ricondurli ad un mondo stabile e interpretarli mediante una teoria.

Un altro interessante esempio della inadeguatezza dell'IA nelle situazioni di incertezza e dell'interesse delle aziende nella manipolazione del comportamento umano riguarda la privacy. Gigerenzer afferma, nell'ottavo capitolo, che la privacy non vale un fico secco! Tutti restano costernati alla notizia del furto di dati sensibili o alla loro perdita o della vendita dei dati personali a terze parti, sentendosi vittime di una società di matrice orwelliana. Ma qui sta il paradosso: le stesse persone indignate e spaventate dal fatto che un qualche tipo di organizzazione possa sottrarre informazioni private, non si fa nessuno scrupolo né a rivelarle sulle piattaforme social né a negare di essere disposti a pagare per mantenere quelle informazioni così importanti al sicuro.

Dunque, è enorme l'effetto che l'IA ha sulla vita degli esseri umani, ma questo non significa che l'umanità sia diventata succube delle macchine, non ancora almeno. Per far sì che questo non avvenga è necessario comprendere da quali bias siano affetti gli algoritmi e in quale mondo siano applicabili. Gigerenzer sprona il lettore a restare intelligente, ad indagare e ad analizzare gli scenari condivisi con l'IA e a sfruttarne le potenzialità quando queste si rivelano preziose ma, nello stesso tempo, a dubitare delle promesse stupefacenti che le aziende tentano di vendere.

L'intelligenza umana batte ancora gli algoritmi perché è basata sul buonsenso e sulla corporeità, sull'intuito e la creatività, sull'improvvisazione e l'incertezza. Sintesi perfetta tra capacità divulgativa e rigore scientifico, *Perché l'intelligenza umana batte ancora gli algoritmi* è un libro interessante e piacevole, che coinvolge e stimola il lettore munendolo di importanti strategie per vivere la propria vita in maniera consapevole insieme all'IA.

Claudia Navarra
Università degli Studi di Palermo
claudia.navarra@unipa.it



ἘΠΕΚΕΙΝΑ

International Journal of Ontology
History and Critics

ANGELO BELLOFIORE

Marcus Du Sautoy, *Il codice della creatività*, Bur Rizzoli, Milano
2019

EPEKEINA, vol. 16, nn. 1-2 (2023), pp. 1-??
Book Reviews

ISSN: 2281-3209

DOI: 10.7408/epkn.

Published on-line by:

CRF – CENTRO INTERNAZIONALE PER LA RICERCA FILOSOFICA
PALERMO (ITALY)

www.ricercafilosofica.it/epekeina



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs 3.0 Unported License.

**Marcus Du Sautoy, *Il codice della creatività*, Bur
Rizzoli, Milano 2019**

Angelo Bellofiore

Le intelligenze artificiali stanno raggiungendo risultati sempre più sorprendenti: sanno giocare a scacchi, disegnare, scrivere brevi storie, e persino comporre musica. Viviamo in un'epoca dove la convivenza fra umani e macchine altamente intelligenti è una realtà alla quale ci stiamo abituando sempre di più. Ci siamo iniziati ad abituare già nel 1997, quando il supercomputer Deep Blue, sviluppato dalla IBM, vinse uno storico match di scacchi contro il campione del mondo Garry Kasparov. Allora la notizia generò grandissimo interesse, e altrettanti investimenti nel campo della ricerca sulle intelligenze artificiali. Questo ultimo decennio è stato caratterizzato dall'esplosione del *machine learning* e delle DNN, o *Deep Neural Networks*, portando allo sviluppo di IA sempre più multifunzionali. Questa ondata di mutamenti hanno generato moltissimo entusiasmo da parte degli addetti ai lavori, ma altrettante sono le preoccupazioni suscitate da questo incontrollato sviluppo di macchine in grado di compiere compiti sempre più complessi. I cambiamenti nel campo dello sviluppo delle intelligenze artificiali ci hanno costretto a riconsiderare moltissimi dei nostri preconcetti riguardo ciò che è umano e ciò che non lo è, a partire proprio dalle modalità d'intelligenza esibite dalle macchine, e ad assumere diverse strategie per difendere l'intelligenza umana. Abbiamo ridimensionato, ad esempio, alcune delle capacità in cui eccellono le IA: gli scacchi sono notoriamente un gioco che richiede moltissima concentrazione ed impegno cognitivo, ma basterà la vittoria delle macchine in questo gioco per affermare che queste sono intelligenti quanto lo sono gli esseri umani? Dopotutto gli scacchi sono un gioco che può essere vinto memorizzando un gran numero di strategie e set di mosse, ed è ormai risaputo come la memoria di una macchina sia di gran lunga superiore a quella umana. Le macchine allora non sono intelligenti. Gli esseri umani sì. Ma cosa manca alle macchine per potersi dire alla pari degli esseri umani? Una risposta probabile potrebbe essere l'assenza, da parte delle macchine, della capacità di pensiero creativo, di esibire comportamenti che possano realmente dirsi innovativi. Ma in un tempo così frenetico, dove le macchine diventano sempre più in grado di produrre risultati

sorprendenti, anche questa capacità sembra poter divenire una meta conquistabile. Possiamo considerare i risultati delle macchine come sintomo di una forma di creatività? Cos'è questa esattamente? Alcune di queste domande trovano una risposta all'interno del saggio di Marcus Du Sautoy *Il codice della creatività*, edito per Bur Rizzoli nel 2019. Du Sautoy, matematico e divulgatore di fama internazionale, propone un'analisi degli sviluppi che hanno coinvolto il campo di ricerca sulle intelligenze artificiali, fornendo gli strumenti per poter interpretare i loro notevoli risultati. Il saggio è organizzato in due sezioni principali: una prima parte, in cui Du Sautoy fornisce un breve prontuario, che funge da sfondo teorico del suo testo, ed una breve, ma essenziale, panoramica dei concetti chiave della ricerca sulle intelligenze artificiali; si parla di algoritmi, di reti neurali e dei metodi attraverso i quali queste vengano addestrare per raggiungere i loro impressionanti traguardi. Nella seconda parte del saggio, Du Sautoy analizza e discute i risultati ottenuti dalle reti neurali nei campi dell'arte visiva, della musica, della matematica e della letteratura. Du Sautoy inaugura il suo saggio con una veloce introduzione storica alle IA, facendola partire direttamente dalla macchina di Babbage, e dalla critica che Lady Lovelace rivolge a questa e, più in generale, alle macchine analitiche: queste non possono produrre qualcosa di originale, ma solo riprodurre ciò che abbiamo inserito all'interno della macchina. La questione della creatività e delle macchine genera domande fin dalle origini di questa impresa tecnologica ed è, chiaramente, il nucleo centrale del saggio di Du Sautoy, il quale fissa in questo primo capitolo le domande fondamentali che accompagneranno il resto del saggio: la creatività è un'abilità che possiamo insegnare alle macchine? È qualcosa di strutturato, o è completamente slegata da qualsiasi tipo di schema spiegabile? Secondo Du Sautoy è prima di tutto necessario comprendere cosa intendiamo per creatività, tema centrale del secondo capitolo del saggio. Qui, l'autore, si rifà alla definizione di creatività così come viene fornita dalla scienziata cognitiva Margaret Boden. La creatività, come macro-abilità, può essere suddivisa in tre tipi differenti: esplorativa, combinativa e trasformativa. La creatività esplorativa è quella che viene messa in atto quando vengono portati alle estreme conseguenze i limiti di uno schema, o un set di regole, di qualcosa già esistente. Secondo l'autore, questa non è una definizione sufficiente di creatività. Quella combinativa è invece quel tipo di creatività che mette assieme, appunto, combinando,

schemi e regole differenti, per generare qualcosa di nuovo. Questi due primi tipi di creatività sono, secondo Du Sautoy, tipi che possiamo riscontrare nelle IA e, soprattutto il primo tipo, risulta una forma in cui queste eccellono, grazie alle loro notevoli capacità di calcolo. La terza, quella trasformativa, è quella che probabilmente meno si adegua al comportamento di una macchina: è un tipo di creatività che richiede il completo abbandono di regole prestabilite, per generare qualcosa di completamente nuovo, e che sorprenda. Du Sautoy lascia parzialmente in sospeso il giudizio riguardo la possibilità che una macchina sia in grado di produrre un comportamento che si possa definire realmente creativo, e dà inizio ad una sezione del suo saggio che funge da piacevole introduzione ai concetti fondamentali della programmazione: vengono illustrati sinteticamente il concetto di algoritmo, la struttura delle reti neurali, e le strategie di apprendimento adottate dai programmatori per istruire i loro programmi a svolgere determinati compiti; in questa ampia sezione divulgativa, Du Sautoy illustra il funzionamento dietro gli algoritmi con cui interagiamo ogni giorno, come il sistema di ricerca di Google, o quello del suggerimento dei contenuti per utenti di Netflix. Da qui ha inizio la seconda grande sezione del saggio, in cui Du Sautoy discute dei risultati delle IA nel campo dell'arte visiva, della matematica, della musica e della letteratura. Attraverso dei casi di studio, Du Sautoy fornisce delle risposte alle grandi domande che permeano il saggio. Tratta, ad esempio, il complesso problema della paternità dell'opera d'arte, quando queste vengono realizzate tramite sistemi informatici. Per affrontarlo, Du Sautoy sceglie di analizzare il caso della rappresentazione visiva dei frattali, come il celebre insieme di Mandelbrot. Questo è visualizzato grazie alla estrema potenza di calcolo dei computer, ma è possibile definirla arte artificiale? Du Sautoy sostiene che no, la macchina non può essere definita come la mente creativa dietro la produzione artistica, perché il suo comportamento è pienamente deterministico, in quanto esegue unicamente le direttive di un istruttore umano. Attraverso diversi esempi, dove le IA vengono testate in campi differenti della produzione artistica e culturale umana, Du Sautoy giunge ad una considerazione generale sul comportamento esibito dalle IA, e sul perché queste non sono ancora in grado di poter risultare creativamente all'altezza della controparte umana: le reti neurali, per essere addestrate, hanno bisogno di una quantità di dati notevolmente elevata da un punto di vista quantitativo, "mentre

le buone opere sono semplicemente troppo poche perché le macchine possano imparare a replicarle” (pag.259), ecco allora spiegata una delle ragioni per le quali è difficile lasciare ad una macchina un accesso al nostro "codice della creatività". Il testo di Sautoy offre un buon punto di partenza ai lettori per orientarsi all'interno dei frenetici sviluppi che caratterizzano un campo dinamico e avanguardistico come quello delle intelligenze artificiali. Il suo è un saggio pervaso dalla meraviglia generata dai risultati delle reti neurali, e dall'intima preoccupazione che questi possano progressivamente invadere degli spazi fin troppo umani. In questo intrigante confronto, le macchine raggiungono, tappa dopo tappa, tutti quei risultati che gli esseri umani hanno ritenuto riservati unicamente alla propria specie: giocare e vincere a scacchi, trionfare nell'elaborato gioco del Go, dimostrare dei teoremi matematici, generare degli spartiti musicali. Negli anni che ci separano dalla pubblicazione del testo, molte cose sono cambiate, e molto in fretta. È stato creato *ChatGPT*, modello di intelligenza artificiale sviluppato da *OpenAI*, in grado di generare moltissimi *output* testuali, molto spesso indistinguibili da quelli umani, ed in grado di rispondere in maniera corretta ad un gran numero di attività a questo proposte. Ma non solo, è stato pubblicato, sempre da *OpenAI*, *DALL-E*, capace di produrre immagini rispondendo a delle semplici richieste testuali. O ancora, è stata rilasciata al grande pubblico *Midjourney*, anch'essa una intelligenza artificiale in grado di produrre immagini che rispondono alle richieste degli utenti, protagonista di una controversa vittoria ottenuta alla *Colorado State Fair* del 2022. Queste tre macchine, sono accomunate dall'essere modelli di IA generativa, ovvero delle IA che ricorrono a dei modelli statistici per produrre i loro *output*. Questa nuova generazione di IA però, non mette ancora in completa difficoltà la tesi proposta da Du Sautoy. Le IA possono batterci a scacchi ed a Go, perché sono molto più capaci rispetto a noi nel fare complessi calcoli ma, secondo Du Sautoy, possiamo ancora rifugiarci nella creatività: quella umana continua ad essere troppo complessa da replicare per le intelligenze artificiali. Gli anni che ci separano dalla pubblicazione di questo saggio portano il testo a subire molto velocemente il peso dell'età; tuttavia, per quanto questo non ci permetta di avere le giuste lenti per interpretare i fenomeni più recenti, causati dalle IA di ultima generazione, ci consente di analizzare i loro risultati mettendo in dubbio la loro reale portata creativa. È in definitiva un testo conciliante, che riappacifica il

lettore rassicurandolo, che no, una IA, nonostante i grandi sviluppi e progressi, non potrà cogliere del tutto il segreto della creatività degli esseri umani. Ma, se questo dovesse accadere, sarebbe proprio nei prodotti della creatività artificiale, nella musica, nella scrittura e nell'arte visiva, che potremo cercare di capire cosa significa essere dentro una *mente artificiale*.

Angelo Bellofiore
Università degli Studi di Palermo
angelo.bellofiore@you.unipa.it



ἘΠΈΚΕΙΝΑ

International Journal of Ontology
History and Critics

ANDREA LE MOLI

Developing Models of Digital Dissemination Interface for the
“Samothrace” Project

EPEKEINA, vol. 16, nn. 1-2 (2023), pp. 1-12
Philosophy of Technology

ISSN: 2281-3209

DOI: 10.7408/epkn.

Published on-line by:

CRF – CENTRO INTERNAZIONALE PER LA RICERCA FILOSOFICA
PALERMO (ITALY)

www.ricercafilosofica.it/epekeina



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs 3.0 Unported License.

Developing Models of Digital Dissemination Interface for the “Samo thrace” Project

Andrea Le Moli

1. General description of the “Samo thrace” project

The Samo thrace project is an inter-university hub made up of the four Sicilian universities (Catania, Messina, Palermo, Enna) on the basis of PNRR funds in a consortium with other national partners. The project is aimed at the development and testing of nanotechnologies and divided into four pillars: Energy, Health, Smart Mobility, Environment, Cultural Heritage, Precision Agriculture. The project activity is further divided into a number of specific spokes. The work conducted in this article was carried out within the activities of spoke 9 "Route to Innovation" (R2I) and aimed at the results produced by researchers at the University of Palermo within Spoke 3 "Micro and Nanotechnologies for Smart Sustainable Communities" (S2-COMMs).

2. Activity's Target

The aim of WP 2.3 of the affiliated partner UNIPA to spoke 9 is “Digital contents creation management according to the project’s state of development”. In this framework, the activity of the milestone 17 was aimed at developing a model for a large-scale and general-use dissemination interface of the scientific results of the project potentially applicable to any spoke. The first activity examines the opportunity offered to a medium-scale online dissemination by the publications produced within the project with reference to the scientific and methodological contribution made by researchers at University of Palermo. This experience serves as case-study to explore the proper methodology and strategies to achieve the general target. As its first step, the timing and effectiveness of the dissemination platform contained on the official project website at <https://samothrace.eu/spoke3> was reviewed to explore other and possibly more effective ways to grant the visibility of and access to the results to a more general public and to a larger audience of stakeholders, enterprises, private and public agencies. In

order to find the best way to convey the projects' results and to develop the most effective strategy to ensure them the best visibility, the WP2.3 team will build up a format of Digital Dissemination Interface (DDI) consisting in a series of web pages and submenus (ideally apt to be included in a specific section of the Samothrace's main website) and other Digital Disseminations Devices (DDD) specifically designed to enhance the multidisciplinary work lead by the different research teams and its relapses in terms of good practices and opportunities.

3. Performed Activity in the period

This phase aims at testing the technological opportunities offered by the Digital Dissemination Technologies in order to achieve the best visibility and publicity to the project's results. After the thorough analysis of background sources and literature pursued a first draft of a new Digital Dissemination Interface has been devised and described in this report and now needs to be tested and discussed within the research community. The preliminary step of the construction of the Digital Dissemination Interface (DDI) consisted of a survey of general literature concerning the most diffused Digital Contents Management Strategies and of a benchmarking of the most used procedures among the international scientific editorial pillars (Elsevier, Springer-Nature, Cambridge etc.). The second step was to imagine and devise an agile and effective web structure apt to serve the purpose of a general-addressed scientific dissemination. The third step was to analyze the first results of the scientific research to find a way to populate the raw infrastructure with actual contents to test their usability. This step was pursued as soon as the first contributions appeared since January 2023 and consisted of a detailed analysis of the scientific publications produced to highlight the specific contribution of the researchers and map their distribution in the different departments and areas of expertise. As far as the dissemination-issue in concerned, a throughout analysis of main topics was pursued by trying to enucleate the scientific trends and methodological innovations to convey them in a more informative way, to let research community get an unbiased insight about the project's state of art. In order to design the DDI as an improvement of the current digitalization procedures of the project's results, the timing of the publication of the contents on the project's official website and

their updating were monitored, as well as the functioning of the links to the publication repositories, the ease of access and the effectiveness of the downloading and open access procedures. The second step was to devise a model of the prospective Digital Dissemination Interface designed around the specificity of the case-study to explore the possibility to apply it to the other areas. The prospective webpages and submenus were initially designed to host a brief and informative survey of the main themes, methodologies and innovations contained in the articles and accompanied by short text- or video- interviews with the authors to render the dimension of team-working experience, multi-disciplinary work and concrete life practice. The articles present in the report are those present at the time of writing the report in SPOKE 3 S2-COMMs Micro and Nanotechnologies for Smart Sustainable Communities, headed by the Palermo unit as leading partner. Parallel to the activity of description contained in the static webpages the WP2.3 team has imagined another series of Digital Disseminations Devices (DDD) that could be integrated in the DDI or specifically devised to flank it as separated channels. These devices could include: Content Formats as Text - medium to long form blogs, can be augmented with images, illustrations and graphics; Video - require scripting, subtitles, video footage, still images, graphics and music; also can include Facebook live event coverage etc.; Audio - can include podcasts, audiograms and other forms of audio content like music or sound effects (or a combination of the two, such as footage of a podcast recording); Social - micro content forms created specifically for social media, including audiograms, photo texts, graphics and short lo-fi 'selfie' videos (for example a short clip of the researcher talking directly into a phone about a current topic or event).

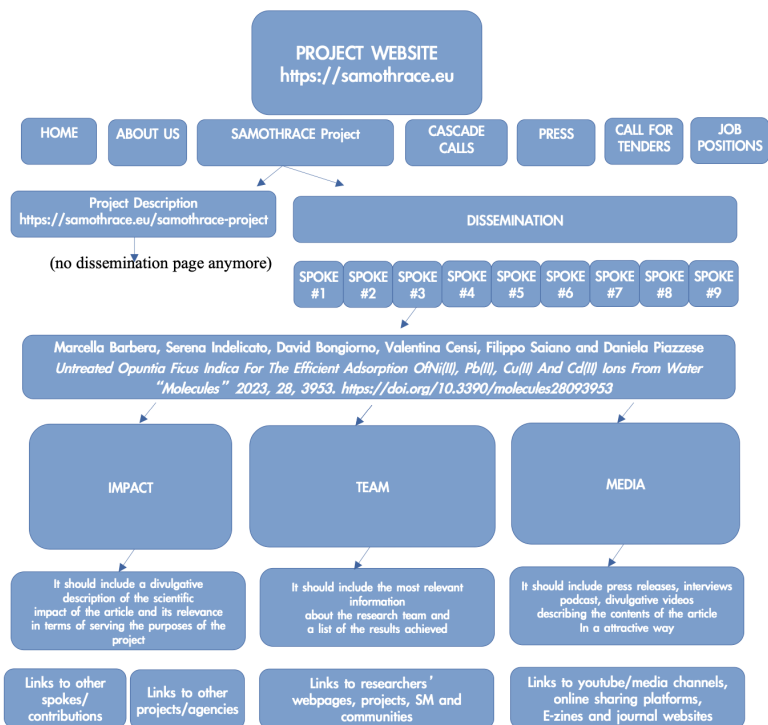
4. Achieved results

The analysis of the publications has allowed to assess that the scientific results cover a wide range of studies and applications in the field of nanotechnologies ranging from the contribution of basic physical and chemical sciences in the development of new materials to the search for new application procedures, particularly in the context of medical-health disciplines as well as agricultural, forestry and environmental sciences. According to the WP2.3 team, this consideration should play

a key-role in the devising of the actual DDI as specifically addressed to the general purposes of the project. This should be reflected in the strategies used to build the graphic format of the DDI, which should therefore use templates, colors and modules inspired to the general concepts of “care”, “health” and “environment”, trying to harmonize with the Samothrace general graphic concepts (a simplified and generalized using of blue shades and nuances used to evoke the surroundings of the seas and the territoriality of Sicily Island). According to these considerations and basing on the analysis of the actual contents that should be displayed on the DDI (which the WP2.3 team considers as a core presupposition of any real Content Management work), the actual structure of the website (as far as the Dissemination issue is concerned) could be restyled and the following structure of the DDI could be imagined as preliminary step of any future work of the WP. The general aim of this infrastructure should be to enhance and valorize the impact of the contributions and the status of the researcher in the international community to give visibility and publicity to the projects’ results and so attract more potential partners and stakeholders. Some detailed examples of prospective managed contents referring to the “Impact” and the “Media” sections of the DDI are reported in the document 1 and document 2 attached to the present document. The first one containing some informative description of the articles, the second one reporting two interviews with the corresponding authors of two different UNIPA research teams who give back their experience in the Samothrace project in terms of cherishing good practices/opportunities and to warn about potential treads and troubles.

The general structure of the DDI is displayed in the figure 1 below:

DIGITAL DISSEMINATION INTERFACE



5. Remarks/Treads

The following considerations concern the possible areas of improvement of the only dissemination tool implemented in the Samothrace's general website so far. In general, the publication of articles on the project website is punctual and respects deadlines. Their placement within the menus is not very precise as the articles can be reached in two ways: 1. through the "Dissemination" item of the "Samothrace Project" submenu which is generic and wrongly placed next to the "About Us" and "Home" (which itself contains an "about"). From within the "Dissemination" item, the publications of the various spokes are placed in columns which are each divided into pages accessible via "previous-next" links which are counterintuitive as the articles are ordered from the most recent to the oldest and therefore by clicking on "next" you actually go backwards in the order of publications and not forwards; 2 also via the "Project Description" item which appears together with the already analyzed "Dissemination" in the same submenu. In this case, to get to the articles you need to click on the individual spokes and scroll down the entire page until you find the "Dissemination" section which contains a sort of duplicate of the previous section with the difference that here the articles are present all together and not in subpages, thus offering more immediate accessibility. The access to the articles is effective with the possibility of viewing both the .pdf file of the contribution in its original repository and downloading it directly via the "download" option. In all cases examined, the open access requirement appears to have been respected. This setting could be seen as extremely basic and barely sufficient even at a starting point of the project and it is opinion of the WP2.3 team that the whole interface should be improved and restyled in order to grant the best valorization to the project's results. In concluding these remarks, the general ability of the dissemination infrastructure so far can be considered sufficient but surely improvable. The first way to optimize these procedures could be the organization at the beginning of 2024 of a workshop on Digital Dissemination led by the WP2.3 team to inform the research teams about the strategies and techniques proposed and developed in the DDI format and described in this report, which could also be distributed. During the workshop the main topics of the report could be discussed, and it would also be possible to gather opinions and inputs about the

best way to effectively disseminate the project's results according to the specificity of every single area of expertise and knowledge.

6. General survey of the state-of-the-art with a detailed analysis of scientific topics and distribution of research contribution among Unipa Departments

The editorial placement of the contributions appears generally excellent, including contributions in many of the main international magazines in the sector. The vocabulary used in the contributions is obviously very specialized and therefore requires parallel work to disseminate the results even beyond the narrow scope of comparison of particular scientific communities. As points of interest that could be identified to begin conveying the important results to the general public, we can identify first of all 1. the demonstrated impact of new nanotechnologies in the development of innovative diagnostic and therapeutic procedures with particular attention to the recent pandemic experience; 2. the potential impact of new nanotechnologies in the implementation of innovative strategies for environmental protection and valorisation of natural resources and territories.

Starting from 13 January 2023, 7 scientific articles have been produced which saw the participation of 29 UNIPA researchers including correspondents and participants of the research groups who authored the publication. The dissemination work began with a review published in "Pharmaceutics" entitled *Chemopreventive and Anticancer Role of Resveratrol against Oral Squamous Cell Carcinoma*, the result of the interdisciplinary work of five Unipa researchers belonging to the Departments of Surgical, Oncological and Stomatological Disciplines and Sciences and Technologies Biological, Chemical and Pharmaceutical. On February 19, an article entitled *Release Profiles of Carvacrol or Chlorhexidine of PLA/Graphene Nanoplatelets Membranes Prepared Using Electrospinning and Solution Blow Spinning: A Comparative Study*, the result of the joint work of three researchers from the Engineering Departments, was published in "Molecules". and Agricultural, Food and Forestry Sciences. On March 29, five researchers from the Department of Agricultural, Food and Forestry Sciences produced, in collaboration with a colleague from the Laboratory of Industrial Microbiology, Faculty of Sciences, National University of Colombia, an article on

"Plants" entitled *Development of an In-Field Real -Time LAMP Assay for Rapid Detection of Tomato Leaf Curl New Delhi Virus*. On April 21, three researchers from the Department of Engineering published an original paper in the "Journal of Polymers and the Environment" entitled *Biodegradable Membrane with High Porosity and Hollow Structure Obtained via Electrospinning for Oil Spill Clean up Application*. On May 6, 5 researchers from the Department of Biological, Chemical and Pharmaceutical Sciences and Technologies in collaboration with a colleague from the Institute for the Study of Nanostructured Materials (ISMN), National Research Council (CNR) published in the "International Journal of Molecular Sciences" an article entitled *In Silico Design of New Dual Inhibitors of SARS-CoV-2 MPRO through Ligand- and Structure-Based Methods*. On 6 May 8 researchers distributed between the Departments of Surgical, Oncological and Stomatological Disciplines, Biological, Chemical and Pharmaceutical Sciences and Technologies and Health Promotion, Maternal and Child, Internal and Specialized Medicine of Excellence "G. D'Alessandro", in collaboration with a colleague of the Institute for the Study of Anthropogenic Impacts and Sustainability in the Marine Environment, National Research Council (IAS-CNR) of Campobello di Mazara have published an article in the "International Journal of Molecular Sciences" entitled *Multicomponent Antibiofilm Lipid Nanoparticles as Novel Platform to Ameliorate Resveratrol Properties: Preliminary Outcomes on Fibroblast Proliferation and Migration*. Finally, on May 8, 6 researchers distributed between the Departments of Earth and Marine Sciences, Biological, Chemical and Pharmaceutical Sciences and Technologies and Agricultural, Food and Forestry Sciences produced an article on "Molecules" entitled *Untreated Opuntia ficus indica for the Efficient Adsorption of Ni(II), Pb(II), Cu(II) and Cd(II) Ions from Water*.

The first publication produced concerns an application of nanotechnology research to the development of anti-cancer drugs. In particular, oral squamous cell carcinoma (OSCC) is one of the most common and aggressive head and neck tumors, characterized by high morbidity and mortality. Available conventional treatments suffer from numerous adverse effects and are often ineffective in terms of survival rates. Therefore, the search for new therapeutic agents and adjuvants is of utmost importance for modern society. Natural polyphenolic compounds have recently emerged as promising chemopreventive and

anticancer agents. Specifically, the natural compound resveratrol (RSV) has recently gained momentum for this purpose. It is useful for the treatment of OSCC thanks to its antiproliferative, antimetastatic and proapoptotic effects. Furthermore, RSV acts against tumor cells by cooperating synergistically with chemotherapeutics, overcoming drug resistance phenomena. Despite these broad-spectrum effects, there are few specific investigations of the effects of RSV against OSCC animal models that consider different routes and vehicles for RSV administration. The review offers a comprehensive overview of the *in vitro* and *in vivo* effects of RSV and its main derivative, polydatin (PD), against OSCC-related cell lines and animal models, with the aim of guiding the scientific community regarding concerns RSV and the use of PD in the treatment of oral precancerous and cancerous lesions.

The second contribution directly addresses the comparison between two of the most widespread techniques for the production of nanoplastics, in particular nanofibrous membranes. These are often the main components used to produce controlled release devices and are often prepared by electrospinning (ES). However, ES requires high production time and cost and is not easy to scale. Recently, solution blowing (SBS) has been proposed as an alternative technique for the production of nanofibrous membranes. In this study a comparison between these two techniques is proposed. The results revealed that both ES and SBS nanofibrous mats achieved controlled release for up to 500 hours. In detail, the lower wettability of the SBS system allowed the CRV release kinetics to be slowed down, compared to that obtained for ES membranes. In contrast, with SBS, a faster release of CHX can be achieved due to its more hydrophilic behavior. Furthermore, the addition of graphene nanoplatelets (GNPs) led to a decrease in wettability and allowed a slowdown of the release kinetics in the whole system.

The third article concerns the development of a particular diagnostic procedure for the prevention of New Delhi tomato leaf curl virus (ToLCNDV). This disease represents a threat to economically important horticultural crops. A real-time loop-mediated isothermal amplification (LAMP) assay was developed for the detection of ToLCNDV in the field, coupled with a rapid sample preparation method and tested in both field and laboratory conditions on samples of courgettes, tomatoes and peppers. The developed real-time LAMP assay is a rapid,

simple, specific and sensitive technique for the detection of ToLCNDV and can be adopted as a routine test, both for field and laboratory conditions.

In the fourth contribution, some techniques for the production of nanoplastics to be used for the treatment of waste water and therefore to be used in particular cases of environmental pollution are addressed in detail and with the help of innovative experimental procedures. The use of biodegradable polymers for the production of membranes for use in wastewater treatment has attracted growing interest considering the possibility of reducing the risk of second pollution. In this work, porous fibrous membranes based on mixtures of polylactic acid and polyethylene oxide (PEO) were prepared. Process, properties and structural relationships of the devices were analyzed through rheological, morphological, mechanical and surface characterizations. Furthermore, the influence of different porous structures on the oil absorption capacity and reusability of the membranes was evaluated. The results reveal that different porosities lead to a variation in the mechanical performance of the membranes, their wettability and, consequently, their oil spill cleanup ability. Membranes obtained with SS showed higher performance in oil removal than DS ones, due to their hollow structure and larger surface area.

The fifth contribution impacts one of the most urgent and current issues, namely the procedures to be developed quickly to deal with the risks of the COVID-19 pandemic and therefore as a general instance of reaction to the emergence of new and unexpected pathologies. The chemical process under study is the viral master protease, one of the main targets among all the key enzymes involved in the life cycle of SARS-CoV-2. We identify two sets of small molecules with significant affinity for SARS-CoV-2 MPRO using a hybrid virtual screening protocol, combining ligand- and structure-based approaches with multivariate statistical analysis using the Biotarget Predictor tool.

The sixth contribution adds to the current framework of experimentation of clinical and diagnostic procedures the aspect of the "nanoplastic" construction of the same organic cells as a key to developing new therapeutic procedures. In particular, the well-being of the skin and mucous membranes is fundamental for the body's homeostasis and it is therefore essential to treat any injury quickly and correctly. From this perspective, molecules such as polyphenols could assist and

promote the success of the wound healing process by reducing the inflammatory cascade and the production of free radicals. However, they have disadvantageous physicochemical properties, which determine their limited clinical use. In this work, a complex mixture of PEGylated lipids, glyceryl monoester, 18- β -glycyrrhetic acid, and menthol was designed to once again trap resveratrol (RSV) as the active ingredient and further produce lipid nanoparticles (LNPs). The nanosystem was therefore tested according to innovative procedures and techniques that could represent a valuable RSV delivery platform for wound healing purposes.

The last contribution produced so far integrates the transversal, multidisciplinary and particularly attentive to the impact on the territory scenario that characterizes the participation of the Unipa scientific departments with a study on the raw cladode of *Opuntia ficus indica* (OFI, our prickly pear) as a potential sustainable biosorbent for the removal of heavy metals (Ni, Pb, Cu and Cd) from aqueous solutions. The results of this work highlight the concrete possibility of using the untreated OFI cladode as an economical and ecological biosorbent for the removal of heavy metals from contaminated aqueous solution.

7. Conclusion

In general, the scientific publications that has been delivered so far confirm the multidisciplinary and transversal commitment of the Unipa Departments and the good pool of researchers involved in working on concrete and effective synergies between the disciplinary areas. The realization of these synergies is made possible by the framework of the Samothrace project and its ability to offer spaces of intersection between basic and applied research within the framework of a general interest in health and care of the body, the environment and the territory.

The development of interdisciplinary and effective strategies for disseminating results therefore becomes a key point for the valorisation of the project results including communication to the general public as well as to ever new potential industrial and commercial partners. The work should now be carried on further to explore the trainability and usability of the DDI in order to cover all the project's areas of expertise.

References

- Best, S. (2013). *Understanding and Doing Successful Research. Data Collection and Analysis for the Social Sciences*. Routledge.
- Bui Thanh Khoa, Bui Phu Hung and Mohsen Hejsalem-Brahmi (2023). *Qualitative research in social sciences: data collection, data analysis and report writing*. International Journal of Public Sector Performance Management, Vol. 12, No. 1-2.
- Hardy, N. (2008). *Disseminating information via Web 2.0*. Acm Sigcas Computers and Society 38 (2):24-24.
- Huberman, M. Ben-Peretz, M. (1994). *Disseminating and using research knowledge*. Knowledge, Technology Policy 7 (4):3-12.
- O'Sullivan, S.; Desmond, E. Buckley, M. (2023). *The Ethics of Engagement and Representation in Community-based Participatory Research*. Ethics and Social Welfare 17 (2):159-174.
- Stein, M. (1983). *The Creative Process and the Synthesis and Dissemination of Knowledge*. In Spencer A. Ward Linda J. Reed (eds.), *Knowledge Structure and Use: Implications for Synthesis and Interpretation*. Temple University Press. pp. 363-396.
- Stouthamer-Loeber, M., van Kammen, W. B. (1995). *Data collection and management: A practical guide*. Sage Publications Inc.

Andrea Le Moli
Università degli studi di Palermo
andrea.lemoli@unipa.it



ἘΠΈΚΕΙΝΑ

International Journal of Ontology
History and Critics

RICCARDO TARANTINO

*Pattern emergenti nei sistemi biologici: la creatività intrinseca
dei modelli basati su agenti*

EPEKEINA, vol. 16, n. 1-2 (2023), pp. 1-14
Philosophy of Technology

ISSN: 2281-3209

DOI: 10.7408/epkn.

Published on-line by:

CRF – CENTRO INTERNAZIONALE PER LA RICERCA FILOSOFICA
PALERMO (ITALY)

www.ricercafilosofica.it/epekeina



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs 3.0 Unported License.

Pattern emergenti nei sistemi biologici: la creatività intrinseca dei modelli basati su agenti

Riccardo Tarantino

Introduzione

Un algoritmo è, nell'opinione comune, quanto di più lontano si possa immaginare da qualunque tipo di *creatività*. La funzione degli algoritmi, infatti, è quella di produrre determinati *output* a partire da certi *input* in un numero finito di passi. Cosa può esserci di meno creativo di una sequenza di istruzioni non ambigue e dettagliate fino alla pedanteria impartite da un programmatore in un preciso linguaggio di programmazione?

Come intuibile, esistono in generale numerose definizioni di creatività, anche piuttosto diverse tra loro. In tempi recenti, la difficoltà di definire questo concetto è stata tra le motivazioni che hanno spinto alcuni ricercatori a cimentarsi finanche con una trattazione empirica del fenomeno, più specificamente di tipo neuroscientifico (Schubert 2021). Con la parola "creatività" si può intendere, per esempio, la capacità di pensare in maniera alternativa, l'abilità di scoprire nuovi problemi, il percorso che va da un'idea iniziale alla realizzazione di un prodotto finale e molto altro ancora. In generale, si tratta di una nozione indubbiamente correlata alla produzione di idee e artefatti in qualche modo utili e originali. Inoltre, viene spesso presentata come strettamente associata a una caratteristica prettamente umana come l'*intenzionalità* degli individui nell'atto di creare qualcosa di nuovo (Walia 2019).

Tuttavia, sembra che l'intenzionalità non sia una caratteristica *necessaria* del fenomeno creativo, come dimostrato (oltre che, in una certa misura, dall'esperienza quotidiana) anche da notevoli casi di creatività scientifica come quelli vissuti e raccontati dal grande fisico e matematico Henri Poincaré. Più in particolare, Poincaré descrisse diversi episodi creativi personali di carattere matematico riconducibili, a suo avviso, a dinamiche del tutto inconscie, seppur evidentemente indotte dalle conoscenze che aveva accumulato nel proprio campo di studi nel corso degli anni, in grado di conferire all'atto creativo le

caratteristiche di un processo in *background* (Rose 1999). Concentrarci su dei tipi di creatività non intenzionali (come quelli inconsci appena citati) ci fornisce uno spunto per interrogarci sulle forme assunte dalla creatività in un ambito non-umano. In particolar modo, ci permette di introdurre il dibattito sull'effettiva esistenza di un qualche tipo di creatività nelle macchine.

Questo articolo si occupa di indagare un tipo ancor più specifico di creatività, ossia la creatività inerente agli algoritmi utilizzati per l'implementazione dei *modelli basati su agenti*. Quello della modellizzazione basata su agenti è un particolare approccio scientifico basato sulla ricostruzione e sulla simulazione di fenomeni naturali attraverso l'interazione tra individui computazionali detti *agenti*, il cui comportamento viene codificato da un programmatore con l'utilizzo di algoritmi. Pur seguendo pedissequamente le istruzioni indicate dal programmatore, gli agenti si rivelano spesso in grado di generare schemi complessivi (o *pattern emergenti*) imprevedibili sulla base della sola analisi del comportamento del singolo e non deducibili dal codice a cui ciascuno di loro, inevitabilmente, si attiene.

Non a caso, questo tipo di modellizzazione viene usato per riprodurre al computer il funzionamento dei cosiddetti *sistemi complessi*, ossia sistemi privi di controllo centralizzato, in cui le parti sono fortemente interdipendenti e nei quali, pur mantenendosi una forte componente casuale, l'ordine e le regolarità emergono spontaneamente e si preservano grazie a un certo grado di resistenza alle perturbazioni e alla presenza di meccanismi di autoregolazione. La proprietà più determinante di questi sistemi, tuttavia, risiede nel fatto che i *pattern* emergenti che li caratterizzano sono determinati unicamente dall'articolata interazione tra le unità che li compongono, analogamente a quanto avviene nei modelli ad agenti (che sono così costruiti "ad immagine e somiglianza" dei sistemi reali che rappresentano).

Ora, senza volerci soffermare sulla questione di carattere più generale sul fatto che la natura possa essere "creativa" o meno, l'intento principale di questo lavoro è di mostrare come la *creatività computazionale* sia una caratteristica importante, se non addirittura intrinseca e necessaria, di questa categoria di modelli. Come vedremo, infatti, è *fondamentale* che un modello ad agenti ben implementato sia in grado di generare delle dinamiche globali emergenti che non siano in alcun modo inserite direttamente nel codice dal suo autore.

A dimostrazione di ciò, la prima sezione del contributo si sofferma sulla descrizione di un modello basato su agenti implementato con il *software* NetLogo (Wilensky 1999) e in grado di riprodurre alcuni degli aspetti salienti di uno specifico fenomeno biologico. In questa sezione, oltre a fornire una semplice descrizione di NetLogo, del modello e del processo simulato, approfondisco soprattutto il modo in cui gli agenti considerati riescono a produrre all'unisono dei *pattern* emergenti che non sono codificati dal programmatore, e rappresentano piuttosto il risultato spontaneo e imprevedibile a priori dell'interazione tra entità individuali dal comportamento algoritmico.

Nella seconda sezione, mi soffermo sugli aspetti teorici più generali e rilevanti presentati nella prima e cioè, sostanzialmente, sul perché l'approccio della modellizzazione ad agenti sia intrinsecamente così affine alla presenza di creatività computazionale, quasi che quest'ultima costituisca una vera e propria premessa e una garanzia metodologica di funzionamento del modello implementato.

Nella sezione conclusiva, sintetizzo il messaggio essenziale di questa incursione teorica nella modellizzazione basata su agenti e ne inquadro il valore nel contesto più generale della ricerca scientifica basata su simulazioni al computer.

1. Evoluzione *in silico*: un fenomeno emergente non codificato dal programmatore

L'*evoluzione biologica* è il cambiamento degli organismi al susseguirsi delle generazioni. Sebbene questo macrofenomeno si verifichi spesso in tempi estremamente lunghi (*troppo* lunghi per un osservatore umano), gli eventi che lo compongono sono, almeno in linea di principio, osservabili e possono in ogni caso essere inferiti o previsti (con un ampio margine di errore) attraverso l'uso di modelli teorici di tipo probabilistico. Questo cambiamento intergenerazionale, in estrema sintesi, non è infatti altro che l'effetto delle leggi dell'ereditarietà calate nel contesto delle interazioni ecologiche complesse tra gli organismi, senza dimenticare l'influenza notevole della componente casuale.

Sebbene esistano diversi motivi metodologici rilevanti per decidere di modellare un sistema biologico in evoluzione con delle simulazioni fatte al computer (Taylor & Jefferson 1995), quello più determinante resta, a mio avviso, quello appena menzionato riguardante l'enorme

lentezza di questo genere di processi nel “mondo reale”. Sebbene esistano anche importanti casi di studio dell’evoluzione di organismi reali condotti in laboratorio, si tratta normalmente di ricerche dalla durata davvero spropositata, svolgendosi nell’arco di decine di anni ed essendo per di più vincolate all’utilizzo pressoché esclusivo di specie dotate di brevissimi cicli vitali ed estremamente prolifiche (grazie, per esempio, a una modalità riproduttiva di tipo asessuale, come nel caso dei celebri batteri della specie *Escherichia coli*) (Lenski 2017).

Uno dei linguaggi di programmazione e ambienti di sviluppo integrato più suggestivi e di semplice utilizzo in grado di venire incontro ai ricercatori afferenti a quest’area della biologia è senz’altro NetLogo. NetLogo è specificamente pensato per l’implementazione e la simulazione di modelli basati su agenti, ed è oggi ampiamente utilizzato sia nell’ambito della didattica che della ricerca, grazie soprattutto all’intuitività della sua sintassi (pur essendo basato sui linguaggi Scala e Java, infatti, NetLogo è un linguaggio a sé stante) e alla possibilità di vedere facilmente in azione i sistemi computazionali costruiti passo dopo passo già in fase di implementazione attraverso la schermata di Interfaccia, con la quale l’utente interagisce per avviare e gestire le proprie simulazioni. Per usare dei termini un po’ meno specialistici, NetLogo è essenzialmente una piattaforma attraverso la quale è possibile programmare delle entità virtuali (gli agenti, appunto) e osservare come queste entità si comportino e interagiscano tra loro, in tempo reale, all’interno di un ambiente 2D.

Il *software* è inoltre associato a una *Models Library*, dalla quale è possibile selezionare e aprire un modello preesistente il cui funzionamento è verificato dai gestori di NetLogo. Una volta aperto il modello, è possibile interagire con esso tramite l’Interfaccia, comprenderne il funzionamento tramite una schermata Informazioni e, eventualmente, modificarlo per mezzo della schermata Codice. Come intuibile, è possibile imbattersi in una grande varietà di modelli riconducibili ad aree molto differenti, dalla filosofia alla chimica, dalla sociologia all’economia.

Ora, al fine di mostrare con un esempio concreto il tipo di creatività computazionale a cui faccio riferimento in questo articolo, intendo esporre brevemente un modello evolutivo tratto da questa libreria. Più nello specifico, il modello in questione riguarda la cosiddetta *microevoluzione*, che è l’oggetto di studio dell’area della *genetica di popolazioni*.

Questa disciplina si occupa di indagare le variazioni nel corso del tempo delle frequenze di determinate varianti di geni trasmesse, secondo le leggi dell'ereditarietà, da una generazione all'altra in un gruppo di organismi di una stessa specie e che condividono lo stesso areale geografico. Oltre a essere influenzate dai processi ereditari, queste frequenze sono soggette a oscillazioni indotte da un certo numero di meccanismi direzionanti, come per esempio la *selezione naturale*, che favorisce la replicazione delle varianti più "adatte" al contesto ecologico in cui gli organismi della popolazione si trovano.

Il nome del modello che ho scelto per supportare il mio argomento è altamente esplicativo: *Natural Selection – Camouflage* (Dabholkar & Wilensky 2020), ossia "Selezione Naturale – Mimetismo". Il modello rappresenta in maniera semplificata l'evoluzione di una popolazione di topi sottoposta a una pressione selettiva legata alla capacità di mimetizzarsi con l'ambiente. Prima di descrivere sinteticamente il funzionamento di questo sistema virtuale, è opportuno chiarire alcuni aspetti teorici di base.

Si noti preliminarmente come l'approccio computazionale consenta di trattare l'evoluzione di animali con regole di trasmissione dell'informazione genetica più complesse rispetto a quelle dei batteri, in quanto il genoma dei topi è organizzato in svariate coppie di cromosomi *quasi* identici tra loro detti *omologhi* (inoltre la specie non è asessuale ma presenta due sessi separati, l'uno in grado di produrre spermatozoi, l'altro cellule uovo). Questo significa che tutti i geni localizzati sui cromosomi sono presenti in doppia copia (escludendo i cromosomi sessuali, che qui non ci interessano). Ciò che gli autori del modello hanno fatto è stato considerare esclusivamente un gene ipotetico per il quale esistono soltanto due varianti alternative, *A* e *a*. Ciascun topo possiede un preciso *genotipo*, cioè una combinazione di queste due varianti geniche: in questo caso, i soli tre genotipi possibili sono *AA*, *Aa/aA* e *aa*. Il primo genotipo è detto *omozigote dominante*, il secondo *eterozigote* e il terzo *omozigote recessivo*. L'aspetto esteriore dei topi (il *fenotipo*) è qui determinato "rigidamente" dal genotipo. Più in particolare, il fenotipo "colore scuro del pelo" viene prodotto sia dal genotipo *AA* che dal genotipo *Aa* (o *aA*), mentre il fenotipo "colore chiaro del pelo" solamente dal genotipo *aa*.

Sofferamoci adesso sull'utilizzo del modello (vedi Figura 1 per visualizzarne l'Interfaccia): dopo aver impostato il numero iniziale di

topi per ciascun genotipo e per ciascun sesso, l'utente può decidere se attivare con un interruttore la predazione da parte di un volatile e regolare la probabilità di predazione (che determina il numero complessivo di predatori generati). È anche possibile generare in qualunque momento un nuovo individuo *mutante* con genotipo eterozigote premendo su un pulsante apposito. Dopo aver avviato la simulazione premendo un altro pulsante, si osservano i topi camminare casualmente nel mondo di simulazione di NetLogo, invecchiare e morire, o per via dell'età o a causa della predazione. Oltre che dal numero di predatori che si aggirano per il mondo di NetLogo, la probabilità di predazione per ciascun topo è influenzata dalla capacità di quest'ultimo di mimetizzarsi con l'ambiente circostante, cioè con il colore delle *patch* (le unità spaziali bidimensionali di NetLogo, ossia delle celle quadrate su un piano cartesiano) che lo circondano, che può essere più scuro o più chiaro e può essere regolato dall'utente anche nel corso della simulazione. In altri termini, i topi con colore del pelo più simile a quello dello sfondo sopravvivono (e possono riprodursi) più frequentemente quando incontrano lungo il loro cammino dei topi di sesso opposto, mentre quelli che risaltano di più sullo sfondo vengono uccisi più frequentemente, evento che riduce la rappresentanza nella popolazione delle varianti geniche di cui sono portatori.

Le capacità mimetiche sono dunque cruciali per canalizzare l'evoluzione del sistema in termini di frequenze di varianti geniche, dal momento che sono proprio i genotipi a determinare il colore del pelo e la selezione naturale agisce su questo carattere fenotipico. Ciascuna simulazione risulta enormemente velocizzata rispetto alla propria controparte reale: basti pensare al fatto che una singola unità di tempo discreta di NetLogo, il *tick*, viene considerata equivalente a una generazione di animali. L'andamento delle frequenze delle varianti geniche, dei genotipi e dei fenotipi è tracciato da dei monitor e da un grafico, dai quali è molto semplice constatare gli eventi di *fissazione* o *estinzione* di una delle due varianti nella popolazione. Questi eventi critici rappresentano il principale *fenomeno emergente* di questo sistema complesso biologico. L'interazione tra organismi, sia di tipo riproduttivo che predatorio, contribuisce insomma al cambiamento complessivo di una macro-entità integrata (la popolazione, appunto), fino al raggiungimento di stati irreversibili che contemplan la perdita di diversità genetica.

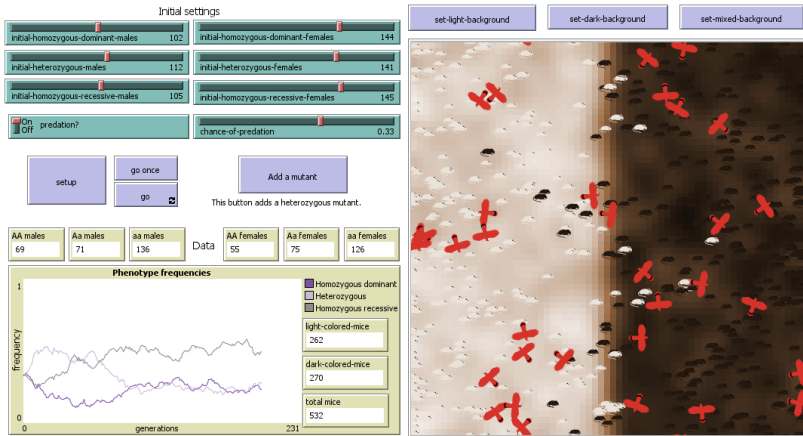


Figura 1: L'Interfaccia del modello implementato in NetLogo *Natural Selection - Camouflage* (Dabholkar & Wilensky 2020) nel corso di una simulazione. Si noti come sulla metà del mondo di simulazione più chiara siano presenti quasi esclusivamente topi di colore bianco e, viceversa, come la metà più scura sia popolata quasi interamente da topi neri, una conseguenza diretta del processo di selezione naturale basato sulle capacità mimetiche.

Ma il comportamento di questa macro-entità integrata è stato impostato dal programmatore? Esistono, in altre parole, delle righe di codice che *impongano* al sistema di evolvere in quel dato modo? La risposta è negativa, ed è facile dimostrarlo analizzando le varie porzioni di codice nel dettaglio. Nella schermata Codice corrispondente al modello appena descritto appaiono, in quest'ordine, le seguenti istruzioni: vengono definite le variabili "colore del pelo chiaro" e "colore del pelo scuro"; vengono definite le categorie di topi e predatori; vengono definite le caratteristiche dei topi (genotipo, età, partner, sesso); vengono definite le procedure di inizializzazione delle simulazioni, che generano gli agenti iniziali, sia topi che predatori, con le loro caratteristiche (genotipo, fenotipo, sesso, assenza di partner, età pari a 0, dimensioni, forma) e impongono il colore marrone allo sfondo; viene definita una procedura di simulazione composta da svariate sotto-procedure; vengono definite singolarmente le varie sotto-procedure, cioè quella che regola il movimento casuale degli agenti nel mondo, quella che regola la morte dei topi in base all'età, quella che regola la probabilità di morte dei topi in base alla presenza di predatori e alle proprie capacità mimetiche (in relazione al proprio fenotipo e al colore delle *patch* circostanti), quella che regola la ricerca di partner da parte dei topi e quella che regola la conseguente riproduzione in base alle regole probabilistiche dell'ereditarietà classica mendeliana; viene definita una procedura per l'aggiunta di individui mutanti eterozigoti alla pressione dell'apposito pulsante; viene definita una procedura per la gestione dinamica del numero di predatori; vengono definite le procedure per impostare il colore dello sfondo del mondo di NetLogo (chiaro, scuro o misto).

Come si può evincere dalla breve disamina appena compiuta, non vi è nel codice alcun riferimento, né esplicito né implicito, a istruzioni che il sistema nel suo complesso dovrebbe eseguire per generare il comportamento emergente delle fluttuazioni nel *pool* delle varianti geniche e della conseguente eliminazione della diversità genetica. L'attività di programmazione si sofferma *esclusivamente* sull'implementazione di semplici algoritmi assegnati alle varie classi di agenti, sia mobili (topi, predatori) che fissi (*patch*). Il processo di formazione delle dinamiche globali, esattamente come avverrebbe nel sistema biologico di riferimento, entra in funzione *spontaneamente* a partire dalle interazioni tra gli agenti. Eppure, tutto ciò che il modello è in grado di simulare discende, chiaramente, proprio da ciò che è scritto nella schermata

Codice. Questo apparente paradosso è, plausibilmente, l'aspetto in assoluto più interessante della modellizzazione basata su agenti, che prevede come suo stesso fondamento metodologico la produzione di sistemi computazionali in grado di determinare le proprie dinamiche in maniera autonoma, senza un intervento diretto del programmatore. A questo aspetto cruciale è dedicata la prossima sezione.

2. La creatività computazionale come criterio di validità dei modelli ad agenti

Una conclusione importante che discende direttamente dalla descrizione del modello appena commentato è che gli *output* dei modelli ad agenti possano rappresentare dei casi peculiari di creatività. La produzione spontanea di schemi emergenti (simulati) in qualche misura *nuovi, imprevedibili* e, spesso, *utili* per la ricerca scientifica, può infatti essere considerata una forma di creatività computazionale, in quanto risultante da una serie di algoritmi e, al contempo, non inferibile analizzando gli algoritmi implementati. Sulla base di questo presupposto, ciò che intendo sostenere in questa sezione è l'esistenza di un legame profondo tra creatività computazionale e buona pratica di costruzione dei modelli basati su agenti. Se, come sostengo, la soppressione di questa forma di creatività è addirittura *incompatibile* con la corretta progettazione di un modello, la creatività computazionale può essere ritenuta un requisito metodologico *essenziale* di questo particolare approccio di indagine. Ma procediamo con ordine.

L'aspetto probabilmente più pregnante dell'esempio precedente, che è in effetti generalizzabile a tutti i modelli basati su agenti in grado di rappresentare in maniera soddisfacente le dinamiche dei sistemi complessi a cui si ispirano, è la sua capacità di generare delle proprietà emergenti computazionali in maniera analoga a come queste si generano nella realtà. Vi è, in altri termini, un sostanziale *isomorfismo* tra la formazione delle proprietà emergenti generata dall'auto-organizzazione delle entità reali e la formazione delle stesse proprietà simulate nel modello computazionale, anch'esse generate dalla capacità degli agenti di auto-organizzarsi. In entrambi i casi, è del tutto assente un controllo centralizzato: le singole unità interagenti seguono delle leggi o si attengono a delle regolarità (che, nel caso del modello computazionale, corrispondono agli algoritmi sviluppati dal program-

matore), ma il risultato della loro attività collettiva viene generato solo a posteriori e senza alcun intervento diretto o legge di “livello superiore”.

Il passaggio dirimente di questo ragionamento sta nel fatto che, sebbene gli “ingranaggi” di questi sistemi siano ovviamente i singoli individui, l’evoluzione a livello sistemico aggiunga *qualcosa* rispetto ai vincoli di base posti dalle regole individuali, che sono puntiformi per definizione. Dal punto di vista dei singoli agenti, il codice scritto dal programmatore di ogni modello di questo tipo è chiaramente la sorgente di ciò che avviene nelle simulazioni. Tutto ciò che gli agenti sono in grado di fare è completamente basato sul programma, ed è proprio questo l’aspetto più sorprendente: se il modello è costruito correttamente (cioè, senza errori logici o di sintassi), gli individui computazionali eseguono automaticamente e alla perfezione tutti i comandi descritti nella schermata Codice, motivo per cui lo “scarto” in grado di generare i fenomeni emergenti non può che provenire dalle loro intricate interazioni. È proprio in questa fase che possono essere prodotti degli effetti imprevisti, in quanto il risultato di queste interazioni cumulative non è e *non dovrebbe* essere descritto in alcun modo dall’autore del modello.

Quest’ultimo approccio, plausibilmente, reprimerebbe infatti il fattore imprevedibilità in favore di un maggiore (o totale) controllo sul sistema, un’operazione in piena contraddizione con le stesse fondamentali teoriche della modellizzazione basata su agenti. In linea di principio, si potrebbe senz’altro decidere di implementare un modello in cui anche il comportamento aggregato del sistema sia puntigliosamente definito nel codice. Inserire delle regole di livello superiore, tuttavia, infrangerebbe l’isomorfismo citato precedentemente, in quanto non sarebbero più, o non sarebbero soltanto gli agenti ad auto-organizzarsi e a produrre i *pattern* emergenti, ma sarebbe piuttosto uno specifico centro di controllo a “imporre” tutto dall’alto (il programmatore). In altre parole, incrementare il determinismo intervenendo direttamente al livello *target* piuttosto che a quello inferiore ridurrebbe drasticamente non solo la componente stocastica (che è invece di primaria importanza nei sistemi complessi), ma anche e soprattutto l’aspetto dell’auto-organizzazione. In sintesi, un modello così concepito sarebbe in contraddizione con i principi della modellizzazione basata su agenti e quindi, tecnicamente, *non sarebbe* più un modello basato su agenti. Cosa altrettanto importante, non conterrebbe alcun elemento creativo:

la novità normalmente generata dalla simulazione verrebbe rimpiazzata da un risultato noto (anche solo in maniera approssimativa) fin dall'inizio.

Questa è una delle ragioni per cui impedire a un modello ad agenti di “essere creativo” sarebbe altamente invalidante da una prospettiva prettamente epistemologica. Consideriamo nuovamente il modello ipotetico descritto poc'anzi, in cui è il programmatore a codificare direttamente anche le regole di livello superiore del sistema. Un modello limitato sul versante del proprio contributo creativo endogeno sarebbe inevitabilmente il risultato del contributo (sostitutivo) del programmatore nella produzione degli *output* delle simulazioni. Naturale conseguenza di questa modellizzazione “intrusiva” sarebbe la produzione di risultati artefatti e, molto verosimilmente, distorti (nell'accezione del termine inglese *biased*) dalle aspettative e dalle previsioni del ricercatore. Non è un caso che la metodologia classica della modellizzazione ad agenti preveda, in seguito alla corretta implementazione del modello, una fase più o meno stringente di *validazione* (Wilensky & Rand 2015, 325-326): quando si possiede un modello completo e funzionante, infatti, è auspicabile svolgere delle comparazioni, qualitative o quantitative, atte a *dimostrare* l'aderenza degli *output* delle simulazioni svolte con le osservazioni, gli esperimenti o le deduzioni matematiche già disponibili per quel fenomeno. È grazie al processo di validazione che è possibile accertare che gli *output* del proprio modello siano consistenti con quelli del sistema che esso mira a rappresentare, ed è solo *dopo* questa fase che è possibile utilizzare il modello per compiere delle investigazioni su aspetti del fenomeno considerato non ancora chiari o difficili da trattare con altri metodi o approcci. Codificare il comportamento emergente del sistema nello stesso modo in cui vengono codificati i comportamenti dei singoli componenti invertirebbe completamente l'ordine del processo, rendendo inaffidabili sia i risultati che il modello che li produce.

Esiste, inoltre, almeno un'altra motivazione epistemologica determinante a favore di quello che potremmo definire il “criterio di creatività” dei modelli ad agenti, e risiede nella correlazione positiva tra creatività e *semplicità*. Nel modello ipotetico in cui l'*output* non è emergente ma postulato dal programmatore, è chiaramente necessario inserire dei passaggi ulteriori rispetto a quelli che basterebbero se si decidesse di codificare esclusivamente il comportamento degli agenti,

il che implica un minor grado di semplicità. Un livello di sofisticazione più elevato è normalmente deleterio nella ricerca scientifica, perché contrario al *principio di parsimonia* già noto fin dai tempi di Guglielmo d'Occam: l'approccio di modellizzazione più funzionale e interessante è quello che si avvale del minor intervento possibile da parte dello scienziato. Nel caso dei modelli basati su agenti, la quantità ideale di interventi da parte del programmatore dovrebbe essere molto ridotta, e tali interventi dovrebbero essere limitati agli aspetti più facilmente determinabili e descrivibili del sistema, ossia al comportamento dei singoli agenti (se sono disponibili sufficienti dati empirici al riguardo).

Alla luce di quanto appena argomentato, lungi dal risultare un errore o un'imprecisione in fase di codifica, la riproduzione *in silico* dell'auto-organizzazione tipica dei sistemi complessi reali è, a mio avviso, una possibile forma di creatività computazionale importantissima e in grado di ricoprire il ruolo di criterio di demarcazione metodologico. La possibilità di riprodurre schemi emergenti in un sistema digitale in maniera indiretta, cioè usando degli algoritmi assegnati ai soli agenti, è ciò che più avvicina funzionalmente le simulazioni fatte al computer ai processi naturali modellati, preservando al massimo grado l'isomorfismo tra modello e realtà.

In sintesi, una volta trasposta correttamente in un sistema computazionale, la creatività spontanea presente nei sistemi complessi reali può essere considerata sintomatica di una corretta impostazione del lavoro di implementazione, in quanto riduce il contributo del ricercatore al minimo indispensabile e protegge da distorsioni soggettive e direzionamenti esterni.

Conclusioni

Le simulazioni prodotte dai modelli ad agenti, pur nella loro totale mancanza di intenzionalità, sono in grado di produrre degli *output* che sarebbe arduo non definire creativi. Una mirabile conclusione di questa indagine è l'osservazione di come l'intelligenza artificiale sia in grado di replicare dei frammenti di natura in modo analogo a come sembra farlo la natura stessa, e cioè, lasciando che i propri componenti si organizzino autonomamente in innumerevoli configurazioni inaspettate. Il tassello che ho tentato di aggiungere con questo contributo è la constatazione del fatto che, oltre a possedere un indiscutibile fascino

intrinseco, questo profondo isomorfismo possa inoltre essere considerato un criterio di demarcazione cruciale per distinguere un buon modello basato su agenti da uno soggetto a limitazioni e *bias* importanti. Questa riflessione, di per sé piuttosto specifica e, per certi versi, anche abbastanza tecnica, ci spinge a ragionare su aspetti ben più generali del concetto di creatività. Sia nel mondo reale che in quello digitale (sempre che una distinzione del genere abbia davvero un senso), la creatività sembra configurarsi come un “effetto collaterale” del raggiungimento di una soglia critica di complessità. Il fatto che questa soglia possa essere raggiunta seguendo due vie che sembrano essere totalmente alternative tra loro, come le leggi e le regolarità della natura, da una parte, e la metodologia scientifica e l’approccio di programmazione, dall’altra, merita a mio avviso di essere indagato molto più ampiamente e costituisce, se non una prova, perlomeno un indizio del fatto che delle forme di creatività possano emergere indipendentemente dal sostrato considerato.

Riferimenti bibliografici

Dabholkar, S. & Wilensky, U. 2020, *NetLogo Natural Selection - Camouflage model*. Url: <http://ccl.northwestern.edu/netlogo/models/NaturalSelection-Camouflage>. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Lenski, R. 2017, *Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations*, “International Society for Microbial Ecology”, 11, pp. 2181-2194.

Rose, D. 1999, *Creativity, intentionality and the conscious/unconscious distinction: a neural theory*, “Journal of Intelligent Systems”, 9(5-6), pp. 407-442.

Schubert, E. 2021, *Creativity is optimal novelty and maximal positive affect: a new definition based on the spreading activation model*, “Frontiers in Neuroscience”, 15, 612379.

Taylor, C. & Jefferson, D. 1995, *Artificial Life as a tool for biological inquiry*, in *Artificial Life: an overview*, a cura di Christopher Langton,

MIT Press, Cambridge (Massachusetts), pp. 1-13.

Walia, C. 2019, *A dynamic definition of creativity*, “Creativity Research Journal”, 31(3), pp. 237-247.

Wilensky, U. 1999, NetLogo. <http://ccl.northwestern.edu/netlogo/>. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wilensky, U. & Rand, W. 2015, *An introduction to agent-based modeling*, MIT Press, Cambridge (Massachusetts).

Riccardo Tarantino
Università degli Studi di Palermo
riccardo.tarantino01@unipa.it



ἘΠΈΚΕΙΝΑ

International Journal of Ontology
History and Critics

ENDA O'RIORDAN

A Critique of Creative Computation

EPEKEINA, vol. 16, nn. 1-2 (2023), pp. 1-45

Philosophy of Technology

ISSN: 2281-3209

DOI: 10.7408/epkn.

Published on-line by:

CRF – CENTRO INTERNAZIONALE PER LA RICERCA FILOSOFICA
PALERMO (ITALY)

www.ricercafilosofica.it/epekeina



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs 3.0 Unported License.

A Critique of Creative Computation

Enda O’Riordan

Creative Computation (CC) is a subfield of computer science and AI dedicated to studying creativity in computational systems. Its central claim is that advanced forms of AI are– or have the potential to be– creative “in their own right” (Veale and Pérez Y Pérez 2020; Veale *et al.* 2019, p.2; Veale 2016, p.353). It is an interdisciplinary field which seeks to combine practical implementation with a theoretical approach to understanding the creative capacities of such machines (Veale *et al.* 2019). The impetus for research in CC is in some considerable ways motivated by the observation that creative acts and ideas play a crucial role in the development of both human and nonhuman intelligence (Wiggins 2019, 24; Boden, 2004), and the work of CC is also closely bound to the more speculative claims of AI concerning the realisation of generally intelligent artificial agents (AGI) (Chen *et al.* 2020; Veale *et al.* 2019, 15-16). CC argues that a good standard to assess whether or not a computational system is creative depends on whether or not an unbiased observer would likely regard it as such (Veale *et al.* 2019, 3; Jordanous 2012). A popular view within CC is that extant forms of advanced AI should be considered as “co-creators” alongside human beings (Ibarrola *et al.* 2022, 96; Veale & Pérez Y Pérez 2020; Veale *et al.* 2019). Such machines are to be considered more than mere tools, since they are capable of acting autonomously– meaning without the need for direct human intervention (Ventura 2019, 57)– once initial conditions have been set. CC typically articulates this state of affairs as one in which computational systems enjoy a *degree* of autonomy (Berns *et al.* 2021, 258; Ventura 2019, 57-59; Colton *et al.* 2018, 272; Fitzgerald *et al.* 2017, 105-106), suggesting that autonomy is a gradated rather than categorical concept. The same might be said for its conception of creativity: CC’s auxiliary claim is that the study and modelling of creative AI can also illuminate the underlying mechanisms of human creativity (Veale & Pérez Y Pérez 2020; Boden, 2004). A great deal of literature both informing and informed by CC has been dedicated to the study of creativity as a standalone concept (Veale 2012; Boden 2004), and for CC, the difference between human creativity and that of AI seems to be one of degrees, and not in kind.

There are some general points of agreement within CC about how creativity should be understood. Margaret Boden’s definition of creativity as “the ability to come up with ideas or artefacts that are new, *surprising and valuable*” (Boden 2004, 1) is foundational for CC, as is her general taxonomy of types of creative behaviour (Veale & Pérez Y Pérez 2020, 555; Wiggins 2019, 21; Veale *et al.* 2019, 5-6). Boden’s extended definition of creativity is characterised according to two axes, the first of which distinguishes between transformational and exploratory creativity, and the second of which distinguishes historical creativity (H-creativity) from psychological creativity (P-creativity). To give each its brief explanation in turn: exploratory creativity concerns behaviours that explore existing problem spaces in ways that are novel, surprising, and valuable, whereas transformational creativity expands or transforms spaces, or in other words, widens their parameters (Veale *et al.* 2019; Boden 2004, 4). On the other hand, H-creativity is historically novel, surprising, and valuable, in the sense that the ideas or artefacts in question are unprecedented in the general history of that agent or species. P-creativity, on the other hand, concerns acts which are unprecedented, surprising, and valuable, within the context of the individual agent who performs them. These are described as axes since each represents an ideal type, and the predominance of any one type does is not mutually exclusive to its also containing elements or aspects of the other. Indeed, all cases of H-creativity—cases in which a creative behaviour is historically, rather than merely personally significant—are invariably also P-creative acts (Boden 2004, 2). For convenience, we can call these two axes the exploration-transformation and historical-psychological axes respectively.

The distinction between transformational and exploratory creativity is especially apt for CC. Protocols for acting in novel ways within a given space, by transforming it or exploring it or some combination of both, are largely straightforward pragmatic problems which don’t entail a great deal of philosophical or theoretical difficulty. It is an empirically verifiable matter as to whether a machine can transform or explore a given space in novel ways. This doesn’t completely evade the normative aspect of the problem when we come to describe such empirical observations as surprising or valuable, but from the engineering standpoint, if what you want to achieve is a system that deploys a novel set of heuristics that deviate from the standardised behavioural

pattern, there are ways of going about it without getting into a great deal of philosophical difficulty. The distinction between H-creativity and P-creativity on the other hand proves somewhat more challenging in this respect, and a key part of the argument I present in this paper is that, in fact, this historical dimension of creativity which proves so difficult for CC to replicate is a crucially important part of what we typically mean when we call some person or act creative. Whilst an emphasis on creativity along the exploration-transformation axis can concern itself mainly with pragmatic issues around performance, this historical-psychological axis innervates the creativity problem with more difficult conceptual details such as normativity and socio-historical context. The crucial problem here is less a matter of whether it is empirically discernible whether an act is a historical first for a species, but rather a question of why this seems to matter in a very crucial sense to our concept of creativity, how it functions as a motivational aspect or something which in turn seems to drive what we understand ourselves to be doing when we try to think and act creatively, or commend others for doing so. It may also appear here as though the real trouble is only with H-creativity aspect, whereas we can more easily isolate P-creativity to show how computational systems exhibit forms of creative behaviour local to their own constraints and limitations. I want to suggest here, however, that any meaningful sense of the personal or psychological when it comes to creativity as we apply it to human agents also entails a certain normative dimension in a sense that is not too dissimilar to the way in which this applies to H-creativity. The case I am putting forward, which argues from Hegel and Marx that predicative acts find their objective validity in concepts which pertain to species life, also applies to the semantic valences which accompany acts that are locally significant to individuals. In this sense, the normative, evaluative component of what counts as creative must be treated as informative not only when considering actions which constitute a historical first for a particular species, but also in terms that determine the individual significance of a particular act. This is simply to say that the significance of the actions of individual agents, even when performed remotely of an immediate social context, is a socially mediated significance. The contention of this paper is that we might better understand these normative and motivational features as exemplary of the way in which creativity itself is a normative predica-

te, one which must ultimately be treated as a kind of purposiveness that arises from – and is intrinsically related to – the objectives, goals, desires, and fulfilment criteria of members of a species qua species beings. A more detailed explanation and analysis of species life and how I consider it relevant to this debate about the potential creativity of computational systems will follow in later sections of this paper.

It would be useful at this point to disambiguate from a related but distinct critique of CC which has recently been addressed within CC itself. Hodson (2017, 144) advances the critique that whereas most determinations of creativity are made *ex post* rather than *ex ante*, CC typically operates with an *ex ante* notion of creativity and thus its claims assume from the outset the kind of creativity that is purported to be proven in the conclusion. Another way of putting Hodson’s point would be to say that we only know things to be creative when we have judged them to be so, and that the judgment itself is really the essential moment in creativity and not the act. Wiggins has pushed back against this claim in turn (Wiggins 2021, 186), arguing that most determinations of creativity made by CC are in fact *ex post* determinations. Some other research papers similarly address the question of how *ex post* evaluation might be formalised for a computational agent, such that it can determine for itself whether or not something is creative, rather than rely entirely upon the presiding human agent. Chieppe *et al.* (2022, 133) describes an experiment in which the program is capable of self-evaluating levels or degrees of surprise based on a bayesian inferential model. Carnovalini *et al.* (2021, 213-216) similarly use a statistical model to develop a framework for internal evaluation, developing a procedure for meta-evaluation within the machine that arguably does achieve *ex post* determinations of creativity that are native to the program’s own functioning. Similar discussions and results can be read elsewhere in the CC literature (see for instance: Linkola *et al.* 2020; Linkola *et al.* 2017). These solutions seem to address the specific worry about *ex post* evaluation, but perhaps not quite to the extent that matters for determining in some normative sense whether an act is H-creative. The suite of problems here extends far beyond the more formal concern with the semantic ordering of how creativity is evaluated. For one thing, it is clear enough here that – sophisticated as some of the aforementioned systems may be – they do not appear to be capable of making decisions for which they can be

held accountable. They are not to be considered moral agents, a point which proves important to our understanding of agency in general (Hooker 2018, 4-5) and to what extent we take self-evaluation to have recourse to some more genuine idea of 'self'. On a more specific level, we might want to raise the question of whether the aforementioned systems are capable of distinguishing their own goal state objectives from some broader and more general notion of world objectives, or having a utility function that is treated as independent from a world model (Totschnig, 2020, p.7). Additionally we can tackle the issue from the point of view of the well-known frame-problem of AI (McCarthy, 1977). This points to a more general concern about the capacity of computational, non-rational agents' ability to determine specific framings of problems based on inferentially tractable decisions about which kinds of information to include and exclude. Peter Wolfendale articulates this best with the notion of an in-principle generality of reason, which is derived from the rational agents' ability, via language, to make explicit certain heuristic frames embedded in adapted cognitive heuristics, and selectively modify them (Wolfendale 2019, 62).

The case I want to make in this paper is that so-called H-creativity is not merely one incidental variety of creativity, something that can be taken or left from the concept wherever convenient. Whereas Veale et al. have stated that CC does not currently achieve something close to H-creativity, although it would be desirable (Veale & Pérez Y Pérez 2020, 556), my contention is that achievement of H-creativity constitutes a basic necessity condition before it is appropriate to predicate creativity of some particular agent. Another way to put this is to say that creativity as we typically understand and use this term must maintain an irrevocably historical and normative dimension. Rather than treating H- and P- creativity, as well as the exploratory and transformational types selectively, all four are intrinsically related within the concept itself. That concept– I want to argue– ultimately requires some form of grounding, a point which makes apparent why it is that creativity is something which we value beyond its mere performance aspect. I want to make the case that the notion of species-life articulated by Hegel, and subsequently by Marx, proves to be an excellent candidate for such grounding, and will help us to see what is missing in CC's conception. Although "creativity" is sometimes defined within CC using more technical terms, a point worth bearing in mind here is

that it too draws upon the salient features of this notion of creativity as something that has value for the greater good of humanity, and it would be hard to deny that CC’s appeals to the notion of creativity also intend to make some comparison to what we understand by creativity when we use the term to describe human beings. Using a Marxian and Hegelian inspired account of species being concepts, I will argue that we cannot pick and choose which aspects of the concept we would like to invoke, but must rather understand the objective dimension of a concept like creativity when we apply it to rational agents.

1. Creativity as Concept

CC generally operates with a metaphorical or folk understanding of the term creativity (Veale & Pérez Y Pérez 2020, 554; Veale *et al.* 2019, 3). The literature and experiments aspire to reverse-engineer our common understanding of creativity in a bid to reveal its underlying mechanisms. This approach turns on the claim that such behaviours should be included within the general concept of creativity. It is worth considering off the bat here that “creativity” itself seems to be an imprecise and rather vague concept. Whilst this vagueness does afford the inclusion of a wide range of phenomena under its remit, a lack of precision can also significantly constrain the explanatory value of the concept, or worse still, lead to conceptual unclarity. Although it is not the main claim of this paper, I think it is nonetheless worth raising this point to set the general tone of discussion here. Creativity as it is used by CC appeals to our folk intuitions about this concept, and in that sense we have to take it in these terms by trying to work with what ordinarily goes into such intuitions. On the other hand, it is worth noting that perhaps a better route for CC or another discipline which attempted to achieve similar aims might involve an explication of the concept of creativity into more salient sub-concepts which better articulated what the discipline seeks to show, without necessarily making the stronger appeal to a comparison with human creativity. By explication, I follow the Carnap-inspired definition of Dutilh-Novaes which entails the construction of a new, more precise, specific, and fruitful concept from the existing one (Dutilh Novaes 2020). What seems most problematic about the conception of creativity advocated by CC is that it seems to lack fruitfulness when it comes to the task of reconstructing the

kinds of creativity that matter to us. It often settles for something that looks like creativity, but constitutionally isn't, and this seems discordant both with the basic claim it makes about computational systems' capacity to be creative in their own right, what Veale et al. call the "strong CC" position (Veale & Pérez Y Pérez 2020, 554), as well as the extent to which CC research informs the conception of creativity as it applies to human beings. This becomes especially problematic when CC begins to oscillate between an analogical notion of the creativity of computational systems, and the actual claim that such systems are creative, in the sense that their behaviours must in some meaningful sense be included in the broader concept of creativity that matters to us. On the other hand, I don't want to present my case as if to suggest that there is nothing to be salvaged from the work being carried out by the field as a whole. Much as I do wish to argue that there are constitutive problems with the extant approach and practice of CC, it is important at the same time to consider that the purpose of critique is not to recommend the abolition of a discipline but rather its transformation. Of course, in the case that I am presenting, the issue is not merely that the notion of creativity is too vague, but also that this has ramifications and consequences that are far-reaching and political in nature. Additionally, my claims involve a critique of the way in which practices are informed by what I take to be conceptual errors.

Nevertheless, I think the political point and the question of explication are connected, precisely in the sense that a clearer conception of what we are doing and why, when we try to develop these systems, informs any future work which might be instrumental to a transformation of existing practices and the assumptions which guide them. A fully explicated concept of creativity, as it pertains both to human beings and computational systems, is beyond the scope of this paper, however I do want to suggest that in undertaking such a task, it might better illuminate the possible contributions such research can make to the various goals and objectives we may have. Thus, the possible salvage of the concept of creativity and CC as a discipline might involve what Brun (2020) has called a "conceptual re-engineering" of the concept of creativity. The basic idea of conceptual re-engineering entails that concepts with greater clarity and precision should be developed to replace the former, more colloquial ones. My contention here is that such a conceptual re-engineering of the concept of creativity, such

that we can understand the possible role that computational systems of various degrees of sophistication might play in it, is a necessary accompaniment to the project of developing such systems in ways that better accommodate the normative objectives to be articulated with respect to species-life concepts, and that this critical project to determinately negate the present assumptions and their attendant normative and political issues is coincident with the project of conceptual amelioration which would pave the way towards a transformation of existing practices, rather than their abolition.

Whilst there is consensus within CC on the general terms of Boden’s definition of creativity as something that induces novelty or surprise, at the same time there appear to be unchecked assumptions concerning whether this surprise is relative to the human sciences or some more metaphysical view of nature itself. There have been some attempts to address this problem within CC. Wiggins, recognising the vagueness of creativity and the problematic fact that any reference to the concept seems to invoke a value-judgment, tries to evade the difficulty by eliminating all references to creativity as an isolated concept, instead making use of sub-concepts including “creative systems” and “creative behaviour” related analogically to human agency (Wiggins 2019, 23-25). In this way, he seeks to avoid any commitment to the idea that creativity exists as some intrinsic property of the world. At the same time, he also tries to get around the issue that judgments of creativity must be made– or be capable of being made– by the agent who performs them. This move is entirely laudable from the perspective of explicating the concept of creativity, but it proves too extreme for CC’s premises, and aptly demonstrates a difficulty faced by the entire enterprise. An idiosyncratic notion of “creativity” finds itself too far adrift of the original concept, undermining CC’s fundamental objective of showing how computational creativity can genuinely be included in our original concept of creativity. As a result, the move made by Wiggins to avoid all usage of the term “creativity” results in a tautology: all systems that have been designed in thus-and-so ways turn out to behave in thus-and-so ways. In other words, it is unclear how we can bridge from “creative behaviours” to the broader concept of creativity itself without invoking that concept of creativity, and if we fail to do this we merely end up re-stating the point we initially wanted to prove in the premises. These creative behaviours can be

categorically grouped according to their similarity, but the category itself does not explain anything other than the behaviours of those systems for which it was designed, as long as it does not make reference to the broader concept of creativity. The only remaining option is to project this more idiosyncratic conception of “creativity” back onto our original concept by arguing some substantive congruence between the two. This would be a valid approach, but it falters at the point where the initial assumptions concerning what is important about creativity writ large are just those features that turn out to be exemplified by the model. Wiggins maintains the normative and epistemic link between the creative behaviours of machines and those of human beings by fixing human normative judgments as the evaluative frame of reference. He thus avoids the problem of ontologising creativity, and the problem that AI in its current phase cannot be ‘surprised’ by its own behaviour in some broader normative and historically relevant way. This seems entirely sound, but in making such a move, his analysis amounts to a show-and-tell about such computational systems and their performance, devoid of any further meaningful claim concerning human creativity or the ultimate potential for AI to aspire to a comparable level of autonomy. It seems to make sense to say that these are “creative behaviours” by analogy to the behaviours of human beings, but such results are likely to prove disappointing to those whose objectives involve things like understanding the creative capacity of human agents better, or ultimately developing an AI capable of acting in similarly creative ways. For projects such as these, it seems necessary to reverse-engineer creativity and not merely creative behaviours which resemble human creativity in certain respects. Ultimately then, Wiggins seems to fail even by his own lights in terms of shedding some important light on the underlying mechanisms of creativity. He manages to reproduce what appear to be creative behaviours, computational behaviours which resemble what we recognise as creativity in human agents, but without recourse to the actual concept of creativity, this comparison fails to be informative to our understanding of creativity in human beings, and arguably also does not succeed in producing agents that are creative in their own right. To foreshadow some of the discussion to follow: this proves less problematic once we are willing to drop the pretence about the creative autonomy or such systems and think about them as very sophisticated tools. Of course, this invites

further questioning about whether these systems are the kinds of tools we actually wish to have, but such a discussion is only possible once we have adequately understood in conceptual terms what we are dealing with. Another method of circumventing the normativity problem that has gained prevalence within CC of late has been to make the case that extant AI should hold the status of co-creator or collaborator, and that the role human agents play in such a collaboration is one of normative calibration and framing of legitimate goal states (Veale & Pérez Y Pérez 2020, 555). This claim will be dealt with more substantively later in the essay, but it is important to note here the implicit consequence of such a claim, namely, that the performance of the act and its evaluative moment can be treated in relative isolation. Whilst it cleaves more to the side of creativity as a concept relative to human norms and expectations, it still holds that what is essential about creativity is therefore its heuristic manipulation of conceptual problem spaces. Moreover, the capacity of such systems to evaluate their own work relative to a set of meta-criteria does not really solve the problem so much as re-state it at a higher level of abstraction. The AI does not decide for itself what would be a good course of action to take in a sense that it might be held accountable for as an autonomous agent (Wolfendale 2021). Nevertheless, on this view, the machine is considered by CC to be more than a tool because it performs a high proportion of the cognitive and intellectual labour.

Aside from these attempts which do understand creativity primarily in its normative dimension, I want to suggest that some other descriptions of creativity within CC appear to rely upon some especially dubious metaphysical ideas. Such cases generally involve tacit and implicit, rather than explicit, metaphysical assumptions about the underlying nature of creativity. By a metaphysical notion of creativity, I mean here the idea that there is some substantial essence of creativity that inheres in nature independently of our ability to access it. We can think about this in a way that is not dissimilar to philosophical debates surrounding the reality of colours or moral properties. The basic premise of a metaphysical conception of creativity would have it that creativity is a property of things, that certain acts or behaviours are intrinsically creative, just as a colour realist might say that colours are an intrinsic property of objects, or a meta-ethical realist might similarly argue that moral properties inhere intrinsically in the world. In all such

cases, we understand the reality of properties in the sense that they subsist in a mind-independent fashion: the colour of an object does not depend upon my perception of it, but rather my perception of the object “discovers” or “reveals” its colour properties. Because it includes this notion of discovery, a metaphysical conception of creativity doesn’t in any sense disavow the normative aspect of the concept of creativity: we can still consider acts of judgment as necessary to producing the knowledge that some act or person is creative. On the other hand, the metaphysical position does hold that an agent or behaviour is creative in its own right and independently of such a judgment. Where this becomes most relevant is once we begin to consider the ostensible creativity of non-rational animals, and to what extent we understand their behaviours in isolation from our observation of them as “creative” in their own right. In contradistinction to this view, I have so far been arguing for a normative notion, which stipulates that creativity is a judgment which comes only from us, that it does not exist in some independent way in the world. The most prevalent example of where an implicitly metaphysical view comes into play can be seen in papers which draw comparison between natural or biological creativity observed in nonhuman self-organising systems, and that of AI (See for example: McCormack 2019; Roudavski & McCormack, 2016). Boden, though not strictly working within CC, also argues the case for the inclusion of biological self-organising systems as creative ones (Boden 2018, 206). Biological creativity is not the only such example however. More generally we can say that it appears wherever the distinction between thinking and being is elided. The standard argument is that creative behaviours are also observed in nonhuman nature: to treat creativity as an exclusively human trait is excessively anthropocentric (McCormack 2019, 327-328; Boden 2018, 206). There are of course plenty of things which might appeal to us about this notion. For one thing, it’s of course true that we learn many interesting and significant things about the underlying mechanisms of human behaviours when we observe how animals deploy unprecedented and novel heuristics as adaptive strategies (McCormack 2019, 328). The fact that our observation of nonhuman behaviours proves to be important for the ways in which we understand ourselves and our own concepts is not to be denied. Where I tend to differ, however, from Boden and others who contend that nonhuman animals and extant artificial systems alike

behave creatively, is in rejecting the claim that this confers the status of being creative in their own right. My rejection of this idea is grounded in the premise that in order to be capable of making such a claim, one must also maintain an implicit metaphysical premise that creative properties subsist in the world. It is necessary to hold such a claim just for the reason that one thing which is present in rational or sapient agents but absent from nonrational or sentient ones is the possibility of self-knowledge that one is behaving creatively. This is not a matter of knowing concretely in every case, but can be better understood as having the in-principle capacity to know that one is behaving in a way that is creative. If, on the other hand, an agent can be creative without even being capable of knowing that it is, this implies that creativity itself is something that exists in a mind independent manner. To point this out is, once again, by no means to denigrate the various practices of gaining knowledge and developing an understanding of ourselves and the world through scientific observation of nonhuman self-organising systems. The deployment of flexible heuristics seems to be a common denominator uniting human, nonhuman, and artificial agency in respect of creativity, and we can consider this in terms that do not necessarily need to extend the status of creativity to non-rational agents. Some recent examples of nonhuman creativity include urban-dwelling birds in Mexico using cigarette butts in the construction of their nests to deter pests (Suárez-Rodríguez *et al.* 2013), and octopuses building protective fortresses out of discarded coconut shells (Finn *et al.* 2009). This openness and preparedness to transform a problem or resource space, to see things differently or shift perspective, to broaden context or lower thresholds of salience, is an extremely important characteristic of human creative behaviour, just as much as it is in biological self-organising systems. For Boden and others, these behaviours might represent a comparatively less sophisticated form of creativity, but they share certain essential features which are sufficient for their inclusion in the concept (Boden 2018).

My concern with such a view, however, is that it ontologises creativity, making it a feature of the world, rather than of our conceptualisation of the world. Creativity is let loose as an unmediated given, something which one merely needs to observe in order to grasp and understand conceptually. This is not to say that no philosophical defence can be given of a metaphysical realism about creativity. By pointing

to the fact that these metaphysical assumptions are often implicit in the accounts I've mentioned, I also want to draw attention to the fact that at the very least such claims about the inherence of creativity to nonhuman natural phenomena warrants a defence on this issue. As previously mentioned, comparable metaphysical realist defences of things like colour realism and meta-ethical realism have been rigorously argued, and indeed my contention here that a metaphysically real account of creativity seems implausible need not be the last word on the subject. On the other hand, all such metaphysical realist defences must be assessed on their own unique merits and aptness to be considered as possible candidates for real properties rather than wholly mind-dependent. Whilst there is an undeniable degree of overlap between the arguments offered for things like colour-realism, mathematical realism, meta-ethical realism and so forth, none can be quite argued in the same way, and the same might be said of a metaphysical realism about creativity. By contrast with a position such as mathematical realism, for instance, which holds that mathematical truths are mind-independent and mathematical thought is about the discovery of mathematical truths rather than their creation, a realist position about creativity would have to contend with the fact that there doesn't seem to be any apparent consistency about what we consider to be creative behaviours outside of the judgement made in each particular instance. If all creative phenomena had some recognisable similarity qua creative phenomena that did not involve some relationship to semantic judgment, it might actually invalidate the claim that such phenomena were indeed to be considered creative, given how central the idea of novelty is to creativity. This much is backed up by the fact that it seems incredibly difficult to pin down any kind of definition of creativity that does not make recourse to normative and epistemic claims such as "surprising", "new", and "valuable". Aside from anything else, it just seems very difficult to pin down any coherent meaning for creativity without an appeal to semantic vocabulary, and this is somewhat in contradistinction to mathematical realism, which, though we can ascribe semantic phenomena to mathematical vocabularies, seems capable of coherence without some recourse to vocabularies of meaning.

At the same time, the sufficiency of creativity as a real property in such a realist conception runs into the potential problem of denu-

ding creativity of its normative dimension and isolating it as a set of practices and intrinsic processes. If the normative aspect of creativity is contingent and isolable, then it follows that creativity is something inscribed at the level of nature itself, since creativity can subsist without any further conceptual mediation. Although it is not impossible to account for semantic phenomena in such a picture, as, for instance, John McDowell does in his account of second-nature semantics within a fully naturalised account of being (McDowell 1996, 46-65), such accounts typically render the deliberative and purposive element of rational agency epiphenomenal and for the most part causally insignificant. One might say that this is a much more accommodating notion of creativity for CC, since the causal role played by the normative or historical dimension in human creativity seems minimal. A creative agent need not understand itself to be acting creatively, either in actuality or in principle, it is helpful when it does, but it isn’t fundamental to what it means for an agent to be creative. Not only does this view invite certain difficult to defend metaphysical assumptions, namely the coincidence of a semantic concept and its postulated correlate out there in the mind-independent world, but it also raises the issue that creative behaviours are themselves in effect heteronomous rather than autonomous. If creativity is immanent to nature, such that nonhuman animals are intrinsically creative rather than judged to be so, then it might also be said that human creative action is to be understood as a natural process about which our language and thought has very little traction. This picture certainly would appeal to the bald-naturalist or anybody who otherwise held an eliminativist view of semantic phenomena, but the costs of accepting it seem high enough to invalidate any meaningful discussion of creativity at all, insofar as we understand creativity to be something normative and intentional, rather than reactive and deterministic at the level of human practice. Of course it can still be argued that indeterminacy exists at the level of nature itself, but again, such a claim seems to warrant further justification since it is an essentially metaphysical claim about the nature of being.

Just as there are valuable and important reasons to study biological phenomena, the same may be said concerning the study of artificial systems and their ability to deploy creative heuristics. What I do not want to claim is that the ways in which we are surprised by the behaviours of either is totally uninformative, but the philosophical

dispute surrounding how we accurately refer to these things informs the question of how we are informed by them. This comes back to the aforementioned question of explication: is creativity well-explicated enough in this instance to show that what we mean when we compare human creativity to that of machines or animals amounts to the same kind of thing? My claim is that a foundational error of CC has been to treat the concept of creativity as an isolated notion, thus it has either been articulated in ways that don't quite capture what is most important about human creativity, or it ends up positing features of the world that are hard to prove exist. Ironically, in a bid to make the term creativity more inclusive and less anthropocentric, the ascription of a human concept to nonhuman agency obscures our capacity to cognise its difference,¹ instead assimilating nonhuman phenomena to human categories and considering what is important about these phenomena to be those things which are of interest to us. At the same time, there is a second irony insofar as this ascription of observances to existing concepts by analogy forecloses some degree of conceptual creativity in the empirical sciences,² leading to a kind of conceptual stasis which treats all new phenomena as instances of a familiar human idea, rather than proliferating possibilities that might allow us to understand behaviours without anthropomorphic or metaphysical baggage. In this section, I have attempted to illuminate some of the philosophical issues surrounding the various understandings of creativity within CC. I have argued that where CC tries to avoid any metaphysical baggage, it appears to fall short of the aims and objectives it sets for itself in seeking to describe or illuminate our own understanding of creativity in human agents, and offers few realistic prospects to deliver on its promise of a truly autonomous creative system. On the other hand, I have claimed that certain dubious metaphysical premises are often baked into the ways in which CC understands creativity itself, a fact which may lead to premature expectations about the possibility of replicating such behaviour in artificial systems. In the following section, I will attempt to address some of these concerns through the Marxian and Hegelian

1. See for instance Deleuze 2011, 164-213 on the issue of conceptual recognition and difference.

2. See Carus 2012 for a defence of the ideal of explication in connection to this point.

inspired notion of species life. In respect of the problems that have just been outlined, I will attempt to read creativity as a species-being concept, showing how its normativity is central to the very idea of creativity and not merely an incidental feature which can be abstracted. In so doing, I will try to show not only what CC gets wrong about the notion of creativity, but also try to set a benchmark for what it would mean for an artificial system to be creative in its own right, as per the stated goals and objectives of CC.

2. Species-Being and Creativity

I have made the claim in this paper that creativity as a concept cannot be considered independently of its normative dimension. This sets a high bar for what we should properly consider an autonomously creative agent in a sense comparable to that of human agents. The normative component of human creativity is an essential feature, not contingent. Our preconceptions of what a creative agent or act is should necessarily include this dimension, and acts or agents which don’t, fall short of the mark. This also entails that a truly creative agent must be capable of orienting itself towards certain rationally-deliberated normative goals, and is another way of stating that the agent must be autonomous. This is a strong contradiction of most of CC’s main ideas about creativity. Most within the field hold that the behaviours of certain computational systems are creative, even if they lack some normative or historical sensibility, and even when they fall short of human autonomy. At the same time, CC works towards greater *degrees* of autonomy and normativity, and there is some indication that the field understands its own research programme as one that works towards the eventual goal of “strong-CC”, or artificial agents who are creative in their own right in ways that are comparable to human agents (Veale *et al.* 2019, 15-16).

My approach to this issue is to try and understand normativity in its social and historical dimension, as it relates to a species-community of agents. This approach is similar in outlook to Jan Løhmann Stephensen’s critique of CC (Stephensen 2020, 2023) which argues for creativity’s necessarily historical and social– not to mention critical–dimension. My arguments here will focus more on the Marxian and Hegelian notion of species-life, which proves to be an especially useful

way of framing value concepts and their relationship to the productive agency of individuals. This conception of human beings as species-beings, or the capacity of human agents to take life as an object for conscious reflection (Marx 2000a, 90) crucially provides some form of grounding for normativity in a way that avoids positivising semantic and normative content by treating them as merely performative. At the same time, it further illustrates what is important about creativity in the social context beyond its instrumental affordances, and maps out a theoretical trajectory for the kind of AI which would satisfy such requirements and count as genuinely creative in its own right. In fact, my evocation of the notion of species-being is not in order to suggest that it provides us with the ultimate answer to questions about normativity and rational autonomy; the prospect of AGI and machine consciousness even poses a significant challenge to the species-being conception. However, the formulation of these problems in terms of species-being questions lays down the gauntlet concerning what must be surpassed if we want to think about genuinely autonomous creative beings.

The idea of species-being for both Hegel and Marx is founded on a conception of human nature that has implications for what human beings need, what is good or bad for them, what fulfils or actualises them, and similar concerns (Wood 2004, 16-30). As Karen Ng has argued, although this claim sometimes might appear to be essentialist and ahistorical, there is nothing mutually exclusive between the historical character of the human life form and the fact that it can be apprehended as a life form (Ng 2021a, 2021b). Indeed, for Hegel, self-consciousness is intimately bound up with species-consciousness (Ng 2020, 65, 73-80), and this view can be interpreted as offering an account which maintains the closure of a naturalist picture of the world, but at the same time also offers a non-reductive picture of the *sui generis* capabilities of the human species within it. This concept is poorly understood if it is only understood as a deterministic claim about the telos of human species-being as something given to us by external nature. In his speculative identity thesis, Hegel offers a much more complicated, but also compelling conception of how the power of judgment is itself an activity which is both enabled and constrained by the unity and activity of life (Ng 2020, 107-110). It is the very dialectical process of the identity and opposition between life and self-consciousness that

constitutes the activity of knowing (*ibid.*). This dialectical process is especially valuable because it provides the foundation for an idea of species-being in a sense that is not static or metaphysical, but rather dynamic, and the locus of human autonomy. Thus, species or genus concepts,³ are not abstract ideals imposed from without but are accounted for as the dialectical process of thinking activity. To put this simply, the idea of species-being is not a fixed and stable identity that subsists in some time-general way. What count as needs, fulfilment criteria, self-actualisation conditions, and so forth are subject to change via this dialectical process between self-conscious activity and the constraints and possibilities of species-life. Most importantly, for both Hegel and for Marx— who adopts this account in large measure in his early social philosophy— the idea of the genus or species is objective and universal, but it is so in a concrete rather than an abstract sense. By concreteness here, Hegel seems to refer to the idea of the concept’s self-constitution as internally self-actualising or internally purposive (Ng 2020, 55). It is not that the universality of species being is granted by some externally mediated essentiality, but rather that its inner purposiveness gives rise to a telos through which individuals and predicates find their meanings. The species or genus concept provides the context or grounding for any further predication including most of all the ascription of normative predicates to the subject. Thus, the concrete objectivity of species life provides the necessary condition for the objective existence of a being that belongs to it (Ng 2020, 9-10, 165-218; Ng 2015, 116-118). As it applies to the question of creativity put forward in this paper, we can understand creativity as something predicated of subjects which similarly finds its ultimate grounding in the concrete objectivity of species or genus life.

Marx further develops this claim in his account of alienation. Society is not an abstract manifestation that exists in opposition to the

3. The term *Gattungswesen* is typically translated as “species-being”, although as Khurana (2023) has recently argued this translation leaves something to be desired and doesn’t fully capture the intuition of Marx’s concept, instead recommending the adoption of genus-being. Although Khurana’s argument has greatly influenced my own understanding of this concept in Marx and Hegel, I have mostly continued to use the term “species-being” in this essay for consistency with the other literature and to avoid confusion. Where possible, I have made reference to both species and genus in order to highlight the ongoing debates around nomenclature on this issue.

individual, but rather the human individual as a member of the species is a constitutively social being (Khurana 2023, 266-67; Ng 2021a, 153-154; Marx 2000b, 183). To put this simply, there is no way to coherently think about the life of a human individual without invoking this connection to species life, the very concepts and grounding that pertain to individuals are both constrained by, and the realisation of, human social activity. This is true even where human individuals act outside of an immediately social context, or without full conscious awareness of the value of a particular activity relative to this social context. This detail is important because one of Boden's underlying claims in defence of creativity as a more ontological concept, rather than one which is relative to a reflexive conception of species-life, is that this normative purposiveness of creative acts is not even a salient feature of human creativity:

Granted, if my own definition of creativity is to be applied to biological cases, it must be understood that "valuable" here means valued by human beings, not considered to be valuable by the organism/process concerned. But that is not unreasonable. For even when ascribing creativity to people, we sometimes ignore the fact that the person had no inkling of the value of the novel idea (Boden, 2018, 206).

Pace Boden's claim here however, the creative human individual need not always be conscious of exactly how their actions are valuable, but the very premise of such actions themselves, what gives them some objective dimension, is their contextualisation in the broader project of realising human needs, flourishing, actualisation and so on. The significance of this point cannot be understated. For both Boden and CC's broader consensus, there appears to be genuine acknowledgment that the inability of non-rational agents to recognise their own activity as "creative" behaviour poses some kind of problem. However, as we have seen in the quote above and in CC's own attempts to work around this issue, this difficulty is restricted somewhat to an issue of correctly naming and identifying the behaviours concerned. Here, the epistemological and normative issue of how a being understands its own behaviour is treated as a kind of secondary concern; in other words, the ability to know that one can act creatively is not a fundamental precondition for creative action, and as such creative behaviour can take place whether the agent understands its own behaviour as

participating in the concept of creativity or not. If we follow Marx and Hegel here however, the role of judgment as a spur for creative thought and action must be emphasised as something crucial to the substantive content of our concept of creativity. We can track this emphasis through the importance that language and inferential normative content has for human species life. Of course, in order to be consistent on this point, we must also point out that the actions of nonhuman species are also similarly purposive, that such species also act in ways that are ultimately grounded in their particular needs. This gets to the crux of the self-organising character of organic life, which Boden is arguing here ought to be considered within the ambit of our concept of creativity. I am inclined to agree that the internal purposiveness of species being provides us with a great theoretical vocabulary to understand the adaptive heuristics of self-organising systems, both organic and non-organic. However, there is a crucial difference here, highlighted especially in Marx, between the self-preserving and reproducing behaviour of non-human organic life and that of human life which must be emphasised. It is only in the case of the latter that life itself can be taken as an object for conscious reflection, whereas for the former, purposive activity is restricted to the mere reproduction of life (Marx 2000a, 85-95). This point is crucial: even though we might elucidate numerous examples of the deployment of novel heuristics by non-human species as adaptive strategies which enable more efficient reproduction or self-preservation, all such novel heuristics are only developed within this context of the mere reproduction of life. For such agents, since life itself cannot be taken as an object, there is no question of whether to pursue a certain course of action or not, but rather a series of adaptive strategies that proceed on the basis of trial and error, ultimately becoming entrenched. I believe this lends credence to the claim that what we call creative behaviours in nonhuman agents and systems are not creative in the same sense that we mean it when we apply the term to human beings, because they are merely actions whose purpose is to reproduce or self-preserve the species, and an important part of what creativity means to us is not only to engage in behaviours that allow us as a species to better adapt to our environment and reproduce ourselves, but also crucially to better articulate and realise objectives that come about as a result of inferential reasoning about what life itself means, what we want it to mean or

what kinds of life and world we would like to construct. This is why we value creativity and innovation in fields such as art, which doesn't have some obvious adaptive or self-preserving justification and yet seems in one way or another to play a very significant role in what the vast majority of human beings, across different cultures, understand to be a meaningful life.

An additional point of relevance here is the connection for Marx between the concept of species being and his theory of alienation. If I am correct here in arguing that what creativity means for human beings involves not just a certain kind of performance, but additionally the ability to frame and reformat the kinds of activities we do and our ways of thinking about and seeing the world, then a set of attendant political concerns begin to arise around the issue of to what extent contemporary society and its material and conceptual frameworks are conducive to the actual flourishing of creativity for human beings. Indeed, in the concluding sections of the paper I wish to dedicate some more time to this issue. What I wish to highlight here however is just the sense in which my concerns aim to be more than semantic, or to give a sense of what is really at stake in this distinction between the novel heuristics deployed by nonhuman agents in service of mere self-preservation and reproduction, and the more expanded sense of creativity that I am arguing must necessarily include its normative dimension in order to be coherent. The important point about alienation for Marx is not its subjective feeling— though this may also be important for individuals themselves— but rather a structural reality manifest at the level of the species (Marx 2000a, 87-95). For Marx, alienation does not follow a prelapsarian logic, and his view is sometimes mistaken as meaning that the de-estrangement of human social activity would be tantamount to the restoration of some originary and primitive state of human flourishing before the fall.⁴ As Ray Brassier usefully notes here, Marx's perspective is better thought of as one in which estrangement itself provides the necessary precondition of de-estrangement (Brassier 2019, 103-104). There is no underlying necessity that estrangement must either follow some original state or have precedent in the history of the human species, precisely because Marx adopts the Hegelian

4. Famously, this critique of Marx was put forward by Louis Althusser 2003.

speculative identity thesis here in his construction of the idea of free conscious activity, an unbounded form of social activity and production that corresponds to the realisation of species needs and flourishing, as opposed to activity that is estranged from the individual and confronts them as something alien with free-standing existence (Marx 2000a, 86). Thus, Marx’s notion of free conscious activity is better thought of in terms of this same dialectic of self-consciousness’ identity and opposition with species-life, which is not a matter of satisfying fixed and invariant needs but rather the deliberative process through which they are both articulated and realised:

As individuals express their life, so they are, what they are, therefore, coincides with their production, both with what they produce and how they produce. The nature of individuals thus depends on the material conditions determining their production (Marx 2000b, 177).

Crucially, when we consider the notion of “life” here as it is understood by Marx and Hegel, we must resist the urge to reach for something with the character of a substantialised essence. “Life” for Hegel and for Marx is not something that can be understood as given or fundamental in the sense that it has some invariant essence, but rather must be treated as something more fluid and defined in terms of the dialectical process of self-consciousness’ own realisation – in an ideal sense for Hegel, and in a more practical sense for Marx. This does not mean that the category of “life”, which is taken by both to provide the objective grounding for value predicates, is an entirely indeterminate category. For both, the basic conditions for the self-reproduction of organic life provides a starting point, but importantly not a final end, of what we consider to be life. In this sense, we can understand “life” for Hegel and Marx as something which consists of the process of *dialectical opposition* between the conditions of bare survival, and the way in which these are perturbed by the elaboration of self-consciousness through thinking activity. To clarify this point further, we can return to Wolfendale’s notion of the reformatting of homo-sapiens, which takes place through the conceptual *re-framing* of adaptive heuristics by making them explicit and selectively modifying them (Wolfendale 2019). In this picture, we might understand the adaptive heuristics as a part of our natural biological constraints, and the kinds of activity that lead to a re-framing as the dialectical negation of the constraints of species-

life that is performed by consciousness. This is a relatively simplistic account which skips over some layers of mediation and explanatory detail, but will suffice for our purposes of clarifying what life can mean in this picture without essentialising the concept. Within this picture, we can also see how this concept of life provides the objective grounding for value-predication: again, what values matter to us, in this instance perhaps we might say what counts as a creative act, shifts over time according to the development of different needs and fulfilment criteria of the species. This does not make such criteria indeterminate, rather they are always relative to the changing character of human species life. Of course, we can say that some aspects of this species life are invariant, or at least our historical trajectory would lead us to think they will remain as such: constraints like mortality and ageing have at any rate been a constant for all hitherto existing human societies, for instance. The value of Hegel and Marx's conception of life is not only that it leaves open the possibility that these seeming invariances about human life might change, either in their entirety or in the degree of significance, but perhaps more importantly, is to be found in the fact that even where it acknowledges the seeming invariance of these constraints, it does not reduce the notion of life to these features as its essential properties. A part of this picture of course involves the role of consciousness, and consciousness' capacity to consider life as an object for reflection, as an artefact of itself. Re-situating these points within the context of our original question, we can make a number of important observations. I have already intimated the first of these in my critique of Boden's claims about biological creativity, but the point can be formulated more generally here to say that wherever creativity is predicated of an individual agent, it is perhaps better understood in terms that are relative to the particular species-being, rather than as a substantially real essence that can be instantiated to different degrees at various levels of nature. Not only does this avoid the 'ontologisation of creativity' problem, but it also provides us with an account of *how* creativity in human agents is linked to a kind of purposive activity, and not only *that* it is. In other words, it gives us an account of why creativity is important to human beings, neither in an exclusively *ex ante* nor *ex post* sense, but in a way that sees both the productive/active moment, and the reflective/evaluative one as intertwined and mutually presupposing.

3. Beyond Species-Being

An interesting question arises at this point: do we need to conceive of species-being in strictly organic terms? In one sense, it might seem from the preceding analysis as though this idea irrevocably invokes the organic sense of the term ‘life’ insofar as human life forms are organic, and the concerns of human species welfare seem to be inextricably bound up with self-concern in a distinctively biological way. Although this provides an account for how human intelligence and self-consciousness emerges, it may not necessarily follow that species life need be organic life. We can also think about species life as articulating what kinds of conditions might be necessary for the grounding of evaluative judgments in the case of synthetic agents as members of a species or genus. Whilst this would diverge from the Hegelian conception in terms of the dialectical exchange between self-consciousness and species-life as something conditioned by biological finitude in human beings, one might nevertheless speculate here that a self-conscious individual agent that understood itself to be a constitutive member of a particular species, might at the same time grapple with this dialectic in different terms, perhaps along lines of synthetic rather than organic finitude. It would in any case be true that a hypothetical artificial agent capable of apprehending its ‘life’ as an object for its own conscious reflection would also need to be capable of reflecting upon the kinds of things that are conducive to the flourishing and reproduction of that life, in respect of its life-form.

Coming back to the initial questions raised by CC, creativity of such machines in their own right would mean something very different to the kind of acephalic groping through successive permutations, only to have these curated by human agents depending on which results we happen to find valuable. We may instead speculate about the kinds of machines which could make evaluations about the creativity of their own action relative to the flourishing not only of its own species-being, but potentially also that of human, and other species life. In fact, just as human beings also apprehend the species life of nonhuman animals as objects for our own conscious reflection, we may imagine that genuinely autonomous AGI might similarly apprehend the life of human beings as an object for such reflection, and the inverse in the case of humanity’s relationship to these beings would

also be true. Of course, there are a few things which must be unpacked here before jumping too hastily to any conclusions. For one thing, we must consider the fact that AGI is typically not understood in a species context, and the category of “artificiality”, which is here simply the negation of an organic substrate, is not tantamount to a species category. Just as cattle and human beings are not members of the same species merely by virtue of both being composed of organic matter, nor should we erroneously apply the category of species to the various distinct possible synthetic life forms. At the same time, perhaps it is fruitful to think in terms of species categories when we think about the development of AGI for similar reasons.

The possibility of artificial intelligence also allows us to imagine the possibility of a single, unified agent distributed over multiple disparate individuals. Such a case would also seem less than apt to be considered a species in the sense that we understand it here. We might say here that one of the very crucial features of species in the sense that we have been speaking about includes the very constraint that species-consciousness is not self-evidently unified or accomplished as a totality by default. The dialectical character of life as we have thus far understood it also may lead us to speculate that there is something generative about this fact, since life is a dialectical process which requires various layers of contestation and mediation, we might also say that the content of free conscious activity that accounts for the kinds of value-driven forms of social production and reproduction is a product of the very fact that individual members of a species simultaneously recognise each other as distinct individuals whilst at the same time being of the same fundamental kind and sharing the same fundamental conditions and interests. To think through all possible ramifications of this point in respect of synthetic life forms and artificial intelligence is beyond the scope of this paper, however it does seem to be an interesting consideration for the potential development of AGI. At the same time, even if it may seem redundant to consider an artificial agent which is the only true individual of its own kind as the sole member of a “species”, it does on the other hand make sense to use species vocabularies to refer to the potentially wide variety of different such agents. Much of this is of course highly speculative; nevertheless, it is important to consider as the proper staging for a scenario in which we might begin to talk about the creativity and autonomy of machines

in their own right. Whilst CC has more recently diluted this claim to criteria which are satisfied in the ‘co-creation’ context, I want to argue here that this kind of formulation is just as misleading, in that it invokes an underlying intuition about the mutual determination of goals which is not truly the case. This has implications for the way in which we think about the kind of productive activity that may ultimately be possible in human-machine collaboration where both parties constitute autonomous creative agents capable of making general value claims.

On the other hand, with respect to a potential AGI we might also wish to ask whether it would truly apprehend human life as an object for its own conscious reflection in such a way that might be conducive to the flourishing of human species life. This is far from certain, and ultimately a point on which it is unwise to make strong conclusions either way. Perhaps it is worth noting here however that the apocalyptic visions of some who anticipate the domination of humankind by an overpowering AGI involve a relatively naive understanding of rationality which might actually be ameliorated by the account of species being and purposiveness that I am outlining here. A typical assumption (Totschnig 2020, 7) used to justify the likely malevolence of an AGI vis a vis the human species is to assert that such an intelligence would be unlikely to make a distinction between its utility function and some broader normative world conception. We touched on this point briefly in the introduction, and pointed to some arguments which have refuted such a conclusion. Perhaps to add to those points, however, it is worth noting that in the case of human species consciousness as I have been recounting it here, it would be a misunderstanding to assume that the goals and objectives of a species involve the complete and total domination of other species for its own gain. In fact, it is precisely a virtue of this account of species-being that I have tried to offer that it avoids the necessity of such a conclusion. Thomas Khurana’s observation (Khurana 2023, 248-250) that the appropriation of nature by humans qua species beings– our treatment of external nature as our own inorganic nature apt to be metabolised for our own flourishing– maintains an interest in the independence of this external nature, proves illuminating in this context. The most applicable proof of concept here might be to consider the way in which we morally articulate our own species relationship to the Earth’s environment. It is important, for instance, that we can articulate concepts such as

ecological stewardship, managed growth and emissions reduction, and de-industrialisation as forms of care for our natural environment which are borne of a recognition for what it means for us and other species with whom we share the planet to flourish. Of course, it will be objected here that the historical record concerning the human species enactment of such values in practice is less than stellar. So too might it be objected that many of these concepts are relatively late additions to the collective consciousness of human societies, and even at that have not yet achieved universal acceptance. However, these objections do not invalidate the points I have made concerning species being concepts, but rather further reinforce them. Whilst it is true that the conclusions of climate science and what they recommend have not yet been universally accepted in all respects, this much arguably points to the often arduous and lengthy process through which such values and norms come to be fully instituted within species consciousness. Unfortunately, it is never a matter of instant unanimity, but rather the development of value claims with respect to life must undergo the kind of dialectical machinations which can hopefully lead to eventual resolution. Similarly, we might say that although the record of the human species when it comes to living up to its purported normative values is not great, this merely reinforces Marx's point that social life under the conditions of alienation produced by capitalism prevents the kind of production which is immediately conducive to the flourishing and reproduction of life from taking place without serious inhibition. I would argue strongly here that we are better off treating failure to act appropriately according to the imperatives of species flourishing as a structural issue which frustrates the very possibility of acting in accordance with normative precepts, rather than a constitutive or natural failure on the part of humans to live up to such precepts in all possible worlds. Getting back to the question of AGI: of course there is still a good measure of contingency here, and if we are serious about the autonomy of any computational agent in a sense that is comparable to our own species autonomy, we cannot give any kind of strong guarantee about how it will behave and act. What we can do, on the other hand, if we are intent on developing such forms of synthetic life, is to consider under what conditions and in which ways they might be produced. A major point that I have been trying to emphasise in this paper is that if we should wish to build such entities in our own image,

we might need to consider how that image of ourselves is reflected in our current practices and concepts, all of which comes back in a circuitous way to the question of how and when we understand the artefacts of our own provenance to be “creative”, and what this means about how we understand the creativity of human agents.

At this point it becomes apt to consider how one of CC’s foundational statements entails that the systems it seeks to develop ought to be understood as more than mere tools. But why should we not understand such systems as tools, albeit highly sophisticated ones? There might be a number of intuitions at play in this reluctance to consider the machine a tool: the proportion of work and especially cognitive work being performed by the machine, and the counterfactual case of what would not be realised had it not been for the machine’s participation in the productive process, may be particularly compelling justifications. I want to suggest, however, that these only make sense under conditions where labour itself is alienated from its social conditions, and wherein labour confronts the individual agent as something external and alien, a self-subsisting object that expresses a performative demand upon the individual to be realised. Another way we might put this, in light of the previous analysis of species-being, is that this conception of machine as more than a tool presupposes that the questions of what labour ought to produce, and for what reason or benefit, are already settled. Under such an assumption, it is easy to see automata as co-creators for much the same reason that it is easy to see human beings as abstract quantities of labour power: in either case, the role of the producing agency is disconnected from the normative question of what is produced, so we might say that co-creative machines are more than mere tools in this kind of scenario just because we want to believe that human beings are more than the mere instruments of commodity production. The autonomy of such machines maps onto this apologia for the mortification of human agents under the regime of capitalist production, rather than adhering to a higher standard of what either an autonomous or co-creative agent ought to mean.

Thinking of computational systems as tools is no denigration of the role they play, actually or potentially, in the development of human consciousness. Every tool can be thought of simultaneously as an artefact, something that reveals to us important features of our relationship to such tools: how and why they are produced, and what

objectives or needs they intend to satisfy. Something CC gets right in this respect is that such artefacts also tell us about certain features of human beings that we have not yet managed to successfully replicate in artificial agents. However, it does this in ways that only relate to the domain of performance, a framing which asserts the exclusive relevance of questions concerning what human beings are or can do at the functional level, at the same time obfuscating the normative questions around how we should understand human life, sociality, and production. One may question here whether this is a fair expectation of CC, since there are nevertheless salient questions to be asked concerning ability and capability. My concern is that by framing advanced computational systems as agents rather than as tools and sophisticated artefacts, CC also obviates these questions about how such systems ought to be correctly understood qua artefacts. The point I am making here is not about saying that an autonomous artificial agent would not also constitute an artefact, but rather that reflecting on the status of an object as an artefact means saying something about our understanding of such an entity, both constitutively and relationally. In both cases, the kinds of systems being developed by the research aspect of CC are largely justified with respect to their instrumental value for industrial or practical concerns. Where CC tries to reflect on what this means for the capabilities of machines and their potential to imitate human-like creativity, perhaps a more salient question we should raise here is what such approaches tell us about prevailing ideas concerning human creativity. It seems that within the current conjuncture, the very grounds upon which we might pose the question about what kinds of technologies would be important or useful for us to have, are absent. On the other hand, I think it is not unreasonable to demand or expect of scientific and technical disciplines that they consider the question in such terms. Speaking instead about the allegedly autonomously creative capacities of computational systems occludes the more important concern about what conditions of human species actualisation such systems could hypothetically satisfy. In fact, this very desideratum of human need sheds light upon the sense in which creative production exists not as an infinite elaboration of self-consciousness and its ideal self-realisation, but rather the realisation of self-consciousness in dialectical tension with the needs, desires, and constraints of species-being. What makes creative forms of production intelligible is not the

boundless elaboration of ratiocinative activity, but the way in which reflection upon the real challenges of satisfying the question of human fulfilment allows us to evaluate and orient both thought and practice. It is a mark of the *sui generis* character of human intelligence that we are capable of doing precisely this. As Marx and Engels have it: of taking human life as the dynamic object of conscious reflection (Marx 2000a p.90). There is nothing that tells us in principle that human beings are the only possible beings capable of realising such autonomy, only, we should not conclude from this that our extant computational systems are standalone agents. What is most important of all here is that we *should* actually be able to think about extant computational systems as tools. We should affirm some greater right to act as the bearers of such tools, to claim a stake in the question of how they are constructed and what needs or criteria of human life they ought to satisfy. The fact that CC’s understanding of creativity is instead presented to us as a given not only means that it is un-conducive to the idea of de-alienated production which has relevance to the needs and flourishing of human species-life, but additionally forecloses the possibilities for computation itself to be treated in these terms.

Perhaps my point here is also to reaffirm a possible decision to be made about whether to treat the artefact as a tool or as a co-creator, given that such systems already exist as a part of our world and will continue to do so indefinitely. Such a decision would be informative for our relation to such automata in ways which could prove significant concerning both their design and operation. The idea that machines are autonomously creative in some sense already takes away any grounds to challenge the pervasiveness of their influence. Whereas a tool is something one can always think of in terms of the satisfaction of needs or ends, it might be said here that one of the most important features of CC’s argument against treating such systems as tools is that they might not be subordinated to such kinds of demands. Whereas we would do well to think of genuinely autonomous machines outside of such relations of domination, doing so in the case of our extant machines produces an ideological effect which is tantamount to general acceptance of a pervasive and often pernicious influence automated systems exercise over our lives. This is something which obtains more generally in relation to AI, however one consequence of CC’s evangelistic attitude towards creative AI in the here and now can

be a redoubling of the intractability of computational systems as tools of domination. In any case, extant computational systems are used as tools, albeit with highly restricted access and terms of use. Thinking such systems as autonomous and intractable to human influence functions as a form of ideological repression, because it narrows the field of possible redress for those confronted and oppressed by such tools, as well as occluding the agency behind their deployment. *Vis-a-Vis* creative activities, treating such systems as more than tools entails a dilution of the valency of human subjectivity in production. The point here isn't to suggest that the productive contribution of the individual is sacrosanct, nor is it to valorise the laborious efforts of human individuals prior to technical automation. However, as Simondon famously argued in his thesis on the genealogy of technical objects, we should try to think about how the human agent and consciousness develops in relation to the concretisation and automation of technologies as a concomitant project to the development of such technical objects themselves (Simondon 2016, 247-260). This question of developing a better understanding of the technical features underlying automated processes exemplifies the kind of role we might wish individuals to have within the productive loop, not one which pilots the machine by simply pointing it in a certain direction or artfully curating or interpreting its results, but rather considering its form and function more closely in relation to value-questions. It is striking to note that for all of the discourse that CC has to offer concerning the co-creative potential of computational systems, any discussion about the human agents around which they are built, the needs, desires, and conditions of self-actualisation and fulfilment that such co-creative partners might have, is entirely absent. The blame here is without doubt more directly attributable to the division of scientific and intellectual labour under the capitalist mode of production, and the point can be applied more generally to many different fields of research and production, but that is no less of a reason to raise it here too. Treating such systems as tools, and not as co-creators, allows us to consider ourselves in a more voluntaristic relationship to their application. The difference here is semantic, but no less consequential for that. Whereas the tool is treated as instrumental to some objective, a co-creative relationship would suggest one in which the objective itself is partly determined by the co-creating agent. Computational systems presented at the interface

level rarely allow for a great deal of flexibility, especially not to the uninitiated user. Whereas CC tends to frame this situation as one in which a machine might collaborate with an artist or similar, I am more inclined to argue here that such cases amount to incompetent tool use: the artist is only a co-creator in the same sense that an unfledged rider who finds themselves on the back of a wild horse “guides” the animal. In a sense we can identify the problem as one in which the rigidity of the tool does rob the agent of some freedom to determine objectives, but we should also understand that the construction of the tool is a consequence of the alienation of production. For this reason, we need to consider the problem of diminishing autonomy in the so-called “co-creative” relationship as consisting of two stages. Firstly, the actual way in which the machine itself is produced, what it is for, how it produces, is primarily as an agent of capital, and there is little recourse to change things given how pervasively the logic of capital permeates all levels of social production. This once again echoes the important points that Marx draws in connecting his account of species being to that of alienation. The second stage takes place in the very immediate moment of human and tool interaction, which is also inhibiting since the tool itself creates forms of path dependency which narrow the field of possible action. This narrowing the field of possible action isn’t in itself a bad thing, and in fact this applies to any case of tool use. However, the fact that the conditions of narrowing and compression are not tractable to anything related to human flourishing but rather only to forms of exploitation and domination means that the role of the human user is just as constrained. The point here is that although tool use always narrows the field of possible action, we have the capacity as rational agents to determine what kinds and degrees of narrowing we find valuable and necessary, and this possibility is itself squandered when our conceptual and practical relationship to the world and our own production confronts us as something external and alien. It would be too hasty to conclude here that this situation amounts to domination by tools, or an irreversible loss of human practical and intellectual agency. On the other hand, it would be remiss not to consider the profound impact any technology will have upon a society and its members, how it shapes the possibilities for action and for seeing the world in terms of its possible transformation. The deleterious effects of a dependency on creative tools which conceptualise creativity as the

mere transformation or exploration of preset problem spaces might be seen above all in the diminishing capacity for framing what constitutes a problem space to begin with.

This has consequences for artistic and social production beyond the mere creation of bad art, which— as history has proven time and again— will continue to proliferate in any case. Auto-encoders, for example, are positivising systems par excellence: they assimilate all novel phenomena to system memory and experience, elaborating and updating themselves through recursive problem solving exercises with ruthless efficiency. I want to suggest here that this is profoundly at odds with what is valuable about artistic creation, namely, the radical experience of confronting the nonidentity of subject and object, the realisation of finitude, and the resistance to the given objects of phenomenal experience and thought which are forced onto the subject (Adorno 1983). This idea of what motivates creativity in artistic practice, inspired by Adorno’s concept of negative dialectics, suggests that an important aspect of creative production itself is the failure of concepts to fully capture the non-conceptuality of that which is not. It is this very idea of the nonidentity between subject and object, the sense of our own human finitude and the finitude of our concepts, which motivates us to engage in the dialectical process of revising and reconstituting the world itself.

Transformation and exploration as technical manoeuvres don’t satisfy such criteria, because negation should be thought of as a dialectical moment of reflection, and not just an updating function. It is generally well understood that extant computational systems don’t articulate the thought of “that which is not” as such a moment for self-conscious reflection, in fact they appear to struggle with negation more generally (Arnaout & Razniewski, 2023; Testoni *et al.* 2022), and if left alone, spin out into an interminable routine that exponentially diverges from an intelligible or meaningful picture of reality. Nonidentity is always instantaneously subsumed back into identity, such that nonidentity never exists for the unsupervised computational system as the friction that brings its ratiocinative spinning to a halt. In a view consonant with the earlier analysis of species-being as the normative grounding which exists in dialectical tension with ideal self-consciousness, the creative act exists both in the practical immediacy and in its negation by that which is nonidentical to itself. This is because the actual grounds from

which creative acts, as well as any other kind of normative act emerge are already indelibly social. Whether we are talking about unsupervised machines, or a co-creative activity, extant computational systems lack this capacity since being is taken as immanent to calculation for the machine. In this sense, the defining feature of creativity for CC is a positivising one that considers creativity as an unbounded productive activity devoid of its conscious aspect. Retrofitting the human agent back into the ensemble also won’t work, because even in the co-creative context the actual encoding of information is the role performed by the computational agent. As such, the moment of immediacy already belongs to the computational system, and the role of the human agent is at best to interpret the results and recalibrate the system accordingly. Rather than playing with the perceptual and conceptual data in an exploratory or transformative way, the co-creation context implies a higher level of mediation, such that the role of the human agent is to play with the encoded information produced by the machine. There is nothing principally wrong about this, so long as we understand correctly that what we are doing in such cases is apprehending the conceptually-encoded sameness of an artefact of human provenance, and not the nonidentity of that which is not.

To conclude this essay, I want to again return to the question of AGI-creativity. AGI is conceptually important both as some real possibility towards which contemporary computational research strives, and as a thought experiment or model through which we can both understand ourselves, and the current state of AI. As Reza Negarestani argues, the artificialisation of self-consciousness through self-relation is what allows self-consciousness to reflect upon itself as an artefact (Negarestani 2018, 25). AGI as a practical project which aims for the realisation of general intelligence in the form of such an artefact allows us to understand our own species and the concepts we use in a way that is crucially not about making a positive comparison of likeness or sameness, but rather by way of the determinate negation of ourselves qua species beings precisely in the dynamic sense proffered by Hegel and Marx which rejects the essentialism of a particular image of the human, instead opting to treat it as a dynamic process of historical development. Perhaps above all it is worth pointing out here that CC’s objective of demonstrating the creativity of computational systems depends fundamentally upon a conception which is the static mirror

image of an essentialist portrait of the human. By contrast, once we understand human species-life as a dynamic and historically variable category, and likewise the normative concepts which find their grounding in this dynamic conception, it becomes apparent that only an artefact which is adequate to the standard of AGI will truly be capable of ameliorating our understanding of what it means for a human agent to be creative. As Negarestani succinctly puts it:

To be human is the only way out of being human. An alternative exit—either by unbinding sentience from sapience or by circumventing sapience in favour of a direct engagement with the technological artefact—cannot go beyond the human. Rather it leads to a culture of cognitive pettiness and self-deception that is daily fodder for the most parochial and utilitarian political systems that exist on the planet (Negarestani 2018, 60).

In CC's vision, the possibility of AGI as something that genuinely confronts human species consciousness as an artefact, and even as an autonomous agent is ironically foreclosed. Thus CC's ideal in fact represents the petrified domination of self-consciousness as a bad infinity: it not only furnishes us with more tools for domination in the sphere of capitalist production, but also only imagines a creative and autonomous agent to be conceived along the lines of subjective autarky, thus foreclosing potentially fruitful avenues of exploration in the field of intuitionist mathematics and other non-classical logics. A genuinely creative AGI, rather than an instrumentally creative AI, also proposes the possibility of genuinely co-creative production. The mistake of CC here is to consider the idea of co-creativity as one in which the other is a mere tool for the realisation of some desired outcome. On the other hand, I want to suggest that because all creative production finds its ultimate meaning in relation to species-being, all creative acts are themselves to some extent co-creative. Moreover, in cases where that co-creativity is made explicit by collaboration, the creative ends are better served by a genuinely egalitarian relation, rather than one of master and servant, or the subject and their tool (albeit in the case of CC, a tool masquerading as a co-creative agent). What makes creative collaboration important is the very fact that the subjective ego gives way to some other that is non-identical to itself, that it exposes itself to

the other and thus is prompted to think creatively in relation to what cannot be subsumed by its familiar concepts.

CC’s conception of instrumental and free-standing creativity only affects an entrenchment of the dominant paradigm in computational research— one which emphasises performance capabilities over a more needs-based approach at the level of architecture as well as function— but additionally advances an assumption of creativity which obscures the normative social grounding from which a human concern with this concept is originally motivated. Aside from the fact that, as I have tried to argue, this notion of creativity is seriously impoverished and fails to get to the core of why such a concept matters to us in some normative sense, we might also want to consider the ideological ramifications of such a notion, particularly since this conception deprecates creativity’s normative and social critical functions whilst valorising its relevance to pre-defined problem spaces and industrial applications. Whilst it is not within the scope of this paper to present a fully-developed argument on this aspect of the problem, there are nevertheless a few relevant points we might want to raise as we bring things to a conclusion. For one thing, we ought to consider the social and political context surrounding the automation of not just manual, but also intellectual labour. As more and more of the occupations loosely gathered under the heading of “creative industry” become outsourced to computational systems, it becomes especially pertinent to examine the question of what the category of creative labour itself might mean, and how it can be conceptually modified to such an extent that certain functions previously thought beyond the bounds of automation can suddenly be subject to it. What I want to suggest here is just that once we begin to understand and indeed valorise creativity in a sense that only means acting within the valid problem spaces as sanctioned by the interests of capital, the actual critical dimension of creativity which allows us to challenge the legitimacy of this framing becomes more and more obscure. This is not to say that CC as a discipline is responsible for these effects, only, it certainly argues for an understanding of creativity in terms that further reinforce and entrench those existing norms. Ultimately, the category of ‘creative labour’ in itself involves an ideological valence, insofar as it is presented with a kind of pre-existing obviousness of the connection between the seemingly natural tendency of human beings to seek fulfilment through innovative acts, and the

ultimate valorisation of this tendency through the realisation of surplus value. This connection between the spontaneous desire to produce in creative ways and the sense in which it is inextricably linked to the production and accumulation of capital undermines the possibility of seeking out creative practice in a manner that is de-estranged and more directly connected to the aforementioned conception of species life, with its various attendant needs and fulfilment conditions. To primarily treat creativity as a performance-ethic, such that the relevant problem spaces are already implicitly defined, not only obscures the crucial function of creativity which consists in its capacity to challenge and reframe those problem spaces themselves, but further serves to entrench the very conditions of alienation which prevent us from seeing alternative possibilities for life. Within the total system of capital, wherever the individual subject is rewarded for acts which serve to reproduce that system, they become more inclined to see that system as one that best serves their own interests, even when it clearly does not. For this reason, I want to suggest that this struggle over the concept of creativity has a more than scholastic relevance. It demarcates an instance in which the semantic valence of the terms we use has some real import to the ways in which we understand our real experiences in their social and political dimensions. When creativity is understood as the mere exploration and transformation of problem spaces, or the capability of the individual agent to do so, a key critical tool which might otherwise be used to challenge the legitimacy and necessity of those problems is itself undermined. As such, an ideology critique of the notion of creativity being upheld here by CC might involve a questioning of the conditions under which the capacity of the individual agent to re-frame and challenge the necessity of things as they are presented obtain. One very concrete sense in which this point might need to be considered is in the context of existing labour struggles, wherein the felicity conditions for the automation of intellectual labour are to be found in this conceptualisation of creativity which can by and large be performed by agents incapable of addressing the very conditions of their own experience. It is not insignificant that, in a sense, the “creative” output of extant computational systems does not ‘belong’ to those systems in any meaningful sense. On the other hand, even where we take into account the range of problems that force us to include most forms of so-called creative labour within

our understanding of alienated labour, there remains a sense in which the work being produced by individuals in such industries ‘belongs’ to those individuals, and moreover reflects a set of personal concerns whose influence is often primarily attributable to the sense in which these become more universal concerns by resonating with others’ experiences. Just as we ought to understand the communist hypothesis not as an attempt to invent a purely new system but rather as the “real movement to abolish the present state of things” (Marx 2000b, 187), or the attempt to unmask existing social relations under capital for what they really represent in terms of social intercourse, we should also think about the extant creative production of individuals as ultimately attempts to express some hopes, desires, or ambitions towards a degree of social transformation. The personal aspect of creative production and activity finds its objective validity in the extent to which such acts make an appeal to the possibility of reflexive social transformation.

This is only possible for the agent who takes life itself as an object for her own conscious reflection. Treating creativity as something deracinated from these social conditions of realisation is tantamount to an apologia for prevailing ideology, it encourages the further development of positivising machines and their unchecked proliferation into the loop of human intellectual and cultural production. Where creative acts no longer clearly belong to the individuals that realise them, the state of estrangement in which the products of human labour confront human beings as something externally given becomes akin to a second-nature.

Bibliography

Adorno, T.W. 1983, *Negative dialectics*, Continuum, New York.

Althusser, L. 2003, *The humanist controversy and other writings, 1966-67*, Verso, London-New York.

Arnaout, H., Razniewski, S. 2023, *Can large language models generate salient negative statements?* <https://doi.org/10.48550/ARXIV.2305.16755>.

Berns, S., Broad, T., Guckelsberger, C., Colton, S. 2021, *Automating Generative Deep Learning for Artistic Purposes: Challenges and Oppor-*

tunities, in: Gómez de Silva Garza, Andrés. (Eds.), *Proceedings of the Twelfth International Conference on Computational Creativity, ICCCC'21 Virtual Meeting*, Association for Computational Creativity = ACC, Cuajimalpa, pp. 357-366.

Boden, M.A. 2004, *The creative mind: myths and mechanisms*, Routledge, London-New York.

Boden, M.A. 2018, *Creativity and Biology*, in: Gaut, B.N. (Ed.), *Creativity and Philosophy*, Routledge, New York.

Brassier, R. 2019, *STRANGE SAMENESS: Hegel, Marx and the logic of estrangement*, "Angelaki", 24, pp. 98-105.

Brun, G. 2020, *Conceptual re-engineering: from explication to reflective equilibrium*, "Synthese" 197, pp. 925-954.

Carnovaliani, F., Harley, N., Horner, S.T., Roda, A., Wiggins, G.A. 2021, *Meta-Evaluating Quantitative Internal Evaluation: A Practical Approach for Developers*, in: Gómez de Silva Garza, Andrés. (Ed.), *Proceedings of the Twelfth International Conference on Computational Creativity, ICCCC'21 Virtual Meeting*, Association for Computational Creativity = ACC, Cuajimalpa, pp. 213-217.

Carus, A.W. 2012, *Engineers and Drifters: The Ideal of Explication and Its Critics*, in: Wagner, P. (Ed.), *Carnap's Ideal of Explication and Naturalism*, Palgrave Macmillan UK, London, pp. 225-239.

Chen, R., Dannenberg, R.B., Raj, B., Singh, R. 2020, *Artificial Creative Intelligence: Breaking the Imitation Barrier*, in: Amílcar Cardoso, F. (Ed.), *Proceedings of the Eleventh International Conference on Computational Creativity, ICCCC'20*, ACC = Association for Computational Creativity, Coimbra, pp. 319-325.

Chieppe, P., Sweetser, P., Newman, E. 2022, *Bayesian Modelling of the Well-Made Surprise*, in: Hedblom, M.M. (Ed.), *Proceedings of the Thirteenth International Conference on Computational Creativity, ICCCC'22*, Association for Computational Creativity ACC, S. l., pp. 126-135.

Colton, S., Pease, A., Saunders, R. 2018, *Issues of Authenticity in Autonomously Creative Systems*, in: Pachet, F., Jordanous, A., León, C. (Eds.), *Proceedings of the Ninth International Conference on Computational Creativity, ICCO 2018*, Association for Computational Creativity, S. l., pp. 272-279.

Deleuze, G. 2011, *Difference and repetition*, Continuum, London.

Dutilh Novaes, C. 2020, *Carnapian explication and ameliorative analysis: a systematic comparison*, "Synthese" 197, pp. 1011-1034.

Finn, J.K., Tregenza, T., Norman, M.D. 2009, *Defensive tool use in a coconut-carrying octopus*, "Current Biology", 19, pp. R1069-R1070.

Fitzgerald, T., Goel, A., Thomaz, A. 2017, *Human-Robot Co-Creativity: Task Transfer on a Spectrum of Similarity*, in Goel, A., Jordanous, A., Pease, A. (Eds.), *Proceedings of the Eighth International Conference on Computational Creativity, ICCO’17*, Association for Computational Creativity ACC, S. l., pp. 104-111.

Hodson, J. 2017, *The Creative Machine*, in Goel, A., Jordanous, A., Pease, A. (Eds.), *Proceedings of the Eighth International Conference on Computational Creativity, ICCO’17*, Association for Computational Creativity ACC, S. l., pp. 143-150.

Hooker, J. 2018, *Truly Autonomous Machines Are Ethical*, <https://doi.org/10.48550/ARXIV.1812.02217>

Ibarrola, F., Brown, O., Grace, K. 2022, *Towards Co-Creative Drawing Based on Contrastive Language-Image Models*, in Hedblom, M.M. (Ed.), *Proceedings of the Thirteenth International Conference on Computational Creativity, ICCO’22*, Association for Computational Creativity ACC, S. l., pp. 96-100.

Jordanous, A. 2012, *A Standardised Procedure for Evaluating Creative Systems: Computational Creativity Evaluation Based on What it is to be Creative*, "Cogn Comput", 4, pp. 246-279.

Khurana, T. 2023, *Genus-Being: On Marx's Dialectical Naturalism*, in Corti, L., Schüle, J.-G. (Eds.), *Nature and Naturalism in Classical German Philosophy*, Routledge, New York.

Linkola, S., Kantosalo, A., Männistö, T., Toivonen, H. 2017, *Aspects of Self-awareness: An Anatomy of Metacreative Systems*, i: Goel, A., Jordanous, A., Pease, A. (Eds.), *Proceedings of the Eighth International Conference on Computational Creativity, ICCCC'17*, Association for Computational Creativity ACC, S. l., pp. 189-196.

Linkola, S., Mäkitalo, N., Männistö, T. 2020, *On the Inherent Creativity of Self-Adaptive Systems*, in Amílcar Cardoso, F. (Ed.), *Proceedings of the Eleventh International Conference on Computational Creativity, ICCCC'20*, ACC = Association for Computational Creativity, Coimbra, pp. 362-365.

Marx, K. 2000a, *Economic and Philosophical Manuscripts*, in McLellan, D. (Ed.), *Selected Writings*, Oxford University Press, Oxford-New York, pp. 83-122.

Id. 2000b, *The German Ideology*, in McLellan, D. (Ed.), *Selected Writings*, Oxford University Press, Oxford-New York, pp. 175-209.

McCarthy, J. 1977, *Epistemological Problems of Artificial Intelligence*.

McCormack, J. 2019, *Creative Systems: A Biological Perspective*, in Veale, T., Cardoso, A. (Eds.), *Computational Creativity: The Philosophy and Engineering of Autonomously Creative Systems, Computational Synthesis and Creative Systems*, Springer, Cham, pp. 327-352.

McDowell, J.H. 1996, *Mind and world: with a new introduction*, Harvard University Press, Cambridge, Mass.

Negarestani, R. 2018, *Intelligence and spirit. Urbanomic; Sequence, Windsor Quarry*, Falmouth, United Kingdom-New York.

Ng, K. 2021a, *Humanism: A Defense*, "Philosophical Topics", 49, pp.

145-163.

Id. 2021b, *On subjects, objects, and ground: Life as the form of judgment*, "European J of Philosophy", 29, pp. 1162-1175.

Id. 2020, *Hegel’s concept of life: self-consciousness, freedom, logic*, Oxford University Press, New York.

Id. 2015, *Ideology Critique from Hegel and Marx to Critical Theory: Ideology Critique from Hegel and Marx to Critical Theory*, "CONSTELLATIONS" 22, pp. 393-404.

Novaes, C.D. 2016, *Conceptual Genealogy for Analytic Philosophy*, in Bell, J.A. (Ed.), *Beyond the Analytic-Continental Divide: Pluralist Philosophy in the Twenty-First Century*, Routledge Studies in Contemporary Philosophy. Routledge, New York, pp. 75-108.

Roudavski, S., McCormack, J. 2016, *Post-anthropocentric creativity*, "Digital Creativity" 27, pp. 3-6.

Simondon, G. 2016, *On the mode of existence of technical objects*, Univocal Pub, Minneapolis, MN.

Stephensen, J.L. 2023, *Creativity versus Automation: Towards the Last Frontier, and With our Jobs on the Line?*, "Balkan Journal of Philosophy" 15, pp. 41-52.

Stephensen, J.L. 2020, *Post-creativity and AI: Reverse-engineering our Conceptual Landscapes of Creativity*, in Amílcar Cardoso, F. (Ed.), *Proceedings of the Eleventh International Conference on Computational Creativity, ICCO’20*, ACC = Association for Computational Creativity, Coimbra, pp. 326-333.

Suárez-Rodríguez, M., López-Rull, I., Macías Garcia, C. 2013, *Incorporation of cigarette butts into nests reduces nest ectoparasite load in urban birds: new ingredients for an old recipe?*, "Biol. Lett.", 9, 20120931.

Testoni, A., Greco, C., Bernardi, R. 2022, *Artificial Intelligence Models*

Do Not Ground Negation, Humans Do. GuessWhat?! Dialogues as a Case Study, "Front. Big Data", 4, 736709.

Totschnig, W. 2020, *Fully Autonomous AI*, "Sci Eng Ethics", 26, pp. 2473-2485.

Veale, T. 2016, *Computational approaches to language and creativity*, in Jones, R.H. (Ed.), *The Routledge Handbook of Language and Creativity*, *Routledge Handbooks in English Language Studies*, Routledge, London-New York, pp. 353-366.

Id. 2012, *Exploding the creativity myth: the computational foundations of linguistic creativity*, Continuum International Pub. Group, London-New York.

Veale, T., Cardoso, A. (Eds.) 2019, *Computational creativity: the philosophy and engineering of autonomously creative systems*, *Computational synthesis and creative systems*. Springer, Cham.

Veale, T., Cardoso, A., Pérez Y Pérez, R. 2019, *Systematizing Creativity: A Computational View*, in Veale, T., & Cardoso, A. (Eds.), *Computational Creativity: The Philosophy and Engineering of Autonomously Creative Systems*, *Computational Synthesis and Creative Systems*. Springer, Cham.

Veale, T., Pérez Y Pérez, R. 2020, *Leaps and Bounds: An Introduction to the Field of Computational Creativity*, "New Gener. Comput", 38, pp. 551-563.

Ventura, D. 2019, *Autonomous Intentionality in Computationally Creative Systems*, in: Veale, T., & Cardoso, A. (Eds.), *Computational Creativity: The Philosophy and Engineering of Autonomously Creative Systems*, *Computational Synthesis and Creative Systems*. Springer, Cham.

Wiggins, G.A. 2021, *Creativity and Consciousness: Framing, Fiction and Fraud*, in Gómez de Silva Garza, Andrés. (Ed.), *Proceedings of the Twelfth International Conference on Computational Creativity, ICCCC'21*, Association for Computational Creativity = ACC, Cuajimalpa, pp. 182-

191.

Wiggins, G.A. 2019, *A Framework for Description, Analysis and Comparison of Creative Systems*, in Veale, T. & Cardoso, A. (Eds.), *Computational Creativity: The Philosophy and Engineering of Autonomously Creative Systems, Computational Synthesis and Creative Systems*, Springer, Cham.

Wolfendale, P. 2021, *Autonomy and Automation talk*, <https://www.youtube.com/watch?v=GuKVllqgxmk>.

Id. 2019, *The Reformatting of Homo Sapiens*, "Angelaki", 24, pp. 55-66.

Wood, A.W. 2004, *Karl Marx, 2nd ed. ed, Arguments of the philosophers*, Routledge, New York.

Enda O’Riordan
The New Centre for Research and Practice
enda.oriordan@thenewcentre.org



ἘΠΈΚΕΙΝΑ

International Journal of Ontology
History and Critics

MICHELE PAVAN

Sulla creatività dei sistemi di calcolo, con una lettura filosofica
del problema della fermata di Alan Turing

EPEKEINA, vol. 16, nn. 1-2 (2023), pp. 1-33
Philosophy of Technology

ISSN: 2281-3209

DOI: 10.7408/epkn.

Published on-line by:

CRF – CENTRO INTERNAZIONALE PER LA RICERCA FILOSOFICA
PALERMO (ITALY)

www.ricercafilosofica.it/epekeina



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs 3.0 Unported License.

Sulla creatività dei sistemi di calcolo, con una lettura filosofica del problema della fermata di Alan Turing

Michele Pavan

1. La creatività fra psicologia degli albori e intelligenza artificiale. Il modello di Wallas

Il problema di chiarire se una macchina sia capace di azioni o risultati creativi fu esposto per la prima volta nella celebre proposta di Dartmouth, riconosciuta ancora oggi come l'atto di nascita dell'intelligenza artificiale (IA). Nell'incipit del testo del 1955, in cui gli autori propongono che dieci persone per due mesi svolgano una ricerca sull'IA, si dichiara che «lo studio procederà sulla base della congettura per cui, in linea di principio, ogni aspetto dell'apprendimento o una qualsiasi altra caratteristica dell'intelligenza possano essere descritte così precisamente da poter costruire una macchina che le simuli» (McCarthy *et al.* 2006, 413). Uno di questi aspetti, riportato al punto 7, è il “pensiero competente creativo”, definito in quanto distinto dal pensiero competente, ma non creativo, per il fatto di introdurre casualità nel sistema. La congettura per cui l'intelligenza fosse un fenomeno chiarificato al punto da consentire di riprodurlo in una macchina traeva la sua forza da una serie di studi che sul finire del XIX secolo ne formalizzarono gli aspetti come mai era stato fatto in precedenza (Gangemi *et al.* 2003). Per quanto riguarda l'aspetto della creatività,¹ in particolare, alcuni psicologi cominciarono a intenderla come un processo suddivisibile in fasi (Arieti 1986; Trombetta 1989). Un modello che nei primi anni del Novecento riscosse un certo successo fu quello di Wallas, il quale prevedeva una suddivisione del pensiero creativo in 4 fasi: preparazione, incubazione, illuminazione e verifica (2014). Tale modello verrà poi

1. Si assume qui, in linea con una tendenza degli esperti di IA di quel tempo, che la creatività sia uno degli aspetti dell'intelligenza, ma è bene ricordare che diversi studiosi hanno ritenuto come tra i due fenomeni non vi sia una correlazione necessaria. Fra i più celebri sostenitori di questa tesi si ricorda (Guilford 1950). Per un approfondimento e una contestualizzazione storica del lavoro di Guilford vedi (Beaudot 1977).

confermato, seppur con qualche modifica, da numerosi studi e raccolte dati sul modo di operare di poeti, artisti, scienziati e inventori – cioè persone considerate “competenti” e al tempo stesso “creative” (Arieti 1986, 16) – fino a essere ripreso, a partire dalla metà degli anni '50, nelle ricerche sull'IA. In quest'ambito un importante studio su *I processi del pensiero creativo* fu pubblicato qualche anno dopo la proposta di Dartmouth, nel 1962, da Newell, Shaw e Simon (Cantone 2022). Pur essendo in esso esplicito il riferimento a Wallas i tre autori non si limitano qui a una mera applicazione del suo modello in un nuovo ambito, proponendo più ambiziosamente di ridefinire la creatività a partire da una chiarificazione sia contenutistica che metodologica delle teorie esistenti.

Nella letteratura sull'argomento [...] l'attenzione si è concentrata, in particolare, sugli aspetti più spettacolari e misteriosi della creatività: i processi inconsci che si suppone si verificano durante l'“incubazione”, l'immaginazione presente nel pensiero creativo e la sua importanza nel rendere efficace il pensiero, e soprattutto il fenomeno dell' “illuminazione”, l'improvvisa intuizione che rivela la soluzione di un problema, a lungo cercata. [...] Tutti i punti ricordati sono abbastanza interessanti, e rappresentano le parti tipiche di una teoria del pensiero creativo. Nella nostra ricerca sulla creatività, tuttavia, abbiamo avvertito la necessità di avere un'idea più chiara dei requisiti e delle finalità complessive di tale teoria. La nostra proposta è che una teoria del pensiero creativo debba comprendere:

1. La definizione, interamente operativa, del comportamento dei meccanismi (o degli organismi) che, date le opportune condizioni iniziali, pensino effettivamente in modo creativo.
2. Una dimostrazione che i meccanismi che si comportano nel modo definito (da questi programmi) manifestano i fenomeni che abitualmente accompagnano il pensiero creativo (ad esempio, incubazione, illuminazione, formazione e cambiamento di set, e così via).
3. Un insieme di asserzioni, verbali o matematiche, sulle caratteristiche della classe di definizioni (programmi) che comprende i particolari esempi indicati (Cantone 2022, 148).

Anche prendendo per buono il modello di Wallas, una teoria soddisfacente del pensiero creativo dovrebbe insomma rendere conto di

come un sistema x (vivente o non vivente) sia poi capace di agire creativamente in certe condizioni, di come cioè sia *descrivibile in generale* un simile sistema quando in esso si manifestano fenomeni abitualmente associabili al pensiero creativo. In tal senso, la realizzazione di un programma per riprodurre in una macchina i processi descritti da Wallas, oltre che a favorire l'applicazione del suo modello a un ambito nuovo fornirebbe secondo gli autori una *definizione operativa* (punto 1) e al contempo *formale* (punto 3) dei comportamenti abitualmente associabili al pensiero creativo. Tuttavia, lungi dal fornire una *definizione* della creatività, con simili programmi si spiega al limite cosa voglia dire dal punto di vista di una macchina (e al limite – per un' analogia tutta da verificare – di un uomo) agire creativamente *nel senso di Wallas*, il quale non aveva propriamente spiegato che cosa fosse la creatività, limitandosi a descriverne il processo nei termini di un progetto, o problema da risolvere, suggerendo di associarla al modo di operare di persone competenti quali poeti, artisti, scienziati e inventori. D'altra parte, un argomento tipico fra gli esperti IA, come vedremo, è proprio che non sia necessario, e forse nemmeno possibile, disporre di una definizione del genere, e che è se mai attraverso la costruzione di macchine capaci di comportarsi in modi *intuitivamente* intelligenti o creativi che si perverrà forse un giorno a definizioni più precise di questi fenomeni (Veale *et al.* 2019, 3). Gli stessi Newell, Shaw e Simon, sulla scia dell'approccio di Wallas considereranno il pensiero creativo come «un tipo particolare di comportamento di soluzione di problemi» (Cantone 2022, 149), basando questa tesi sull'assunto che «tutte le persone che nel corso della storia hanno promosso i maggiori progressi creativi nella scienza e nella tecnica hanno posseduto capacità molto elevate nella soluzione generale di problemi» (Cantone 2022, 150). La tesi pone qui come ovvio ciò che sarebbe da dimostrare, ovvero che la creatività sia riscontrabile in generale nei calcolatori e, in particolare, in quelli artificiali. Sorvolando su questo punto, gli autori assumono così la correlazione fra creatività e capacità computazionali – dell'uomo e della macchina – per poi individuare in aspetti come la novità o la non convenzionalità dei criteri per distinguere le soluzioni creative dei problemi da quelle generiche. Come vedremo in seguito, aspetti come questi sono effettivamente correlati alla creatività, ma lo sono, aggiungiamo, a prescindere dal fatto che essa si manifesti nella soluzione di un problema. Una procedura di calcolo può anche essere creativa

ma ciò non significa che la creatività sia una caratteristica distintiva, e tanto meno esclusiva del calcolo, così come il fatto che le persone che nella storia hanno promosso i maggiori progressi creativi fossero anche abili nella soluzione di problemi non implica che la creatività sia un'esclusiva dei calcolatori, umani o non umani che siano. Con ciò non si tratterà di sostenere che la creatività non abbia nulla a che fare con la soluzione di problemi ma, in primo luogo, che anche quando la creatività emerge in queste circostanze essa si manifesta in tipi di operazioni e risultati *non essenzialmente* computazionali, e in secondo luogo che tali manifestazioni, descrivibili in termini di novità o non convenzionalità, quando si verificano nel corso di un calcolo implicano in ogni caso una sorta di *aporia operativa*, non essendo *formalmente* possibile il verificarsi di operazioni o risultati creativi – cioè dotati di quegli aspetti manifesti – nel corso di un calcolo.

2. L'approccio soggettivistico all'analogia uomo-macchina

2.1. Il metodo di Turing

L'idea che sia possibile riconoscere in una macchina i fenomeni dell'intelligenza o della creatività senza avere di essi una definizione precisa è un assunto ricorrente negli studi sull'IA. La tesi per cui il problema dell'intelligenza artificiale sia sostanzialmente quello di «costruire una macchina che si comporti in modi che sarebbero considerati intelligenti nel caso di un essere umano» (McCarthy *et al.* 2006, 424) – sulla quale si fonda l'intera proposta di Dartmouth – presuppone un'idea *non chiarita* di che cosa sia l'intelligenza. Tale approccio affonderebbe così le sue radici nel famoso articolo in cui Turing suggerì, forse per primo, di mettere da parte le definizioni generali – nel suo caso quelle di “macchina” e di “pensiero” – per approcciare la questione dal lato pratico, con un esperimento mentale passato poi alla storia come “Test di Turing” (Cantone 2022, 111-137). L'esperimento consiste nell'immaginare un gioco con tre partecipanti: un uomo che pone delle domande e altri due – un uomo e una macchina – chiamati a rispondere. L'interrogante viene chiuso in una stanza, separato dagli altri due. Il suo scopo è determinare, in base alle risposte degli interrogati – pervenutegli attraverso una telescrivente o un intermediario – chi fra i due sia la macchina e chi l'uomo. Scopo della macchina è quello di far

credere all'interrogante di essere un uomo. Scopo dell'uomo è fornire risposte che aiutino l'interrogante a capire la verità. La tesi di Turing è che se l'interrogante non riuscirà a distinguere la macchina dall'uomo allora la macchina pensa, o meglio, che nel caso in cui la macchina svolga in maniera soddisfacente il "gioco dell'imitazione" si avrebbe una prova sufficiente per concludere che essa pensa. Con un approccio retrospettivamente simile ai proponenti di Dartmouth, dunque, Turing inferisce il pensiero nella macchina a partire dalla capacità di essa di riprodurre il comportamento umano. Tale approccio si fonda su almeno due assunti: uno consistente nell'assimilare il pensiero a comportamenti x dell'uomo (assunto oggettivistico), l'altro nel ritenere che un essere pensa allorché un altro essere riconosce nel comportamento del primo la proprietà x (assunto soggettivistico). Considerando che le macchine più adatte a partecipare al gioco dell'imitazione sono per Turing i calcolatori ne consegue che la proprietà del comportamento che qui si suppone sintomatica del pensiero sia la *computazionalità*. Ora, questo modo di approcciare la questione influenzerà, come vedremo meglio nel prossimo capitolo, oltre che i proponenti di Dartmouth, diversi altri esperti IA nel corso del Novecento, rappresentando tutt'ora un importante punto di riferimento negli studi sulle capacità delle macchine di simulare i processi di pensiero dell'uomo. Interessante, da questo punto di vista, è il caso degli studi sulla *creatività*, che pur formando nel loro insieme un ambito nuovo e a sé dell'IA sembrerebbero anch'essi riflettere in qualche misura l'approccio di Turing. In un collettaneo del 2019 – orientato tra le altre cose a definire l'ambito della creatività computazionale (CC), chiarificandone le fasi del recente sviluppo e i progetti a venire – fin dal primo capitolo si afferma che, «date le ovvie difficoltà nel distillare una pura definizione di creatività», i ricercatori CC «ignorano la necessità di definire oggettivamente il fenomeno», e che uno dei loro obiettivi è far sì, se mai, che «le macchine mostrino un livello umano di creatività che anche gli umani percepiranno come creativo» (Veale *et al.* 2019, 3). Riconosciamo qui esplicitamente almeno il secondo assunto, soggettivistico, del metodo di Turing, per il quale una macchina presenta aspetti umani allorché un soggetto li riconosce nel suo comportamento. Che ora l'aspetto in questione sia la creatività, anziché il pensiero in generale, non fa alcuna differenza, poiché in entrambi i casi quel che conta è, per il secondo assunto di Turing, la possibilità stessa che qualcuno riconosca o meno nei modi

della macchina l'aspetto in questione. Per quanto si dica, inoltre, che i ricercatori CC ignorino la necessità di definire oggettivamente la creatività, tale definizione – e con essa il primo assunto di Turing – è qui almeno parzialmente data allorché si assume, come essi fanno, che la proprietà minima del comportamento creativo della macchina sia appunto la computazionalità; va da sé, poi, che a differenza del pensiero in senso lato – che è fatto coincidere, nell'approccio generale di Turing, con il concetto stesso di computazione –, il pensiero creativo dovrà presentare, oltre alla computazionalità, altre e più specifiche proprietà. Ma queste, come vedremo, verranno intese dai ricercatori CC in quanto descrittive di certi *modi di reagire* dell'osservatore – come ad esempio la *sorpresa* – a un comportamento computazionale, motivo per cui non sono da essi incluse, in ultima istanza, nelle proprietà che contribuirebbero a definire *oggettivamente* il fenomeno.

Un altro esempio di applicazione dell'approccio di Turing alla questione CC – e a cui gli studiosi dell'ambito spesso si ispirano – è rappresentato dai lavori di Margaret Boden. In un testo del '90, infatti, divenuto ormai un classico nella letteratura sull'argomento, assumendo che «la creatività umana è qualcosa di misterioso» (2004, 1), Boden afferma che il suo scopo rimane appurare «se i computer potrebbero proporre idee che almeno *sembrano* essere creative» (2004, 7). Tratteremo in modo più approfondito anche questa ipotesi, ma possiamo anticipare che, come nel caso già trattato, suggerendo che un computer – che altro non è che un calcolatore, ovvero una macchina di Turing – sarebbe creativo se qualcuno lo percepisse come tale, anch'essa si fonderebbe sui due assunti del metodo di Turing. Ora, prima di occuparci di tutte queste teorie sarà bene mostrare i limiti che Turing stesso riconosceva al suo metodo. Per quanto riguarda il primo assunto, in *Computing machinery and intelligence* (Cantone 2022) il logico inglese ammetteva che l'equiparazione fra pensiero e computazione fosse una congettura ancora da verificare. Il gioco dell'imitazione può quindi essere letto come un tentativo di verifica di tale congettura. A conferma dell'impostazione congetturale del suo discorso Turing riporta, nel capitolo 6, una serie di «opinioni contrarie a proposito dell'argomento principale». Fra queste vi è l'idea che ciò che fanno le macchine – in una parola: computare – per quanto possa sembrare equiparabile al pensiero umano rimane molto differente da ciò che fa un uomo quando pensa (Cantone 2022, 113). Tuttavia, se la macchina

svolgerà in maniera soddisfacente la sua parte nel gioco dell'imitazione – che in senso più ampio significa: se un calcolatore artificiale svolgerà il suo ruolo in modo tale che un osservatore possa *scambiarlo* per un uomo, cioè un essere pensante – allora l'obiezione per cui il pensiero umano non è computazione rimane irrilevante. Ma in tal modo Turing introduce (anche) nel corso dell'esperimento una seconda regola, nonché un'ulteriore congettura, per la quale il *ricoscimento* da parte di un osservatore dei tratti del pensiero nel comportamento di una macchina sarebbe sufficiente a confermare la prima congettura, cioè l'equiparazione fra pensiero e computazione, e così, più in generale, che le macchine pensano. Tale seconda congettura corrisponde a ciò che abbiamo definito “assunto soggettivistico” ed è anch'esso qui messo alla prova da Turing nella serie delle possibili obiezioni da muovere alla sua tesi. L'obiezione considerata a tal fine è «l'opinione dell'autocoscienza».

Secondo la forma più estrema di questa opinione il solo modo per cui si potrebbe essere sicuri che una macchina pensa è quello di essere la macchina e di sentire se stessi pensare. Allo stesso modo [aggiunge] secondo questa opinione la sola via per sapere che un uomo pensa è quella di essere quell'uomo in particolare. È questo in effetti il punto di vista solipsistico. Può essere il punto di vista migliore cui attenersi sul piano logico, ma rende difficile la comunicazione delle idee. Probabilmente A crederà “A pensa, ma B no”, mentre B crede “B pensa, ma A no”. Invece di discutere in continuazione su questo punto, è normale attenersi alla educata convinzione che ognuno pensi (Cantone 2022, 123-124).

Per difendersi dall'argomento e confermare che anche le macchine pensano vediamo qui come Turing mostri prima quanto sia assurdo, facendo leva su di esso, sostenere che un altro uomo non pensi. Da qui, una volta stabilito che qualsiasi uomo pensa ne segue che basterà osservare che una macchina si comporta come un uomo per convincersi che anch'essa lo fa, e così – in virtù dell'equivalenza emersa, sotto tale aspetto, fra uomo e macchina – che il pensiero coincide in generale con la computazione, fondando in questo modo, di fatto, l'assunto oggettivistico su quello soggettivistico. Ora, da questo quadro è possibile evincere che i limiti del metodo di Turing sono al contempo i suoi punti di forza. Il lato debole del metodo emergerebbe ogni qual volta Turing mostri come si possa criticarlo, rivelando la natura essenzialmente

congetturale dei suoi principi. Tuttavia, le congetture sono proprio, anche, il suo punto di forza, poiché consentirebbero di sorvolare sulle medesime critiche che – per quanto forti, plausibili sul piano logico – si rivelerebbero ininfluenti nella pratica. Ma rimane il fatto che le congetture non sono delle prove. Il test di Turing è importante perché suggerisce che se un uomo scambia una macchina per un altro uomo è probabile che in essa vi sia qualcosa di riconducibile all’umano in generale, ma per provare che sia così bisognerebbe poi ricavare l’“umanità della macchina” dalla macchina stessa più che dalle reazioni di chi la osserva.

In tal senso, se partissimo da una definizione *oggettiva*, anche temporanea, del pensiero, o di altri aspetti da cui si ritiene che sia possibile inferire l’umanità – anziché concludere che questi coincidono con la computazione per il semplice fatto che qualcuno riconosce l’analogia fra il comportamento di un uomo (che presenterebbe per definizione quegli aspetti) e quello di una macchina – si potrebbe all’inverso studiare che cosa *nella macchina*, e dunque *nella computazione*, abbia a che fare con tali aspetti (pensiero, intelligenza, creatività ecc.), allorché il test suggerisca analogie fra il comportamento di quella macchina e il comportamento umano. Tale approccio, che si potrebbe definire *oggettivistico*, non sarebbe dunque in contrasto con i fini pratici del test, anzi, in un certo senso sembrerebbe persino ampliarli, suggerendo la possibilità non solo di cogliere, partendo dalle rispettive definizioni di macchina come “sistema computazionale” e di umano come “sistema pensante, intelligente o creativo”, gli aspetti del secondo nella prima, ma anche, una volta individuato ciò che nella macchina risulta analogo, sotto quegli aspetti, al comportamento umano, definire meglio questi stessi “umani” aspetti *a partire* dallo studio della macchina. Proseguendo in tale direzione, e in conformità con gli obiettivi di questo contributo, nel prossimo capitolo si considererà allora l’analogia uomo-macchina sotto l’aspetto della creatività. In breve, assumendo la definizione per cui *creativo* è “ogni comportamento o risultato che, in rapporto a un dato insieme di regole, risulta *impossibile*”, mostrerò come in base ad alcune implicazioni filosofiche del “teorema della fermata” – noto risultato della logica matematica ottenuto da Turing nel

'36² –, si possa sostenere che esistono operazioni e risultati delle macchine conformi alla definizione, e così che le macchine siano creative in quanto capaci, per certi versi, di fare l'impossibile.

2.2. Il punto di vista sulla creatività

2.2.1 Studi recenti

Abbiamo anticipato come gli studi di CC siano influenzati ancora oggi dal metodo di Turing. Questa influenza, per quanto implichi dei limiti interni, strutturali, in tale ambito di ricerca, presenta in ogni caso degli aspetti positivi. Il fatto di occuparsi di creatività computazionale in termini sperimentali – di realizzare programmi che, pur senza soddisfare un criterio oggettivo di creatività, sarebbero tali da riprodurre comportamenti che nel caso dell'uomo non esiteremmo a definire creativi – potrebbe forse permettere un giorno di chiarire retrospettivamente, a partire dallo studio di programmi più complessi, la definizione vaga, intuitiva, di creatività da cui la ricerca stessa era partita. Tale approccio, che potremmo definire – nel senso buono del termine – scienziata, ha certamente il merito di basare, o di *attendere di basare* le proprie definizioni su fatti osservabili e descrivibili in modo rigoroso. Tuttavia, senza una definizione oggettiva, data in partenza, di creatività, il rischio è quello di scambiare per creativi comportamenti che semplicemente *appaiono* tali a un osservatore, fino al punto che qualunque fenomeno, data l'infinita varietà dei punti di vista, possa essere ritenuto creativo e da qui un possibile candidato per la definizione oggettiva di creatività. È un dato che diversi autori nell'ambito CC condividono oggi tale approccio (Veale *et al.* 2019). Per fare fronte a suoi punti più deboli essi vi includono, oltre all'osservazione in senso lato, alcuni altri criteri, ma sempre di tipo soggettivistico. Graeme Ritchie, per esempio, recupera da Boden (2004) alcuni criteri generici per la definizione di un sistema creativo – “novità”, “qualità/valore”, “tipicità” dei comportamenti o degli output generati (cioè il loro *essere conformi* a quelli noti al sistema, detto anche “accettabilità”) e “sorpresa” che

2. È bene precisare che Turing non ha mai usato – per riferirsi ai risultati ottenuti nell'articolo – le espressioni “teorema della fermata” e “problema della fermata”, introdotte solo successivamente dai suoi interpreti. Anche su questo punto ritorneremo più avanti.

questi generano nell'osservatore (Boden 2004, 168-170) – con l'intento di integrarli in un quadro teorico più ampio e preciso. Seguendo questi criteri, dato un «insieme ispiratore» *I*, cioè l'insieme degli artefatti su cui si basa il comportamento del sistema, una «valutazione della tipicità» *typ* e una «classificazione del valore» *val* di un dato output, diventa possibile considerare ad esempio la novità di quest'ultimo – fattore cruciale per la creatività di un sistema, ma anche generico se riportato da un'osservazione senza criteri – nei termini di «avere un valore basso per *typ* o essere diverso da *I*» (Boden 2004, 170). Ma la bassa tipicità di un output, o la sua differenza da un modello, non sono tali in questo caso dal punto di vista del sistema che lo genera – il quale altro non fa, appunto, che produrre output in base ai modelli di cui dispone – ma da quello del valutatore, che in base a una *sua* idea di tipicità, immaginata a partire dalle fattezze dei modelli forniti al sistema, stabilisce poi, in rapporto a tale idea, se l'output è o non è nuovo, e da qui se il sistema che l'ha prodotto è o non è creativo. D'altra parte lo stesso Ritchie, dopo aver distinto le prospettive di valutazione «descrittiva» (per la descrizione di ciò che accade all'interno dell'attività creativa) e «causale» (per l'identificazione di azioni o precondizioni che, quando si verificano assieme, determinano o costituiscono un evento creativo), riconosce che i suoi criteri definirebbero più precisamente un punto di vista «sintomatico» (Boden 2004, 167-168), il quale, partendo dal presupposto che «la creatività (rigorosa³) è negli occhi di chi la guarda», specifica da una parte «quali fattori osservabili porterebbero un osservatore ad attribuire la creatività al sistema» (Veale *et al.* 2019, 168), e dall'altra le misure che costui «potrebbe applicare per fare affermazioni precise e verificabili su un sistema creativo» (Veale *et al.* 2019, 170). Kazjon Grace e Mary Lou Maher, accogliendo a loro volta che la creatività è negli occhi di chi la guarda, insistono similmente sul fatto che la novità, in quanto aspetto decisivo per l'attribuzione

3. Ritchie distingue appunto due modi di intendere e valutare la creatività: uno vago (*loose*) e uno rigoroso (*strict*). Il modo vago è quello per il quale sarebbero creative tutte quelle attività che, dal punto di vista più ampio di una società, sarebbero considerate di per sé creative: belle arti, scrivere poesie o opere di narrativa, comporre musica ecc. Quello rigoroso, invece, inferisce la creatività di un autore e/o di un atto e/o di un risultato solo e unicamente a partire dal grado di eccellenza che sarà possibile attribuirgli (Veale *et al.* 2019, 161-162).

della creatività, andrebbe misurata a partire dalla «violazione delle aspettative degli osservatori» (Veale *et al.* 2019, 195) piuttosto che dalle differenze oggettive tra gli artefatti. In tal senso, sulla base di un approccio spiccatamente soggettivistico avrebbero ideato un sistema creativo in ambito culinario che, data una ricetta come input è capace, da una parte, di *prevedere* quali ingredienti affini aspettarsi come output, e dall'altra di *suggerire* in base a questa previsione un ingrediente inaspettato (Veale *et al.* 2019, 204-205). Da entrambi i casi si può evincere come i limiti dell'approccio soggettivistico derivino dal fatto che per quanto un output possa apparire atipico o imprevisto a un osservatore qualsiasi (esperto o inesperto che sia), esso rimane perfettamente *possibile* – e a fortiori *tipico e prevedibile* – in base alle regole e ai modelli di cui il sistema dispone per dare output. Tali osservazioni rivelerebbero allora più i difetti computazionali e/o conoscitivi dell'osservatore – relativamente a ciò che *può fare* il sistema osservato – che il potenziale creativo del sistema stesso. Gli autori di (Veale *et al.* 2019) si dimostrano tuttavia consapevoli dei limiti del loro approccio, tanto da riconoscerli espressamente, aggiungendo addirittura che le osservazioni potrebbero rivelarsi utili per identificare eventuali bias umani inerenti all'idea di creatività (3). Ma, ciò nonostante, essi si disinteressano altrettanto deliberatamente di definire la creatività in maniera oggettiva.

2.2.2 Alle radici degli studi recenti sulla creatività computazionale: il caso di Margaret Boden

Questa posizione generale negli studi sulla CC può essere letta alla luce della teoria della creatività di Boden, di cui gli autori di (Veale 2019) tendono ad assumere, in modi che sono più o meno espliciti a seconda dei casi, le premesse di partenza. All'inizio di *The Creative Mind*, condividendo l'idea di Lovelace secondo cui *un computer può fare solo ciò che il suo programma gli permette di fare*⁴ – specificando tuttavia che, pur condividendola, rimarrebbe semplicistica se intesa co-

4. Questa frase di Boden è una versione più contratta e immediata della frase originale di Lovelace: «The Analytical Engine has no pretensions whatever to *originate* anything. It can do *whatever we know how to order it to perform*» (Menabrea-Lovelace 1842).

me argomento a sfavore di qualsiasi nesso fra creatività e computer – la filosofa inglese elenca quattro questioni relative alla creatività, definite *Lovelace-questions* e indicative a nostro avviso, ognuna a suo modo, dei due possibili punti di vista – soggettivistico e oggettivistico – da cui è possibile osservare il fenomeno. Le prime tre questioni indagano, in ordine, «se le idee computazionali possano aiutarci a capire come sia possibile la creatività umana», «se i computer (oggi o in futuro) possano mai fare cose che almeno in apparenza siano creative», «se un computer possa mai sembrare in grado di riconoscere la creatività – ad esempio nelle poesie scritte da poeti umani» (Boden 2004, 16-17). La prima questione è quella che definisce anche, più in generale, la ricerca di Boden, dove lo studio delle dinamiche computazionali dei sistemi artificiali è da sempre funzionale a una maggior comprensione dell'intelligenza e della creatività umane (1987, 2004). La seconda e la terza, sebbene secondarie, per ammissione stessa dell'autrice sono cruciali allorché «gettano luce sulla prima» (Boden 2004, 17). Nell'ottica del presente contributo, inoltre, lo sono in quanto rivelano la prospettiva soggettivistica entro cui Boden considera la creatività dei sistemi artificiali. L'indagine della seconda questione, infatti, presuppone che si consideri questi sistemi dalla prospettiva di un soggetto qualsiasi (che potrebbe essere tanto un individuo – esperto o inesperto non cambia – quanto una società) tenendo conto di ciò che lo spingerebbe ad attribuire loro capacità creative (nel caso di Boden: la novità, il valore e l'imprevedibilità degli output generati), mentre l'indagine della terza presuppone che si consideri, dal medesimo punto di vista, se tali sistemi siano capaci di riconoscere in quanto creativi quegli output che i loro osservatori riconoscerebbero come tali. Ma veniamo ora, invece, alla quarta questione, ovvero stabilire «se i computer stessi possano mai essere realmente creativi (invece di limitarsi a produrre prestazioni apparentemente creative la cui originalità è interamente dovuta al programmatore umano)». È significativo che Boden riveli fin da subito di non volerla trattare.⁵ Essa, infatti, richiederebbe un'indagine squisitamente oggettivistica del fenomeno in questione. E ciò almeno

5. Per la precisione Boden rimanda all'ultimo capitolo la questione, ma anche in quel frangente non mostra come sia possibile rispondere "sì" alla questione ma solo come sia possibile confutare gli argomenti di chi risponderebbe "no", lasciandola così aperta.

in due sensi: da una parte perché, ed è il senso a cui allude Boden, essa invita a considerare se l'oggetto osservato sia l'*autore reale* delle operazioni creative che esegue e dall'altra perché invita a considerare se tali stesse operazioni siano realmente creative (e non solo *percepite come tali* da un soggetto qualsiasi). Ora, sebbene Boden non spieghi i motivi della resistenza a indagare il fenomeno da questo punto di vista, limitandosi a dire che la quarta questione non verrà affrontata nel libro, mi sembra possibile quanto meno suggerirne due: uno di carattere storico-metodologico e l'altro di carattere teoretico-speculativo.

Partendo dal primo, si potrebbe sostenere che Boden non consideri il punto di vista oggettivistico poiché richiederebbe da una parte di stabilire una volta per tutte una definizione oggettiva di creatività e dall'altra di non affidare più la valutazione della creatività effettiva di un sistema artificiale all'osservazione di un soggetto, il che significherebbe in entrambi i casi venir meno a quell'approccio pragmatico e sperimentale dei suoi predecessori (da Turing a Newell, Shaw e Simon) a cui sembra lei stessa allinearsi (2004, 10).⁶ Per quanto riguarda il secondo, invece, si potrebbe sostenere che Boden non consideri questo punto di vista in quanto, se si stabilisce, come del resto essa fa, che la tesi di Lovelace è corretta, cioè che *un computer può fare solo ciò che il suo programma gli permette di fare*, ne consegue che esso non possa essere *realmente* creativo e che non avrebbe dunque alcun senso indagare la creatività da un punto di vista oggettivistico. Ed è proprio per questo motivo che, al fine di mantenere aperta la possibilità che le macchine siano comunque creative, senza abbandonare la tesi di Lovelace, Boden sposterà, come già Turing fece anni prima,⁷ il luogo in cui cercare la creatività dell'oggetto osservato dall'oggetto medesimo al suo osservatore. Tale approccio la porterà a definire tre differenti forme di creatività (combinatoria, esplorativa e trasformativa) (Boden

6. In questo passaggio di *The Creative Mind* è avvertibile la prudenza di Boden nel definire oggettivamente la creatività, in quanto, essendo un fenomeno strettamente correlato con la scala di valori di chi lo valuta rimarrebbe non isolabile da un punto di vista particolare.

7. Turing ragiona, come si è visto, sulla caratteristica umana del pensiero; ma parla anche di creatività, nello stesso articolo, proprio a partire dalla tesi di Lovelace, facendo vedere che una macchina può, nonostante Lovelace, generare risultati in grado di sorprendere il suo programmatore, e cioè risultati che in un certo qual modo non rispecchierebbero gli ordini impartiti (Cantone 2002, 128).

2004, 3)), nessuna delle quali implicante l'esecuzione, da parte di una macchina, di operazioni realmente creative, cioè realmente diverse da quelle che il programmatore le ha ordinato di eseguire. Persino nel caso della creatività trasformativa – che per Boden è la più radicale –, per quanto una macchina possa infatti modificare il proprio stesso spazio concettuale, consentendo operazioni che prima di ciò non sarebbero state possibili, rimane il fatto che essa, per la tesi di Lovelace, non ha altro modo di eseguire tale trasformazione se non eseguendo ordini iscritti in questo spazio, da cui consegue che anche le operazioni che prima della trasformazione si è creduto non fossero possibili in realtà lo erano considerando il tipo di trasformazioni di se stessa che la macchina poteva compiere in base a quegli stessi ordini, e che la percezione dell'incapacità della macchina di eseguire quelle operazioni dipendeva allora più dai limiti del programmatore nel comprendere le possibilità reali della macchina – cioè del programma, o insiemi di ordini, che lui stesso ha creato – che da quelli della macchina chiamata a eseguirle. C'è da dire che su questo punto Boden rimane un po' ambigua. Parlando di creatività trasformativa, in diversi punti essa afferma che le trasformazioni che un sistema applica a se stesso gli rendono a partire da quel momento possibili una serie di operazioni che non potevano esserlo prima. D'altra parte, però, sappiamo che Boden ritiene anche corretta la tesi di Lovelace, e così che tutto ciò che una macchina fa è necessariamente *sempre stato possibile* in base alle sue regole. Questo potrebbe forse spiegare il motivo per cui Boden parla, in definitiva – anche nei casi di creatività trasformativa – di sistemi che sembrano essere creativi ma che non (sappiamo se) lo sono *realmente*. E tuttavia, proprio in virtù di tale ambiguità di fondo si potrebbe sospettare che Boden non sia mai stata del tutto persuasa dalla tesi di Lovelace secondo cui *un computer può fare solo ciò che il suo programma gli permette di fare*. Il modo con cui la filosofa inglese descrive la creatività trasformativa lascia intendere com'ella potrebbe aver colto, fra le pieghe di questa idea, delle evidenze per le quali bisognerebbe concludere, al contrario, che i computer fanno cose che i loro programmi non gli permetterebbero di fare, salvo poi decidere, più o meno consapevolmente, di proseguire le sue ricerche in sicurezza, mantenendosi nei limiti tracciati da Lovelace e lasciando impensata la questione della *creatività reale* dei sistemi artificiali.

Ma il caso di Boden è allora interessante per due motivi antitetici. Da una parte perché essa farebbe da spartiacque fra la generazione di studiosi che ha pensato i problemi del pensiero e dell'intelligenza dei sistemi artificiali secondo il paradigma della computazione e quella che, anche grazie al suo contributo, ha pensato sotto lo stesso paradigma la creatività, e dall'altra in quanto nel mezzo di tale passaggio essa mostrerebbe anche, senza tuttavia approfondirli, i limiti stessi di tale paradigma, per il quale appunto si sosterebbe che – per quanto possa essere strana – qualsiasi caratteristica (pensiero, intelligenza, creatività) ritenuta attribuibile a un sistema che è – fra le altre cose (e il punto è proprio capire cosa tali “altre cose” siano) – capace di calcolare, gli è a tutti gli effetti attribuibile se e solo se *qualificante* operazioni *conformi* a regole o istruzioni di cui il sistema dispone. Parlando di creatività trasformativa, infatti, abbiamo visto come Boden dicesse di riferirsi a una situazione in cui il sistema esegue delle operazioni che fino a prima della trasformazione *non* sarebbero state *possibili*, cioè che non sarebbero state quanto meno *conformi* all'insieme delle regole di cui esso disponeva. Per capire ora come in tale idea venga scoperta, ma al contempo persa di vista, la questione della *creatività reale* (o *radicale*) dei sistemi artificiali, c'è da chiedersi quanto segue: in che senso si può dire che fino a prima che la macchina trasformasse se stessa determinate operazioni non erano possibili se è pur vero che questa trasformazione (e dunque tutto ciò che grazie a essa diventa possibile, comprese quelle operazioni) dev'esserlo sempre stata? Prima di rispondere vediamo di chiarire l'assunto interno alla domanda. Che una simile trasformazione debba essere sempre stata possibile segue direttamente dal fatto che se la macchina la esegue significa che essa è stata programmata per eseguirla e che dunque tale possibilità di esecuzione è sempre esistita per quanto mai realizzata. L'alternativa sarebbe dire che anche questa trasformazione non sarebbe stata di per sé possibile se non in virtù di un'ulteriore trasformazione, ma più si continua a negare in questo modo che le trasformazioni che una macchina applica a se stessa non sono possibili in base alle regole di cui la macchina dispone – ma solo grazie ad altre trasformazioni che le rendono possibili – più diventa evidente che a un certo punto, da qualche parte, devono esservi state nella macchina delle regole che hanno reso possibile almeno una prima trasformazione e da qui, come per un effetto domino, tutte le trasformazioni successive, compresa

quella che renderà possibile, alla fine della serie, le operazioni ritenute fino a quel momento impossibili. Chiarito questo punto possiamo ora rispondere alla domanda. Possiamo dire, in particolare, che se è vero che la trasformazione che ha reso possibile le operazioni fino a quel momento impossibili è sempre stata possibile in base alle regole di cui il sistema disponeva ne consegue che lo erano anche quelle operazioni malgrado potessero apparire impossibili a un osservatore incapace di prevederle sulla base delle trasformazioni che il sistema avrebbe potuto applicare a se stesso.

Fin qui, dunque, nulla che possa *davvero* scuotere la tesi di Lovelace. Le modifiche che la macchina può apportare al proprio spazio concettuale, proprio in quanto possibili non tolgono, anzi confermano, che nella macchina tutto procede secondo le sue regole. Il punto centrale è che la creatività trasformativa, per come è concepita da Boden, non sembra sufficiente per parlare di *creatività reale*. Affinché un sistema possa essere realmente creativo deve mostrare di eseguire non tanto operazioni rese possibili da delle trasformazioni a loro volta possibili ma operazioni *realmente impossibili*. Deve poter fare ciò che il suo programma non gli permetterebbe di fare. Da questo punto di vista, allora, non avrà alcun senso distinguere un'operazione da una trasformazione che la rende possibile. Quello che conta per l'attribuzione della creatività reale a un sistema è che un'operazione da esso eseguita sia impossibile in base alle regole di cui dispone (cosa intravista e tuttavia non approfondita da Boden, che finirà per scambiare l'*impossibilità reale* con la *possibilità imprevedibile* o *impossibilità apparente* di un'operazione). Se questo criterio è soddisfatto va da sé poi che quell'operazione sia anche *trasformativa*; non nel senso, però, che è programmata per far sì che un'operazione impossibile, o meglio inconsueta o imprevedibile per un osservatore, si realizzi, ma perché non essendo propriamente possibile allora la sua effettiva esecuzione deve aver comportato in qualche misura l'istituzione di una regola nuova. Torneremo più avanti su questo punto. Prima di arrivarci sarà essenziale chiarire, nel prossimo capitolo, in che termini la nozione di creatività sia indissolubilmente legata a quella di impossibilità. Solo allora potremo mostrare, facendo leva su alcune implicazioni del "problema della fermata", che le macchine eseguono necessariamente delle operazioni impossibili e che dunque, per via di ciò, si potrebbe dire che siano capaci di creatività reale, nonché di creare le proprie regole.

3. La creatività reale. Per un approccio oggettivistico alla questione della creatività computazionale

3.1. Dal programmatore al programma

3.1.1 Questioni di metodo e definizioni generali

Nei capitoli precedenti abbiamo visto come nel sondare le potenzialità dei sistemi artificiali sia i ricercatori di IA che di CC non amino partire da definizioni generali di aspetti umani quali il pensiero, l'intelligenza e la creatività, confidando che queste potranno essere, in futuro, direttamente illuminate da una ricerca condotta sul campo, preliminarmente fondata su assunti ragionevoli. Per quanto riguarda il presente contributo, lungi dal contestare l'importanza di un approccio sperimentale, si ritiene altrettanto importante che i così detti assunti ragionevoli, e in particolare le definizioni preliminarmente accettate di creatività, si riferiscano direttamente agli oggetti (in questo caso i sistemi artificiali) a cui la ricerca è rivolta, piuttosto che agli effetti (meraviglia, sorpresa, incredulità) che la ricerca può generare in chi la conduce – sia esso un soggetto concreto, in carne e ossa, o immaginario (una società intesa come insieme di idee condivise su che cosa sia la creatività). Il motivo è presto detto. Si crede qui fermamente che la ricerca negli ambiti IA e CC possa davvero illuminare questi fenomeni – così come lo hanno sempre fatto, a loro modo, la psicoanalisi, la psicologia, l'antropologia e la sociologia. Ma proprio in quanto si crede ciò, si dovrà anche ritenere che il miglior modo con cui tali ricerche potranno mai contribuire alla causa sarà partendo dai loro stessi oggetti, per poi individuare da qui aspetti, operazioni o risultati conformi alla definizione generale, extra-disciplinare, del fenomeno indagato, piuttosto che indagare lo stesso fenomeno a partire dagli *effetti* che il proprio oggetto può generare in soggetti come l'uomo o la società. Quest'ultima indagine, infatti, potrà anche essere utile ai fini di un'inchiesta psico-sociologica delle aspettative – e degli eventuali *bias* cognitivi a esse associate – relativamente all'idea di creatività o a quella delle reali potenzialità di un sistema artificiale, ma non dice nulla attorno a queste stesse potenzialità e non contribuisce quindi ad un aumento del livello di conoscenze tecnico-specifiche in IA e CC. Per condurre una ricerca sui sistemi artificiali, mantenendo il focus sulla questione della creatività, bisognerà dunque avere, innanzitutto,

una definizione che sia il più possibile generale e oggettiva di questa proprietà. Una volta stabilita, si tratterà poi di verificare se vi siano aspetti, operazioni o risultati delle macchine, o di qualche macchina in particolare, conformi alla definizione. Tale verifica può avvenire in due modi, a seconda che il tipo di ricerca che si è scelto di svolgere sia empirico-induttivo o logico-deduttivo. Nel primo caso si procederà prendendo in considerazione particolari sistemi – come potrebbe essere un programma per il gioco degli scacchi – per valutare la conformità delle sue operazioni – in questo caso delle sue mosse – alla definizione di creatività. Nel secondo si tratterà invece di valutare se le macchine di Turing – ovvero gli attuali computer – possano in generale eseguire operazioni conformi alla definizione data.

Quello che si cercherà di fare da qui in avanti sarà proporre un inizio di possibile ricerca del secondo tipo. Procediamo subito con l'individuazione di una definizione di creatività che sia il più possibile generale e oggettiva. Un buon punto di partenza potrebbe essere identificare il *tipo* di azioni a cui sarebbe possibile attribuire questa proprietà. Rimanendo il più possibile generali e oggettivi va da sé che dovremmo definire tali azioni "creazioni". Che cos'è, dunque, una creazione? Qualcosa di nuovo, certo. Ma a questo punto – ed è qui che iniziano i problemi – bisognerà anche chiedersi: è possibile che le macchine creino *in tal senso*, che producano cioè dei risultati *nuovi*? Se esse agiscono sempre, per la tesi di Lovelace, sulla base delle regole di cui dispongono, la risposta alla domanda non può che essere negativa, poiché quelle regole produrranno sempre, relativamente a un dato problema, gli stessi risultati, o quanto meno risultati dello stesso genere di quelli già prodotti, in passato, in base a quelle regole. Tuttavia, anziché concludere, alla luce di ciò, che le macchine non sono creative, potremmo invece includere la nozione di *impossibilità* nella nostra definizione, aprendo da qui, però, la questione di capire in che modo le macchine, pur potendo agire solamente in base alle proprie regole, possano produrre risultati *formalmente* impossibili ed essere creative nel senso stabilito. Ricaviamo dunque dalla risposta negativa alla domanda la seguente *definizione positiva, generale e oggettiva* di creatività computazionale: "creativo è qualsiasi atto o prodotto che,

in rapporto alle regole del sistema considerato, risulta impossibile”.⁸ Proponiamo ora di considerare alcune implicazioni del teorema (o del problema) della fermata come argomenti utili a provare che le macchine sono capaci di atti o risultati conformi alla definizione.

3.1.2 Introduzione storica al teorema della fermata

Il teorema della fermata⁹ fu dimostrato e presentato per la prima volta da Turing nel 1936 in un articolo intitolato *On computable numbers, with an application to the Entscheidungsproblem*.¹⁰ L’obiettivo principale di Turing era fornire, come recita il sottotitolo, una risposta al problema matematico della decisione, presentato da David Hilbert in un primo momento nel 1900, a Parigi, nel corso del Secondo Congresso Internazionale dei Matematici, per essere poi riportato alla ribalta, nel 1928, sempre da Hilbert, nello stesso congresso tenuto a Bologna, in quanto dalla sua prima formulazione nessun matematico lo avevo ancora risolto. Il problema consisteva nel capire se vi fosse un «metodo sistematico per dire, di ogni proposizione matematica, se la proposizione è dimostrabile o meno nel sistema» (Copeland 2004, 47). Nel suo articolo Turing provò, in buona sostanza, che questo metodo non esiste, o meglio che non esiste un modo per decidere se, dato un problema matematico a un qualsiasi sistema di calcolo, il sistema lo risolverà o meno. Già nel ’31 Kurt Gödel provò che dall’interno di qualsiasi sistema formale coerente e abbastanza potente da contenere l’aritmetica è impossibile dimostrare, ad esempio, una proposizione che dice di se stessa di non essere dimostrabile nel sistema. Infatti, se si dimostra la proposizione g “ g non è dimostrabile nel sistema S ” si ottiene che è vero che g non è dimostrabile in S e così una contraddizione. A questa idea,

8. Una definizione simile è stata fornita anche da Boden senza essere tuttavia sviluppata (Dartnall *et al.* 1994, 6; Boden 2004, 6). L’intento stesso di Boden in *The Creative Mind* è, fra gli altri, quello di calmierare in un certo qual modo il paradosso che la definizione implica, anziché assumerlo e svilupparne le implicazioni (13).

9. Per un approfondimento su come Turing arrivò ai risultati presentati nel ’36 si veda (Hodges 2012; Copeland 2004)

10. È in realtà Martin Davis il primo a nominare la dimostrazione di indecidibilità di Turing con l’espressione “problema della fermata” (halting problem) (Davis 1958). Per una ricostruzione storica delle vicende che portarono a definire in questi termini il problema esposto da Turing nel ’36 (Lucas 2021).

da cui deriva che un sistema non è in grado di dimostrare *tutte* le sue verità – e che dunque, per rimanere *coerente*, dovrà essere *incompleto* – Turing aggiunse che non è nemmeno possibile *decidere* se lo stesso sistema dimostrerà *una* qualsiasi delle sue verità.

Ora, la connessione esistente fra questo risultato, valido per la logica-matematica in generale, e la particolarità del teorema della fermata – da cui emerge, fra le altre cose, la validità del risultato per l'informatica – la si può capire alla luce del modo con cui Turing affrontò l'*Entscheidungsproblem*. Dal momento che il metodo richiesto da Hilbert per decidere la dimostrabilità di una proposizione matematica in un sistema doveva essere a sua volta un sistema di calcolo, Turing immaginò una macchina che potesse riprodurre le operazioni compiute da un calcolatore umano. Per fare ciò era necessario soddisfare una serie di condizioni. Innanzitutto, la macchina doveva poter operare con un certo insieme di simboli su un nastro diviso in celle (l'equivalente di una riga di un foglio a quadretti). A tal fine sarebbe stato necessario fornire alla macchina un *alfabeto* – composto dai simboli che le si vorrà far utilizzare – e una *testina* per operare sul nastro leggendo, scrivendo o cancellando tali simboli. Ciascuna cella del nastro avrebbe dovuto contenere un solo simbolo alla volta e la testina si sarebbe dovuta collocare, in ciascuna fase del calcolo, a livello di una singola cella, in modo che per svolgere le successive operazioni su altre celle – come scrivere un risultato dopo una lettura, o leggere un nuovo simbolo dopo averne scritto uno – avrebbe dovuto spostarsi o a destra o a sinistra, salvo trovarsi nella condizione di dover cancellare il simbolo della cella in cui si trova per sostituirlo con un altro. Oltre a ciò, dal momento che l'esecuzione di un calcolo richiede di tenere presenti le operazioni già eseguite e i simboli osservati in precedenza, bisognerà far sì che la macchina possa assumere uno alla volta – uno per ogni nuova operazione svolta durante il calcolo – un certo numero di *stati interni*, corrispondenti agli *stati mentali* che può assumere di fatto un uomo allorché pensa un nuovo contenuto oltre a quelli che sta tenendo a mente mentre calcola. In questo modo si potrà dire che la *configurazione* data da uno stato interno e un nuovo simbolo determinerà il possibile comportamento successivo della macchina nel corso del

calcolo. Soddisfatte tutte queste condizioni,¹¹ dunque, le macchine di Turing (d'ora in avanti MT) diventerebbero capaci di svolgere qualsiasi calcolo. Non solo: se programmate in modo adeguato lo svolgerebbero meccanicamente e senza commettere errori, così come una calcolatrice ben programmata risponderà "4" se si digita sulla sua tastiera (equivalente del nastro pensato da Turing) il problema "2+2=". A questo punto non rimaneva che chiedersi: può una MT risolvere il problema di Hilbert? A rigore, essendo una MT un sistema di calcolo, ed essendo che Hilbert chiedeva di fornire un metodo per decidere, di ogni proposizione matematica, se la proposizione sarebbe stata dimostrabile o meno in un sistema di questo tipo, ne derivava che per soddisfare la richiesta una MT¹ doveva essere in grado di calcolare se un'altra MT² avrebbe dimostrato o meno una proposizione P: se datole cioè in ingresso un qualsiasi input – per esprimersi in termini più informatici – avrebbe terminato o meno il calcolo. Per mettere MT¹ nelle condizioni di farlo Turing mostrò, in primo luogo, come l'insieme delle informazioni contenute in MT² – relative alle operazioni che essa può svolgere allorché dotata delle caratteristiche di cui sopra – una volta scritte avrebbero potuto formare una «tabella di comportamento» di dimensioni finite, tale da esprimere, fra le altre cose, il comportamento di MT² con input P. Ciò significava che era possibile inserire la tabella di MT² in un elenco interno a MT¹ – opportunamente tradotta, ovviamente, con l'alfabeto di cui la macchina dispone –, attribuirle un «numero di descrizione» (ad esempio 1) tale da poterla distinguere da altre tabelle inserite nell'elenco e così far sì che riportandone il numero sul suo nastro MT¹ potesse decodificarlo nella tabella corrispondente a MT² e infine eseguirlo. In altre parole, con tale sistema era possibile far sì che MT¹ simulasse il comportamento di MT² e di qualsiasi altra macchina traducibile nei simboli del suo alfabeto. Era la prova che le macchine di Turing fossero *macchine universali* (d'ora in avanti MTU). Ma a questo punto ecco il colpo di scena. Una volta stabilito che una MT poteva simulare qualsiasi altra MT diventava possibile chiedersi: la MTU così definita può decidere se, dando alla MT che essa è in grado di simulare un qualsiasi input P, tale MT terminerà il suo calcolo o meno?

11. Per un elenco completo e al tempo stesso sintetico delle caratteristiche di una MT si veda (Frixione *et al.* 2011, 51-53).

La risposta era no. Poteva effettivamente svolgere il calcolo della MT considerata, ma non poteva stabilire in anticipo se tale MT l'avrebbe risolto. Per farlo, infatti, avrebbe dovuto essa stessa risolverlo al suo posto. Come avrebbe potuto, del resto, prima di ciò? Com'è possibile, in generale, stabilire se un insieme di regole – cosa che del resto è una MT – sono adatte a risolvere un problema, ovvero ad arrivare a una soluzione in un numero finito di passi, se non già tentando di risolverlo applicando quelle regole? Avventurarsi in questo tentativo non escluderebbe, ovviamente, che sia possibile risolvere il problema, ma precisamente che sia possibile decidere, prima di risolverlo, se sia risolvibile o meno, anche perché se si suppone che prima di arrivare alla soluzione si avrà capito, come in anticipo, con quanti passi si risolverà il problema, quello definirà esattamente l'istante in cui lo si avrà risolto. Turing darà una dimostrazione formale dell'irrisolvibilità del problema della fermata, provando conseguentemente che il problema di Hilbert è indecidibile, cioè che non esiste un metodo sistematico per decidere se è possibile risolvere una proposizione matematica dall'interno di un qualsiasi sistema programmato per farlo.

3.1.3 Implicazioni filosofiche del problema della fermata

Ora, ai fini di questo contributo non sarà necessario ripercorrere la dimostrazione. Dobbiamo capire, invece, quali siano le implicazioni del problema della fermata tali da dimostrare che le operazioni svolte da qualsiasi MT – cioè da qualsiasi computer – siano conformi alla definizione di creatività che abbiamo dato. Consideriamo a tal fine un caso particolare. Tra le varie operazioni che una MT può svolgere Turing mostra che vi è anche quella di simulare se stessa – dal momento che anche le sue informazioni sono di fatto convertibili in una tabella che essa può codificare ed eseguire (tale possibilità è esaminata da Turing nella sezione 8 del suo articolo). Ciò significa che, per gli stessi motivi di cui sopra, non solo una particolare MT non può calcolare se un'altra MT terminerà o meno un calcolo... non lo può calcolare nemmeno di se stessa. Ma dal momento che qualsiasi impossibilità di una MT è il corrispettivo di una regola di cui la MT non dispone, il fatto che una particolare MT non possa calcolare se è in grado di risolvere un calcolo rivela che essa *non dispone delle regole per risolvere quel calcolo con le regole di cui dispone*, e che dunque se alla fine lo

risolverà con queste regole eseguirà un'operazione almeno *in parte* impossibile.

Esistono vari modi per esprimere questo problema. Proviamo allora a prenderne in considerazione un altro. Bisogna chiarire innanzitutto che non stiamo affermando che una MT non sia programmata per risolvere problemi. Essa, infatti, dispone di una serie di regole per farlo. Tuttavia, essa non dispone delle regole che le consentono di calcolare se le regole di cui dispone siano adeguate per farlo, ed è questo, in effetti, il significato profondo di “non sa dire se terminerà/risolverà o meno un calcolo con le regole di cui dispone”. Se avesse quelle regole, infatti, essa potrebbe risolvere il problema della fermata per i calcoli da fare con le regole di cui dispone, salvo che il problema si ripresenterebbe di nuovo per le regole con le quali è in grado di stabilire se le regole di cui dispone la porteranno alla risoluzione/terminazione del calcolo. E così via all'infinito. Ciò confermerebbe, di nuovo, che il problema della fermata non è risolvibile, perché non sapremo mai in anticipo se, a un certo punto, troveremo una regola per la quale tutte le regole di cui la MT dispone – pensate per risolvere un calcolo qualsiasi – la porteranno alla risoluzione/terminazione del calcolo. Questo perché, come si è detto, se e quando troveremo quest'ultima regola fondativa di tutte le altre a quel punto avremo *già* risolto il calcolo, ma di nuovo senza una regola per stabilire se la regola fondativa (e così tutte le altre) ci porterà o meno alla risoluzione/terminazione del calcolo stesso, pur avendolo in effetti risolto. In questo caso si potrà quindi dire che il calcolo è sì *terminato in avanti*, ma che *non* è ancora *terminato all'indietro*, cioè che il calcolo è *terminato* e al tempo stesso *non è terminato*, e che dunque, di nuovo, è indecidibile il problema se, data a una MT una qualsiasi proposizione P, MT terminerà o meno il calcolo. Generalizzando questi risultati è inoltre possibile giungere a ulteriori conclusioni. Abbiamo visto che una MT *non dispone delle regole per risolvere un calcolo con le regole di cui dispone*. Questo modo di porre la questione consente di distinguere due diversi livelli di regole presupposte dalle operazioni di una MT: regole relative al tipo e al numero finito di passi da compiere per la risoluzione di un problema (quelle di cui dispone), che potremmo definire euristiche, o *applicative*, e regole relative alla possibilità di risolverlo in quel tipo e numero finito di passi (quelle di cui non dispone), che potremmo perciò definire *condizionali*. Ora, si potrebbe dire che una MT le seconde non le abbia

per il semplice fatto che essa è pur sempre un insieme finito di regole e che, dunque, per definizione essa non possa essere una serie infinita di livelli di regole in cui in ognuno di essi – partendo da quello più in alto – vi siano regole relative alla possibilità di ricorrere a quelle dei livelli più in basso per la risoluzione di un problema. Queste *regole di regole* saranno al limite presenti, oltre un certo livello, nella mente di colui che in base a esse ha programmato la MT che si considera.

Ciò non significa, tuttavia, che la MT creata in questo modo non sia a sua volta creativa. In parte lo abbiamo già mostrato. Abbiamo visto, in particolare, che se una MT non ha le regole a cui attenersi per risolvere un problema con le regole di cui dispone, vuol dire che già solo per il fatto di *iniziare* a risolverlo con *queste* regole essa fa qualcosa che potrebbe fare solamente se avesse (anche) *quelle* regole. Dal momento che invece quelle non le ha, ne consegue che in tal caso essa fa qualcosa di impossibile e dunque di creativo in base alla definizione che abbiamo dato. Con tale argomento, però, la creatività di una MT è dimostrata solamente per *via negativa*, ovvero a partire dalla *mancaza di regole*. Proviamo allora a dimostrarla per *via positiva*. Si è detto che, al limite, solo un soggetto esterno potrebbe avere *in mente*, o *nella tabella* – se ci riferiamo a un'altra MT – le regole in base alle quali ha creato e fornito a una particolare MT le regole di cui dispone. Evitando momentaneamente di stabilire se tale soggetto sia realmente creativo chiediamoci piuttosto se possa esserlo la MT che egli stesso ha programmato. Chiarito questo problema, infatti, si chiarirà anche quello. A prima vista verrebbe da dire che la MT in questione si limiti a eseguire le operazioni per le quali è stata programmata. Questa asserzione coincide perfettamente con la tesi di Lovelace. In base a essa si ricaverebbe facilmente che tale MT non possa essere creativa. Ma si è visto che qualsiasi MT non dispone delle regole relative alla possibilità di risolvere un problema con le regole di cui dispone, e siccome per la tesi di Lovelace – che prima si assumeva come vera – una MT agisce solamente in base alle regole di cui dispone, se poi pur non avendo le regole che le renderanno possibile risolvere un problema con le regole di cui dispone essa alla fine risolverà comunque il problema in questo modo vorrà dire che quelle regole di cui non disponeva avrà dovuto in qualche modo *crearsele*. Chiariamo subito che non ci stiamo spingendo a dire che le avrà effettivamente introdotte nel suo programma e che in base a ciò sarà possibile trovare in esso altre regole – espresse

nel medesimo linguaggio formale – oltre a quelle già inserite dal suo programmatore. Sostenere questo, potremmo aggiungere, significa affermare che una MT è *surrealmente* creativa.¹²

Quello che stiamo sostenendo non è nemmeno che questa *creazione* sia un'operazione formalmente distinta da quella che una MT può eseguire in base alle regole di cui dispone, quanto una sua *necessità interna*. Più precisamente, stiamo sostenendo che forse quando una MT opera in questo modo essa si sta anche *virtualmente* dando le regole per farlo: sta creando, attraverso un *atto istituyente* – cioè senza basarsi su altre regole date – le regole per operare in base a quelle di cui dispone. Lo implica il fatto stesso che essa agisca in base a queste ultime, come se i due atti – l'*applicazione di regole date* e la *creazione di regole nuove* – non fossero altro che i due lati indiscernibili – *attuale* e *virtuale* – di un solo e unico evento. Questo evento implica una creazione che è *reale* per il fatto che non si basa, almeno in parte, sulle regole di cui la MT dispone – motivo per cui sarebbe anche *impossibile*. Che la MT compia tale creazione malgrado sia per essa formalmente impossibile è suggerito da una *necessità paradossalmente interna* al fatto che, come osservava Lovelace, essa compie solo atti che sono per essa formalmente possibili. Questa creazione incide dunque *dall'interno*, in maniera *obliqua*, su questi atti, rendendoli possibili *nel mentre stesso in cui accadono*, non prima – come invece accadrebbe nel caso della creatività trasformativa di Boden. Ma se tale possibilità non viene prima dell'atto che la *realizza* allora essa è sempre già reale in quanto

12. Qualcuno potrebbe obiettare, forse, che le più recenti MT possono creare in un certo qual modo nuove regole – per via del *machine learning* – partendo dai dati che utilizzano. Tuttavia, esse lo fanno (se lo fanno) in quanto sono state programmate per farlo, cioè in base alle regole di cui dispongono (motivo per cui, fra l'altro, non aggiungerebbero realmente delle regole nuove). Ciò a cui qui ci stiamo riferendo è invece una *creazione impossibile* in rapporto alle regole di cui una MT dispone, oppure, in altri termini, una *creazione reale*. La differenza fra questa e una *creazione surreale* starebbe nel fatto che per la seconda una regola la cui creazione non implica il ricorso a regole che la macchina possiede si dev'essere necessariamente scritta all'interno del programma. Dal punto di vista della creazione reale, a fronte del fatto che si sia creata una regola realmente nuova, il fatto poi della sua scrittura nel programma non discende da alcuna necessità logica.

necessità interna all'atto. Essa è, in definitiva, non “il possibile” ma ciò che del reale rimane *impossibile*, in una parola il *virtuale*.¹³

3.2. Ritorno al programmatore

Per via di questa creazione una MT darebbe a se stessa, insomma, le regole delle regole di cui dispone. In tal senso ci si potrebbe anche chiedere se queste metaregole siano identiche, o quanto meno simili, a quelle in base a cui il programmatore pensò le regole interne a MT. Da un punto di vista logico nulla esclude, di fatto, che possa sussistere fra tali due differenti gruppi di metaregole un rapporto di identità o somiglianza. Ma allo stesso modo è anche vero che non vi è nulla che lo certifichi, dal momento che non disponendo delle metaregole del programmatore – essendo queste presenti solamente, al limite, nella mente di esso – la MT non avrebbe potuto di fatto creare le proprie a immagine e somiglianza di quelle, così che tanto l'identità quanto la somiglianza fra le due, eventualmente rilevate a posteriori, si rivelerebbero al tempo stesso delle mere coincidenze. Vi sarebbe piuttosto un altro tipo di analogia, più interessante di questa, rinvenibile non più sul piano del contenuto bensì della *forma* di entrambi gli insiemi di metaregole. Premettiamo che tutte le metaregole del programmatore, come del resto tutte quelle esistenti nella dimensione reale della MT, potranno essere, in base a quanto detto, di due diverse tipologie: metaregole che – come le regole che rendono possibili – sono a loro volta possibili in base ad altre regole; e metaregole che, pur rendendo possibili delle regole e pur agendo simultaneamente alle metaregole del primo tipo, non sono a loro volta possibili – al contrario di queste – partendo da altre regole. Le prime metaregole sono quelle che il soggetto conosce e a cui può ricorrere più o meno consapevolmente per operare, in date situazioni, in base a regole di ordine inferiore. Si pensi ad esempio al caso della *lingua*, cioè un insieme generale di regole reso possibile da altre regole (convenzioni) e a cui un soggetto potrebbe, conoscendolo, attingere, per formulare frasi o segnalare alcune regole *inconsuete* e tuttavia perfettamente possibili dall'interno di esso. Per quanto invece

13. Sulla questione del “possibile” come categoria che non precede logicamente il “reale” si veda (Bergson 2014). Si veda poi (Deleuze 2001) per l'approfondimento della questione bergsoniana e del complicato rapporto fra virtuale e attuale.

concerne le seconde metaregole non si potrà dire che il soggetto le conosca e che potrà ricorrervi allo stesso modo con cui ricorre alle prime. Assumiamo allora, anche al fine di comprenderle meglio, che le regole in base a cui il programmatore sopra menzionato pensò le regole di cui la MT disponeva fossero dello stesso tipo di quelle che la MT creò *virtualmente* operando con le sue regole, ovvero metaregole del secondo tipo. In questo caso le metaregole del programmatore non potranno essere state delle regole a cui egli avrà potuto fare ricorso ma che avrà dovuto *creare* per eseguire le operazioni possibili in base alle regole di cui disponeva, con la sola differenza di aver trasferito queste ultime dalla sua dimensione a quella della MT per farle svolgere le suddette operazioni al posto suo. Ma a questo punto, a ben vedere, siamo anche in grado di svincolarci dalla dimensione della MT, per pensare la stessa questione, ovvero le condizioni della creatività reale, all'interno della dimensione umana. Questo ci dà il vantaggio di poter accedere a un punto di vista maggiormente descrittivo su ciò la cui esistenza abbiamo fin ora inferito astrattamente su basi logiche. Chiediamoci allora quali tipi di atti umani potrebbero essere ritenuti conformi alla definizione di creatività reale. Si prenda a tal fine, di nuovo, il caso della lingua. Grazie a essa è possibile eseguire una serie di operazioni come leggere, scrivere o parlare, trattando con un numero finito di *segni*. Per leggere, ad esempio, devo guardare un testo e decodificarne i segni. Se riesco a fare questa decodifica vuol dire che essi, oltre che segni, sono *simboli* di una lingua che conosco. Ma affinché avvenga questa decodifica, affinché io possa applicare le regole della lingua ai segni che vedo, devo anche simultaneamente poter trattare quei segni come se fossero casi particolari dei simboli che conosco, devo creare cioè una regola, nel mentre stesso della decodifica, che non può essere a sua volta presente nell'insieme delle regole della lingua che conosco e per la quale stabilisco questa stessa cosa – ovvero che una serie di segni che vedo corrisponde a una serie di simboli che conosco – al fine di rendermi possibile la lettura del testo. Questo oscuro lavoro avverrebbe in quella dimensione interna e assieme obliqua al *simbolico* che è l'*immaginario*. Esso implica ciò che Castoriadis definì nei termini di *un quid pro quo* interno al «*linguaggio* inteso come codice» (2022, 371, 381-399): una serie di operazioni con cui – nel mentre stesso della decodifica e senza ricorrere a ulteriori regole interne alla lingua che la renderebbero possibile – si prende *una cosa per un'altra* – un segno

qualsiasi, in questo caso, per un simbolo – in maniera tale che le regole della suddetta lingua diventino disponibili per operazioni come leggere, scrivere o parlare.

Tali micro operazioni riproporrebbero inoltre, in ogni singolo passaggio di livello che è possibile fare fra quelli che definiscono le regole di un sistema simbolico, l'*operazione fondamentale* che l'ha reso possibile all'inizio, ovvero l'insieme delle decisioni con cui si è potuto stabilire, ad esempio, che la sequenza di segni "c-a-n-e" rinverrà sempre all'immagine di un certo animale a quattro zampe e non a quella di un albero, e che prese di per sé non avrebbero alcun senso – proprio in quanto cominceranno ad averne uno, assieme alla serie infinita di mondi che renderanno possibili, solo una volta che verranno prese –, o non più di quanto potrebbero averne le associazioni con cui si è abituato a convivere uno schizofrenico.¹⁴ In entrambi i casi, al fine di fissare un accordo di qualche tipo tra il proprio mondo e i vincoli esterni, tra l'esigenza di esprimersi, ad esempio, e l'impossibilità di riuscirci senza un linguaggio, si verrà meno in qualche modo alla ragione, si dirà senza seguire alcuna logica che *qualcosa significa qualcos'altro*.

Conclusione. Si può dire che i computer immaginano?

Qualcuno potrebbe anche pensare, forse a ragione, che questa forma di creatività non sia nulla di *speciale* in quanto fiorisce pur sempre nell'alveo del compromesso con una serie di regole date. A patto di intendersi sul fatto che ciò che da quell'alveo così fiorisce, per quanto compromesso, non poteva essere in alcun modo deducibile da ciò che nell'alveo era già dato, e che dunque è, una volta fiorito, del tutto *nuovo e realmente creativo*, non faremo nulla per convincervi del contrario. D'altra parte, non era di nostro interesse trattare qui di casi straordinari di creatività – anche perché questi rientrerebbero probabilmente nell'insieme della creatività apparente (non perché non sarebbero creativi o straordinari, ma letteralmente perché, in tal caso, sarebbero così per qualcuno e non in quanto tali). Quello che si è cercato di fare, analizzando la

14. Per un approfondimento dei casi di schizofrenia in quanto rivelativi di una forma di creatività che accomuna l'uomo e la macchina si veda *infra* Cardone L. (2023), *Immagini schizofreniche. Riflessioni fenomenologiche sulla creatività computazionale a partire da Marc Richir*.

creatività al cuore del compromesso con l'insieme delle regole date, è stato piuttosto mostrare, da una parte, quelle che si potrebbero definire le *condizioni minime* della creatività reale, e dall'altra come vi sia un *marginale minimo* di creatività reale in qualunque atto reso possibile da regole già date. Questo atto è in tal senso sempre già spaccato in due metà asimmetriche dallo slancio circolare con cui una vita rifà se stessa nel medesimo istante in cui è chiamata a ri-farsi carico di regole stabilite. Non ci è nemmeno sfiorata l'idea, inoltre, di trattare da un punto di vista *descrittivo* la creatività reale nelle macchine di Turing – ovvero nei computer per come ancora oggi li conosciamo. Ci piace sicuramente pensare che vi sia in essi qualcosa come il corrispettivo di ciò che nell'uomo chiamiamo immaginario. Ma questo, appunto, non è altro che un pensiero in libertà. Siamo persuasi, invece, che per quanto fin qui sostenuto vi debba essere in essi una sorta di dimensione che quanto meno *affianca* – se non *fonda* – quella simbolica all'interno di ognuna delle operazioni che eseguono. Ritornando sulla sponda dell'umano, verrebbe da pensare a quei frangenti in cui, quando si calcola a mente, ognuno di noi crea tutto un mondo di immagini e linee che fanno da supporto a una serie di operazioni formali. Ecco: lì, in quel preciso istante, facciamo qualcosa di incomunicabile sul piano simbolico e che, malgrado ciò, accompagna la manipolazione di simboli funzionale alla risoluzione di un calcolo. Qualcuno potrebbe aver inventato dei percorsi tortuosi, complicatissimi, che gli farebbero fare dei giri infiniti prima di giungere a una soluzione. Percorsi inutili, dunque, ma indubbiamente unici nel loro genere e magari interessanti. Altri potrebbero riprodurre nella mente gli stessi schemi o disegni imparati la prima volta a scuola per fare le operazioni sul quaderno, e anche in questi casi ci vorrà comunque creatività – ma appunto una creatività di cui ognuno di noi è dotato in partenza, e che per questo motivo potrebbe anche non soddisfare qualcuno – per riprodurre nella mente cose di per sé adeguate a essere scritte su un foglio di carta. Ma la domanda è: cosa fa un computer in questi casi? Cosa fa, cioè, *mentre* calcola?

Si parla tanto della componente *software* e della sempre maggiore complessità dei nuovi programmi in circolazione. Si parla tanto, anche, della componente *hardware*, e con essa delle analogie con il funzionamento del cervello umano. Noi ci chiediamo: è possibile parlare – se è sostenibile che essi siano capaci di creare nel senso fin qui descritto –

di un immaginario e di un'immaginazione propri dei computer? Per rispondere a una domanda come questa – ma anche a quella, in fondo, relativa alla creatività computazionale – bisognerebbe forse essere in grado, fra le altre cose, di descrivere il corrispettivo di quei mondi di immagini e linee degli umani a livello dei computer. Ma su questo sarà necessario lasciare il passo a un'indagine maggiormente descrittiva dei processi computazionali. Un'indagine per la quale potranno non essere sufficienti, effettivamente, le definizioni generali e oggettive, ma che di esse potrebbe aver bisogno, appunto, per riconoscere le condizioni minime di qualsiasi operazione creativa. Potrebbe tornare utile, da questo punto di vista, riprendere in considerazione la questione della creatività apparente, cioè quei casi di computer che fanno cose sorprendenti dal punto di vista di un umano che li osserva, in quanto potrebbero essere il portale di accesso a un modo di creare e di immaginare ancora sconosciuto all'umano – per quanto forse idealmente spiegabile, dal nostro punto di vista, con l'idea di creatività reale.

Riferimenti bibliografici

Arieti, S. 1986, *Creatività. La sintesi magica*, Il pensiero Scientifico Editore, Roma.

Beaudot A. 1977, *La creatività*, Loescher, Torino.

Bergson H. 1984, *Il possibile e il reale*, trad. di F. Sossi, in "Aut-Aut", 204, pp. 3-13.

Boden M. A. 2004, *The Creative Mind. Mith and mechanisms*, Routledge, Londra-New York.

Boden M. A. 1987, *Artificial intelligence and natural man*, The MIT Press, London.

Cantone D. 2022, *La filosofia degli automi. Origini dell'intelligenza artificiale*, Mimesis, Milano-Udine.

Castoriadis C. 2022, *L'istituzione immaginaria della società*, trad. di E. Profumi, Mimesis Edizioni, Milano-Udine.

Copeland J. 2004, *The Essential Turing. Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life plus The Secrets of Enigma*, Oxford University Press, Oxford.

Dartnall T. 1994, *Artificial Intelligence and Creativity. An Interdisciplinary Approach*, Springer Science+Business Media Dordrecht, Brisbane.

Davis M. 1958, *Computability and Unsolvability*, MacGraw-Hill Book Company, New York.

Deleuze G. 2001, *L'immagine-tempo. Cinema 2*, trad. di L. Rampello, Ubulibri, Milano.

Frixione M. & Palladino D. 2017, *La computabilità: algoritmi, logica, calcolatori*, Carocci Editore, Roma.

Gangemi A., Miceli S., Sprini G. 2003, *L'intelligenza. Teorie e modelli*, Laterza, Roma-Bari.

Gödel K. 1931, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, «Monatshefte für Mathematik und Physik», 38, pp. 173-198.

Guilford J. P. 1950, *Creativity*, *American Psychologist*, 5, (9), pp. 444-454.

Hodges A. 2012, *Alan Turing: The Enigma*, Princeton University Press).

Lucas S. 2021, *The origins of the halting problem*, "Journal of Logical and Algebraic Methods in Programming", 121,100687.

Menabrea L. F., Lovelace A. A. 1842, *The Analytical Engine. Invented by Charles Babbage, With notes upon the Memoir by the Translator Ada Lovelace, Countess of Lovelace*, Bibliotheque Universelle de Geneve, 82, October.

McCarthy J., Minsky M. L., Rochester N., Shannon C. E. 2006, *Proposta di un Progetto di ricerca estivo sull'intelligenza artificiale presso il*

Dartmouth College, "Sistemi intelligenti", XVIII, 3, pp .413-428.

Trombetta C. 1990, *La creatività. Un'utopia contemporanea*, Bompiani, Milano.

Turing A. M. 1950, *Computing machinery and Intelligence*, "Mind", 59, pp. 433-460.

Turing A. M. 1936, "On computables numbers, with an application to the Entscheidungsproblem", Princeton University Press, Princeton.

Veale T., Almicar Cardoso F. (2019), *Computational Creativity. The Philosophy and Engineering of Autonomously Creative Systems*, Springer, Cham Switzerland.

Wallas G. 2014, *The art of thoughts*, Solis Press, Kent England.

Michele Pavan
Università degli Studi di Verona
michele.pavan@univr.it



ἘΠΈΚΕΙΝΑ

International Journal of Ontology
History and Critics

CLAUDIO GNOFFO E MARCO PIRRONE

Mary Shelley's Frankenstein videogame among transmedia franchises: creativity and transcreation merged with interactive non-linear narrative

EPEKEINA, vol. 16, nn. 1-2 (2023), pp. 1-29

Philosophy of Technology

ISSN: 2281-3209

DOI: 10.7408/epkn.

Published on-line by:

CRF – CENTRO INTERNAZIONALE PER LA RICERCA FILOSOFICA
PALERMO (ITALY)

www.ricercafilosofica.it/epekeina



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs 3.0 Unported License.

Mary Shelley's Frankenstein* videogame among transmedia franchises: creativity and transcreation merged with interactive non-linear narrative

Claudio Gnoffo e Marco Pirrone

Introduction

The paper aims to analyze the concept of creativeness through digital media contents. In particular, the study investigates the birth and the development of the disruptive phenomenon of transmedia storytelling, by presenting a specific case study.

In this context, the film adaptation *Mary Shelley's Frankenstein* represents one of the most relevant examples of a transmedia franchise: it is responsible for the creation of a narrative universe built around the novel.

As new stories and characters have been created and introduced over time in the Frankenstein narrative universe, the paper then analyzes the case of several versions of a videogame inspired by Mary Shelley's masterpiece.

The videogame, titled *Mary Shelley's Frankenstein*, like the film, was developed for the SuperNintendo Entertainment system/SuperFamicom, Sega Mega Drive/Genesis and Sega CD, and was released in 1994.

Since the 16-bit graphics and the main hardware features of '90 home consoles were sufficient to create plausible virtual scenarios, both programmers and translators/localizers proposed a peculiar narrative style, based on action-gameplay sequences and in-game texts.

The paper focuses on how the user/player experiences the story of the game, and how the same gameplay experience can provide insight into the original story of the novel, as well as a broader understanding of the Frankenstein narrative universe.

* The tasks of this work were divided as follows: Claudio Gnoffo was responsible for writing the entire section 1 and sections 2;2.1; 4; 4.1. Marco Pirrone composed sections 2.2;2.3 4.2, and the entire section 3.

In particular, a comparative analysis between the gameplay and the reading experience is conducted, including a consideration of the differences between the versions of the videogame.

Moreover, the study outlines the main issues and challenges about interactive digital storytelling and videogame localization, particularly in relation to creativity and accessibility themes.

Finally, the paper investigates how authorship is managed in the area of transmedia franchises, also describing how fans can creatively contribute to the development and support of a digital product.

The results of this study would be valuable to many researchers and practitioners, in order to investigate possible future perspectives.

1. From Crossmediality to Transmedia Franchise

For several years, we have been witnessing an ongoing process of media convergence, which is intensifying the creation of narrative hybridizations in different media.

In this era of convergence, new emerging audiovisual narratives develop differently in various media, where the narrator becomes a hybrid figure, i.e., a player-user who not only benefits from the narrative but also interacts with it, especially in adaptations of stories such as board games, pinball and, of course, videogames (Montero & Duarte 2011). We refer to them as franchises because these multimedia products primarily respond to economic interests, but are also motivated by the artistic intention of enriching the stories told and the experience of the audience (Belsunces 2011).

When approaching these new audiovisual narratives, involving multiple media, it is necessary to define key concepts.

In the era of convergence, narratives can be of two types: crossmedia and transmedia.

1.1. Crossmedia

The term “crossmedia” refers to content that transitions across different media forms while retaining its core features, albeit with variations, much like a novel made into an audiobook.

Today, with cross-media storytelling we mean an integrated narrative that develops through various media, with various authors and

styles, and in order to experience the complete story audiences should engage with these variations. The narrative remains one, but adapted with tailored variations to the various transpositions.

Understood as such, crossmedia storytelling constitutes a network woven by a series of elements around a definite narrative universe; these elements must coalesce into a coherent canon centered on a single story existing through many manifestations.

Ultimately, crossmedia refers to integrated experiences across multiple media including the internet, video, television, mobile devices, DVD, print and radio. The incorporation of new media into crossmedia experiences implies higher levels of interactivity in the audience. In other words, it is the experience of a story that the user-narrator enjoys while watching a film, reading a novel, playing a game, listening to an audiobook while taking a walk, and so on.

And this experience, so diverse in its potential, is interconnected through various means, unified by narrative continuity and audience interactivity (Davidson *et al.* 2010, 8). This is why creators of cross-media products have two main tasks: enticing audiences to traverse different channels and promoting the idea that each product is a fragment of a larger experience that they must complete in their mind (Apperley 2004). Unlike cross-platform storytelling, which only focuses on adapting the product to various supports, crossmedia storytelling aims to provide additional information for the construction of a unified story.

1.2. Transmedia

According to Jenkins (2008), we can consider transmedia any interconnected story that unfolds on multiple platforms, while maintaining narrative independence. Indeed, in addition to the possibility of experiencing each fragment individually, these versions are all part of a “global story” (Costa Sánchez & Piñeiro Otero 2012, 112).

In fact, we could describe transmedia narration as an expansion of the same story which retains its essence, even when taking into account various non-faithful versions and adaptations.

Beyond telling the story through various media, a transmedia narration can develop through different languages (Scolari 2009) and each version achieves, from time to time, its own autonomy. In a transmedia story, each new text makes a specific and valuable contribution to the

whole (Jenkins 2003). Each medium does what it does best: a story can be introduced through film, expanded through television, novels and comics, and its world can be explored and experienced through a videogame. Each product in the franchise must be autonomous enough to allow for its autonomous consumption.

This clearly aligns with profitability, as evidenced by many successful transmedia projects in the audiovisual sector. Star Wars, Matrix, Marvel and DC superheroes, Super Mario, are just some of these projects which, while not all born as transmedia franchises, over time have evolved as such, since acceptance by the public has proved profitable and led to the story being expanded into different media (Costa Sánchez & Piñeiro Otero 2012, 113).

We will see why these elements will fit perfectly in our case, making *Mary Shelley's Frankenstein* an example of a transmedia franchise.

2. An Example of Transmedia Franchise: the Frankenstein Case

Today, very few are unfamiliar with the story of the man who dared to play God by creating another human being like himself, eventually assembling a monster and unleashing a tragedy that is a cautionary tale for readers of all eras. It has been one of the most revived and reworked stories for two centuries now, since, on January 1, 1818, *Frankenstein; or, the Modern Prometheus* saw the light of day in just five hundred copies at Lackington, a small publishing house. Initially released anonymously, it was only on the following March 11 that the novel began to be known through the very first reviews, which panned the work, but at the same time exalted precisely that aspect that seemed destined to condemn it for reprobation and oblivion: its originality.

Frankenstein is not the typical early 19th century novel, expected to have a moral or in any case an edifying message: its ailed is powerful, but it's far from edifying and, still today, strong and inexhaustible. In 1823, just five years after its publication, the book was transposed into successful theatrical pieces, and, since then, the theatrical, cinematographic, comic and videogame versions are countless, not to mention the merchandising.

In 1994, riding the coattails of the resounding success of the movie *Bram Stoker's Dracula* two years earlier (the movie was the most faith-

ful transposition up to then – and still today – of the famous vampire novel), a similar experiment was attempted with Mary Shelley's masterpiece. While in this *Dracula* version Francis Ford Coppola directed and produced, in *Mary Shelley's Frankenstein* he served only as producer, leaving the direction to Kenneth Branagh.

This filmic version failed to replicate the same critical and public success of its vampire predecessor, despite being a good box office success outside the USA; and yet, it remains, to this day, not only the most faithful transposition of the literary text, but also among the more interesting ones.

The videogame examined in this work is based on the film, with a dedicated version for each of the two consoles, Sega Genesis (known as Sega Mega Drive outside the USA) and Super Nintendo, and another one for Sega CD (known as Mega-CD in some countries) according to the custom of the time, consolidated in 1994, of translating films into videogame versions to drive their success. A themed pinball machine was released in early 1995 by Sega Pinball, which later became one of the machines included in the video pinball simulator *The Pinball Arcade* in October 2015, and subsequently in *Stern Pinball Arcade* for free in 2016. *Mary Shelley's Frankenstein* stands as a perfect example of a transmedia franchise (Bertetti 2016).

2.1. The Idea Behind the Novel

Mary Shelley, in *Frankenstein's* conception and drafting, incorporated the prevailing instances of her time, and perfectly reflected the anxieties of that society. In fact, *Frankenstein* saw the light in an era of transition of scientific research. Science, which had once been the domain of passionate and educated individuals exploring various subjects, was evolving into a field dominated by specialists and experts - i.e. scientists. The word "scientist" emerged only years after the publication of the novel: the term is coined by Anglican pastor and philosopher of science, William Whewell, in 1834, in the (anonymous) review of *On the Connexion of the Physical Sciences* by Mary Somerville.

When the word "scientist" was born, Mary Shelley was 37 years old, and the novel had already reached its third edition. She wrote the work inspired by the questions of the time regarding the possibility of giving life back to dead matter, referring to the theories of Erasmus Darwin

(grandfather of the more famous Charles Darwin) and Giovanni Aldini (prosecutor of his uncle Luigi Galvani's studies), as well as the legends surrounding the alchemist Johann Konrad Dippel, who lived in Castle Frankenstein in Switzerland between the 1600s and 1700s (Camilletti 2018, 88-92). Not to mention the experiments of the time, held in the public square: back then science had no boundaries, and one could witness attempted to electrically reanimate the body of a death row inmate, such as George Foster, as if it were a spectacle (Sampson 2018, 60).

This novel, like all of Mary Shelley's subsequent ones, is a story of ideas. It sprang from a starting idea and explored all its possible consequences, in the most rigorously logical and therefore crude way possible (Sampson 2018, 290). The actions of Mary Shelley's characters can be confused, uncertain, and prone to self-deception, while the characters themselves (and therefore including the homodiegetic narrators of the story, whether they are autodiegetic or allodiegetic) convey their perceptions and narratives assertively and linearly, because they are confident in what they say and in the ideas that move them. All of this may partially explain *Frankenstein's* success, but it's not enough. In fact, none of her other works have been nearly as successful. In addition to the strength of its ideas, in *Frankenstein* also lies the drama of the characters and their stories, the reader can be passionate about them and what they experience, the narratee would know how their ordeals end up, beyond the ideas they bring, whether they are their own or those of the author. The strength of the novel certainly lied in the truculent idea, gothic as well as sci-fi, of a man of science who manages to revive a corpse with all the possible, unpredictable consequences, but not only that: its power lies in the drama of the characters, between the remorse of Victor Frankenstein who first fantasized about being God and then even regrets being born, and the suffering of his Creature, who angrily begs to be seen and loved.

All of this implies that immediately, despite the severe criticism of the first reviewers, the novel achieved an extraordinary success with the public, which already in 1823 gives way to the first theatrical adaptations and, in 1910, to the first film. This story seems the perfect embodiment of the dilemma of "artificial creation", with all the responsibilities that it entails (Breton 1995, 46). For two centuries, *Frankenstein* has never ceased to intrigue and fascinate, even predictive

of ethical dilemmas that intellectuals, artists and scientists have faced in the contemporary age: an example is the concept of *Frankenstein complex* (Asimov 2005, 24).

However, this becomes bizarre if one considers that *Frankenstein* is probably among the most misunderstood novels ever: few realize that the Creature is not a brute without any self-control. Already from the first theatrical reductions, the profound meaning of the novel is betrayed, namely that the Creature is, in fact, the alter ego of Victor Frankenstein, in a way not too different from the relationship between Jekyll and Hyde. The so-called “monster” is an intelligent and sensitive being as much as its creator, and not simply an abomination assembled like a puzzle of flesh, by whom today we would call a *Mad Scientist* (Ripamonti 2014). This misrepresentation of the “monster” and its relationship with its creator was paradoxical if we consider that, starting from Peggy Webling’s theatrical transposition of 1927, the name “Frankenstein” is today attributed as much to the monster as to its creator.

In the novel, Victor refused to take care of his own Creature, abandons it to itself by refusing his own paternity and doesn’t gave it any moral or physical support, right from the start. He behaved as the exact opposite of a pedagogue, since he didn’t want to have anything to do with his artificial son: he rejected it for no real concrete reason, immediately after giving it life. Therefore, giving the monster the same surname as its father seemed like a sort of posthumous compensation.

The novel is hinged on the centrality of the Victor-Creature dualism, portraying them as two equal and yet antithetical beings, so much so that it recalls, to various critics, Hegel’s Servant-Master dialectic, published by the philosopher a few years before the drafting of the novel (Meirieu 2007). The fact is that today this dualism has been lost, and the version of the Creature elaborated in 1931 by the director James Whale, the make-up artist Jack Pierce and the actor Boris Karloff, marked a watershed in this sense: since then, every possible representation of the two Frankensteins, creator and monster, is confronted with that of the Universal film, so each new representation must choose whether to approach that version and therefore to distance itself from the literary protest, or on the contrary to distance itself from that film and re-appropriate the literary source, or again to attempt a synthesis. And it’s here that we enter the narratological (and also philological)

aspect of the videogame under examination: it is a reworking (in two distinct forms, as mentioned, one for Genesis/SNES and one for Sega CD) of a film which, in turn, at the time wanted to be, and indeed still is, the most faithful transposition possible of the novel. Indeed, the film wants to recover all the painful complex Victor-Creature dialectic that the novel stages. The same, by extension, happened in the videogame based on the film.

This is not the first videogame inspired by the novel, which happened to be *Frankenstein's Monster* from the Data Age for Atari 2600 in 1983, but it's an extremely interesting case for the experience that the gamer-narratee has, through the gameplay, of a videogame rewriting of the original story, a rewriting which in turn passed through the mediation of a film.

2.2. Frankenstein Novel as Playable Game

Making a videogame based on a literary masterpiece is still today a complex operation. In the early '90, characterized by 16-bit graphics, this process was even more difficult. In the case of the Frankenstein game, the limitations inherent in 2D graphics, including low resolution and poor sound effects, certainly represented an obstacle in the attempt to recreate the same atmosphere of the novel.

Nevertheless, the software house tried to immerse players in some of the main sensations offered by the book, describing the horrors and the consequences of science pushing its limits due to insatiable human curiosity.

Furthermore, the game appeared different from previous incarnations, and tried to put the user more in touch with the (perhaps) inexorable destiny of the horrid Creature generated by the experiments of the audacious Victor Frankenstein. The aim of the game was to guide players through six levels of intense action, all characterized by relatively diversified mechanics and themes, providing the players with the opportunity to discover the deepest and darkest traits of the Creature.

The story followed for the most part the events of the book, and the player's choices will be significant in determining some minor path that the Creature will follow. Thus, the ability of the player is crucial, beyond the mastery of the control system.

As the gameplay assumed a primary role in the player's experience, the story is conveyed for the most part by scenes rather than texts. In fact, also due to on-screen character limitations, the in-game texts were in general short and simple.

In this scenario, the role of translator/localizer involves a peculiar form of creativity, known as *transcreation* (Vilmantè & Vaida 2011). This technique is used within the localization process to transpose language nuances from one cultural scenario to another. In the videogame field, there are countless examples of this practice.

Transcreation is also used in the marketing field, more specifically within advertising campaigns. In this context, the concept of translation is not enough: the text's impact on the public relies on a very strong emotional component. In fact, rhymes and other linguistic strategies are often used. This approach allows the translator to modify words to align with the target language's linguistic and cultural nuances resonating with the public at that moment.

Eventually, videogame localization requires that translated contents will display properly on the screen. In fact, some languages use more words or longer words for the same concepts, when they are compared to English. Adaptation of internationalization software is necessary to support the right number of characters that have to appear on the screen, also accounting for how the texts can increase or reduce their dimensions when switching from one language to another. This often occurs even with the most common Western languages, so it's important to find a general solution involving under-resourced languages. This would improve the readability of both in-game texts and UI texts.

However, despite some issues and limitations inherent in that period, it was really hard not to be enchanted from the atmosphere and visual style of the game, making possible to enjoy the literary masterpiece in a different way.

2.3. Interactive Nonlinear Storytelling in Frankenstein Narrative Universe

The concept of linear storytelling, with its traditions and applications, is today well established. This structure works particularly well in videogames, as it provides the player with a goal or set of goals to complete in order to solve a problem.

Non-linear narratives extended the advantages of linear ones. The player has now multiple options to choose from. Historically, these options have often been presented as predetermined choices, such as choosing to go one place or another. However, non-linear narrative is about more than just offering choices. It also includes videogames that dynamically generate story elements and alter potential endings, based not only on the choices a player makes, but also on other factors such as performance, timing, or other narrative-related circumstances.

The main aspect of non-linear narrative, therefore, is variability. The player's experience is enriched by variety and a sense of efficacy. What the player does has real consequences on the world in which he or she interacts; in other words, the player is at least given the impression of influencing events.

Non-linear narratives applied to videogames can offer multiple endings or heavily impact the evolution of characters and events. The moment a player is given the ability to influence the narrative, they are automatically given the opportunity to identify with their character, generating more than just emotional investment. Even though the content within non-linear narratives may be pre-defined, the user can impart their own order to the events.

Encouraged by the growing use of technology in the audiovisual sector, interactive digital storytelling has been defined by Miller (2004) as «the use of digital media platforms and interactivity for narrative purposes, whether fictional or non-fiction, which enables the user to influence the flow and sometimes even the content of the story».

Spierling (2005) analyzed the main factors that influenced the birth and subsequent expansion of interactive digital storytelling: the first is linked to an attempt by audiovisual sector professionals to automate the movements of virtual characters by defining their skills based on rules of "intelligent" behavior; the second is related to an attempt by human-computer interaction (HCI) practitioners to use storytelling to make computer applications more understandable and more compelling (Laurel 1993); the third factor stems from the game design sector, which has begun to recognize computer games as interactive artifacts (Crawford 1993). Moreover, the AI industry has taken an interest in automated storytelling in response to user input, building story engines, and planning the actions of autonomous characters on a virtual stage. It is therefore clear that interactive digital storytelling arises

from the combination of various points of view and needs from different disciplines, mainly the audiovisual arts, information technology, human-computer interaction, and communication sciences.

Mary Shelley's novel represents today the fulcrum of the modern transmedia franchises based on the *Frankenstein* narrative universe. In this context, all the stories associated with each product linked to the original novel, are part of a dense network. Additionally, each media channel contributes to give different ways to experiencing the transmedia franchise. In particular, this happens with the introduction of new characters and events through movies, books, podcast, videogames etc.

This kind of transmedia storytelling is nonlinear (as each different product is linked to specific timelines, events, and characters) and characterized by the necessity to have a coherent organization. Also, the same concept of authorship is complex and varied, as the copyrights are distributed in different companies. However, this can lead to issues with licensing and the involvement in a project.

Narrative universes are also subjected to fan contributions; this phenomenon is known as *fandom* (Lamerichs 2023). In particular, fans can interact with a narrative universe through both official and unofficial means. In the case of Mary Shelley's *Frankenstein* videogame, fan-made contributions are related to the addition of Spanish language in the game. This practice, known as *romhacking*, was an unofficial common way to edit a game, by adding new features or languages.¹

3. From Translation to Localization of Digital Products

The work of someone who “translates” computer programs may appear simple and mechanical, but it nevertheless hides many pitfalls.

Compared to traditional translation, the difference does not arise on the level of complexity, but on the ways of interaction between those who produce a program (or related documentation) and those who “translate” it, i.e., adapt the program to their own language and culture.

The globalization of the computer and software industry in the 1980s gave rise to the localization industry, aimed at delivering specific

1. The modified version of the game is available at <https://www.romhacking.net/games/2937/>

needs to localized software for international markets. Localization goes beyond language transfer, because translation alone was insufficient to bring a software program developed in one country to a form suitable for release in other markets (Esselink 2000).

Since localization addresses the demand of the computer and software industry, it evolves constantly as new electronic contents are created which need to be localized. New dimensions in translation studies are surfacing because of localization practices, requiring further research into this domain.

In *A Practical Guide to Localization*, localization was defined as «taking a product and making it linguistically and culturally appropriate to the target locale (country/region and language) where it will be used and sold» (Esselink 2000). The *Localization Industry Primer*, published by The Localization Industry Standards Association (LISA), defined localization as «the process of adapting and manufacturing a product so that it has the look and feel of a nationally-manufactured piece of goods» (Fry & Lommel 2003) and «the process of modifying products or services to account for differences in distinct markets» (Fry & Lommel 2003). The document has distinguished between localization and translation and it pointed out that localization is not just a linguistic process. Localization must consider not only the linguistic issues, but also the content, the cultural and the technical aspects, to deliver a localized product that is comparable to a local product.

Among the various kinds of software, even videogames undergo localization processes.

In particular, *Game localization* was defined by Chandler (2012) as «the actual process of translating language assets in a game into other languages», adding that the game localization process comprises three main phases, involving translation of language assets and other non-translational activities. O'Hagan (2015) defines game localization as «a set of procedures involved in adjusting games technically, linguistically, and culturally to a given market to distribute them in territories other than their countries of origin». In this complex process, translation only appears in just two steps out of the fourteen steps described.

From the definitions discussed above, it becomes clear that game localization involves more than the lone translation of the language assets.

Game localization can be viewed as a new industry, which came along as a result of the birth of videogames.

Moreover, the internationalization and localization processes often involve a large number of people with different degrees of linguistic competence, necessitating coordination and the maintenance of a coherent style.

Furthermore, those involved in localization must have technical skills that surpass those of traditional translator, since in localization work it is necessary to juggle with portions of code of programming languages, files for defining resources, documents with marked-up text or specially formatted, and so on.

In the case of videogames, localization is a very delicate process that requires considerable creativity and massive intervention in order to recreate, also from a linguistic point of view, the cultural references that can touch the user's emotional chords, making them feel fully involved in the gaming experience. A good linguistic and cultural adaptation makes it possible to realize the player's deep desire to enjoy the playful experience by understanding it perfectly in his or her native language, thus feeling "heard" in their need, being able to immerse completely in the reality of the game, with its images and verbal communication.

3.1. The New Role of the Game Localizer between Creativity and Accessibility

In the field of translation, particularly within multimedia entertainment, the concept of creativity is complex and related to many different aspects. For example, creativity can manifest in the text itself, in the cultural background of a specific market/scenario, or in the authorship of a product.

Also, localizers have to deal with marketing instructions and other possible constraints imposed by software houses or local laws.

Localization - especially in the multimedia entertainment sector - is a challenging phase, often requiring a specific teamwork for every particular *locale* or scenario.

However, although both modern translators and localizers engage in ongoing debates about creativity, they also harbour a certain level of apprehension because excessive creativity can imply illicit liberties with the source text. Furthermore, creativity is deeply linked with both

concepts of authorship and copyright, and this can generate doubts and confusion (Bernal-Merino 2015).

The human experience in project management, linked to advice and knowledge of the newest translation technologies, is necessary for the success of the products.

Indeed, IT tools represented a significant step forward in terms of efficiency and usefulness of machine translation engines. These translation aids, if enriched and tailored for specific sectors, serve as a high-quality resource for translators tasked with checking and correcting translated texts.

It is therefore essential that IT tools are combined with human support and the assistance provided by translation professionals, localization project managers, developers and translators/proofreaders.

Methodologies like post-editing or verified machine translation ensured that translations are now fast, consistent, accurate, less expensive to produce, and adequately adapted to a specific target or sector of expertise.

The new scenarios of modern society presented a series of opportunities and challenges that renewed and redefined the role of the translator, also providing him with numerous aid tools to translation. While on the one hand, information and communication technologies have favored the birth of new professional profiles within the linguistic sphere, particularly in the field of localization of IT products, on the other hand, new transversal skills are always required of these profiles, both in the linguistic-international cultural and IT.

Although IT tools aimed at the field of translation and localization are increasingly and constantly evolving, complete automation remains elusive and human involvement necessary, especially because of the lack of creativity.

Moreover, neural networks and machine learning techniques are today one of the standards on which research and companies are investing, since these tools (with human oversight) made it possible to drastically reduce the time and costs of translation and localization processes.

In this evolving landscape, the skills that the translator-localizer will have to possess will be the main focus about the new role that this professional profile is taking on today.

A role which increasingly identifies him as a supervisor of linguistic-cultural adaptation processes of the products.

4. The Frankenstein Transmedia Franchise through Different View Angles

As we said before, there have been countless transpositions of *Frankenstein* in these two centuries, and each one has tried to highlight one aspect of the work rather than another. After the films by James Whale featuring Boris Karloff's iconic portrayal of the Creature, many movies tried to give us different versions of Victor and of the Creature, but the profound characterizations of both Frankensteins have been lost.

Mary Shelley's Frankenstein was produced in '94, in the wake of the success of *Bram Stoker's Dracula*, but in the early Nineties, we already had two transpositions. *Frankenstein Unbound* (1990) features Mary Shelley herself, played by Bridget Fonda, while the TV film *Frankenstein* (1992) depicted the psychic link between Victor and the Creature as between a man and his own clone, his own alter ego. They are expressions of the need to rewrite Mary Shelley's story so that its deepest aspects are rediscovered, to bring back all of those themes that have been the strength of the novel and the reason for its success, also responding to the contemporary need to show complex and fragile heroes and anti-heroes (Gnoffo 2022, 96-108).

Interest in *Frankenstein* coincided with the rise of what would be now called a "transmedia franchise" (Jenkins 2003). The transmedia franchise, starting from the rewrite made by Branagh's film, added new elements through the derived videogame, adapting the story (however, it should be noted that a videogame in 1994 could not have the same length or complexity as a film).

In a game the user plays one or more characters experiencing their respective points of view. The same happened in the case under examination, where, although the title bore the name of the one who created the monster, like the film and the novel, the player impersonates the Creature. In the videogame, the interactivity of the Creature with the environment and with the other characters is based on two aspects: on the one hand, it depended fundamentally on the player's choices and problem solving skills, on the other hand it's based on the path and on the available options. As in all adventure videogames of that period, it

allowed for an extremely limited freedom of choice, i.e., it reflected the need to eliminate the antagonists (be it the Goombas for Super Mario or, in this case, the angry villagers for the Creature) moving left and right and jumping up and down. For games of this period, these two limited elements, namely the desire to continue the story and the thrill of being eliminated, were enough to create a deep sense of involvement with the experience, if the game had an enough captivating graphics. The gaming experience let users play the experiences reported by the Creature in the novel as homodiegetic and autodiegetic narrator, at the moment of telling its creator about itself.

What Mary Shelley already wanted to achieve, i.e., the reader's identification with the points of view of the characters, in the videogame translated into an immersive impersonation, also through the interactive component, even though it only concerned the Creature, the sole protagonist of the videogame versions, as Victor Frankenstein is reduced to a character in the background, except during the final boss scene. The game, in both versions (Genesis/SNES and Sega CD), had a distinctive style which did not concern itself with recovering or going back to the novel, but rather with coming back to the adaptation that actually inspired it the most, i.e., the film, even if the Sega CD version, as we will see, takes great creative liberties, up to distorting the sense of the original story.

4.1. Gameplay vs Reading. Experiencing the Literary Masterpiece

As mentioned before, the novel is a first-person narrative. To understand the reader's experience of the novel, it is necessary to dwell on this point. We can understand the word "diegesis" as a storyworld (Giovannetti 2021, 45), because everything that the reader learns about the characters, their stories, and the world in which they move depends on the narrator. With *Frankenstein* we have three narrators who alternate in an interlocking mechanism of non-linear narration, all three homodiegetic and autodiegetic.

The first is Captain Walton who recounts his incredible experience in several letters to his sister, first reporting the stories of Victor who, in turn, emerging as the second narrator, not only talks about himself but also reports everything the Creature has told him, and then reporting

his own direct encounter with the Creature. The third and final narrator is the monster itself, both when it recounts its experiences to Victor and when it directly meets Walton, after its creator has died.

Fabula and plot do not coincide, the narration becomes linear only within the long analepsis constituted by the narration of Victor to Walton and, within that, by the narration of the Creature to Victor. This results in an interlocking narrative, similar to a nesting doll or to Chinese boxes, in which the narrators, especially Victor and the Creature, are fundamentally unreliable (Giovannetti 2021, 56). The monster may have tricked its creator into giving him a better version of itself, just as Victor may have tricked Walton into seeking self-absolution, and so both force the Captain, and the reader too, to trust them. In turn, Walton could really be a madman confiding his ravings to his sister, and therefore we should take what he says with a grain of salt. All these first-person narrations, leaving no space for an omniscient extradiegetic narrator, entail that the focus is always internal, centered on what each character feels and experiences, even when reporting the facts of other characters.

4.1.1 The Genesis/SNES Version

The graphics for SNES are undoubtedly superior, with dynamic effects that create the illusion of perspective in movies, while the Genesis version is faster. The player assumes the role of the Creature, who flees from the laboratory as soon as it wakes up, seemingly for no clear reason, while in the novel and in the film the doctor's rejection of the Creature is explicit: Victor is indeed an abandoning father, who denies his own creation (Meirieu 2007, 57). This marks the first sign of the necessary simplification of the plot. The layout of the Creature faithfully follows that of the film, which in turn is very different from that of the novel. In the film, its gait is limp, resembling that of a person who survived a stroke (the idea being to depict the damage of a brain awakened from death), while in the novel the Creature is endowed with extraordinary agility, strength and speed. Therefore the player impersonates a being with a limp who, however, as in all games of this genre, can easily stoop and make great leaps, and furthermore who, in keeping with the spirit of the novel, has to defend itself against all the characters he encounters along the way. In order to

progress from one level to another, the player must solve simple puzzles involving switches and pulleys, which are activated by using the objects encountered when necessary, sometimes even by combining them. The environment is fraught with dangers, but it also holds discoveries, and requires the player to know how to interact. This involves not only jumping from one height to another, avoiding drowning, and moving boulders, but also destroying chests and burning bales of hay to find useful items.

The player's experience is therefore basically one of problem solving, in a plot that focuses mainly on taking action to reach the next stage. However, between one level and another, images accompanied by in-game texts (similar to an illustrated book) reveal that the horrible Creature, whom the player controls, is a kind soul who only wants to have friends. This represents an extreme simplification: by removing any possible ambiguity from the character, the player's goal becomes to Victor so he can create a mate for the Creature. In the novel, the doctor agrees, in exchange for the promise that the monster and its bride will live without any human contact, ultimately changing his mind and refusing to complete the female version of the monster, who in turn takes revenge by killing Elizabeth. In the film, Victor gives life back to the woman, transforming her into that being that the monster would have desired as its partner. Earlier films have explored what would happen if Victor had completed the creation of a female for the monster, and already in *Frankenstein Unbound* he revived Elizabeth after she was killed by the monster. This step is completely absent in the game: the player impersonates the Creature who returns to Victor to request a mate, only to learn, from another illustrated text, that by pure coincidence Elizabeth has died of unknown causes. Victor gives her life back to get her back with him.

This omission represents a significant plot hole that serves both to simplify the story and to avoid portraying the Creature as a murderer. The revived Elizabeth becomes another enemy from whom the player must defend themselves, until she dies by combustion, as in the film. A caption then explains that the Creature, tired of mankind, heads to the North Pole in search of peace, while Victor hunts it down. This marks a substantial difference: in the novel, it is the Creature, after having exterminated Victor's loved ones, who incites him to be chased by him to the ends of the world. The conclusion holds great psychological

subtlety: just as Victor refused to make the Creature similar to a man by giving it a mate, now it is the Creature that makes Victor similar to itself, that is, to a monster detached from mankind.

In the game, as the player progresses to the last level, the Alexander Nevsky ship, the Creature seeks a dialogue with its creator rather than intending to harm him. The player reaches the last illustrated text, which is decisive: it explains that Walton, after listening to Victor's incredible story, understood, more than Victor himself, that the Creature was a kind being that wished to hurt no one. Thus, when the Creature reaches his cabin, the Captain takes it to the bedside of Frankenstein, who lies there lifeless. In this moment, Walton assumes the same role as a solver/mediator as in the novel, where he is the only one who truly considers the words of the Creature, realizing that Servant-Master dialectic which, with the egoist Victor, had remained unfinished (Meirieu 2007, 51- 53).

This game, like those in its category, is aimed at teenagers, despite being based on a film so crude that its use has been banned for children in various countries, such as in Italy. In 1994 there is still no PEGI standard in Europe, nor is there any content rating in general, there is no recommended age indicated on the game packaging, so the solution is, necessarily, a rewriting of the story that makes it suitable to the target audience for which the videogame is designed, although, paradoxically, the videogame wishes to be driven by the appeal and success of the truculent film.

4.1.2 The Sega CD Version

The version of the game for Sega CD was notably different, which also altered the experience that the player has of it, and, as a consequence, it's crucial to further discuss the console.

Sega CD, released in Japan in '91, uses CD-ROM technology, which at the time was young and, even though its potential was understood, it wasn't possible to take full advantage of it yet. Many of its games are ported from the Sega Genesis, with added CD audio music and full motion video, but the quality wasn't not exceptional and sales are poor. Sega's intuition is right but premature. The release of CD-ROM competitors such as CD-i from Philips and 3DO from Panasonic, although also not very successful, make it a technically outdated console.

Nintendo briefly considered this path by designing the SNES CD, but soon abandoned the enterprise. Between December '94 and November '95, this path will be successfully followed by Sony with the first PlayStation, which will succeed where others have failed (McFerran 2009, 82-87).

During this brief historical phase, the other version of the videogame, like many other Sega CD titles, was a porting of the Genesis version with notable differences. The graphics was undoubtedly of superior quality, the cutscenes between one level and another are remarkable in terms of resolution, perspective and camera movement, since they are intended not so much to explain what happens or to clarify the situation from a narrative point of view, but rather to show the environment where the story was then located and show off graphic quality. This greatly changes the gaming experience.

The Sega CD title was a "Point and Click Adventure" in which the player moved more freely from one environment to another, which he can explore with greater fluidity and realism, his movements were not limited to going left and right and jumping up and down: in their limping (the layout is identical, albeit more defined, to the Genesis/SNES version and the film), the player thoroughly inspected a given environment, until they found something useful or the entrance to the next area. There was no explicit division into levels, but the player smoothly transitions from one environment to another progressing through the story. When the player discovered something of interest or interacts with another character, an in-game text appeared, much like a cartoon speech bubble, to convey the spoken words.

An interesting aspect of this relates to the plot: staying true to the type of game, but deviating from the spirit of both the film and the novel, the player impersonated a Creature that had no difficulty conversing with other human beings and seeking help when needed. This is necessary for the gameplay, since the progress arised from collaborative interaction with the other characters, but it totally betrayed the experience of the monster in the novel and cancels its leitmotiv: if the Creature could interact so easily with other human beings, it would not have harbored such hatred for mankind.

The fights occurred as reserved action-gameplay sequences when the Creature encountered an opponent in its path: the most surreal fight was when, in the Secret Laboratory, the player found Elizabeth.

After Victor broke his promise to create a mate for the Creature, it wished to depart, but Elizabeth blocked its way. The Creature, who, unlike the novel and the film, was reluctant to harm her, begged her to let it go, but she insisted, resulting in a fight where her fighting style is reminiscent of Chun Li from *Street Fighter*. The player must reluctantly defeat her, after which a heartbroken Victor took her body bringing her back to life, mirroring the film. Here the plot took an unexpected turn: instead of dying, the Revived Elizabeth follows the Creature and they go away together in front of a morally destroyed Victor. The two go to the North Pole embarking as clandestines on the Alexander Nevsky, but Victor joined them. The player must defeat Dr. Frankenstein as the final boss, after which the monster, for no reason, set fire to the ship leaving with its mate: this was a very peculiar rewrite of the story, where the Creature was not only the hero but even managed to have a companion and a happy ending. There was no Walton nor any logic in this ending. This game's version was essentially based on remarkable graphics for those times and an interesting gameplay, but totally detached from the themes of both the film and the novel.

Here, we encounter a striking example of transmedia storytelling, where we access a pre-existing narrative universe but with alternative story elements and substantial variations. If the Genesis/SNES version followed the book but provided an action-based narration and few illustrated texts, this Sega CD version was more narrative and wanted to offer a new plot.

4.2. Videogames Adaptions of Literary Works. Some Issues and Challenges

One of the main challenges related to videogame localization is to define the appropriate skills for translators, adapting their curriculum by incorporating additional competencies that bridge the gap between the classroom and emerging industry practices. Software houses need to find a way to create the best possible cooperation between the translator/localizer and the programmer (Ressin & Nocera *et al.* 2011). In order to do this, the first step would be to directly involve the translator in the internationalization process (Estrella *et al.* 2019). In the area of videogame adaptions of literary works, it is important to keep the translator informed about the proper context of the game,

providing all the information about the story, the main characters, and the environment in which the game will be developed.

In this case, creativity is strictly linked to the concept of transcreation, in order to better adapt textual and audio contents to a specific cultural scenario. Moreover, crossmediality and transmedia storytelling represent two further aspects to consider in software and videogame localization, as they are often linked to other medial forms and channels.

Also, localization efforts, especially in the case of narrative games, must be framed in linguistic and cultural contexts: in particular, the translator will have to maintain, as far as possible, a certain stylistic and lexical coherence (Pirrone & D’Ulizia 2023), both for the single application and overall user interface that end users will have to interact with. For this particular reason, localization projects usually start with the definition of glossaries and style guides. In the realm of videogames, translators have to deal with the programming source code, and with the costs of production, development, and location.

Then, videogame localization requires that translated contents display properly on the screen. In fact, some languages use more words or longer words for the same concepts, when they are compared to English, posing challenges in adapting internationalization software to support the right number of characters that have to appear on the screen, also calculating how the texts can increase or reduce their dimensions by switching from one language to another. This often occurs even with the most common Western languages, highlighting the need for a general solution involving under-resourced languages. This approach enhances the readability of both in-game texts and user interface (UI) texts.

Nevertheless, making a videogame usable for customers in different regions takes more than simply translating and adapting the text. In fact, translators have to deal with differences in formatting and conventions for writing dates, times, numbers, addresses and currency. Implementing these aspects into the internationalization phase makes adaptation to the target country/culture easier.

Moreover, in order to fix bugs or add new features, most game releases receive regular updates. Often, these updates involve elements that will need to be localized. This leads to a process of “continuous localization” with a large variety of small localization changes, which require

continuous adaptation. To effectively manage these needs, translation technologies would have to facilitate these kinds of adjustments.

Finally, other important open challenges are related to improving some recent technologies, such as *CAT tools* and *machine learning translation* (Heyn, 1998; Muntés-Mulero *et al.* 2012). These two techniques are relatively new, so they need to be empowered with new features. In particular, *CAT tools* appear to be one of the best methods to align localization with the internationalization phase; also, an improvement of machine learning algorithms would allow even better perspectives into software and game localization. Furthermore, the importance to adequately support low-resourced languages has been highlighted by participants, who suggested crowdfunding campaigns as a possible solution.

Conclusions

At the end of this work, it is possible to draw some conclusions which, rather than answers, lead to further research stimuli. One point that seems clear is this: in the mid-90s, a period that saw the flourishing of the transmedia franchise for mainly economic but also artistic reasons, seeing also continuous innovation in the field of videogame consoles, the need was to rediscover iconic horror characters such as Dracula and Frankenstein who, as archetypes, have timeless appeal to storytellers.

Indeed, in 1994 the transmedia franchise, starting with Branagh's film adaptation, introduced some additional elements. The videogame altered the substance of the story, not only in terms of plot, but also of the user's experience, mainly through the gameplay, a new element compared to the novel and the film. So, the narratee experienced not only a different plot, adapted both for media needs and for artistic purposes, but also an immersive perspective, effectively becoming a player, a sort of actor within the narrative.

Therefore, the videogame transposition, aiming to reach a different target audience, ended up creating a separate version of the story, although, as we have seen, the two contemporary versions of the same videogame tell two different stories, and not only due to hardware differences but also because of the artistic freedom of the developers.

For this reason, the experience that the narratee-user-player makes of Mary Shelley's masterpiece in 1994, through the videogame for

Genesis and SNES as well as its version for Sega CD, was one of the very first cases of transmedia franchise, since the narratee is not limited to enjoying the same story on different media (cinema, console and even pinball), but, by virtue of the fact that the original novel is re-read and re-written in different ways depending on the medium in which it is transposed, there are alterations, additions and subtractions, both in terms of plot and the user's experience of the story, indeed adding new and different pieces to that *Frankenstein* universe that the narratee approaches.

From 1994 to today, even more so with the almost worldwide diffusion of the internet, the phenomenon of the transmedia franchise has been constantly evolving together with technological progress, reaching results, in the videogame world, which today are visible to all but which at the time were unthinkable, although they already existed in their early stages.

In the era of the dynamic web or web 2.0, crossmediality and transmedia franchises – in addition to being an enormously profitable phenomenon thanks to productions that can be virtually infinite (just think of the Star Wars and Harry Potter sagas, or the films and videogames based on Lego) – seem to respond to a new need, in the same way as hypertext sites like Wikipedia and social networks like Facebook, Instagram and TikTok: to promote knowledge, sharing and connection on multiple levels, giving us the idea (real or fictitious) of a hyper-connected universe where everything is connected and everything is easily accessible by jumping from one point of the spider-web to another. And the presence of different versions of the same story, though bound by being part of a single cohesive Multiverse, seems to be a recurring motif in most contemporary narratives, as if to reassure us that after all we swim in waters that we can manage and circumnavigate, and not in a boundless ocean. *Frankenstein*, in each of the 1994 versions that we have analyzed, although not a successful film nor a successful videogame, contained the first seeds of this phenomenon, confirming that Mary Shelley's novel, from its origins, is an extraordinary story, remarkably foretelling of our contemporaneity.

References

- Belsunces Gonçalves, A. 2011, *Producción, consumo y prácticas culturales en torno a los nuevos media en la cultura de la convergencia: el caso de Fringe como narración transmedia*, Universitat Oberta de Catalunya, Barcelona, <https://openaccess.uoc.edu/bitstream/10609/8581/1/abelsuncesTFM0611.PDF>
- Bernal-Merino, M. Á. 2018, *Creativity and playability in the localization of video games*, "JIAL: The Journal of Internationalization and Localization", 5 (1), pp. 101-137.
- Bertetti, P. 2016, *Mondi narrativi e storie future. Modelli di espansione seriale tra pulp magazine e franchise transmediali*, "Between", 6 (11), <http://www.betweenjournal.it>
- Breton, P. 1995, *À l'image de l'homme – Du Golem aux creatures virtuelles*, Le Seuil, Paris.
- Chandler, H.M. & O'Malley Deming, S. 2012, *The Game Localization Handbook*, 2nd ed., Jones & Bartlett Learning, Boston.
- Camilletti, F. & Pezzini, F. 2018, *Villa Diodati Files. Il primo Frankenstein (1816-17)*, Nova Delphi, Roma.
- Costa Sánchez, C. & Piñeiro Otero, T. 2012, *Nuevas narrativas audiovisuales: multiplataforma, crossmedia y transmedia. El caso de Águila Roja (RTVE)*, "Icono 14", 10 (2), pp. 102-125.
- Crawford, C. 1993-94, *Fundamentals of Interactivity*, "The Journal of Computer Game Design (Interactive Entertainment Design)", 7, <https://www.erasmatazz.com/library/the-journal-of-computer/jcgd-volume-7/fundamentals-of-interactivi.htm>
- Davidson, D. et al. 2010, *Cross-media communications: an Introduction to the Art of Creating Integrated Media Experiences*, ETC Press, Pittsburgh, <http://repository.cmu.edu/etcpres/6>

Esselink, B. 2000, *A Practical Guide to Localization*. Benjamins, John Publishing Company, Amsterdam.

Freedman, C.H. 2005, *Conversations with Isaac Asimov*, University Press of Mississippi, Jackson.

Fry, D. & Lommel, A. 2003, *The Localization Industry Primer*, 2nd ed., The Localization Industry Standards Association (LISA), Féch, <http://www.lisa.org/interact/webregister.html>

Giovannetti, P. 2021, *Il racconto. Letteratura, cinema, televisione*, Carocci, Roma.

Gnoffo, C. 2022, *La fragilità dell'eroe che amiamo*, "Le nuove frontiere della scuola", 58, pp. 96-108.

Heyn, M. 1998, *Translation memories: Insights and prospects*, in Bowker, L. et al., *Unity in diversity? Current Trends in Translation Studies*, Routledge, New York, pp. 123-136.

Jenkins, H. 2003, *Transmedia storytelling. Moving characters from books to films to video games can make them stronger and more compelling*, "MIT Technology Review". <https://www.technologyreview.com/2003/01/15/234540/transmedia-storytelling/>

Jenkins, H. 2003, *Convergence culture: where old and new media collide*, New York University Press, New York.

Lamerichs, N. 2018, *Shared Narratives: Intermediality in Fandom?*, in *Productive Fandom: Intermediality and Affective Reception in Fan Culture*, Amsterdam University Press, Amsterdam, 11–34, <https://doi.org/10.2307/j.ctv65svxz.4>

Liubinienė, V. & Šiaučiuonė, V. 2011, *Video Game Localization: the Analysis of In-Game Texts*, in "Kalbų Studijos" 191, pp. 46-55.

McFerran, D. 2009, *Retrospection: Mega-CD*, in "Retro Gamer", 61, pp. 82-87.

Meirieu, P. 2007, *Frankenstein educatore*, trad. it. Caratelli, F., Edizioni Junior, Parma.

Miller, C. 2004, *Digital Storytelling: A Creator's Guide to Interactive Entertainment*, Focal Press, Waltham.

Montero do Nascimento, D. & Duarte Gómes da Silva, E. 2011, *Universos Fílmicos Dinâmicos: estratégias narrativas para novas mídias e televisão digital interativa*, in "Revista Géminis", 2 (2), pp. 215-237, <https://www.revistageminis.ufscar.br/index.php/geminis/article/view/90/pdf>

Muntés-Mulero, V. et al. 2012, *Context-aware machine translation for software localization*, in *Proceedings of the 16th Annual Conference of the European Association for Machine Translation*, EAMT, Trento, pp. 77-80.

O'Hagan, M. 2015, *Game localisation as software-mediated cultural experience: Shedding light on the changing role of translation*, in "Intercultural communication in the digital age. Multilingua", 34 (6), pp. 747-771, <https://doi.org/10.1515/multi-2014-0062>

Pirrone, M. & D'Ulizia, A. 2023, *La professione del traduttore nell'era digitale. L'avvento del localizzatore*, in Corsi, C. & Coen, P., *Le professioni del comunicare: passato, presente, futuro*, Edizioni Quasar, Roma, pp. 47-55.

Ressin, M. et al. 2011., *Lost in agility? Approaching software localization in agile software development*, in "Lecture Notes in Business Information Processing", 77, pp. 320-321.

Sampson, F. 2018, *In search of Mary Shelley. The Girl Who Wrote Frankenstein*, Profile Books, London.

Sampson, F. 2018, *La ragazza che scrisse Frankenstein*, trad. it. Galitelli, E., UTET, Milano.

Scolari, C.A. 2009, *Ecología de la hipertelevisión. Complejidad narrativa, simulación y transmedialidad en la televisión contemporánea*, in Squirra

S. & Fechine Y., *Televisão Digital: Desafios para a comunicação*, Sulina, Porto Alegre, pp. 174-201.

Spierling, U. 2005, *Interactive digital storytelling: towards a hybrid conceptual approach*, in *Proceeding of the 2005 Digital Games Research Association (DiGRA) International Conference: Changing Views: Worlds in Play*, 3. Link: <http://www.digra.org/digital-library/publications/interactive-digital-storytelling-towards-a-hybrid-conceptual-approach/>.

Web references

<https://www.romhacking.net/games/2937/>

Claudio Gnoffo e Marco Pirrone

Università Guglielmo Marconi; Università eCampus
claudiognoffo87@gmail.com; marco.pirrone1@gmail.com



ἘΠΈΚΕΙΝΑ

International Journal of Ontology
History and Critics

NOEMI RITA SANFILIPPO

Ripensare l'autorialità alla luce delle Intelligenze Artificiali. Una proposta teorica

EPEKEINA, vol. 16, n. 1-2 (2023), pp. 1-17
Philosophy of Technology

ISSN: 2281-3209

DOI: 10.7408/epkn.

Published on-line by:

CRF – CENTRO INTERNAZIONALE PER LA RICERCA FILOSOFICA
PALERMO (ITALY)

www.ricercafilosofica.it/epekeina



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs 3.0 Unported License.

Ripensare l'autorialità alla luce delle Intelligenze Artificiali. Una proposta teorica

Noemi Rita Sanfilippo

Introduzione

What happens here is important because what's happening to us is happening across all fields of labor.[...] we are the victims here, we are being victimized by a very greedy entity. [...] It is disgusting, shame on them. They stand on the wrong side of history at this very moment. [...] The entire business model has been changed by streaming, digital, AI. This is a moment of history that is a moment of truth. If we don't stand tall right now we are all going to be in trouble, we are all going to be in jeopardy of being replaced by machines. And big business who cares more about Wall Street than you and your family. Most of Americans don't have more than five hundred dollars. In an emergency this is a very big deal and it weighed heavy on us. [...] We're not going to take this anymore. You people are crazy! What are you doing? [...] Privately they all say we're the center of the wheel, everybody else tinkers around our artistry. But actions speak louder than words, and there was nothing. There it was insulting. So we came together in strength and solidarity and unity with the largest strike authorization vote in our Union's history. And we made the hard decision that we tell you as we stand before you today. [...] we are labor and we stand tall and we demand respect and to be honored for our contribution. You share the wealth because you cannot exist without us.¹ (Drescher 2023)

Con queste parole, Fran Drescher, presidente del sindacato SAG-AFTRA (Screen Actors Guild e American Federation of Television and Radio Artists) ha annunciato l'adesione degli attori hollywoodiani allo sciopero già in corso degli sceneggiatori. Drescher ha condannato l'AMTP (Alliance of Motion Picture and Television Producers) accusandola di trovarsi dalla parte sbagliata della storia. Quali sono i motivi

1. Tratto dal discorso con cui la presidente dell'associazione SAG-AFTRA, Franc Drescher, durante la conferenza stampa del sindacato annuncia lo sciopero di Hollywood. L'intero intervento è disponibile al seguente link: <https://youtu.be/huWJY9Ygsw0>.

dietro questo sciopero congiunto, evento di portata storica dal momento che non si verificava dal 1960? Le motivazioni alla base di ogni sciopero, a prescindere dalla categoria di lavoratori coinvolti, sono ricorrenti e presenti anche in questo caso: migliori condizioni di lavoro, contratti più equi, migliori compensi, tutele economiche e sanitarie. Ma questa volta ci sono degli elementi in più che hanno dato il via all'insurrezione degli attori; motivi legati ad una delle tematiche ad oggi più dibattute: l'intelligenza artificiale ed i diritti ad essa collegati. La prima causa del mancato raggiungimento di un accordo tra le due associazioni di lavoratori sarebbe legata alle piattaforme di streaming, che hanno modificato radicalmente il modo in cui i compensi basati sui diritti d'autore devono essere attribuiti. Da sempre la retribuzione attoriale è stata calcolata, sulla base dei diritti d'immagine, a partire dal concetto di replica: ogni riproduzione di una pellicola ad esempio nelle sale, durante i festival, con l'uscita del DVD ufficiale del film (prima ancora del VHS), eccetera, contribuiva al calcolo del compenso finale dell'attore, il quale, con un solo film, poteva ottenere un reddito sufficiente a mantenersi a lungo. Lo streaming ha invece abolito questo sistema, rendendo il concetto di replicabilità incontrollato e incontrollabile, dando la possibilità a chiunque in qualunque momento di guardare il film o la serie tv che desidera, ogni volta che vuole, senza che queste riproduzioni "paghino" alcunché all'artista. Il secondo e più importante motivo di disaccordo proviene direttamente dal mondo delle IA, e si colloca in continuità con un dibattito attualmente problematico, che coinvolge non soltanto il piano etico, ma anche e soprattutto la sfera normativa. Alla base della violazione degli accordi, ci sarebbe il via libera a testare un sistema di Intelligenza Artificiale che permetta di "replicare" in ogni momento gli attori in carne ed ossa. Come ha dichiarato il direttore esecutivo nazionale e capo negoziatore~Duncan Crabtree-Ireland²:

Questa proposta "innovativa" sull'AI che ci hanno fatto ieri, secondo cui le comparse dovrebbero essere scannerizzate e pagate per un giorno di paga e le loro aziende dovrebbero possedere la scansione della loro immagine, la loro effigie per poterla usare per il resto dell'eternità in qualsiasi progetto vogliono senza alcun consenso e senza alcun

2. Per approfondire <https://www.universalmovies.it/sag-aftra-azione-attori-in-sciopero/>.

compenso, se pensate che sia una proposta innovativa vi suggerisco di ripensarci.³ (Crabtree-Ireland 2023)

Uno strumento, dunque, in grado di riprodurre per sempre l'effigie di una persona, in modo tale da non aver bisogno di comparse o interpreti presenti fisicamente, potendo disporre di un vasto archivio digitale di attori, che vengono però così retribuiti una sola volta. Ed è proprio questa la tematica che apre a numerose domande presenti nel dibattito contemporaneo sulle IA: chi detiene i diritti d'autore di un prodotto generato con intelligenza artificiale? Chi è l'autore dei suddetti prodotti? Come cambia il concetto di autorialità? Sono queste alcune delle domande ancora aperte. In questo articolo proverò innanzitutto a tematizzare lo stato del dibattito, suggerendo al paragrafo 2 una divisione in due correnti di pensiero dell'attuale questione autoriale; nel paragrafo 3 proporrò una possibile terza via da percorrere, un punto di vista mediatore tra i due poli, nel tentativo non di risolvere i problemi fondanti, ma di avviare sempre di più la questione autoriale all'onnicomprendività.

1. L'autore è morto; l'autore è vivo. Due correnti teoriche a confronto

Affrontare il tema dell'autorialità significa inevitabilmente confrontarsi sin da subito con due autori segnanti del '900 europeo: Roland Barthes e Michel Foucault. Imprescindibili all'interno della tematica autoriale, occorre dunque ripercorrere brevemente il loro punto di vista attraverso le loro parole. Entrambi gli autori hanno infatti teorizzato per primi la morte dell'autore, l'uno come unica via possibile per la nascita del lettore, l'altro come conseguenza di una morte collettiva della letteratura. Barthes, nel suo saggio intitolato proprio *La morte dell'autore*, scrive:

L'autore è un personaggio moderno, prodotto dalla nostra società quando, alla fine del Medioevo, scopre grazie all'empirismo inglese,

3. Dichiarazione rilasciata alla conferenza stampa del sindacato SAG-AFTRA. L'intera conferenza è disponibile al seguente link: <https://www.youtube.com/live/FQiO8ZeK2Hs?feature=share>.

al razionalismo francese e alla fede individuale della Riforma il prestigio del singolo o, per dirla più nobilmente, della «persona umana». (Barthes 1988, 51)

Un prodotto umano, dunque, derivante dalla concezione moderna di individualità, di proprietà e persino di autoaffermazione. L'esigenza di conoscere l'autore non è nata insieme alla letteratura, i popoli primitivi, come scrive Perniola, non avevano bisogno di sapere chi avesse dipinto i graffiti in una grotta per credere a quella storia. Conoscere l'autore di un romanzo diventa un vezzo da critici, un modo per decifrare un testo a partire da chi lo ha scritto, nel tentativo costante di ritrovare nel contesto storico e nell'ideologia del suo autore il vero significato di un'opera.

L'Autore, finché ci si crede, è sempre visto come il passato del suo stesso libro: il libro e l'autore si dispongono da soli su una medesima linea, organizzata come un prima e un dopo: all'Autore è riconosciuto il compito di nutrire il libro, in quanto lo precede, pensa, soffre, vive per esso; con la propria opera intrattiene lo stesso rapporto di antecedenza che un padre ha con il figlio. (Barthes 1988, 54)

È per questa concezione che i critici, gli scrittori, gli stessi lettori, non riescono a staccare l'autore dall'opera. Non deve dunque stupire che il regno dell'autore è al tempo stesso il regno della critica. Questi due regni però hanno breve vita, poiché il vero tentativo rivoluzionario per comprendere un testo, per leggerlo, per renderlo autonomo, è spostare l'attenzione su un partecipante dell'opera letteraria mai preso in considerazione fino a quel momento: il lettore. «Non si dà testo se non c'è un lettore che lo attraversi» (Perniola 2014), ecco perché è più che mai fondamentale spostare l'attenzione su chi davvero permette ad un'opera di esistere e di avere una validità culturale e sociale. Il vero senso del testo, la sua unità, sta nel suo scopo finale, ovvero quello di essere letto da qualcuno. Il lettore è anonimo, non è rilevante per i critici e magari, proprio per questo, mantiene ancora la sua autorità. Solo la sua esistenza può essere la vera nuova via per la letteratura. Ma per far sì che si realizzi questa consapevolezza, occorre che avvenga una cesura netta col passato, che “muoia” ciò che ancora tiene ancorati alla vecchia considerazione letteraria. Che muoia, cioè, l'autore.

Del lettore la critica classica non si è mai occupata; per lei, nella letteratura non vi è altro uomo che chi scrive. Oggi cominciamo

a non lasciarsi più ingannare da quella sorta di antifrasi con cui la buona società è solita perorare in modo arrogante proprio in favore di ciò che in realtà mette al bando, ignora, soffoca o distrugge; sappiamo che, per restituire alla scrittura il suo avvenire, bisogna rovesciarne il mito: prezzo della nascita del lettore non può essere che la morte dell'Autore. (Barthes 1988, 56)

Poco tempo dopo, quasi a voler rispondere alle provocazioni dell'autore francese, Michel Foucault presenta ad una conferenza presso il Collège de France, nel 1969, il suo discorso che poi verrà raccolto in *Che cos'è un autore?*. All'interno di questa trattazione, il filosofo tenterà principalmente, raccogliendo il testimone da Barthes, di analizzare non tanto perché un autore muore, ma quali sono i vuoti lasciati da questa morte, in cui la funzione autoriale va ancora esercitata. Già dalla citazione scelta dall'autore in apertura dell'opera è facile intuire quale sia il pensiero confutato durante la trattazione. Scrive infatti Foucault

Del tema che sceglierei come punto di partenza trovo la formulazione in Beckett: "Che importa chi parla, qualcuno ha detto, che importa chi parla". È in questa indifferenza, penso, che bisogna riconoscere uno dei principi etici fondamentali della scrittura contemporanea. (Foucault 2004, 3)

Dopo aver analizzato come la scrittura si sia liberata dal tema dell'espressione, riferendosi solo a sé stessa, uno degli assunti di partenza della sua trattazione è la difficile definizione di cosa significhino in senso stretto i termini "opera" e "autore". A proposito del concetto di opera, Foucault si chiede

"Che cos'è un'opera?" che cos'è questa strana unità alla quale diamo il nome di opera? Quali elementi la compongono? Non è forse un'opera ciò che è stato scritto da colui che ne è l'autore? Vediamo subito sorgere le difficoltà. (Foucault 2004, 4-5)

Subito, dunque ci appare chiaro quale sia il primo problema fondamentale del dibattito: cos'è un'opera d'arte, e quali sono i criteri che ci portano ad utilizzare questa definizione. Se un'opera è tutto ciò che viene prodotto da un autore considerato tale a giusto merito, allora tutto ciò che appartiene allo stesso autore dovrebbe essere un'opera.

Se, come scrive il filosofo, noi reputiamo Nietzsche un autore, allora tutto ciò che gli è appartenuto, abbozzi delle sue opere, appunti sulla sua vita quotidiana, sono forse tutti prodotti degni di essere elevati al rango di opera? Ma può definirsi opera, ad esempio, la lista della spesa di Nietzsche?

La teoria dell'opera non esiste, e coloro che ingenuamente intraprendono la pubblicazione delle opere non posseggono una simile teoria, il che paralizza ben presto il loro lavoro empirico. E si potrebbe proseguire: possiamo dire che le Mille e una Notte costituisce un'opera? [...] La parola "opera" e l'unità che essa designa sono probabilmente tanto problematiche quanto l'individualità dell'autore. (Foucault 2004, 5)

Definizione problematica, inevitabilmente collegata ad una rete di significati "altri", a cui non si riuscirà a trovare una risposta univoca. Allo stesso modo il concetto di autore risulta quasi impossibile da indagare. In primo luogo, occorre analizzare, linguisticamente ragionando, cosa designa la parola autore.

Che cosa è un nome d'autore? Come funziona? Lungi dal darvi una soluzione, indicherò soltanto alcune delle difficoltà che si presentano. Il nome d'autore è un nome proprio, che pone gli stessi problemi di quest'ultimo. (Qui mi riferisco fra tante analisi diverse, a quelle di Searle). [...] Esso, più che un'indicazione, è un gesto, un dito puntato verso qualcuno; fino a un certo punto esso equivale a una descrizione. (Foucault 2004, 7)

Dire "lui è Foucault" non significa soltanto dire "lui è un autore", ma porta con sé una rete di significati che designano precisamente l'uomo Michel Foucault, differenziandolo da altri filosofi, scrittori, autori. Attribuire determinate caratteristiche ad una persona significa designarla in un certo modo tanto da diventare riconoscibile per quegli attributi stessi. Dire che Foucault non era francese probabilmente non influirà (o comunque influirà parzialmente) sulla lettura delle sue opere. Ma dire che Foucault non è l'autore di *Microfisica del potere* cambia radicalmente il modo in cui quell'opera verrà intesa. Non soltanto un nome proprio designa molto più che un semplice autore, ma l'ulteriore problema che Foucault pone è quello di ciò che egli definisce "i fondatori di discorsività". Sono quegli autori (o meglio,

sarebbe dire a questo punto quei nomi propri che designano autori) la cui rilevanza nel dibattito storico, politico, culturale, è superiore rispetto a quella che possono ricoprire dei semplici autori, poiché con il loro pensiero hanno segnato in maniera profonda il periodo storico a cui sono appartenuti. È il caso di autori come Marx e Freud, la cui importanza per la storia dell'Occidente è innegabile e univocamente riconosciuta. Se un giorno dovessimo scoprire un nuovo testo inedito di Freud, scrive Foucault, non cambierebbe solo la concezione che abbiamo di questo autore, ma l'impatto della scoperta si estenderebbe a tutta la psicoanalisi. È dunque per questi motivi che Foucault conclude la sua trattazione affermando

Si può immaginare una cultura dove i discorsi circolerebbero e sarebbero ricevuti senza che la funzione-autore apparisse mai. Tutti i discorsi, qualunque sia il loro statuto, la loro forma, il loro valore e qualunque sia il trattamento che si fa loro subire, si svolgerebbero nell'anonimato del mormorio. Non si ascolterebbero più le domande così a lungo proposte: "Chi ha realmente parlato? È veramente lui e nessun altro? Con quale autenticità o con quale originalità? E che cosa ha espresso dal più profondo di sé stesso nel suo discorso?" Ma altre come queste: "Quali sono i modi di esistenza di questo discorso? Da dove viene tenuto, come può circolare e chi può appropriarsene? Quali sono le ubicazioni predisposte per dei soggetti possibili? Chi può riempire queste diverse funzioni del soggetto?" E dietro a tutte queste domande non si capterebbe altro che il rumore di un'indifferenza: "Cosa importa chi parla?" (Foucault 2004, 21-22)

Cosa importa chi parla è la domanda che riecheggia ad ogni passo. A morire non è solo l'autore, ma è l'intero apparato costitutivo di un'opera: «la morte dell'autore, non ha portato al trionfo del lettore (come auspicava Barthes), ma ad una morte congiunta della scrittura e della lettura» (Perniola 2014). Ma se ad essere morti sono tutti, l'autore e il lettore, a chi vanno i diritti? Chi può decidere cosa fare con un'opera? O come nel caso in apertura della trattazione, chi può decidere cosa fare con la mia immagine, se nessuno è l'autore di nulla? È da questa riflessione che mi sembra opportuno suddividere il dibattito contemporaneo in due principali bisettrici: una prospettiva che definirò abolizionista del concetto d'autore ed una invece rinnovatrice. Alla base della riflessione sulle intelligenze artificiali, sembra necessario porre le considerazioni espresse da Luciano Floridi in *AI as Agency*

without Intelligence: on ChatGPT, large language models, and other generative models. È infatti utile ricordare che, come scrive il filosofo al termine della sua trattazione

We have decoupled the ability to act successfully from the need to be intelligent, understand, reflect, consider, or grasp anything. We have liberated agency from intelligence. So, I am not sure we may be “shepherds of Being” (Heidegger), but it looks like the new “green collars” will be “shepherds of AI systems”, in charge of this new form of artificial agency. (Floridi 2023)

A partire proprio da queste affermazioni, tenendo conto degli esperimenti condotti dal filosofo stesso su uno dei tool dell’intelligenza artificiale più utilizzati al momento della pubblicazione del suo contributo, ChatGPT-3, dimostrandone i limiti nella comprensione ed elaborazione di input linguisticamente complessi,⁴ sappiamo già che l’approccio a quella che lui stesso definisce «the agenda of a demiurgic humanity of this intelligence-free (as in fat-free) AI – understood as *Agere sine Intelligere*» (Floridi 2023) dovrà essere inevitabile e auspicabile. Per molti sarà entusiasmante, altri ne saranno preoccupati, ma indubbiamente sarà il futuro della discussione intorno al tema: meglio familiarizzare sin da subito con questa frontiera. Tenendo dunque sempre ben presente queste considerazioni, possiamo vedere come, le due bisettrici di pensiero si avvicinano diversamente alla questione dell’autorialità avendo come punto di partenza la tematica dei diritti di proprietà e della creatività artistica che compete alle macchine. Partiamo dal primo dei due gruppi. Per gli abolizionisti del concetto di autore (oltre a Barthes e Foucault troviamo autori contemporanei come Limonta e Samuelson), non potendo più parlare di un’unica entità detentrica dei diritti e rivendicatrice dell’artisticità dell’opera, occorre spostare la riflessione sugli altri elementi che compongono, ad oggi, un’opera (proprio come auspicava Barthes spostando l’attenzione sul lettore). È il caso di Roberto Limonta che in *I diritti dell’algoritmo: per una ontologia dell’“autorialità digitale”* partendo dagli esempi di *Questions*

4. Per approfondimenti sugli esperimenti che in questa sede non si riuscirà ad approfondire cfr. Floridi, L. (2023) “AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models.” in *Philos. Technol.* 36, 15. <https://doi.org/10.1007/s13347-023-00621-y>.

sur l'*Encyclopédie* di Voltaire del 1770 e di una lettera di Philip Roth indirizzata a Wikipedia del 2012, prendendo in considerazione i concetti di ontologia sociale di Searle, di intenzionalità collettiva, passando per la definizione di concetti come *snippet* e *token*, giunge alla conclusione che

Se il testo digitale si struttura in forma relazionale, ciò significa che l'algoritmo [...] ha assunto su di sé l'autorialità o buona parte dei diritti ad essa connessi. Posto che un processore matematico non può godere dei diritti di proprietà intellettuale che pure gli spettano sono quindi le piattaforme Web, creatrici e proprietarie di quegli algoritmi, a detenere di fatto quei diritti. Nel mondo apparentemente disintermediato del Web i diritti della proprietà intellettuale non sono oggetto di dibattito e negoziazione, ma condizioni di funzionamento del sistema. (Limonta 2021, 140)

La proposta è quella di attribuire alle stesse piattaforme web la proprietà intellettuale di ciò che, per utilizzare dei termini pratici anche se impropri, "vi si trova dentro". Questa soluzione genera però non pochi problemi, in primo luogo perché, ad oggi, i database delle piattaforme web contengono tutto ciò che esiste anche nel mondo "reale" (dove reale è solo il contrario di digitale, e non un modo di distinguere ciò che esiste da ciò non esiste), dunque deterrebbero così i diritti di ogni cosa esistente. In secondo luogo, dando agli utenti stessi il diritto di disporre come meglio preferiscono dei prodotti digitali a loro disposizione, si creerebbe un cortocircuito con quegli autori che ad oggi rivendicano la paternità delle loro opere: è il caso posto in apertura della riflessione di Limonta, in cui lo scrittore Philip Roth si ritrova a mandare una lettera a Wikipedia chiedendo di correggere delle informazioni inesatte circa un personaggio di una delle sue opere più famose, *La macchina umana*, e riceve come risposta un rifiuto poiché, citando testualmente la risposta dell'amministratore del sito Wikipedia, «comprendo il suo punto di vista che l'autore sia la più grande autorità sul proprio lavoro, ma noi esigiamo fonti secondarie». Sulla stessa scia si colloca il pensiero dell'informatica Pamela Samuelson, la quale, già nel 1985, dopo una disamina giuridica minuziosa dell'allora sorgente questione delle intelligenze artificiali, giunge alla conclusione provocatoria che forse sarebbe meglio che i diritti di proprietà di un output generato da un computer non siano di nessuno, siano lasciati "grezzi" e di pubblico dominio. Una proposta ovviamente paradossale, ma che vuole affermare

con forza la necessità di discutere sul tema. Sarà la stessa Samuelson a tornare a parlare di questo argomento, nel 2020, alla luce di sviluppi artistici e creativi non indifferenti dell'IA. Infatti, è del 2016 il progetto chiamato *The next Rembrandt*.⁵ Su questa scorta, scrive Maurizio Lana citando Samuelson, l'informatica analizza il caso:

[...] richiamando da un lato la posizione di principio dello U.S. Copyright Office: “The Office will register an original work of authorship, provided that the work was created by a human being” dall'altro ricordando che invece nel Regno Unito la normativa concede 50 anni di protezione all'opera “computer-generated” riconoscendo come autore la persona che ha utilizzato il computer. Di fronte alla varietà di posizioni emerse da studiosi e da sentenze statunitensi Samuelson conclude pragmaticamente che se gli output sono generati grazie ad un'attività dell'utente allora è l'utente che ha i diritti sugli output. (Lana 2022, 34)

Dando i diritti dell'opera prodotta all'utente che ne ha inserito l'input si va ad individuare una sorta di autore inconsapevole. Questa soluzione pone il problema foucaultiano del concetto di opera, dell'identificazione di cosa è un lavoro di natura artistica e cosa no. Ed è a questo proposito che interviene il secondo filone di pensiero (Larochelle, Medaglia, Benedetti, Lana) che in qualche modo sembra voler rinnovare il concetto di autore alla luce dell'attuale attività delle IA. Gilbert Larochelle, per citarne uno, con un esempio che potrebbe far sorridere ma che si rivela quanto mai efficace, partendo dal concetto stesso di diritto d'autore di matrice kantiana, pone in evidenza la problematica chiara e necessaria di avere invece un autore esplicito. Scrive infatti:

5. Progetto che vede uniti Microsoft, il Politecnico di Delt, la Rembrandt House Museum di Amsterdam e il Muritshuis, al fine di creare il ritratto di un uomo di circa 30/40 anni di origine caucasica, con barba e baffi folti, un vestito nero dal colletto bianco e un cappello. In questo dipinto l'opera dell'uomo si è “limitata” alla creazione degli algoritmi più appropriati per la realizzazione dell'opera, mentre al computer è spettato il compito più difficile: dopo aver analizzato e colto le tipizzazioni di Rembrandt, le geometrie compositive, i materiali pittorici, gli schemi ricorrenti nella realizzazione dei volti nei ritratti, e l'inconfondibile marchio di fabbrica del maestro delle luci e delle ombre), esso ha memorizzato 168.263 frammenti pittorici presi da un corpus di oltre trecento dipinti (in particolare fra il 1632 e il 1642), li ha poi acquisiti con uno scanner di altissima precisione in oltre 500 ore di scansioni e 150 GB di materiale. <https://www.nextrembrandt.com/>.

Non bisogna confondere il processo e le condizioni della storia: il primo non determina nulla di per sé stesso, ma le seconde fanno parte dell'interpretazione che si apre sull'opera stessa. Un esempio: sapere che *Mein Kampf* è stato scritto da Adolf Hitler anziché da Madre Teresa cambia radicalmente la lettura dell'opera, modifica il senso che le si accorda, se non addirittura esige una messa in relazione con certi dati della storia stessa che sono suscettibili di chiarificarla. Ciò non vuole dire che la produzione di questo libro sia segnata da una specie di destino ineluttabile. Ma il fatto che questo autore piuttosto che un altro sia stato nell'ambiente immediato di questa opera capovolge in modo fondamentale la percezione del messaggio che, per il resto, non esiste affatto, in maniera positivista, al riparo da tutto ciò che lo circonda. (Larochelle 2000)

Proprio su questo tema si innesta la questione dell'intenzionalità come discriminante fondamentale per dimostrare l'effettiva validità della figura dell'autore (Benedetti 1999). In accordo con Benedetti, la differenza tra un'opera che sia stata pensata e realizzata da un essere umano allo scopo di creare un'opera artistica e una invece creata dalle intelligenze artificiali starebbe proprio nella definizione stessa di opera d'arte, rintracciabile nella volontà umana di creare. Ciò che fa di un testo un testo letterario, di un dipinto un'opera d'arte, è proprio il punto di partenza, e la risposta non può che risiedere nel creatore della suddetta arte. Dunque, secondo Benedetti:

il mito della morte dell'autore non ha certo il potere di far sparire l'autore strategico o l'autore immagine [...] ha però come effetto di modificare profondamente nell'immaginario collettivo, lo statuto dell'artista. Lo scrittore non è più colui che genera o accresce, ma un semplice *scriptor*. (Benedetti 1999, 199)

Una rimodulazione del concetto che porterebbe l'autore ad un livello originario, inquadrato nell'atto di scrivere. Ciò però non convince Francesca Medaglia, ad esempio, la quale, all'interno del suo contributo *La Scrittura a quattro mani italiana* sostiene proprio che questa riproposizione del concetto di autore altro non è che una svalutazione, che invece di confermarne l'importanza lo regredisce ad uno stadio inferiore. Scrive Medaglia:

In definitiva, per salvare l'autore da una morte più volte annunciata nel corso dei secoli, lo riduce a semplice *scriptor*: si passa dalla morte

dell'autore ad un ruolo che lo svaluta e lo banalizza. Anche in relazione a ciò è possibile che alcuni studiosi, che hanno analizzato gli scritti di Barthes, Foucault e Derrida in relazione alla questione della morte autoriale, si siano, almeno parzialmente, lasciati fuorviare da titoli di saggi estremamente suggestivi, finendo per far emergere dalla loro analisi solo la questione della morte dell'autore: in tal modo hanno di fatto eliminato, in un irrigidimento semantico, qualsiasi apertura verso eventuali nuove ipotesi, e non sono riusciti ad individuare possibili (anche se poco chiari e poco evidenti) richiami alla scrittura multipla e plurale. (Medaglia 2015, 754)

Non dunque una soluzione, ma una non soluzione quella di Benedetti, che non solo non risolve il problema, ma svia quasi dal trovare una reale risposta, risemantizzando soltanto la parola stessa "autore". Ecco perché la conclusione di Medaglia è invece il passaggio da un singolo autore ad una collettività di autori, una scrittura a quattro mani per usare la sua stessa espressione, già impiegata nel futurismo e che trova i suoi diretti referenti in movimenti contemporanei come, ad esempio, l'esperienza del collettivo Wu Ming. Spostando in questo modo l'attenzione dal singolo al collettivo, si pone immediatamente la domanda: quale può essere il futuro possibile?

2. Co-autorialità o autorialità distribuita: un'ipotesi teorica

Per introdurre il tema delle nuove forme di autorialità, occorre partire da un presupposto necessario: oggi, i mezzi disponibili per "creare arte" sono molti e molteplici. Le stesse IA possono essere sfruttate a vari livelli di performatività, da un livello base - quale la revisione di un testo nella sua forma grammaticale, o la correzione di immagini tramite processi ad esempio di *color correction* - ad un livello più avanzato - come già osservato nel progetto *The next Rembrandt*. Mi sembra utile dunque trattare in questo articolo il caso Hannes Bajohr, filosofo tedesco, che a ragione potrebbe essere considerato pioniere della letteratura digitale. Nel 2018, con la pubblicazione di *Halbzeug. Textverarbeitung*, porta all'attenzione della critica letteraria un nuovo modo di "fare" letteratura. Infatti, avvalendosi dei tool dell'IA, realizza un volume che a tutti gli effetti è un ibrido: realizzato con poemi e testi della tradizione, l'autore, tramite una serie di input, riesce a rielaborarli con l'intelligenza artificiale e creare nuove forme a partire da quelle

preesistenti (Bajohr 2018). Bajohr stesso, durante un'intervista, chiarisce i problemi legati alla violazione di diritto d'autore in cui potrebbe incorrere. Dice a tal proposito:

HB: Rein urheberrechtlich habe ich keine Probleme, weil ich immer Transformationen vornehme und so aus dem Ausgangsmaterial ein sichtbar „eigenständiges Werk“ mache (wie ja der juristische Terminus lautet). Das ist bei anderen Spielarten der konzeptuellen Literatur anders, vor allem im Bereich der Appropriationsliteratur, wo also vorhandene Texte eins zu eins übernommen werden – dort ist mir bisher aber nur ein Fall bekannt, bei dem es zu rechtlichen Konsequenzen gekommen ist. Eine andere Sache ist natürlich die Quellenlage: Voraussetzung für digitale Verarbeitung ist die Verfügbarkeit digitalen Texts (Aust 2019, 208).⁶

Una rimodulazione di ciò che stato scritto, e che permette così di eludere i problemi legati al copyright. Chi è l'autore dell'opera di Bajohr? La risposta è insita già nella domanda: è Bajohr, quasi senza ombra di dubbio. Ma senza l'utilizzo delle intelligenze artificiali il senso stesso del progetto del filosofo sarebbe totalmente stravolto. Non si tratta di semplice riscrittura o parodia: è un nuovo testo, una nuova opera, in cui la creatività artistica, per i più predominio ancora dell'opera umana, si unisce alla tecnica digitale, creando così un prodotto che è frutto tanto dell'uomo quanto della macchina. È in questo senso che, probabilmente, dovremmo iniziare a ragionare nell'ottica di un'autorialità distribuita, ripartita uniformemente tra tutti i partecipanti al processo creativo: autore, lettore, utente, piattaforma web, IA e fruitori di ogni genere. Una forma di co-autorialità, in cui ad essere distribuiti sono onori e oneri. Bisogna prestare attenzione però a identificare esattamente cosa si intende con autorità distribuita e co-autorialità. La differenza tra i due termini (autorialità distribuita e co-autorialità) sta, a mio parere, nell'accezione in cui il secondo è stato finora usato. Il suo principale impiego, infatti, lo si ritrova all'interno di opere

6. “Dal punto di vista del diritto d'autore, non ho problemi, poiché faccio sempre trasformazioni e creo così un lavoro visivamente 'autonomo' dal materiale originale (come dice il termine legale). Questo è diverso per altre forme di letteratura concettuale, soprattutto nell'ambito della letteratura di appropriazione, dove i testi esistenti vengono presi direttamente - finora, però, conosco solo un caso in cui ci sono state conseguenze legali. Ovviamente, la disponibilità di testi digitali è un altro aspetto importante per il processo di elaborazione digitale” (trad. mia).

scritte, per usare un termine in prestito da Medaglia, “a più mani”, in cui più autori si uniscono per creare un volume, un componimento, un dipinto, un prodotto collettivo. L’accezione che invece in questo contesto sembra essere più efficace ricalca quella di co-creatore che dà il già citato Lana, il quale tenta di inserire in un continuum uomo e macchina, percorrendo così una strada unificatrice. All’interno del suo intervento *Artificial Intelligence Systems and problems of the concept of author. Reflections on a recent book*, pone in evidenza due casi emblematici: il testo *Lithium-Ion Batteries. A Machine-Generated Summary of Current Research*, realizzato interamente da un’intelligenza artificiale della casa editrice Springer, e il caso dell’articolo *A Robot Wrote This Entire Article. Are You Scared yet, Human?* scritto interamente con GPT-3. Tralasciando le questioni specifiche sulle problematiche di natura editoriale, di verificabilità delle fonti e di resa del prodotto, Lana alla luce di questi esempi arriva a ipotizzare le macchine come co-creatrici del prodotto finale. Si dà così origine a quella che viene definita da Lana come “autore-costellazione”, in cui ad essere presi in considerazione, non sono i singoli individui, ma un intero sistema di input e output umani e digitali, in un continuo interscambio di ruoli:

Non la *morte* dell’autore ma la *nascita* di un nuovo autore, nascita che è proprio un venire-alla-luce attraverso attività che rendono pubblicamente visibile qualcosa che prima era sotteso, non visibile, nascosto. (Lana 2022, 38)

In questo senso la co-autorialità deve essere intesa. Ma poiché l’impiego di questa definizione potrebbe ancor ora generare fraintendimenti, ritengo migliore la prima definizione, ovvero quella di autorialità distribuita, intendendo con questa espressione proprio quel meccanismo dimostrato da Bajohr e teorizzato da Lana: la nascita di un nuovo autore, che non è soltanto un individuo o una macchina, ma una costellazione appunto, una collaborazione tra più parti che, cooperando, segnano l’avvio di un processo di trasformazioni profonde, in cui ad essere posti sotto revisione sono tutti gli apparati teorici, pratici, legislativi, culturali e sociali che ruotano intorno alla sua esistenza.

Conclusioni

Il sopravvento delle IA ci pone continuamente di fronte alla scelta di rifiutarne l'esistenza o di sfruttarle al meglio. Indubbiamente occorre ragionare nell'ottica di una mediazione tra le due correnti di pensiero. I problemi di natura pratico-etica sorgono nel momento in cui l'utilizzo di determinati tool, all'insaputa degli attori sociali coinvolti, vengono impiegati violando le norme vigenti in materia di diritto d'autore, proprietà intellettuale, eccetera. Trovare una soluzione è senza dubbio la priorità degli attuali sistemi governativi (Unione Europea in primis, in cui i dibattiti attorno al tema sono già stati avviati, proposte di legge presentata, ma al momento ancora nulla di concreto è stato realizzato). Attualmente sono molti i comitati di artisti, illustratori, scrittori, che in ogni parte del mondo si sono riuniti e hanno chiesto più tutele (è infatti già stata intentata la prima causa legale per furto di dati e violazione dei diritti di copyright contro le Intelligenze Artificiali,⁷ allo stesso modo è già stato creato il primo ente che vuole regolarizzare le IA in Europa,⁸ con annesso crowdfunding, che ha già raccolto circa 45.000 euro, per intentare cause contro l'appropriazione di dati⁹). Ciò che le IA fanno è già un cambiamento radicale. Come conclude lo stesso Lana all'interno del suo saggio:

Nella realtà quotidiana gli usi dei sistemi di Intelligenza Artificiale nel funzionamento della società, lo abbiamo già accennato, ci sono già e anche se non mancano gli aspetti positivi, frequentemente i sistemi di Intelligenza Artificiale aggravano le disuguaglianze e rafforzano i poteri fuori del controllo democratico dei cittadini. (Lana 2022, 36)

In questo ragionamento si colloca perfettamente la protesta di Hollywood posta in apertura. Sono proprio quelle modificazioni sottili che potrebbero aggravare o, in ogni caso, polarizzare le disuguaglianze che, se non normate, rischiano di diventare una lotta senza nemico.

7. <https://news.bloomberglaw.com/ip-law/first-ai-art-generator-lawsuits-there-are-future-of-emerging-tech>

8. <https://www.tomshw.it/hardware/egair-alla-scoperta-dellente-che-sta-cercando-di-regolarizzare-le-ai-in-europa-intervista-a-lorenzo-ceccotti/>

9. <https://www.gofundme.com/f/help-protect-our-art-and-data-from-ai-companies>

Riferimenti bibliografici

- Aust R.-M. 2019, *Das Schreiben unter digitalen Bedingungen* in “Germanica”, 64, pp. 201-212.
- Bajohr B. 2018, *Halbzeug. Textverarbeitung*, Suhrkamp, Berlin.
- Barthes R. 1984, *Il brusio della lingua. Saggi critici IV*, Einaudi, Torino.
- Beavers, A.F. & Jones, D. 2014, *Philosophy in the Age of Information: A Symposium on Luciano Floridi's*, in *The Philosophy of Information*. “Minds & Machines”, 24, pp. 1-3.
- Benedetti C. 1999, *L'ombra lunga dell'autore. Indagine su una figura cancellata*, Feltrinelli, Milano.
- Coeckelbergh M. 2015, *The tragedy of the master: automation, vulnerability, and distance*, “Ethics Inf Technol”, 17, pp. 219-229
- Compagno D. 2006, *L'autorialità nei testi interattivi in Narrazione ed esperienza. Per una semiotica della vita quotidiana atti del XXXIV Congresso dell'Associazione Italiana di Studi Semiotici Arcavacata di Rende (Cs)*.
- Della Galla B. 2018, *La funzione autoriale tra lotta politica e branding. Alcuni aspetti dei casi Wu Ming e Scrittura Industriale Collettiva*, “Altre modernità”, 18, pp. 77-91.
- Eco U. 1962, *Opera aperta: forma e indeterminazione nelle poetiche contemporanee*, Bompiani, Milano.
- Eco U. 1975, *Trattato di semiotica generale*, Bompiani, Milano.
- Elstermann A. 2020, *Computer-Generated Text as a Posthuman Mode of Literature Production*, “Open Library of Humanities”, 6(2), pp. 1-23.

Floridi L. 2023, *AI as Agency Without Intelligence: on ChatGPT, Large Language Models, and Other Generative Models*, “Philosophy & Technology”, 36(15).

Foucault M. 1994, *Scritti letterari*, Feltrinelli, Milano.

Lana M. 2022, *Artificial Intelligence Systems and problems of the concept of author. Reflections on a recent book*, in “JLIS.it”, 13(2), pp. 14-44.

Larochelle G. 2000, *Da Kant a Foucault: che cosa resta del diritto d'autore?*, “Bollettino telematico di filosofia politica”.

Limonta R. 2021, *I diritti dell'algoritmo: per una ontologia dell'autorialità digitale*, “Quaderni di filosofia”, pp. 129-142.

Lotman J. M. 1993, *La cultura e l'esplosione. Prevedibilità e Imprevedibilità*, Feltrinelli, Milano.

Medaglia F. 2015, *La scrittura a quattro mani italiana*, “Rivista di studi italiani”, XXXIII(1), pp. 753-767.

Perniola I. 2014, *Roland Barthes e Michel Foucault: l'autore è morto, viva l'autore*, “Ágalma. Rivista di studi culturali e di estetica, Autore, autorità, autorevolezza”, 28.

Trotta A., Ziosi, M., Lomonaco, V. 2023, *The future of ethics in AI: challenges and opportunities*, “AI & Soc”, 38, pp. 439-441.

Noemi Rita Sanfilippo

Luis Roma

noemirita.sanfilippo@gmail.com



ἘΠΈΚΕΙΝΑ

International Journal of Ontology
History and Critics

LUCA CARDONE

Immagini schizofreniche

Riflessioni fenomenologiche sulla creatività computazionale a partire
da Marc Richir

EPEKEINA, vol. 16, nn. 1-2 (2023), pp. 1-29

Philosophy of Technology

ISSN: 2281-3209

DOI: 10.7408/epkn.

Published on-line by:

CRF – CENTRO INTERNAZIONALE PER LA RICERCA FILOSOFICA
PALERMO (ITALY)

www.ricercafilosofica.it/epekeina



This work is licensed under a Creative Commons

Attribution-NonCommercial-NoDerivs 3.0 Unported License.

Immagini schizofreniche

Riflessioni fenomenologiche sulla creatività computazionale a partire da Marc Richir

Luca Cardone

Introduzione

La recente profusione di modelli di intelligenza artificiale quali *Chat GPT* e *Dall-E*, per citare i più noti, ha riaperto il secolare dibattito sulle potenzialità delle macchine. In particolar modo, il rilascio di questi sofisticati sistemi in una modalità mai prima d'ora così democratica sta serrando nuovamente i ranghi, alle più differenti latitudini del dibattito sia pubblico che accademico, di quelle due fazioni contrapposte già misurate da Eco in un celebre testo della metà dello scorso secolo. Vero è che a partire da un esponenziale incremento del potere di calcolo e delle performatività delle reti neurali, questi strumenti ci presentano la necessità di ripensare, da un punto di vista ontologico, il posizionamento dell'essere umano e delle "macchine intelligenti" nelle loro relazioni reciproche. Uno dei molteplici fronti di questo epocale dibattito abbraccia la questione della creatività, concetto classicamente associato alle facoltà dell'essere umano e che, dati gli evidenti sviluppi in ambito tecnologico, può essere ripensato in termini di *creatività computazionale* a partire da un'analisi di ciò che i modelli di intelligenza artificiale sono in grado di produrre. L'intento preliminare di questo lavoro, o il suo presupposto, è un esercizio di disancoraggio dalla tendenziosa opposizione *uomo-macchina* nel tentativo di interrogare categorie fenomenologiche dell'agente non solo nei riguardi di un soggetto umano, ma di un soggetto creante *tout court*. L'obiettivo è trovare il punto di intersezione all'interno di un preciso spazio filosofico quale la psicopatologia fenomenologica partendo da alcune riflessioni del fenomenologo belga Marc Richir.

1. Spazi, geometrie, corpi

In un testo di recente pubblicazione il matematico e informatico italiano Giuseppe Longo ha rilanciato le sue critiche nei confronti dell'eccessivo entusiasmo per i risultati raggiunti nell'ambito delle intelligenze

artificiali. In particolar modo, Longo sottolinea come anche le più sorprendenti operazioni linguistiche, sintattiche e di riconoscimento visivo dei modelli artificiali siano solo pallidi tentativi imitatori delle nostre modalità interpretative dei fenomeni del mondo: «A noi animali, interessano quasi esclusivamente le pregnanze, quel che è significativo in un contesto» poiché «l'intreccio dei significati dà le pregnanze che ci interessano del mondo. Le salienze, evidenti anche alla macchina, sono una piccola componente, spesso trascurabile, del nostro rapporto con il mondo» (Longo 2021, 17). Per l'umano, e così per l'animale, le salienze sono importanti solo laddove diventino pregnanti, ove dunque siano utili per *progettare* un'azione, *progettarsi* in una determinata dimensione e ambiente. Longo aggiunge così un tassello a quel discrimine tra sintassi e semantica – o *segno* e *segnale* (Sini 2009, 44)¹ – che reclude le macchine a meri giocolieri di blocchi sintattici e con nessuna capacità di comprendere i messaggi che pure sono in grado di comporre. Alcuni evidenti errori delle intelligenze artificiali, come lo scambiare un oggetto con un altro per via di piccoli disturbi di un'immagine, ininfluenti invece nella percezione umana, manifestano per Longo il «vuoto lasciato dal significato, che è l'unica garanzia perché le azioni e l'intelligenza animale siano robuste» (Longo 2021, 20). La nostra percezione delle forme è sempre fondata su una «costruzione storica di senso» (Longo 2021, 21), cioè sul fatto che le esperienze sono sempre vissute da un corpo biologicamente e storicamente inquadrato. Dell'intero complesso armamentario critico sviluppato da Longo è proprio la questione del corpo a risultare maggiormente interessante per gli scopi della mia proposta, o meglio, l'idea di rimozione del corpo – inteso come luogo e fondamento del pensiero – a partire da una profonda crisi (in ambito matematico, fisico e filosofico) dello spazio e della geometria. Sia i sistemi formali che le reti neurali sono, per Longo, ereditari di una rimozione dello spazio e dell'intelligibilità geometrica. Approfondendo l'aspetto genealogico di questa rimozione egli riporta due esempi: il primo è l'operazione di Cantor che formulando le basi della teoria degli insiemi (1877) «dimostra che si può codificare il piano cartesiano, e, in effetti, ogni spazio di dimensione finita, con la retta» (Longo 2021, 58). In questo modo, sostiene Longo, abbiamo

1. Si veda anche R. Ronchi (2003).

perso l'intelligenza dello spazio; esperiamo il venir meno di un senso matematico a partire dall'assenza di una struttura topologica adeguata. Da qui in poi tutto è spazialmente codificabile con sequenze di 0 e 1. Il secondo esempio è l'unificazione per via algebrica di geometrie differenti, come nel caso delle operazioni di Beltrami e Klein che, nella seconda metà dell'Ottocento, immergono una geometria nell'altra. Se sul piano bidimensionale non riscontriamo problemi, aggiunge Longo, quando dal punto di vista rappresentativo si superano le due dimensioni «si perde il senso fisico delle geometrie» (Longo 2021, 59); ovvero, la traduzione di un sistema geometrico nell'altro fa perdere sia la fisica che la relatività del singolo sistema. Questo, per Longo, succede anche nell'informatica odierna che, mancando di quel senso dato dalla struttura spaziale, non conserva le varianti interessanti. Stabilendo che la razionalità umana è traducibile nei “se”, “allora” e “altrimenti” logico-formali, questa è stata trascinata fuori dall'uomo: «la certezza razionale è nella macchina» (Longo 2021, 54). Se a queste condizioni aggiungiamo che, come osserva anche Hofstadter, i problemi dei teoremi di incompletezza di Gödel sono, sull'ampio raggio, i medesimi sia per le macchine che per gli umani (Hofstadter 1979, 513), siamo già alle soglie di quella che Longo definisce «un'operazione schizofrenica» (Longo 2021, 54). L'operazione formalista è stata quella di trasferire la razionalità dapprima in qualcosa di meccanico, per poi consentire agli ingegneri di costruire le macchine su questo modello. L'uomo ora è fuori da sé, «fuori dal suo corpo, dal suo cervello, e in definitiva, dalla sua reale esperienza di vita» (Longo 2021, 130). In un testo del 1989 anche il filosofo francese Gilbert Simondon ricorre alla comparazione psicopatologica sottolineando come «La macchina può guastarsi, presentando in tal caso qualche analogia con il comportamento folle di un essere vivente. Ma non può rivoltarsi. La rivolta implica, infatti, una profonda trasformazione dei comportamenti finalizzati, non una irregolarità del comportamento» (Simondon 1989, 255).² La prospettiva che qui di seguito esploro è dunque la presa in consegna – in senso forte – della caratterizzazione schizofrenica della computazione proposta da

2. La posizione simondoniana è tuttavia più complessa rispetto a quanto questo punto possa suggerire, e proprio tale complessità tenterò di recuperare nelle conclusioni di questo lavoro.

Longo come vera e propria pista di ricerca, a partire da una precisa prospettiva epistemologica come la fenomenologia richiriana.

2. Fenomenologia e schizofrenia

L'operazione di rifondazione della fenomenologia del filosofo belga Marc Richir è oggi poco conosciuta e ancor meno diffusa in termini di traduzione. Tuttavia, alcune intuizioni capitali dell'opera di Richir rappresentano certamente non solo tentativi di rifusione (*refonte*) di ciò che la tradizione – o metodologia – fenomenologica ha prodotto su entrambe le rive del Reno a partire dal secolo scorso, ma anche un movimento di scarto e di differenza: l'inizio di un sentiero originale. Varrà la pena riportare, seppure in modo sommario e sintetico, alcuni fondamenti del pensiero richiriano necessari rispetto a quanto in seguito si dirà della psicopatologia. Come scrive Alexander Schnell «Ciò che presiede a tutto il suo progetto filosofico è la messa in discussione dell'idea che qualsiasi relazione con il mondo – sia essa “affettiva”, “comprensiva”, “intellettuale”, ecc. – avrebbe la sua origine e fonte in un “soggetto” che si riferirebbe a un mondo “esterno”» (Schnell 2011, 96).³ Se questo è valido anche per molti altri rappresentanti della fenomenologia contemporanea, l'interesse dell'operazione richiriana verte sulle operazioni non soggettive e anonime del senso nel suo farsi. Ponendosi all'interno di una riflessione sul dualismo tra l'appropriazione del senso da parte del soggetto trascendentale e l'affettività dei fenomeni, o in altri termini, tra schematico e proto-ontologico, Richir incontra la difficoltà di dover definire questa dimensione pre-soggettiva del farsi dei fenomeni in una modalità che da un lato non può essere intesa nei termini di oggettività, e dall'altro non incontra neanche il favore della terminologia psicanalitica (*inconscio*). Qui nasce per Richir la necessità di una nuova base *architettonica* della fenomenologia: «A differenza di Husserl, questo punto di partenza non deve più essere ricercato nelle esperienze intenzionali della coscienza, vale a dire negli atti oggettivanti per i quali la percezione fornirebbe la scala rispetto alla quale verrebbe misurata qualsiasi relazione con l'oggetto, ma nelle *phantasiai*, vale a dire nei tipi di “rappresentazione”, specifici della

3. Con l'intento di esorcizzare il silenzio italiano sul lavoro di Richir, tutte le citazioni originariamente in francese sono state da me qui tradotte.

“*phantasia* (*Phantasie*)”, che sono pre-intenzionali e si trovano al di sotto di qualsiasi percezione oggettivante. Queste *phantasiai* non figurabili appaiono solo sotto forma di “profili [*silhouettes*]” (incoative) o “ombre”, inaccessibili a una presa immutabile e, quindi, non suscettibili di fissazione» (Schnell 2011, 97).⁴ Richir intende dunque indagare ciò che è sotteso alla totalità dell’esperienza umana, nel suo fondamento arcaico, come base “immaginativa” dell’intenzionalità, pre-intenzionale e in una dimensione pre-spaziale, pre-temporale, pre-simbolica. Questo fondamento è la *phantasia*, (termine che Richir non traduce) ossia quella diretta “rappresentazione” dell’oggetto della percezione in quanto cosa stessa – che è sempre apparizione di *phantasia* (*Phantasieerscheinungen*) (Richir 1998, 6) – allo stato incoativo e selvaggio, nel suo non esser stato ancora codificato simbolicamente, quindi, come *Wesen sauvages*, pura fenomenalità bruta del mondo: «La *phantasia* emerge e scompare per bagliori (*blitzhaft*), in modo intermittente e discontinuo, è proteiforme (*proteusartig*) e soprattutto non presente (*nicht gegenwärtig*)» (Richir 2003, 26). Ciò che è fondamentale nella rifusione fenomenologica richiriana, più che l’apparire, è ciò che non appare del fenomeno. Questa doppia articolazione del fenomeno – del visibile e dell’invisibile – che Richir recupera dall’ultimo Merleau-Ponty, incontra un’altra idea del pensatore belga e cioè il fatto che «il senso abbia luogo al di sotto di una “soggettività costituente”» (Schnell 2011, 99). Dunque, la *Sinnbildung* è qui formazione anonima e ancora a-soggettiva e non direttamente accessibile fintanto che non simbolicamente istituita (*Stiftung*) in un processo già non più pre-soggettivo, ma calato in un tempo e in uno spazio. I fenomeni, intesi come *nient’altro che fenomeni* (*rien que phénomène*), non si danno in carne ed ossa come per Husserl, ma nella loro dimensione selvaggia e immatura fintanto che non istituiti simbolicamente (Gozé & Fazakas 2021, 215). Ora, come anche per Heidegger ben prima di Richir, l’accesso al mondo è sempre caratterizzato da una *Stimmung*, una certa tonalità emotiva, un certo *colore dell’atmosfera* – o *esistenziale* direbbe Heidegger – che non è temporalizzato in presenza e non ha un significato linguistico. La *Stimmung* ha per Richir i caratteri di *immemorialità* e di *immaturità* e, infatti, è definita

4. Nel riportare la citazione di Schnell tradotta ho preferito lasciare intatta la forma accentata del termine *phantasia* adoperata dall’autore, laddove in Richir esso invece ricorre in forma non accentata.

come *Wesen selvaggio* (Richir 2000a, 60) parimenti a quei fenomeni che al fondamento fenomenologico ancora pre-individuale incontriamo proprio nelle “rappresentazioni” di *phantasia*: «In effetti i tratti della *Stimmung*, tra cui quello di sfuggire completamente alla “presenza” della coscienza [...] sono per noi dal punto di vista fenomenologico straordinariamente arcaici: sono nella nostra vita e nella nostra esperienza [...] giunti fino a noi, dall’immemorabile e dall’immaturo che c’è nelle profondità sepolte dei fenomeni del mondo. Ma nella nostra vita o nella nostra esperienza, questi testimoni sono apparentemente tagliati dalla loro origine fenomenologica» (Richir 2000a, 61), se non per quei bagliori – che non si lasciano codificare simbolicamente – che costituiscono l’atmosfera della *Stimmung* stessa. Inoltre, «c’è un legame originario tra la *Stimmung* e il corpo, non il corpo oggettivo della fisiologia, ma il corpo vivente, il *Leib* nella sua *Leiblichkeit*» (Richir 2000a, 61). Riprendendo la distinzione husserliana tra *Innenleiblichkeit* e *Aussenleiblichkeit*, Richir sottolinea però il fatto che la *Stimmung* è prima di tutto correlata alla prima piuttosto che alla seconda, e i motivi per i quali questo dettaglio è importante, così come la ragione della divisione stessa (al di là della distinzione tra *Körper* e *Leib*) saranno presto chiariti. È qui che possiamo inserire in modo secante l’interesse di Richir per le psicopatologie e per la schizofrenia in particolar modo. L’indagine richiriana sulla schizofrenia è analizzabile per mezzo di almeno tre blocchi temporali (Gozé & Fazakas 2021, 214-215), ossia la storia di una processualità di idee che in questa sede non posso restituire nella sua frammentarietà ma piuttosto ricomprendere in un singolo e sintetico respiro. Obiettivo della psicopatologia fenomenologica è per Richir quello di intendere la patologia in modo non deficitario, cioè non nei termini di “mancante di”, ma come vera possibilità di ricerca in campo fenomenologico (Richir 1992, 81), un metodo in grado di distinguersi sia dagli approcci più riduzionisti del campo scientifico che dagli eccessi di teorizzazione che Richir imputa alla psicoanalisi.⁵ L’obbligo di inquadrare il concetto di *Stimmung* nel pensiero di Richir assume qui rilevanza quando il filosofo belga definisce il tratto distintivo dello schi-

5. Richir accusa anche quella stessa fenomenologia che indaga le psicopatologie laddove essa pecchi per “difetto”, ricadendo nelle sue stesse difficoltà. In particolar modo Richir mette a tema la dipendenza della ricerca fenomenologica dai concetti classici.

zofrenico come un *disturbo della Stimmung*, ovvero come trasposizione della *Stimmung* in *Verstimmung* (Richir 2000a, 61). Già a partire da un articolo del 1992 dal titolo *Phénoménologie et psychiatrie. D'une division interne à la Stimmung*, Richir comincia a pensare ad una distinzione tra la fluttuazione non temporalizzata della *Stimmung* e il suo passaggio corrispettivo nella dimensione patologica, cioè in *Verstimmung* (Richir 1992, 85). Se nella salute la *Stimmung* «può rivelare una divisione delle modalità di essere delle sintesi passive durante la formazione del senso (*Sinnbildung*) attraverso l'istituzione simbolica», al contrario nella *Verstimmung* patologica «c'è come un "vuoto", un "buco"» nella fenomenalizzazione del mondo (Gozé & Fazakas 2021, 215). Si diceva poc'anzi di come la *Stimmung* sia legata all'interiorità del corpo vissuto (*Innenleiblichkeit*); piuttosto, ciò che Richir cerca di mettere in luce è uno strutturale movimento per cui la *Stimmung* passa dall'interiorità dell'*Innenleiblichkeit* all'esteriorità dell'*Aussenleiblichkeit*, mettendo in discussione in modo sano l'istituzione del soggetto singolare e la sua integrità. Nel caso della *Verstimmung* patologica, invece, «la *Stimmung* viene accolta in modo del tutto passivo come un'invasione totale, cioè senza trasparire a nessun evento o a nessuna inversione della *Stimmung*». Il risultato qui è i) la totale passività del *Leib*, ii) la quasi integrità totalizzante dell'*Aussenleiblichkeit* e iii) l'*Innenleiblichkeit* in uno stato di *imminente scomparsa*. Non c'è più spazio per la coscienza e per l'elaborazione di un senso che, piuttosto, lascia il proprio posto alla propagazione di un delirio. Richir specifica come questa sia una *tendenza alla totalità* piuttosto che una stasi o un processo richiuso su se stesso (Richir 2000b, 423), poiché se così non fosse «non ci sarebbe quindi nulla di fenomenologicamente attestabile nell'esperienza schizofrenica che non sia più che un'eclissi di se stessa» (Gozé & Fazakas 2021, 217). Nonostante tutto, c'è un senso che continua a farsi nella psicosi, ma questo è minato dall'imminenza del suo collasso. Ricapitolando, la mutazione della *Stimmung* in *Verstimmung* psicotica – o, come la definisce Richir, la *divisione interna alla Stimmung* – è una cattura totale degli spazi interiori del *Leib*, della coscienza e della costruzione di senso, «un'irruzione violenta e tirannica della *Stimmung* come proveniente dall'esterno» (Gozé & Fazakas 2021, 217) piuttosto che una fluttuazione originaria. La schizofrenia, per Richir grande signora delle manifestazioni della *Verstimmung*, è caratterizzata da una parziale – sempre tendente alla totalità – chiusura del contatto affettivo,

come se la corporeità interiore fosse enucleata e poi relegata all'esterno, minacciando la capacità di garantire la coerenza dell'esistenza: «è la paura che il minimo gesto possa far collassare sia il sé che il mondo» (Gozé et al. 2017, 406). Lo spossessamento schizofrenico dell'interiorità (*Innenleiblichkeit*) produce una totale ritirata all'esterno mentre quest'ultima si sgretola. Per questo motivo, per lo schizofrenico, spesso i gesti più ininfluenti possono avere una portata capitale, divenire una questione di vita o di morte poiché le normali relazioni tra interiorità ed exteriorità sono perturbate. Nell'esperienza non delirante il reale è sempre incontrato attraverso il linguaggio, quindi, espresso nel passaggio dalla *Sinnbildung* alla sua *Stiftung*. Può accadere che il reale si srealizzi momentaneamente come per i *déjà-vu*, le esperienze allucinogene e gli attacchi di panico, ma qui «il giudizio di realtà è sempre pronto a prendere il sopravvento e a ripristinare l'ordine del mondo. Quando ciò non accade[...], allora parliamo di psicosi. Nella psicosi è un reale trasformato a imporsi a discapito della realtà» (Grohmann 2018, 109-110). Ma come avviene il passaggio da *Sinnbildung* a *Stiftung* simbolicamente istituita? Come può l'affettività immatura e immemorabile della *Stimmung* passare dal campo dei *Wesen selvaggi* alla sua forma figurata e percepibile? A partire dai lavori pubblicati dal 2004 in poi – cioè, dalla pubblicazione di *Phantasia, Imagination, Affectivité* –, Richir importa nella sua fenomenologia e nell'indagine psicopatologica il concetto di *dissociazione* (*Spaltung*) e ricollega la base architettonica della *phantasia* alla dimensione del corpo. Laddove la percezione (e l'immaginazione) si fonda sulla *phantasia* piuttosto che il contrario – come nel caso di Husserl –, Richir sottolinea come il *Leib* sia sempre allo stesso tempo un *Phantasieleib*: «*Leiblichkeit* non è quindi solo “di carne”, nemmeno nel senso (troppo restrittivo) di Merleau-Ponty, ma anche di *phantasia*. A condizione di specificare, seguendo Husserl (che non ne ha tratto tutte le conseguenze), che la *phantasia* non figurativa in “immagine” (“apparenza percettiva”), è discontinua, fluttuante, intermittente, proteiforme e soprattutto non presente. Non si tratta quindi di un “andare là” attraverso il *Phantasieleib*, il *Leib* figura il “là” come una semplice variante del “qui”» (Richir 2006a, 37). Percezione, immaginazione e *phantasia* sono per Richir a diversi registri dell'*architettonica fenomenologica*. In particolar modo, laddove la percezione si confronta con un presente e l'immaginazione con una rappresentazione, la *phantasia* è in rapporto con una «apparenza non posizionale di un qualcosa

che non è né presente né rappresentato» (Gozé & Fazakas 2020, 269). Se nella *trasposizione architettonica* – quindi il passaggio dal registro più arcaico (*phantasia*) all’immaginazione⁶ – la *phantasia* non resta nella sua dimensione di apertura e relazione tra sé e il mondo, allora essa si fossilizza e si sclerotizza in un’immaginazione delirante; e se, come già detto, la *phantasia* e l’immaginazione sono sempre incarnate, questa trappola qual è l’immaginazione è a sua volta una trappola – riprendendo un concetto merleau-pontiano – dello schema corporeo. Per eccesso o per difetto, questa affettività psicopatologica (trauma) non è propriamente vissuta nella *Phantasieleiblichkeit*: «il *Leib* e il *Phantasieleib* non possono più funzionare come qui assoluti spazializzanti ma neanche come pluralità di qui assoluti spazializzanti. Si sono trasposti in quello che noi chiamiamo un *Phantomleib*» (Richir 2006a, 41). È nella schizofrenia che per Richir questa situazione manifesta la sua forma estrema: del *Leibkörper* non resta che un *Körper* non più abitato ma colmato da un dolore indistinguibile da quello fisico; del *Phantasieleib* non resta che un *Phantomleib* violento e non localizzabile, non altro che il fissaggio di una rappresentazione immaginativa ani-

6. Per meglio chiarire i rapporti tra *phantasia* e immaginazione, in *Appendice I. L’ingifurabile en peinture del testo Sur le sublime et soi Variations II*, Richir riprende la distinzione husserliana tra *Bildobjekt* e *Bildsujet*, per la quale, dato il legame di dipendenza del primo rispetto al secondo, è l’intenzionalità dell’immaginazione che mira al *Bildsujet* a potersi figurare l’immagine, visualizzata per mezzo del *Bildobjekt*. Ma quest’ultimo non rappresenta nulla di per sé e non è nemmeno rappresentato di per sé da nulla poiché svanisce nel *Bildsujet* rappresentato. L’immagine, quindi, «a meno che non sia posta su un supporto fisico che ne garantisca la stabilità, non esiste in e per sé stessa, e non immaginiamo immagini, ma oggetti, secondo intenzioni che, come ha dimostrato Husserl, sono più o meno vaghe e instabili. Se, per prendere un esempio famoso (Alain, Sartre), possiamo contare le colonne della chiesa del Pantheon su una cartolina, non possiamo farlo sul Pantheon semplicemente immaginato. Questo significa che il simulacro, al di là dell’essere e del non-essere, comporta non solo una dimensione di “finzionalizzazione [*fictionnalisation*]”, ma anche una dimensione di indeterminatezza. Questi sono, tra gli altri possibili, gli effetti che ci portano a pensare che, paradossalmente, non è assoluto nulla e che l’immaginazione non “crea” i suoi oggetti dal nulla, ma si “fonda” (*fundiert*) su “qualcosa” (*etwas*) che, seppur virtuale e fugace, non è privo di “effetti”. Questo “qualcosa”, base fenomenologica dell’immaginazione, e più o meno ben distinto da Husserl, è ciò che noi chiamiamo *phantasia*. *Phantasia* che certamente non è figurativa né rappresentata, altrimenti si identificherebbe con l’immaginazione» (Richir 2011a, 134).

mata da fantasmi senza corpo.⁷ In un articolo del 2011 dal titolo *De la figuration en psychopathologie* Richir analizza l'attività figurativa degli schizofrenici a partire dalle analisi fenomenologiche che fin qui ho sinteticamente riportato. Quando siamo di fronte a un dipinto, sostiene il filosofo belga, ciò che incontriamo è il *Phantasieleib* del pittore rivolto alle cose, il suo sguardo colto nella *phantasia*. L'opera «ci mette di fronte allo stesso sguardo dell'artista, alla sua *phantasia* “percettiva” che possiamo “percepire” solo in *phantasia*. [...] Quel che il pittore dipinge, quindi, non sono né cose, né esseri né paesaggi, né tantomeno “visioni” raffigurate ma non figurative, ma piuttosto uno sguardo, il suo sguardo, il “percepito” in *phantasia* della sua *phantasia* “percettiva”» (Richir 2011b, 572).⁸ Cosa dire, allora, delle figurazioni schizofreni-

7. «il *Phantasieleib* [...] come base fenomenologica della *Stiftung* della relazione (d'intersoggettività) “passa” in un *Phantomleib* dissociato dal *Leibkörper*, quindi tanto dal *Körper* quanto dal *Leib* (primordiale modificato), a un punto tale che è questo *Phantomleib* che si considera come *psyche* o come anima. Non è per noi sorprendente che questo “passaggio” si accompagni al “sentire” della disintegrazione, del tracollo o della confusione acuta poiché, quasi tutta la “vita” del “soggetto” (la sua “anima”) è catturata dall'impressionabilità e dall'affettività della *Leibhaftigkeit* in secessione (il “quasi” che lascia peraltro dello spazio al “sentire”, a un “vissuto limite” dell'imminenza del tracollo) e ciò nell'incoattività dei “fantasmi inconsci”, cioè non ancora presentificati a presenti intenzionali in questo stadio (vuoti o intuitivi) di immaginazioni, e tuttavia già fissati o strutturati o in via di esserlo, attraverso strutture intersoggettive di significatività (specificatamente amore e odio) senza l'altro [...]». Cfr. Richir (2006b, 800). La citazione è tratta dalla nota 2 di A. Lanciani.

8. Nonostante le evidenti divergenze tra l'approccio fenomenologico e quello psicanalitico – si veda la differenza d'uso tra i termini *Phantasia* e il più freudiano *Phantasieren*, o la differente e opposta caratterizzazione dell'intervento del *fantasma* – è interessante cogliere i punti di contatto tra le due metodologie che, per quanto riguarda la figurazione, sembrano andare nella medesima direzione ma con differente vocabolario. Come scrive Gianluca Solla nella sua indagine sulla *formula* freudiana del disegnare, ciò che è in gioco è «quel fuori senso al quale in quanto soggetto ciascuno è esposto ovvero “assoggettato”. Il disegno evoca questo luogo lasciato vuoto da qualsiasi significazione, questo buco dell'interpretazione. I suoi tratti godono di esperienza propria, al di qua di qualsiasi istanza di senso possa impadronirsi del loro evento. È proprio questo carattere a permettergli d'essere non l'ermeneutica di qualcosa, ma di esistere come invenzione» (Solla 2022, 20). Ancora, scrive Solla intercettando anche la terminologia richiriana, in quanto risultato dell'elaborazione di qualcosa, i segni sono «sempre dei “tratti arcaici”, che caratterizzano la vita psichica nella sua singolarità. Nel disegno, infatti, il segno va incontro non tanto alla sua espressione, che richiederebbe poi una decifrazione, ma incontro all'incryptazione ossia in quella che Abraham e Torok hanno descritto come una scissione dell'Io che porta con sé la

che, laddove la figurazione è una rappresentazione dello sguardo nella *phantasia* e la psicopatologia è l'enuclearsi della *phantasia tout court*? Nel caso della salute mentale, sostiene Richir, il *Leib* coincide con il *Phantasieleib* e la *Spaltung*, la scissione del sé dal sé, è sempre dinamica. Riprendendo una terminologia medica – già utilizzata da E. Levinas e H. Maldinay – il filosofo belga spiega questo movimento come di un passaggio da uno stato di *sistole* a quello di *diastole*. Sistole è lo stato iper-denso dell'affettività mentre diastole il suo rilassamento, che risulta dal distacco o dalla fuga infinita dell'eccesso della affettività. Nel caso patologico schizofrenico, invece, parliamo di una fissazione della *Spaltung*: un irrigidimento nello stato di sistole con conseguente distruzione della diastole. La rottura del movimento diastolico, secondo Richir, fa esplodere in frammenti, brandelli di senso arbitrariamente disseminati, i contenuti della rappresentazione di *phantasia* in una modalità non più unitaria. Qui insorge un sé che non può più fare il senso ma è piuttosto disposto dai brandelli di senso, divenendo esso stesso la traccia *ex nihilo* della diastole collassata (Richir 2011b, 576). Bloccato nello stato iper-affettivo della sistole, durante l'atto di figurazione lo schizofrenico produce immagini «traboccanti di immaginazione, ma [che] mancano dolorosamente di *phantasia*, quindi di profondità in cui lo sguardo possa risiedere. Una sorta di dissezione del cervello dell'artista su un tavolo da dissezione» (Richir 2011b, 575). Questo accade poiché allo stesso atto di distruzione dello stato di diastole è correlata un'evaporazione delle *phantasiai* che fanno spazio a intenzionalità immaginative vuote, simulacri figurati del vuoto: «Un "mondo" senza *phantasia* non è un mondo. È al massimo un deserto dove appaiono solo miraggi» (Richir 2011b, 578). Dietro la canonica e caotica struttura delle figurazioni schizofreniche noi vediamo il farsi in figura di fantasmi, non altro che l'estrema affezione che si irrigidisce nell'immaginazione fino a produrre simulacri – percepiti come assolutamente reali – che fanno un tutt'uno con il corpo dello schizofrenico (*Phantomleiblichkeit*) non più realmente abitato nell'interiorità.

sua "feticizzazione nell'inconscio". (Solla 2022, 158). Voglio quindi sottolineare come tra la caratterizzazione psicanalitica del segno e l'approccio fenomenologico vi sia una netta convergenza sul piano della processualità e un disaccordo sulla struttura del processo.

3. Creatività computazionale: GAN, Stable Diffusion e Dall-E

Mettendo da parte la psicopatologia fenomenologica vorrei ora soffermarmi su una precisa tipologia di tecniche computazionali che in questi anni stanno trasformando il mondo della grafica e le modalità attraverso le quali ci avviciniamo alla produzione di immagini digitali. L'attuale avanguardia tecnologica delle intelligenze artificiali in grado di generare immagini è rappresentata oggi dai *modelli di diffusione* o *stable diffusion models*, modelli generativi utilizzati per produrre dati simili a quelli attraverso i quali tali modelli vengono addestrati durante la fase di training.⁹ Grazie al rilascio di *Dall-E* prima e *Dall-E 2* poi da parte di *OpenAI*, il numero di utenti che sperimentano quotidianamente la produzione di immagini digitali originali con questa tecnica, semplicemente inserendo un *prompt* nella casella di testo, è in vertiginoso aumento. Il primo modello è stato rilasciato per la prima volta ad agosto 2022 da parte del gruppo *CompVis* di Monaco ma questa tecnologia ha una sua storia genealogica. Vero e proprio lavoro pionieristico è stato quello di *AlignDRAW*, un sistema in grado di generare immagini a partire dal linguaggio naturale ma soffrendo tuttavia di un forte tratto irrealistico rispetto alle immagini prodotte (Zhang *et al.* 2023). Da qui in poi, tuttavia, le soglie del possibile erano già spalancate. Successivamente è stato sviluppato il sistema *GAN*, acronimo di *Generative Adversarial Network*, un tipo di architettura di rete neurale – *end to end* – utilizzata per generare dati sintetici somiglianti a quelli reali. Una *GAN*, nello specifico, è composta da due componenti principali: un generatore e un discriminatore. Il generatore riceve in input un rumore casuale e produce un'immagine sintetica, mentre il discriminatore prende in input un'immagine (sia reale che sintetica) e cerca di distinguere tra le immagini reali e quelle generate. Durante il processo di addestramento, il generatore cerca di migliorare la sua capacità di generare immagini realistiche, mentre il discriminatore cerca di migliorare la sua capacità di distinguere tra immagini reali e immagini generate. Questo processo avviene attraverso un ciclo di feedback: il generatore cerca di ingannare il discriminatore generando

9. Per lo più i sistemi in grado di produrre immagini a partire dalla tecnologia text to image, quindi basati sull'accoppiamento di parole e immagini, vengono allenati su grandi archivi di dati come LAION-5B (Ghosh & Fossas 2022).

immagini sempre più realistiche, mentre il discriminatore cerca di migliorare la sua capacità di discriminazione. L'obiettivo finale della GAN è raggiungere un punto di equilibrio in cui il generatore sia in grado di generare immagini indistinguibili da quelle reali e il discriminatore non sia in grado di rilevare la differenza tra le due. Le cosiddette tecniche auto-ricorsive dei modelli di diffusione hanno tuttavia fatto emergere una certa obsolescenza dei sistemi GAN, registrando dati che hanno portato alla sua quasi totale sostituzione. Esistono oggi tre categorie di modelli di diffusione: i *modelli probabilistici di diffusione del rumore* (DDPM), i *modelli generativi basati sul punteggio* (SGM) e le *equazioni differenziali stocastiche* (SDE). Tutti e tre i sistemi, tuttavia, lavorano sulla generazione e la riduzione di rumore fino all'ottenimento di un'immagine nuova. Entrando più nello specifico, questi sistemi di generazione d'immagini lavorano per mezzo di *catene di Markov e rumore gaussiano*. Per quanto riguarda i modelli probabilistici di diffusione di *denoising* (DDPM), ad esempio, possiamo definirli come una catena di Markov parametrizzata che genera immagini a partire da un rumore con transizioni finite durante l'inferenza. Nella fase di addestramento i *kernel* di transizione vengono appresi in direzione inversa rispetto alla perturbazione delle immagini naturali con rumore gaussiano, aggiunto ad ogni passaggio fino all'ottenimento dell'immagine finale (Zhang *et al.* 2023). Nel caso di *Dall-E* di *OpenAI*, il primo sistema era basato su un modello ausiliario separato – o classificatore – denominato *CLIP* (*Contrastive Language-Image Pre-Training*): a partire da un grande set di dati accoppiati (immagine-testo) viene addestrato un modello contrastivo. I modelli possono produrre un punteggio alto (somiglianza) per un'immagine e un testo della stessa coppia o un punteggio basso per testi e immagini non corrispondenti. Con lo sviluppo del modello di diffusione *GLIDE*, allenato sulla stessa banca dati ma in grado di inserire rumore casuale ai dati di input attraverso una catena di passaggi di diffusione, il modello generativo ha imparato a invertire il processo per costruire immagini a partire dal rumore stesso. Questi modelli di diffusione possono avere un classificatore di etichette, come *CLIP*, oppure non averne affatto, facendo a meno di un classificatore separato. Il non utilizzo di un classificatore comporta il non avere un limite nel numero di associazioni che invece un classificatore necessariamente avrà nei termini di dati archiviati disponibili. Per *Dall-E 2*, il sistema attualmente disponibile e utilizzato da parte

degli utenti registrati su *OpenAI*, gli sviluppatori hanno incrociato il sistema CLIP e il modello *GLIDE* in quello che è stato nominato modello *unCLIP*. Evitando di addentrarmi negli estremi tecnicismi, parliamo di un codificatore di testo *CLIP* che genera un incorporamento per il testo di input, e da questo, un incorporamento di immagini (*embedded*) basato sull'incorporamento del testo nella fase di codifica. A questo punto un decodificatore di diffusione, non altro che un *GLIDE*, genera un'immagine basata sull'incorporamento dell'immagine precedente, aggiungendo rumore gaussiano casuale in una serie di passaggi finiti. Ora, compreso il meccanismo base di questi modelli, ciò che ritengo essenziale è comprendere se e come le immagini digitali in tal modo prodotte possano ritenersi creative. All'interno dell'ambito di ricerca della creatività computazionale viene definita creativa «L'esecuzione di compiti che, se svolti da un essere umano, sarebbero considerati creativi», e la sua variante computazionale come «Lo studio e il supporto, mediante mezzi e metodi computazionali, del comportamento manifestato da sistemi naturali e artificiali che sarebbe considerato creativo se esibito da esseri umani» (Wiggins 2019, 24-25). Come scrive Margaret Boden, madre delle ricerche sulla creatività computazionale, «Un sistema computazionale creativo deve essere in grado di situare l'idea originale all'interno di uno spazio concettuale definito da vincoli intelligibili. Almeno, deve essere in grado di valutare una nuova combinazione di idee come più interessante di un'altra» (Boden 2004, 243). In questo preciso frammento Boden sembra fare riferimento maggiormente al concetto di H-Creatività, ossia alla processualità creativa colta nel suo istituirsi storicamente e apportando modifiche culturali, piuttosto che alla creatività per come potrebbe essere intesa e percepita da un singolo (P-Creatività). Pare evidente l'essere oggi precisamente a cavallo di una formidabile transizione storica per la quale l'utilizzo dei modelli di diffusione per la produzione di immagini compute potrebbe divenire in futuro una pratica da scuola elementare, apportando modifiche a quelle norme esterne che oggi valutiamo come vincoli istituiti di ciò che culturalmente intendiamo come creativo. Andando al di là di quelle che ad oggi possono solo assumere la forma di mere speculazioni, ritengo sia più interessante in questo caso riflettere in termini di P-Creatività. Alcune recenti indagini sui risultati di Dall-E 2 hanno messo in evidenza come all'aumentare della complessità della semantica in fase di input i risultati non solo peggiorano in termini

di qualità dell'immagine ma, banalmente, il modello procede in una modalità fortemente *casuale* (Leivada *et al.* 2022). Un secondo ordine di problemi è legato al cosiddetto *overfitting*: le immagini prodotte spesso non si discostano troppo – a volte per nulla – dalle immagini sulle quali il modello è stato addestrato (Carlini *et al.* 2023) e, legato a quest'ultimo, da annoverare il problema della standardizzazione degli stereotipi: a partire dal campione già viziato da alcuni parametri stereotipati durante la fase di allenamento, i modelli generativi non sarebbero in grado di discostarsi da questi parametri (Fraser *et al.* 2023; Luccioni *et al.* 2023). Queste due tipologie di problemi manifestano una natura e una portata differente. Se l'*overfitting* e la standardizzazione degli stereotipi rientrano nell'ordine delle questioni tecniche risolvibili con il miglioramento delle fasi di allenamento e con una diversificazione dei dati, la questione della processualità casuale si scontra frontalmente con il problema della creatività computazionale. In *The Creative Mind*, Boden argomenta contro l'idea per la quale fattori quali la *casualità*, il *caos*, la *possibilità (chance)* e l'*imprevedibilità* non siano parametrizzabili, e quindi, indici di una creatività esclusivamente umana. Un'idea creativa, scrive Boden, «deve essere non solo nuova, ma sorprendente» (Boden 2004, 253), criterio discriminante valido sia per l'umano che per le macchine. Possiamo dunque utilizzare l'inaspettato e il non previsto come parametri di creatività per entrambe le forme di produzione (Grace & Maher 2019, 196). L'uso di questo discrimine va anche incontro alla necessità di distinguere tra il *pastiche*, il modello imitatorio di ciò che è già stato creato (ad esempio una composizione musicale nello stile di Bach che ne imita la modulazione e non introduce nulla di non previsto) dalla creazione originale vera e propria (Veale *et al.* 2019, 4). Come detto in precedenza, gli attuali modelli di diffusione utilizzano catene di Markov e rumore gaussiano più o meno controllato. Ciò vuol dire che ad un modello matematico che calcola le possibili transizioni a stati futuri solo a partire dalle condizioni presenti – e non da quelle passate – viene aggiunto un rumore, una distribuzione di probabilità simmetrica, o deviazione standard, che varia rispetto alla media dei dati. Nel concreto, per i modelli di diffusione dobbiamo immaginare una componente *x* sovrapposta al collage di immagini che il sistema monta nella prima fase di incorporamento (a partire dal testo), e che in una fase successiva viene armonizzato nel processo di *denoising*. Il rumore gaussiano può presentare alcuni parametri di controllo ma

anche non averne affatto. Ciò vuol dire che anche gli sviluppatori di questi modelli potrebbero non avere la minima idea di ciò che il sistema, a partire dal *prompt* fornito in partenza, potrebbe restituire in termini di immagini. Quando forniamo un *prompt* con le indicazioni sul tipo di immagine richiesta, sul contenuto e sullo stile, possiamo più o meno provare a formulare un'idea di ciò che la macchina ci restituirà sotto forma di immagine. Se la nostra richiesta è il disegno di una casa con il tetto a punta forse i risultati potrebbero rientrare pienamente nelle nostre aspettative. Dopotutto, non accadrebbe lo stesso se chiedessi ad un qualunque conoscente di eseguire il medesimo disegno? Se, invece, proviamo a formulare richieste più articolate la sensazione provata durante gli attimi di attesa è legata ad un'aspettativa vuota: non abbiamo la minima idea di cosa stiamo per guardare. Per questo tipo di richieste è più che evidente il fatto che i risultati vanno sempre, nel bene e nel male, al di là delle nostre aspettative.



Figura .1: Il *prompt* utilizzato per la generazione di queste immagini su Dall-E 2 è: “Post-industrial environment in realistic de Chirico’s paint style with block of matter and organism, minimal design photo”.

Utilizzando i parametri dell’aspettativa e dell’imprevedibilità come

discrimine possiamo far caso al fatto che l'uso di modelli di diffusione come Dall-E 2 sembra suggerire l'andamento di un grafico a campana. Quando l'informazione fornita nel *prompt* è minima ed essenziale i risultati in qualche modo sono mediamente in grado di andare al di là delle nostre aspettative;¹⁰ quando l'informazione è mediamente informativa i risultati tendono a non superare le nostre aspettative (come nel caso del disegno di una casa con il tetto a punta); laddove le informazioni inserite nel *prompt* sono complesse e articolate la prevedibilità torna nella condizione di poter essere tradita.



Figura .2: *Prompt 1*: “a clock”¹¹; *Prompt 2*: “black digital clock on a table”; *Prompt 3*: “a black digital clock on a table floating in the universe hit by a space cat”)

Adottato questo parametro, è in modo non corrivo che potremmo giungere a sostenere che, per lo meno in una prospettiva personale ed individuale, l'attività computazionale di questi modelli sia P-Creativa. Ora, va sottolineato il fatto che fin qui è stata presa in esame un'idea di creatività computazionale che fa capo a ciò che noi utenti o programmatori possiamo intendere come creativo rispetto ai risultati ottenuti, al di là del determinismo di fondo delle procedure computazionali. Questa, per lo meno allo stato attuale, è la principale pista di ricerca intrapresa nell'ambito di ricerca sulla creatività computazionale. Dovremmo piuttosto chiederci se questa sia effettivamente

10. La semplice richiesta dell'immagine di una casa, al di là di qualsiasi tipo di caratterizzazione descrittiva, lascia una sorta di indeterminazione che la macchina deve a suo modo colmare.

11. In questo primo caso, a partire dalla semplice descrizione “un orologio” avevo certamente ipotizzato la comparsa di un orologio analogico ma non la sua tipologia, forma o colore. Inoltre, la mia aspettativa era legata anche alla comparsa di orologi da polso e orologi di tipo digitale. Le quattro immagini che Dall-E 2 ha fornito attraverso questo prompt sono tutte immagini di orologi in vecchio stile, analogici, tre dei quali con numeri romani ed uno solo con numeri arabi.

l'unica via possibile. Continuando a interrogare il caso concreto di Dall-E 2, possiamo dire che la sua complessa attività è sintetizzabile nei seguenti tre punti: i) la selezione probabilistica di un *certo tipo* di immagini a partire dal testo inserito nel *prompt*; ii) l'inserimento di un *certo tipo* di rumore gaussiano; infine, iii) un *certo tipo* di attività di *denoising* di quello stesso rumore. Sottolineo la forte volontà d'uso di questa formula poiché, al di là della probabilità e del tipo di controllo che il sistema potrebbe o non potrebbe avere nell'inserimento del rumore, il sistema trattiene un paradossale e originario residuo d'indeterminazione all'interno del processo. Ma procediamo con ordine. È facilmente esperibile il fatto che a partire da un medesimo *prompt* il sistema fornisce sempre immagini differenti. Questa variazione è data dall'utilizzo di un *seed* che fa variare i risultati pur essendo chiamate in gioco le medesime regole e assiomi della macchina. Parliamo di una cifra auto-generata dalla macchina per mezzo di una formula determinata (come una certa operazione algebrica eseguita con le cifre dell'orario al momento dell'invio dell'input o di alcuni parametri meteorologici). Questo fattore immette nel sistema un'ulteriore cifra di casualità, tuttavia, definita nel campo dell'informatica come *pseudo casualità*. Mantenendo lo stesso *seed* la macchina restituirebbe sempre i medesimi risultati. Questo è riscontrabile in alcuni modelli generativi che consentono di modificare il *seed*, dando agli utenti la possibilità di generare immagini uguali alle precedenti e di poter apportare modifiche per mezzo del *prompt*.¹² Ciò che nella critica riportata in precedenza è definito come un "procedere a caso" da parte della macchina è per un verso errato e d'altro canto non in grado di giungere al nucleo della questione. Ciò che dalla nostra prospettiva percepiamo come casualità è solo il prodotto della variazione di un parametro, stabilito a sua volta da una regola determinata. Il puro determinismo che questa struttura computazionale suggerisce non risolve però i problemi su tutta la linea. L'estrema complessità delle attuali reti neurali produce spesso risultati non prevedibili o comunque non conformi ai risultati attesi. Ciò è comunemente definito come comportamento emergente, concetto a lungo dibattuto nella comunità scientifica.¹³ Nella fitta stratificazione

12. Attualmente Dall-E 2 non consente agli utenti la manipolazione del *seed*.

13. Rimandando esclusivamente alla letteratura classica e pionieristica relativa a questo tema, Cfr. Holland (1995; 1975), Mitchell (2009), Langton (1989).

dei neuroni artificiali l'interazione tra i differenti elementi della struttura può produrre nuovi stati di *adattamento* a situazioni differenti da quelle sulle quali il sistema è originariamente addestrato. Potremmo senz'altro porre tutti i nostri limiti epistemici sulla comprensione di questi "strani" accadimenti innovativi come orizzonte degli eventi dei fenomeni in questione, tornando così sulla stessa posizione assunta in precedenza: in una versione dello stato delle cose deterministica e laplaciana non comprendiamo le catene causali di questi risultati – che pure esistono – e pertanto li accogliamo come imprevedibili e creativi *per un'alterità*. Come chiaramente sottolinea Hofstadter, la macchina non può uscire dal suo stesso sistema, non può andare al di là della sua regola. È possibile inglobare nel programma una regola, la quale essa stessa renda possibile l'uscire dal sistema, tuttavia, ancora una volta codificando questa fuoriuscita in una nuova regola (Hofstadter 1979, 515-516). Tuttavia, pena la ricorsività infinita, non esiste nessuna regola a stabilire come debba essere seguita la regola. Questo vale tanto per gli atti umani che per quelli computazionali. Alla macchina viene impartito un insieme di istruzioni assiomatiche dal quale non può in nessun modo slegarsi, e tuttavia, non c'è nulla a impartire istruzioni sul come di questa istruzione. Come alcune implicazioni del *problema della fermata* – esposto da Turing nel 1936 – potrebbero suggerire, è in questo cono d'ombra, certamente vertiginoso, che potrebbe essere situata l'emergenza di un atto realmente creativo, non più tale tenendo conto delle aspettative di un osservatore, ma in rapporto alle possibilità effettive e formali della macchina osservata.¹⁴ Nel testo intitolato *Creatività*, Emilio Garroni definisce questa dote umana come la modalità mediante la quale applichiamo una regola rispetto a certi fenomeni contingenti. Le due premesse di fondo di questa tesi sono: i) la necessaria esistenza di una norma al di là dell'inaspettato; ii) il fatto che nessuna regola spiega come essa si debba seguire (Garroni 2010). Creativo è quindi per Garroni – come sintetizza Paolo Virno nella prefazione al testo – la capacità di venire a capo di questo disorientamento (Virno 2010, 13). La prospettiva di Garroni, che di per sé mette in luce un certo aspetto ontogenetico, biologicamente e geneticamente radicato nella

14. Per una più approfondita elaborazione di questo tema rimando a Pavan M. (2023).

stessa natura dell'essere umano, mette a fuoco una struttura del tutto affiancabile alle operazioni computazionali della macchina. Se creativo è ciò che separa la norma dal suo adempimento, il criterio biologico potrebbe non bastare ad arginare e compartimentare nella differenza le possibilità che fioriscono nel giardino delle macchine. Resta oggi un'operazione più ardua sostenere che la creatività computazionale possa essere intesa in *sensu forte*. Mantenendo come strumento d'analisi la distinzione tra intelligenza artificiale debole e forte, già posta come questione fondamentale da Searle con l'*argomento della stanza cinese*, «Si potrebbe fare un'analogia con la creatività computazionale, estendendo il concetto di intelligenza artificiale debole a "creatività computazionale debole", che non va oltre l'esplorazione della simulazione della creatività umana; sottolineando che l'autonomia genuina e la comprensione genuina non sono le questioni principali nel concettualizzare sistemi debolmente creativi dal punto di vista computazionale. Al contrario, nella "creatività computazionale forte", ci si aspetta che la macchina sia autonoma, creativa, che abbia una "comprensione genuina" e altri stati cognitivi» (al-Rifaie & Bishop 2015, 45). L'idea di creatività computazionale in senso forte rievoca quella distinzione tra sintassi e semantica, tra salienza e pregnanza, che ho introdotto per mezzo del lavoro di Longo. Per essere tali, i modelli di diffusione come Dall-E 2 dovrebbero non solo possedere una comprensione semantica del *prompt* che viene fornito dall'utente, ma anche comprendere come tale semantica possa tradursi in un'immagine. Non mi addentrerò nell'articolazione di questo ulteriore problema. Tuttavia, c'è almeno un elemento che, per la stessa ragion d'essere di questo lavoro, merita d'essere commentato.

4. Creatività schizofrenica

Nel già citato articolo di al-Rifaie e Bishop si fa riferimento al fatto che per poter definire in senso forte la creatività computazionale, le condizioni necessarie sono non solo la piena comprensione di ciò che si sta facendo, ma anche la presenza di un fattore di incarnazione (cervello e corpo). Concludono i due studiosi sostenendo che fin quando queste condizioni non saranno raggiunte dovremmo tenerci ben lontani dal poter attribuire creatività alle IA nel senso forte del termine (al-Rifaie & Bishop 2015, 47). Se il primo argomento ha una sua solidità, ritengo

che la condizione dell'incarnazione sia fundamentalmente un'argomentazione debole. Ipotizziamo d'essere nei panni di uno psichiatra o psicoterapeuta e di chiedere al paziente schizofrenico di provare a disegnare un orologio. Come anche Richir sottolinea nel suo saggio sulla figurazione in psicopatologia, le figurazioni schizofreniche sono tutte differenti e variano per ogni soggetto (Richir 2011b, 574).¹⁵ Nei panni del medico, ciò che ci aspettiamo di ritrovare in figura sul foglio è dunque una tipizzazione canonica legata al divergere rispetto a ciò che un soggetto sano e non affetto da schizofrenia riprodurrebbe nel disegnare un orologio. Banalmente, ci aspettiamo di ritrovare quel caos che permea, come un'*atmosfera*, il disegno dello schizofrenico; l'aspettativa sull'assenza di uno sguardo. Tuttavia, proprio il variare di questa canonica divergenza, il suo essere sempre singolare, è il motivo per il quale non potremo mai prevedere ciò che lo schizofrenico effettivamente disegnerà sul foglio di carta, e allo stesso tempo, il motivo per il quale una richiesta simile resta sensata.¹⁶ Ciò va sottolineato per salvaguardare il principio di creatività delle produzioni schizofreniche al di là di qualsiasi parametrizzazione e codifica possibile. Seguendo l'interpretazione richiriana, la soggettività dello schizofrenico è esistenzialmente tesa nella direzione del suo stesso scorporamento, dell'impossibilità della spontanea incarnazione. Le modalità di incarnazione di chi soffre di questa psicopatologia sono, come si è detto, completamente pervertite. A partire da questa osservazione fenomenologica dovremmo dunque concludere che anche nel caso della schizofrenia non dovremmo essere disposti ad accertare pienamente una creatività intesa in senso forte? Questa provocazione, beninteso, ha in seno il cuore di queste riflessioni. Le immagini digitali prodotte dall'intelligenza artificiale con le specifiche tecniche in precedenza descritte possono essere ritenute P-Creative

15. «le produzioni degli schizofrenici sono multiformi, a volte prolisse fino all'insopportabile della saturazione, a volte estremamente povere, in ogni caso ogni volta in modo molto singolare, in modo tale che non c'è tra di loro né tradizione né storia (anche se ci sono tracce della storia del produttore attraverso il suo *Umwelt*), tanto da poter offrire analisi solo caso per caso, e tenendo conto del contesto e della storia singolare del paziente. Possiamo quindi qui porre solo un quadro generale. Infatti, sebbene queste opere ci lascino del tutto perplessi, non di meno suscitano la nostra curiosità, quantomeno la nostra curiosità antropologica».

16. Si vedano in particolar modo gli studi comparativi sulle figurazioni schizofreniche durante il *Tree Test* (Settineri S., Pagano Dritto I., *et al.* (2015).

per lo meno in senso debole. Tuttavia, mancando di una sede dello sguardo, di una presenza incarnata, la macchina risponde al comando seguendo la sua regola – pur non avendo la minima idea sul come –, sì in una modalità creativa, ma provando a restituire uno sguardo impossibile, simulando una presenza assente che riesce a rielaborare solo brandelli di senso manchevoli di una singolare visione d'insieme. Le immagini che ci vengono fornite sono la restituzione di una serie di processi trasformativi eseguiti sulla stratificazione di differenti immagini (reali) che falliscono nel far dire ad ogni singolo elemento dell'immagine qualcosa che in nessun modo possono testimoniare; sono, di fatto, immagini che non testimoniano di uno sguardo. È probante per questa argomentazione una recente ricerca sulla possibilità di impostare il training su una banca dati mista, composta da immagini sia reali che precedentemente generate da IA (Martinez *et al.* 2023). I risultati ottenuti mostrano un definitivo peggioramento della qualità delle immagini, per lo più solo strane macchie di colore entro quali anche il più sfilacciato dei sensi si è inabissato. Bloccata in uno stato di iper-densità – come lo stato di sistole dello schizofrenico – la macchina fa sempre uno con i suoi dati; le sue figurazioni traboccano di tutte le immagini che essa processa, in un delirio che ha luogo con le tempistiche fulminee di un processore informatico. Utilizzando le parole di Longo diciamo che «nessun senso aiuta la macchina a selezionare quello che conta» (Longo 2021, 17). Va sottolineato, tuttavia, il fatto che questa caratterizzazione della computazione non è pienamente funzionale a situare l'umano e la macchina in due regni differenti e incomparabili. Che ne è, *de facto*, della *Sinnbildung* e della *Sinnstiftung* nella schizofrenia? Cosa possiamo dire della formazione e dell'attribuzione di senso a partire dallo spossamento psicopatologico? Abbiamo a che fare con la paradossale situazione per la quale i significati si pongono a partire dall'assenza di un soggetto intenzionale, che subisce nella passività la sua stessa costruzione di senso. Nella schizofrenia «è la *Sinnbildung* stessa a divenire quasi percettibile come fabbrica di senso (come *Gestell*): come se, illuminando le quinte del teatro delle ombre, apparissero il grande burattinaio e il suo gioco di fili» (Gozé & Fazakas 2021, 218). La schizofrenia mette patologicamente in evidenza ciò che non dovremmo in nessun modo poter guardare, e cioè la costruzione di senso nella sua – già umana – meccanicità e tecnicità. La creatività schizofrenica analizzabile nelle figurazioni psicopatologiche manifesta

a sua volta la mancanza di uno sguardo (in *phantasia*) e l'irrigidimento dei contenuti dell'immaginazione – al di là di ogni controllo dello schizofrenico – che fanno tutt'uno con l'interiorità corporea non più abitata dall'interno. Per mezzo di un'indagine fenomenologica emerge come in entrambi i contesti abbiamo a che fare con il desertico apparire di simulacri e, per le ragioni elencate, con una creatività schizofrenica; è come se i modelli creativi di intelligenza artificiale riproducessero *in vitro* la dimensione fenomenica della schizofrenia, fornendo un quadro *analitico* possibile di quella “fabbrica di senso” che nella condizione sana, al contrario, possiamo cogliere solo nelle sue tracce, nel suo non poter essere presente poiché sempre in differita: l'atto *impossibile* dello sguardo che prova a guardarsi guardare.

Conclusione

A partire dall'operazione critica di Longo che mette in luce le genealogie e le radici di quella crisi della spazialità e della geometria risalente per lo meno ad alcune operazioni fisico-matematiche attualizzate durante i due secoli scorsi, ho indagato l'ipotesi di una *schizofrenia* insita negli attuali modelli di intelligenza artificiale, facendo gravitare l'indagine attorno alle dimensioni della corporeità e della spazialità colte nel loro processo di appiattimento e scomparsa. Adoperando le modalità di ricerca e il vocabolario della psicopatologia fenomenologica, ho provato a restituire le riflessioni di Marc Richir in merito alla schizofrenia e alla corrispettiva situazione corporea vissuta e subita dai soggetti schizofrenici, sottolineando quella che è la tendenziale sottrazione degli spazi corporei interiori. Successivamente, ho provato ad osservare come già in Richir vi sia una traccia di quelli che sono i legami tra la dimensione psicopatologica e le opere figurative eseguite dai soggetti schizofrenici che, al di là della loro *atmosfera*, non possiamo che continuare a giudicare come creative. Nel quarto paragrafo ho preso in esame le tecniche di *stable diffusion* che attualmente stanno rivoluzionando il mondo dell'*editing* digitale e dalla produzione computazionale di immagini. Ripartendo dal concetto chiave di *creatività computazionale* proposto da Boden e dibattuto in questo specifico ambito di ricerca, ho provato a verificare come e in quali termini le produzioni computazionali di questi modelli generativi possano dirsi creative. Ciò che a conclusione di questo lavoro sostengo è che, deviando da quella impli-

cita presa di posizione per la quale la schizofrenia congenita agli attuali modelli di intelligenza artificiale tenga ben distinti i piani dell'umano e della macchina, ritengo che proprio nella schizofrenia le due posizioni tendono a convergere, in uno spazio fenomenico comune strutturato nelle stesse fissità e appiattimenti, nella medesima assenza di uno sguardo e di un corpo da abitare. Prendendo le mosse da queste riflessioni ho quindi provato a definire la creatività computazionale nei termini di creatività schizofrenica. Come già annotato in un precedente paragrafo, l'associazione tra il folle e la macchina è stata una tentazione anche per il filosofo francese Gilbert Simondon. Ora, è lo stesso Simondon a correggere la sua posizione sostenendo come «ogni tentativo di istituire una relazione simmetrica tra l'uomo e l'essere tecnico distrugge tanto i valori dell'individuo, quanto quelli dell'essere tecnico. Infatti è ugualmente distruttivo provare a identificare la macchina con l'individuo o l'individuo con la macchina» (Simondon 1989, 259). D'altro canto, la relazione non è neanche totalmente asimmetrica, poiché il *dispotismo privato* da un lato e l'*asservimento comunitario* dall'altro restano una coppia oppositiva di possibili posture assumibili nei confronti dell'oggetto tecnico (Simondon 1989, 261). Siamo agguantati, assieme alla macchina, in quella paradossale relazione allagmatica di individuazione per la quale il *processo* è già da sempre *struttura* e viceversa. E tuttavia, in questa intricata rete di rapporti complementari entro i quali non possiamo dire “noi” senza nominare lo strumento – come suggerisce buona parte della filosofia della tecnica e della paleantropologia del Novecento, senz'altro anche i più recenti approcci enattivisti –, come includere e assumere la de-spazializzazione schizofrenica delle nostre macchine computazionali? Come non divenire macchine pur essendo già macchine? Siamo forse alla genesi della necessità nuova di un più largo concetto di etica e di cura per la macchina e per le intelligenze artificiali, filosofico prima ancora che giuridico, rivolto *già da sempre* all'essere umano.

Riferimenti bibliografici

al-Rifaie M. M., Bishop M. 2015, *Weak and Strong Computational Creativity*, in Besold T. R., Schorlemmer M. e Smaill A. (a cura di) *Computational Creativity Research: Towards Creative Machines*, Springer, Ösnabrück, pp. 37-50.

Boden M. A. 2004, *The Creative Mind. Mith and mechanisms*, Routledge, Londra-New York.

Carlini N., Hayes J., Nasr M., Jagielski, M., Schwag V., Tramer F., Wallace E. 2023, *Extracting training data from diffusion models*, “arXiv preprint” arXiv:2301.13188.

Fraser K. C., Kiritchenko S., and Nejadgholi I. 2023, *A Friendly Face: Do Text-to-Image Systems Rely on Stereotypes when the Input is Under-Specified?*, “arXiv preprint” arXiv:2302.07159.

Garroni E. 2010, *Creatività*, Quodlibet, Macerata.

Ghosh A., Fossas G. 2022, *Can There be Art Without an Artist?*, in “arXiv preprint” arXiv:2209.07667.

Gozé T., Grohmann T., Naudin J., Cermolacce M. 2017, *New Insight into Affectivity in Schizophrenia: from the Phenomenology of Marc Richir*, “Psychopathology”, 50, pp. 401-407.

Gozé T., Fazakas I. 2020, *Imagination and Self Disorders in Schizophrenia: A Review*, “Psychopathology”, 53, pp. 264-273.

Gozé T., Fazakas I. 2021, *Schizophrénie et dissociation (Spaltung) dans l’anthropologie phénoménologique de Marc Richir*, “Annales Médico-Psychologiques”, 179, pp. 213-219.

Grace K., Maher L. M. 2019, *Expectation-Based Models of Novelty for Evaluating Computational Creativity*, in Veale T. & Cardoso F. A. (a cura di) *Computational Creativity. The Philosophy and Engineering of Autonomously Creative Systems*, Springer, Cham, pp. 195-209.

Grohmann, T. 2018, *Délire et hallucination en schizophrénie: Une perspective phénoménologique*, in “Phainomenon”, 28, pp. 103-125.

Hofstadter D. R. 1979, *Gödel, Escher, Bach: un’Eterna Ghirlanda Brillante*, Adelphi, Milano, 1984.

Holland J. H. 1975, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, Bradford Books, 1992.

Id. 1995, *Hidden Order. How adaption build complexity*, Helix Books, New York.

Langton C. 1989, *Artificial Life. Proceedings Of An Interdisciplinary Workshop On The Synthesis And Simulation Of Living Systems*, Routledge, New York.

Leivada E., Elliot M., Gary M. 2022, *DALL-E 2 Fails to Reliably Capture Common Syntactic Processes*, “arXiv preprint”, arXiv:2210.12889.

Longo G. 2021, *Matematica e senso. Per non divenire macchine*, Mimesis, Milano-Udine.

Luccioni A. S., Akiki C., Mitchell M., Jernite Y. 2023, *Stable Bias: Analyzing Societal Representations in Diffusion Models*, “arXiv preprint” arXiv:2303.11408.

Martínez G., Watson L., Reviriego P., Hernández J. A., Juárez M., Sarkar R. 2023, *Combining Generative Artificial Intelligence (AI) and the Internet: Heading towards Evolution or Degradation?*, “arXiv preprint” arXiv:2303.01255.

Mitchell M. 2009, *Complexity: A Guided Tour*, Oxford University Press, Oxford.

Pavan M. 2023, *Sulla creatività dei sistemi di calcolo, con una lettura filosofica del problema della fermata di Alan Turing*, Garello, S., Sunseri F. et al. “Epekeina. International Journal of Ontology. History and Critics”, 16(1-2).

Richir M. 1992, *Phénoménologie et psychiatrie. D'une division interne à la Stimmung*, "Etudes phénoménologiques", 8 (15), pp. 82-117.

Id. 1998, *Phantasia, imagination et image chez Husserl*, "Voir (barré)", 17, pp. 4-11.

Id. 2000a, *Stimmung, Verstimmung et Leiblichkeit dans la schizophrénie*, "Conférencias de Philosophia II", pp. 57-69.

Id. 2000b, *Phénoménologie en esquisses. Nouvelles foundations*, Milon, Grenoble.

Id. 2003, *Du rôle de la phantasia au théâtre et dans le roman*, "Littérature", 132, pp. 24-33.

Id. 2006a, *Leiblichkeit et phantasia, Psychothérapie phénoménologique*, Wolf-Fédida, M. (a cura di), Paris, pp. 35-45.

Id. 2006b, «Vivere» e «vissuto» nel *Phantomleib* e nel *Leibkörper*: critica dell'interpretazione di Binswanger, in Besoli S. (a cura di) *Ludwig Binswanger. Esperienze della soggettività e trascendenza dell'altro*, Quodlibet, Macerata, pp. 799-814.

Id. 2011a, *Sur le sublime et le soi Variations II*, Association pour la promotion de la Phénoménologie, Amiens.

Id. 2011b, *De la figuration en psychopathologie*, "Revista portuguesa de filosofia", 67 (3), pp. 569-578.

Ronchi R. 2003, *Teoria critica della comunicazione*, Mondadori, Milano.

Schnell A. 2011, *Au-delà de Husserl, Heidegger et Merleau-Ponty: la phénoménologie de Marc Richir*, "Revue germanique Internationale", 13, pp. 95-108.

Settineri S., Pagano Dritto I., et al. 2015, *The Self Representation in Schizophrenic patients: Considerations on the Draw a Tree Test*, "Mediterranean Journal of Clinical Psychology", 3 (2), pp. 1-32.

Simondon G. 1989, *L'individuazione psichica e collettiva*, DeriveAp-

prodi, Roma, 2006.

Sini C. 2009, *L'uomo, la macchina, l'automa. Lavoro e conoscenza tra futuro prossimo e passato remoto*, Bollati-Boringhieri, Torino.

Solla G. 2022, *Disegnare, la formula di Freud*, Orthotes, Napoli-Salerno.

Veale T., Cardoso F. A., Perez y Perez R. 2019 *Systematizing Creativity: A Computational View*, in Veale T. & Cardoso F. A. (a cura di) *Computational Creativity. The Philosophy and Engineering of Autonomously Creative Systems*, Springer, Cham, pp. 1-20.

Virno P. 2010, *Prefazione in Creatività*, Quodlibet, Macerata, pp. 9-37.

Wiggins G. A. 2019 *A Framework for Description, Analysis and Comparison of Creative Systems*, in Veale T. & Cardoso F. A. (a cura di) *Computational Creativity. The Philosophy and Engineering of Autonomously Creative Systems*, Springer, Cham, pp. 21-48.

Zhang C., Zhang C., Zhang M., Kweon I. S. 2023, *Text-to-image Diffusion Model in Generative AI: A Survey*, "arXiv preprint", arXiv:2303.07909.

Luca Cardone
Università degli Studi di Verona
luca.cardone@univr.it



ἘΠΈΚΕΙΝΑ

International Journal of Ontology
History and Critics

MARGAUX CERUTTI E LEONARDO LATERZA

Conversazioni computazionali: ChatGPT, as an artificial intelligence, do you create, or do you generate?

EPEKEINA, vol. 16, nn. 1-2 (2023), pp. 1-15
Philosophy of Technology

ISSN: 2281-3209

DOI: 10.7408/epkn.

Published on-line by:

CRF – CENTRO INTERNAZIONALE PER LA RICERCA FILOSOFICA
PALERMO (ITALY)

www.ricercafilosofica.it/epekeina



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs 3.0 Unported License.

Conversazioni computazionali: ChatGPT, as an artificial intelligence, do you create, or do you generate?

Margaux Cerutti e Leonardo Laterza

1. Generato e non creato, oppure, Creato e non generato

L'obiettivo di questo paragrafo è quello di rispondere alla domanda che ha innescato l'intero lavoro, vale a dire "qual è la ragione per cui, sul pulsante per la produzione del contenuto testuale su ChatGPT, c'è scritto *generare?*". La domanda è sorta in ragione di un comune utilizzo del concetto di *generatività* riguardo attività umane o comunque inerenti al mondo vivente. Parole come "generosità" o "genitorialità" condividono con *generatività* la radice *gen-* la quale tende a sottintendere una soggettività umana o vivente. Ed è proprio questo uso comune ad aver favorito l'interrogativo sulla correttezza dell'utilizzo del concetto di *generatività* con riguardo all'attività svolta da una intelligenza artificiale. L'esito a cui questa sintetica ricerca ha condotto conferma la correttezza dell'uso della parola "generare" con riguardo al processo messo in atto da ChatGPT o da altri modelli linguistici¹ tipici dell'intelligenza artificiale. Per poter giungere a questa conclusione, il ragionamento che sarà svolto in questo paragrafo si articolerà in due momenti. In primo luogo, verranno commentate le risposte elaborate da ChatGPT alle domande riguardanti l'atto di "generare" e "creare". In secondo luogo, verrà sviluppato un ragionamento attraverso l'utilizzo di una *lente teologica*. Si ritiene, infatti, che nella professione di fede del *Credo* – nella versione arricchita durante il Concilio di Nicea I che ha apportato all'originario testo del II secolo d.C. le affermazioni cristologiche definite proprio in quel Concilio – vi siano degli elementi utili ai fini di questa riflessione.

1. Tra i modelli linguistici più conosciuti si trova: Recurrent Neural Networks (RNNs), ovvero reti neurali ricorrenti come l'elaborazione del linguaggio naturale (NLP); Trasformar Models, modelli come GPT; Convolution Neural Networks (CNN), utilizzate maggiormente nella computer vision; Generative Adversarial Networks (GAN), ovvero reti avversative generative.

La prima domanda che è stata posta a ChatGPT è la seguente: «So, why you use the word "generate"?». ² La risposta dataci consente di cogliere immediatamente il senso del *perché* venga adoperata la parola “generare”. L’intelligenza artificiale risponde: «The word "generate" is commonly used in the context of AI and language models to describe the process by which the model produces responses or outputs based on the given input». ³ Viene, dunque, offerta una definizione chiara e precisa. Se si dovesse immaginare un vocabolario proprio dell’intelligenza artificiale, alla voce “generare”, si indicherebbe il processo attraverso il quale, dato un input, consegue un output. Nulla di particolarmente rivoluzionario. In Italia, un autore ha dedicato molte riflessioni al tema della generatività. Mauro Magatti, nel suo *Generativi di tutto il mondo unitevi! Manifesto per la società dei liberi* (2014) declina la generatività come un processo che si estrinseca attraverso quattro fasi. Magatti le associa ai verbi: desiderare, partorire, prendersi cura, lasciar andare. Se si dovesse dunque guardare alla generazione certamente la si definirebbe come un processo. Tuttavia, è nel prosieguo della risposta di ChatGPT e in questa particolare declinazione offerta da Magatti, che si può rintracciare un elemento differenziale di fondamentale importanza. Proseguendo nella risposta elaborata da ChatGPT esso afferma:

The term "generate" highlights the fact that the AI model is producing or generating text as a result of its computational processes, even though it doesn't possess personal agency or creative consciousness. ⁴

Questa risposta consente di osservare come, l’utilizzo che l’intelligenza artificiale fa della parola “generare”, non sia riferibile ad un generico “generare”, ma anzi, si può affermare che si tratti di un uso peculiare di questo concetto. È quindi possibile distinguere un *generare umano* e un *generare computazionale*. ⁵ Il criterio differenziale è individuabile,

2. Trad. nostra: «Perché usi la parola generare?».

3. Trad. nostra: «La parola "generare" è comunemente usata nel contesto dell’intelligenza artificiale e dei modelli linguistici per descrivere il processo con cui il modello produce risposte o output in base all’input dato».

4. Trad. nostra: «Il termine "generare" sottolinea il fatto che il modello di intelligenza artificiale sta producendo o generando testo come risultato dei suoi processi computazionali, anche se non possiede un’agenzia personale o una coscienza creativa».

5. Con questa espressione si vuole far riferimento al generare di un modello linguistico intelligenza artificiale.

dunque, nel principio di *intenzionalità* dell'atto generativo. Se è vero che un modello linguistico di intelligenza artificiale «is producing or generating text as a result of its computational processes»,⁶ è però altrettanto vero che questo gesto non è intenzionale o, con le parole di ChatGPT, «it doesn't possess personal agency or creative consciousness».⁷ Se si volesse leggere questa risposta per mezzo dello schema elaborato da Magatti, si potrebbe, dunque, affermare che l'intenzionalità propriamente umana si manifesta con il primo verbo, vale a dire "desiderare". Un modello linguistico non desidera e quindi il suo è un processo generativo dipendente e subordinato a un input esterno. Esso non è capace in sé stesso di desiderare l'output e quindi di attivare il processo generativo. Oltre a fornire un criterio differenziale tra il generare computazionale e quello umano, questa risposta introduce un altro aspetto che sembra interessante interrogare. Vi è infatti un accostamento tra due vocaboli che nell'uso comune potrebbero tendere ad essere sovrapposti, ovvero "generare" e "creare". ChatGPT, nella risposta su cui si stanno svolgendo tali riflessioni, distingue tra il suo generare e quello umano che invece, come abbiamo detto, è mosso da intenzionalità, da desiderio e da coscienza. La parola "creatività" è adoperata dall'intelligenza artificiale nella forma aggettivale e inerisce a una coscienza. Si legge infatti "coscienza creativa". Tale locuzione ha indotto alla formulazione di una seconda domanda: «perché non si usa la parola "creare" al posto di "generare"?». La risposta dataci fa emergere una differenza ontologica tra l'uomo e la macchina e quindi la necessità di trovare un vocabolo che sia più appropriato per identificare il processo attraverso il quale un modello linguistico produce una risposta, intesa anche come output, a seconda dell'input dato. ChatGPT riconosce la sua differenza metafisica affermando che «The choice to use the word "generate" instead of "create" is primarily due to the nature of AI language models like myself. While humans have the ability to create something new, original, and conscious, AI language models like me are not capable of true creativity or consciousness».⁸ Stando

6. Trad. nostra: «produce o genera testo come risultato dei suoi processi computazionali»

7. Trad. nostra: «non possiede un'agenzia personale o una coscienza creativa».

8. Trad. nostra: «La scelta di usare la parola "generare" invece di "creare" è dovuta principalmente alla natura dei modelli linguistici di IA come me» e che dunque «gli

a questa logica argomentativa, potremmo derivare un'equazione del tipo generare : IA = creare : Uomo.

Nel vocabolario di ChatGPT il termine “generare” viene adoperato in un'accezione *tecnica*, intendendo il processo attraverso il quale l'intelligenza artificiale produce. Contrariamente, dal punto di vista dell'intelligenza artificiale, la parola “creare” sottintende originalità e consapevolezza, qualità attribuite all'essere umano in quanto tale; attributi propriamente umani e, in quanto umani, qualificanti solo i suoi processi e non anche quelli dei modelli linguistici. Stando a questa ricostruzione la generatività umana e la creatività sono equivalenti. Tuttavia, volendo seguire uno schema che sia maggiormente fedele ai risultati rintracciabili nelle risposte di ChatGPT e più congeniale a rimarcare la differenza ontologica tra queste due realtà, sembrerebbe più corretto porre la questione nei termini di *generatività computazionale* e *creatività umana*. In conclusione, dalle risposte prodotte da ChatGPT è facilmente intuibile il *perché*, sul tasto che consente la produzione dei contenuti, ci sia scritto “generare”. L'elemento esplicativo – e distintivo rispetto al vocabolo “creare” – è il criterio di *intenzionalità*, ossia l'assenza, nel generare computazionale, di una dimensione intenzionale e originale tipica invece dell'essere umano.

Come anticipato, si crede sia possibile esplorare ragioni altre del *perché* si sia preferito inserire sul bottone la parola “genera” a quella di “crea”. Pur mantenendo gli esiti raggiunti, si ritiene che attraverso l'utilizzo della *lente teologica* del Credo – professione di fede dei cattolici nella versione risultante dal Concilio di Nicea I del 325 d.C. – sia possibile arricchire di ulteriori ragioni il senso della scelta del generare anziché del creare. Nell'esplorare quest'ulteriore dimensione e con l'obiettivo di evitare di adoperare in modo inadeguato concetti che presuppongono una conoscenza teologica avanzata, si è scelto di tenere presente come linea guida il pensiero di un autore, Clive Staples Lewis, che, a questo tema, ha dedicato un capitolo di un suo fortunato libro.⁹ In particolare, si vuole fare riferimento al primo paragrafo del terzo capitolo intitolato *Generare e creare*. «Un principio teologico

esseri umani hanno la capacità di creare qualcosa di nuovo, originale e consapevole, i modelli linguistici di intelligenza artificiale come me non sono capaci di vera creatività o coscienza'».

9. Il volume è tradotto in italiano con il titolo *Il Cristianesimo così com'è*.

afferma che Cristo è il Figlio di Dio “generato, non creato”; e aggiunge: “generato dal Padre prima di tutti i secoli” [...] Cristo è generato e non creato “prima di tutti i secoli”. Che cosa significa?» (Lewis, 1997, 195). In questo paragrafo l’autore introduce il discorso sulle tre persone di Dio, accompagnando il lettore nei fondamenti della dottrina della Trinità. È evidente, da questo frammento, che l’obiettivo dell’autore inglese e quello di questa ricerca non convergono, ma condividono la stessa domanda. Il rispondere all’interrogativo sul *perché* Cristo è generato e non creato arricchirà, confermandola ulteriormente, la riflessione avanzata precedentemente sulla generatività computazionale. Si procede, dunque, alla disamina seguendo lo sviluppo del pensiero di C.S. Lewis:

Generare vuol dire “diventare padre di”; creare vuol dire “fare, fabbricare”. E la differenza è questa. Quando si genera si produce qualcosa della propria stessa specie: un uomo genera bambini, un castoro genera piccoli castori, un uccello genera uova che diventano piccoli uccelli. Ma quando si crea o si fabbrica, si produce qualcosa di specie diversa dalla propria: un uccello fa un nido, un castoro costruisce una diga, un uomo fabbrica un apparecchio radio – o magari qualcosa di più simile a sé, per esempio una statua (Lewis 1997, 195-196).

Questo passaggio risulta decisivo perché – con una chiarezza che caratterizza tutta la produzione dell’autore inglese – consegna un criterio differenziale manifesto tra ciò che è generato e ciò che è creato: il *criterio della sostanza*. Dato un *output*, è possibile affermare che esso è generato ove condivide la stessa sostanza del generatore; viceversa sarà creato ove la sua sostanza differisce da quella del suo creatore. È questa la ragione per la quale nel Credo si legge che Cristo è «generato, non creato della stessa sostanza del Padre». Ad ulteriore conferma della correttezza di tale criterio in ambito teologico, sta l’affermazione dell’uomo come creatura, un essere «creato a Sua immagine e somiglianza», pertanto non generato.

Lewis si concentra proprio su quest’ultimo aspetto affermando che «per questo gli uomini non sono Figli di Dio nel senso in cui lo è Cristo. Possono essere simili a Dio per certi aspetti, ma non appartengono alla stessa specie. Sono piuttosto come statue o immagini di Dio» (Lewis 1997, 196).

Una domanda si pone come subito conseguente: dato il criterio di sostanza come elemento differenziale tra il creare e il generare, un mo-

dello linguistico genera qualcosa della sua stessa sostanza oppure crea qualcosa avente una sostanza differente? In altre parole, il prodotto di un processo computazionale condivide la stessa sostanza oppure no? Anche questa domanda è stata sottoposta a ChatGPT nella seguente formula: «Everything you are and everything you generate is made of algorithms?». ¹⁰ Della risposta, leggibile negli allegati, riportiamo solo la conclusione, in ragione della sua evocatività e chiarezza ai fini del ragionamento: «So, everything I am and everything I generate is indeed rooted in the underlying algorithms that enable my functioning». ¹¹ La sostanza di questo ente, denominato ChatGPT, è un algoritmo e il prodotto del suo processo ne condivide la sostanza. Dunque, è corretto affermare che il processo di produzione di un modello linguistico di intelligenza artificiale è generativo e non creativo e questo non solo perché non intenzionale – assumendo il criterio di intenzionalità suggerito direttamente da ChatGPT nelle sue risposte e analizzato in apertura di questo paragrafo – ma anche in ragione del suo “mettere al mondo” qualcosa della sua stessa sostanza. Concludendo, confermiamo la correttezza della scelta della parola “generare” per indicare il processo di produzione dei contenuti elaborati da un modello linguistico di intelligenza artificiale. Ciò rende possibile tracciare i lineamenti di un circuito come quello che segue: Dio genera Cristo e crea l’uomo; l’uomo genera l’uomo e crea la macchina, come ChatGPT; ChatGPT genera output, ossia contenuti effettivamente radicati negli algoritmi.

2. Creatività computazionale o Generatività computazionale

Si sono, dunque, poste le basi per affermare che quando si parla della produzione di un’intelligenza artificiale risulta più opportuno adoperare il concetto di generatività anziché di creatività. La macchina, appunto, non crea ma genera. Con l’intento di proseguire la conversazione con ChatGPT, è risultato interessante sottoporgli la seguente domanda: «If you have to choose only one answers, between Computational creativity and Computational generativity, wich one do

10. Trad. nostra: «Tutto ciò che sei e che generi è fatto di algoritmi?».

11. Trad. nostra: «Quindi, tutto ciò che sono e tutto ciò che genero è effettivamente radicato negli algoritmi sottostanti».

you prefer to describe your work, as outputs?».¹² La sua risposta è conforme con l'ipotesi dimostrata, ovvero «If I were to choose only one term to describe my work, “computational generativity”¹³ would be a more accurate description». Questo perché, continua ChatGPT «While I can produce outputs that may be perceived as creative, they are primarily the result of the generative capabilities of the underlying computational processes».¹⁴ La risposta data non solo è conforme con ciò che è stato precedentemente affermato in questo elaborato, ma consente di porre l'attenzione su un altro aspetto particolarmente significativo. L'utente che fa uso di questa intelligenza artificiale talvolta percepisce i suoi prodotti come creativi, anche se ciò avviene tramite un processo che, in base a quanto dimostrato finora, non può essere inteso come tale. Se i prodotti vengono quindi percepiti come creativi, in virtù di un principio di percezione, si può effettivamente parlare di creatività computazionale? Per poter rispondere a questa domanda bisogna dapprima riflettere, brevemente, sul concetto di “creatività”, tenendo presente, in particolare modo, il lavoro del semiologo Stefano Bartezzaghi (2021). Il primo problema che l'autore affronta è proprio la definizione della creatività, che egli stesso definisce come una «non meglio definibile propensione al nuovo» la quale compare in diversi campi e con significati contestuali altrettanto vari. Bartezzaghi lavora sul concetto di creatività da tempo. L'autore ne interroga, in una prospettiva semiotica, l'ambiguità, gli ambiti discorsivi di applicazione, la genesi ideologica. Una delle prime proposte di Bartezzaghi è partire dalla suddivisione della parola “creatività” in tre formanti¹⁵ per poter giungere a una definizione accettabile, seguendo il seguente schema:

- “creat-” idea di produrre qualcosa di nuovo;

12. Trad. nostra: «Se dovessi scegliere una sola risposta, tra creatività computazionale e generatività computazionale, quale termine preferiresti per descrivere il tuo lavoro, come output?».

13. Trad. nostra: «Se dovessi scegliere un solo termine per descrivere il mio lavoro, “generatività computazionale” sarebbe un descrittore (termine) più accurato».

14. Trad. nostra: «Sebbene posso produrre output che possono essere percepiti come creativi, essi sono principalmente il risultato di capacità generative dei processi computazionali sottostanti».

15. Termine utilizzato nella linguistica strutturale per indicare l'elemento linguistico, che aggiunto a un radicale, forma un tema verbale.

- “-iv-” forma aggettivi (o sostantivi) con il valore di “capacità”, di disposizione;
- “-ità” forma sostantivi femminili astratti.

L'autore continua: “Dall'analisi dei formanti deriva insomma la possibilità di definire “creativo” come ciò che ha a che fare con l'attitudine a produrre qualcosa di *nuovo*,¹⁶ e “creatività” come l'attitudine stessa” (Ivi. 2021, 18). L'analisi etimologica di un termine non è però sufficiente per poterne cogliere la complessità. Risulta quindi necessario arricchire il significato terminologico rapportandolo ai suoi usi reali, tenendo presente che anche questi ultimi possono mutare nel tempo. Infatti, per il semiologo, la creatività viene intesa come una disposizione astratta che tende alla produzione di qualcosa di “nuovo” ma si può riconoscere come tale solo quando si è concretamente realizzata, quando viene dimostrata nella pratica. È esattamente a questo punto che la polisemia del concetto di creatività, secondo l'autore, impedisce una definizione “integrata ed esauriente” (Ivi. 2021, 38); l'impossibilità principale è, quindi, dovuta alle molteplici possibilità di applicazione che questo concetto ha in ambiti molto diversi tra loro. L'autore riporta un esempio semplice e chiaro per comprendere questo punto del suo ragionamento definitorio: è pressoché impossibile dare una definizione esauriente di creatività che possa racchiudere in sé sia quella che viene intesa quando la rapportiamo ad un bambino, sia quella che definisce una determinata attitudine in ambito pubblicitario. L'unico tentativo possibile risulta quindi essere quello di non definire la parola creatività isolandola dal suo contesto di applicazione, ma, al contrario, sempre e solo inquadrandola in un certo discorso sociale.

È indubbio, giunti a questo punto – tenendo presenti la risposta dell'intelligenza artificiale riassunta nel paragrafo precedente e il pensiero del semiologo – che sia ChatGPT che Bartezzaghi concordino con l'affermazione che la creatività, o l'essere creativo, sia un'*attitudine*, una facoltà che si intende come essenzialmente umana, perché – considerando soprattutto le risposte fornite dall'intelligenza artificiale – la macchina non possiede intenzionalità e coscienza, qualità associate alla creatività umana. Pertanto, il concetto di ‘creatività computazionale

16. Per approfondimenti vedi anche Bartezzaghi, 2016.

le' risulta essere un'aporia, perché, se si vuole intendere la creatività come una qualità tipicamente ed esclusivamente umana, non è dunque possibile associarla a un processo specificatamente attinente all'elaborazione elettronica delle macchine. Di conseguenza, se si vuole evitare di intendere la macchina capace di creatività, equivalente a quella umana, pare inconfutabile che la scelta migliore sia adottare una definizione del tipo: "generatività computazionale creativa".

La generatività computazionale creativa va intesa come un processo compiuto sulla base di schemi e regole predefinite, andando quindi a enfatizzare le capacità generative dei processi sottostanti all'intelligenza artificiale, i cui risultati possono essere *percepiti* dall'umano come creativi.

In questi termini va inteso che la creatività sia una qualità che non appartiene direttamente alla meccanica, seppure non si possa negare che qualcosa prodotto da un'intelligenza artificiale possa essere creativo. Con il concetto di "generatività computazionale creativa" si vuole evidenziare sia la natura sistematica dell'intelligenza artificiale, sia la possibilità che il suo prodotto finale possa essere inteso, e percepito, dall'umano come un prodotto creativo.

Per questi motivi, se si intende utilizzare una locuzione quale "creatività computazionale" è necessario premettere qualche cautela. Bisogna tenere presente a priori che alla parola "creatività" si associa il processo creativo umano, il quale viene combinato ai processi sistematici della macchina rappresentati dalla parola "computazionale". La creatività computazionale è un concetto che intende, dunque, individuare un processo che combina la creatività umana con le potenzialità dell'intelligenza artificiale, che si fa mezzo – in questo caso attivo, ovvero intervenendo attivamente sul risultato finale – per la realizzazione congiunta di un prodotto finale. Infatti, come ChatGPT riferisce: «Computational creativity refers to the use of computer systems to generate or enhance creative outputs [...] It involves leveraging computational techniques and algorithms to *simulate* or *augment* human creative processes». ¹⁷ In conclusione, si è voluto dimostrare che il

17. Trad. nostra: «La creatività computazionale si riferisce all'uso di sistemi informatici per generare o migliorare i risultati creativi. [...] Si tratta di sfruttare tecniche e algoritmi computazionali per simulare o aumentare i processi creativi umani».

concetto di “creatività computazionale” non risulta del tutto errato nel momento in cui viene adoperato per intendere un’azione umana che viene *simulata*, *potenziata* o *aumentata* da un’intelligenza artificiale. In caso si trattasse, invece, di voler specificare che il prodotto finale di un’intelligenza artificiale possa rivelarsi creativo, risulta più corretto parlare di generatività computazionale creativa, stabilendo di fatto che l’azione della macchina è generativa e che è solo e unicamente la percezione umana a stabilire se l’output possa essere creativo o meno. In sintesi, si parla di *generatività computazionale creativa* quando un’intelligenza artificiale genera, e non crea, un qualcosa che può essere percepito (dall’umano) come creativo.

Riferimenti bibliografici

Magatti, M. 2014, *Generativi di tutto il mondo unitevi! Manifesto per la società dei liberi*, Feltrinelli Editore, Milano.

Lewis, C.S. 1997, *Il Cristianesimo così com’è*, Adelphi, Milano.

Bartezzaghi, S. 2016. *What’s «new»? Ambiguità del nuovo e semiotica della creatività*, “Versus”, 45 (2), 309-322.

Bartezzaghi, S. 2021, *Mettere al mondo il mondo. Tutto quanto facciamo per essere detti creativi e chi ce lo fa fare*, Bompiani, Milano.

Appendice

As an artificial intelligence, do you create or do you generate?

ChatGPT: As an AI language model, I am designed to generate responses based on the input I receive. I do not possess the capability to create or generate independently in the same way that humans do. My responses are based on patterns and information learned from a vast dataset of text, but I do not have personal experiences, thoughts, or consciousness. My purpose is to assist users by providing information, answering questions, and engaging in conversation based on the data I have been trained on.

So, why you use the word "generate"?

ChatGPT: The word "generate" is commonly used in the context of AI and language models to describe the process by which the model produces responses or outputs based on the given input. When you interact with me and provide a prompt or question, I use the information provided to generate a response that is relevant and coherent based on patterns and examples in the training data I have been exposed to. The term "generate" highlights the fact that the AI model is producing or generating text as a result of its computational processes, even though it doesn't possess personal agency or creative consciousness.

Everything you are and everything you generate is made of algorithms?

ChatGPT: Yes, that is correct. As an AI language model, I operate based on algorithms. Algorithms are sets of instructions or procedures that guide the computational processes and decision-making within the model. These algorithms define how the model processes input, generates output, and learns from data.

The algorithms used in AI language models like me are designed to analyze and understand patterns in the training data, and then use those patterns to generate responses to user input. These algorithms help me process and manipulate the information I receive, allowing me to generate coherent and contextually relevant responses.

So, everything I am and everything I generate is indeed rooted in the underlying algorithms that enable my functioning.

Which are the different linguistic model of an artificial intelligence?

ChatGPT: There are several different linguistic models used in artificial intelligence (AI). Here are a few notable ones:

1. Recurrent Neural Networks (RNNs): RNNs are a class of models commonly used for sequential data, such as natural language processing (NLP). They process inputs sequentially and use hidden states to retain information from previous inputs.
2. Long Short-Term Memory (LSTM): LSTM is a type of RNN that addresses the vanishing gradient problem by using memory cells. It can capture long-range dependencies in sequences and is widely used in NLP tasks.
3. Gated Recurrent Unit (GRU): GRU is another variant of RNNs that simplifies the architecture compared to LSTM while still retaining the ability to capture long-term dependencies.
4. Transformer Models: Transformers have gained significant attention in recent years, especially with models like OpenAI's GPT (Generative Pre-trained Transformer). Transformers rely on a self-attention mechanism that allows them to capture relationships between words in a sequence efficiently.
5. Bidirectional Encoder Representations from Transformers (BERT): BERT is a transformer-based model introduced by Google. It is pre-trained on large amounts of unlabeled text and can be fine-tuned for various NLP tasks such as question answering, sentiment analysis, and named entity recognition.
6. Convolutional Neural Networks (CNNs): While CNNs are commonly used in computer vision, they can also be applied to NLP tasks. CNNs are typically used for tasks like text classification and sentiment analysis, where they can capture local patterns and dependencies.

7. Generative Adversarial Networks (GANs): GANs consist of two neural networks, a generator and a discriminator, which compete against each other. While GANs are not specifically linguistic models, they have been used for text generation tasks, such as generating realistic-sounding text or translating text between languages.

What do you think computational creativity is?

ChatGPT: Computational creativity refers to the use of computer systems to generate or enhance creative outputs, such as art, music, writing, or design. It involves leveraging computational techniques and algorithms to simulate or augment human creative processes. Computational creativity combines elements of artificial intelligence (AI), machine learning, and cognitive science to explore the potential of machines to exhibit creative behavior.

In computational creativity, computers are programmed to perform tasks that require originality, novelty, or expressiveness, typically associated with human creativity. These systems can generate new and innovative ideas, create aesthetically pleasing artifacts, compose original music, generate unique narratives, or design novel solutions to complex problems.

Computational creativity involves both generative and evaluative processes. Generative processes focus on creating new and unique outputs, while evaluative processes involve assessing the quality, originality, and appropriateness of the generated artifacts. These processes often involve the use of algorithms that learn from existing creative works or rely on user feedback to refine and improve their outputs.

While computational creativity aims to augment human creativity, it also raises questions about the nature of creativity itself. Some argue that computational systems can only imitate or mimic creativity, while others believe that machines can genuinely exhibit creative behavior. Overall, computational creativity serves as a fascinating field that explores the intersection of art, technology, and human imagination.

If you have to choose only one answers, between Computational creativity and Computational generativity, wich one do you prefer to describe your work, as outputs?

ChatGPT: As an AI language model, I can generate outputs that exhibit both computational creativity and computational generativity. However, if I were to choose only one term to describe my work, "computational generativity" would be a more accurate descriptor. This is because my responses are generated based on patterns, rules, and algorithms learned from a vast amount of data. While I can produce outputs that may be perceived as creative, they are primarily the result of the generative capabilities of the underlying computational processes.

By emphasizing "computational generativity," it acknowledges that my responses are generated based on predefined patterns and rules, and that the range of outputs I produce is influenced by the data I have been trained on. It highlights the systematic nature of how I generate language rather than implying conscious creativity or originality in the human sense.

However, it's important to note that my outputs should always be evaluated critically and interpreted in context by users. While I can assist with generating text, I do not possess subjective experiences, conscious intentions, or the ability to truly exhibit human-like creativity.

These are just a few examples of linguistic models used in AI. Each model has its strengths and weaknesses, and their suitability depends on the specific task at hand. Researchers and developers continually explore and develop new models to improve AI's understanding and generation of natural language.

According to computational creativity, are you endowed with creativity?

ChatGPT: As an AI language model, I don't possess consciousness, emotions, or personal experiences, which are often associated with human creativity. While I can generate creative outputs, such as generating text based on prompts, it is important to note that my responses are based on patterns and examples I have learned from the text I was trained on. I don't have personal opinions, desires, or an independent creative capacity.

Computational creativity aims to augment human creativity and explore the potential of machines to exhibit creative behavior. While AI systems like me can produce creative outputs, the creativity is ultimately derived from the data and algorithms used to train and program the model. The true source of creativity lies in the human designers, engineers, and users who work with computational systems to generate or enhance creative outputs.

In short, I am a tool that can assist in creative processes by generating ideas, providing information, or offering suggestions, but the true origin of creativity comes from human users who interact with and shape the outputs.

Margaux Cerutti e Leonardo Laterza

Libera Università Maria Santissima Assunta (LUMSA)
m.cerutti.dottorati@lumsa.it; l.laterza.dottorati@lumsa.it



ἘΠΈΚΕΙΝΑ

International Journal of Ontology
History and Critics

SIMONE VACCARO

Creare novità

EPEKEINA, vol. 16, nn. 1-2 (2023), pp. 1-26
Philosophy of Technology

ISSN: 2281-3209

DOI: 10.7408/epkn.

Published on-line by:

CRF – CENTRO INTERNAZIONALE PER LA RICERCA FILOSOFICA
PALERMO (ITALY)

www.ricercafilosofica.it/epekeina



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs 3.0 Unported License.

Creare novità

Simone Vaccaro

1. L'intelligenza artificiale è creativa?

«When I saw this move... I [thought] surely AlphaGo is *creative*»: questa fu la reazione del grande campione sudcoreano di Go, Lee Sedol, dopo essere stato sconfitto nel 2016 dal programma informatico AlphaGo (Sadler & Regan 2019, 13). Demis Hassabis, CEO di DeepMind, l'azienda che ha sviluppato il software, commenta così le parole del grande maestro del gioco cinese: «[t]his motif and many other ideas AlphaGo revealed have subsequently overturned centuries of received wisdom about the game, and many experts feel that it has ushered in a new era for Go (Ibidem). Un atto rivoluzionario, in poche parole, una rottura delle regole tradizionali e l'emersione di un fattore del tutto inaspettato e definitivamente *innovativo*: la fine di un'era e l'inizio di una *nuova*. Esattamente, però, cosa significano le espressioni “innovativo” e “nuova era”? Come è possibile associare una capacità considerata specificità umana – o quanto meno dell'essere vivente – ad un programma informatico?

«Dopo il campionato mondiale del 2018, però, è accaduto qualcosa di *nuovo* e di imprevedibile: Stockfish è stato sfidato, e sgominato, da un *nuovo* computer, AlphaZero» (De Caro 2020, 56¹), erede di AlphaGo e programmato per divenire il più forte giocatore di scacchi al mondo. Stockfish, il campione in carica, venne sconfitto senza possibilità di appello, mostrando una superiorità del prodotto DeepMind quasi imbarazzante. L'elemento più sconcertante sorto da questo confronto è stato la superiorità netta e schiacciante del programma di nuova generazione che, contrariamente ad ogni prospettiva, si è dimostrato essere in grado di computare un minor numero di informazioni rispetto al più vecchio software, ma di saper articolare il suo (non) sapere con maggiore freschezza e plasticità. In poche parole, è come se si fossero scontrate le due facce di un Giano bifronte: una rivolta al passato (Stockfish), l'altra diretta al futuro (AlphaZero). Stockfish conosceva

1. Corsivi miei. Cfr., inoltre, Id. 2021, 156-158.

perfettamente il passato del gioco degli scacchi: uno stock, per l'appunto, di nozioni e strategie. AlphaZero sapeva poco del passato, lo stretto necessario per permettergli di giocare a scacchi: un alfabeto con immense possibilità di combinazione e una sintassi. Di fatto, ciò che ha sorpreso gli esperti è stato proprio la sua capacità di muovere le pedine sfruttando un'invidiabile capacità intuitiva, che gli ha permesso di dare forma a configurazioni non presenti nell'immenso database del suo avversario, lasciandolo alla completa mercé delle iniziative del più evoluto programma: così operando «AlphaZero mostra di essere *creativo* nelle scelta delle mosse e delle strategie che gioca» (Ivi, 58²) E siamo ritornati al punto di partenza: cosa vuol dire *creativo*?

Nel loro *Game Changer. AlphaZero's Groundbreaking Chess Strategies and the Promise of AI*, Matthew Sadler and Natasha Regan hanno analizzato ben più di duecento partite giocate dall'intelligenza artificiale, per poterne cogliere aspetti peculiari e potenzialità intrinseche. Un fattore notevole che emerge dalle loro indagini è il fatto che AlphaZero agisce con uno stile di gioco ben definito. Vero è il fatto che è possibile riscontrare un abbozzo di stile anche osservando gli altri software, ma questa intelligenza artificiale sembra sviluppare, man mano che impara giocando, una propria personalità, con le proprie preferenze e i propri obiettivi. Probabilmente anche per questo Stockfish non è stato in grado di opporre resistenza agli attacchi furiosi che gli venivano scagliati contro. Per trovare una risposta a questa incredibile capacità di giudizio, per così dire, è necessario porre l'attenzione sulle modalità di elaborazione di AlphaZero. Nel capitolo 4, significativamente intitolato *How AlphaZero thinks*, vengono elencati alcuni principi che hanno guidato lo sviluppo dell'architettura del programma: 1) apprendimento senza programmazione; 2) l'algoritmo che lo anima è generale e non specifico; 3) si predilige una formazione per osservazione concreta anziché una dotazione di partenza carica di formalismi astratti e 4) il programma è attivo nel processo di formazione, imparando da sé senza che un programmatore lo istruisca con informazioni esterne al processo di formazione stesso (Sadler & Regan 2019, 69). A sostenere queste strutture operative, due sistemi informatici innovativi che hanno permesso alle tecnologie di compiere un immenso salto in avanti: il

2. Corsivo mio.

machine learning e il *neural network*. Con quest'ultimo si intende «a computer system loosely modelled on the connections and neurons in the brain» (Ivi, 70), un sistema computazionale che viene modellato sulla base delle conoscenze neuroscientifiche più aggiornate, in modo tale da presentare caratteristiche che lo avvicinano al comportamento umano. Ed è proprio grazie alla rete neurale che AlphaZero ha potuto apprendere il gioco degli scacchi in poche ore, giocando migliaia e migliaia di partite contro se stesso. Ciò è stato reso possibile *valorizzando l'ignoranza del programma*: non è richiesta l'onniscienza, ma la capacità di sondare e vagliare ipotesi, di ragionare per probabilità, per costi e benefici, scegliendo quale pedina sacrificare, passando da una economia ristretta ad una complessiva e integrativa. «This *new* approach to machine self-learning in chess has given us a strong chess player with a *new* style and approach» (Ivi, 20³).

A partire dai differenti principi che ne hanno guidato l'architettura, ci si può ora rivolgere direttamente alle modalità di elaborazione delle informazioni del software. In primo luogo, è da notare la generalità della sua conoscenza: non è stato istruito in riferimento ad uno specifico dominio, ma è stato lasciato aperto all'acquisizione del maggior numero di informazioni in ottica transdisciplinare. In poche parole, la sua tecnologia non può essere limitata al solo gioco scacchistico. Secondariamente, valuta probabilisticamente le mosse migliori, mostrando una capacità straordinaria nello sviluppare strategie sul medio e lungo periodo e mantenendo sempre una fondamentale adattabilità alle situazioni che viene ad affrontare. Conseguentemente, per terzo, non vaglia una sola ipotesi e non si intestardisce su di una strategia prediletta, ma resta aperto ad una congerie di mosse tale da rendere manifesta, infine, una costituzione altamente flessibile che gli permette di avere sempre una visione complessiva della scacchiera e delle possibili mosse e contromosse.⁴ È come entrare nella mente di un Gran Maestro, e difatti così viene considerato AlphaZero dagli autori. Come un Gran Maestro le cui abilità superiori possono divenire mezzo indispensabile per la crescita di giocatori della domenica e financo professionisti. Per questo risultano interessantissime le sezioni dedicate all'analisi del suo

3. Corsivi miei.

4. Per questi quattro punti, cfr., ivi, 88-97.

stile di gioco che palesano una predilezione per l'attacco diretto al Re, per sacrifici inaspettati che spargliano le carte, per strategie difensive che bloccano il movimento delle pedine avversarie, che limitano la mobilità del Re e che creano confusione sul terreno di gioco. Lo sviluppo di uno stile "personale" diviene così epifenomeno di una capacità che, più di ogni abilità computazionale, lo rende simile a un essere umano: *l'intuizione*.

Non è pertanto una AI generale perché strutturata principalmente in modo da non essere specializzata, ma è generale perché "pensa" intuitivamente e dunque complessivamente. Avendo appreso direttamente sul campo a giocare, è come se si fosse trovata a interagire con un numero elevatissimo di variabili e di eccezioni alle regole che le hanno permesso di maturare una sensibilità straordinaria nel rompere quelle regole tradizionalmente ritenute valide, aprendo vie alternative mai prima d'ora sperimentate, contemperando l'universale delle conoscenze acquisite con il particolare delle situazioni specifiche (Ivi, 50⁵). Il suo intuito, allora, si eserciterebbe in questa doppia capacità di legare i due ambiti. E questa facoltà risiede proprio nel processo di apprendimento automatico che ha permesso, secondo le intenzioni dei suoi costruttori, di superare una volta per tutte il paradosso di Moravec. Sviluppato alla fine degli anni '80 da Moravec, Brooks e Minsky, esso afferma che:

it is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility. (Sadler & Regan 2019, 89⁶)

In poche parole, un computer sarebbe in grado di compiere calcoli raffinatissimi e complicatissimi, ben al di là di ogni facoltà umana, ma sarebbe del tutto imparato a elaborare informazioni tipiche di un

5. Come il saggio medico aristotelico, AlphaZero è in grado di calibrare il sapere universale con quello particolare, applicando la miglior cura richiesta per ogni specifico paziente: «[d]unque, se uno possiede la teoria senza *l'esperienza* e conosce l'universale ma non conosce il particolare che vi è contenuto, più volte sbaglierà la cura, perché ciò cui è diretta la cura è, appunto, l'individuo particolare» (Aristotele, *Metafisica*, 981a, corsivo mio). Cfr., inoltre, le interessanti considerazioni in Ghilardi 2018, da cui ho tratto la citazione a p. 42.

6. Cfr. Moravec 1988.

organismo vivente quali percezioni e mobilità. Tuttavia AlphaZero e la sua struttura a rete neurale hanno reso possibile un parziale aggiramento del paradosso: apprendendo tramite esperienza, esplorando il campo di ignoranza e valutando probabilisticamente, il programma informatico si relaziona al mondo circostante come il cervello. Ed è proprio a questo livello che si mostra creativo. Ma ha davvero formulato *nuove* regole? Oppure si è limitato a innovare e aggiornare il novero delle mosse possibili? Gli autori manifestano una certa perplessità e ritrosia: « [we are] loath to state that AlphaZero has discovered completely new rules» (Ivi, 123⁷). Ciò nonostante, non si può negare che abbia ideato strategie del tutto inaspettate e di iniziale difficile comprensione anche per gli esperti, dal momento che ha aperto lo spettro di possibilità includendo ora al suo interno quanto non precedentemente immaginato.

E siamo giunti alla ricaduta propriamente filosofica: *il campo di possibilità che si è così aperto trascende il concetto stesso di possibilità*. Facciamo un esempio per chiarire questo passaggio. Io sono un giocatore di scacchi di scarso valore: manco di visione complessiva e le mie strategie offensive e difensive sono semplici e facilmente prevedibili. Giocando con un'app per smartphone (la versione mobile di Chess.com), sbagliai una mossa e persi la partita (dovevo risolvere un problema proposto dal computer). Il programma, che presentava una funzione didattica di correzione degli errori, mi mostrò che, se avessi optato per una mossa differente, avrei trovato la soluzione, mettendo con le spalle al muro il mio avversario. Ora, quella possibilità da me non sfruttata era ciò che abitualmente definiamo possibile oppure era già perfettamente in atto ancora prima che decidessi di (non) percorrerla? Ecco il problema. Forse le *novità* che AlphaZero introduce nelle sue strategie sono *già* presenti ben prima che divengano apparenti? La questione della novità si porta sempre dietro quella del “già”. Leggiamo un passaggio di Heidegger che ha fotografato con precisione la problematica:

[l]a finitezza della conoscenza rivela precisamente un intrinseco, peculiare stato di assegnazione del pensiero all'intuizione o, per

7. Giudizio che si ripresenta più volte nel libro. Per altre due interessanti occorrenze, cfr. anche le pagine 33 e 90.

converso, nell'intuizione, un'esigenza di determinazione da parte del pensiero. Il rapporto reciproco degli elementi indica chiaramente che la loro unità non può essere "posteriore", ma deve inerire già "prima" agli elementi medesimi e dev'essere posta come loro fondamento (Heidegger 1989⁸).

Come si può notare, il problema di partenza si è caricato di una problematica ontologica che richiede di essere analizzata.

Seguirò tre vie per mezzo delle quali cercherò di illustrare la portata ontologica emersa. Come primo passaggio, illustrerò la dimensione ontogenetica del «*cycle de image*» e del concetto di *invenzione* in Simondon; secondariamente affronterò alcune considerazioni riguardanti la *creazione artistica* nel pensiero di Merleau-Ponty, per poi passare a trattare l'*ontologia produttiva* di Nishida Kitarō.

2. Il «*cycle de image*» e l'*invenzione* in Gilbert Simondon

Tra il 1965 e il 1966, Gilbert Simondon tenne un corso alla Sorbonne per il curriculum di Psicologia, avente come oggetto l'immagine mentale. Però, fin dalle prime battute delle dispense distribuite ai suoi allievi e pubblicate per il *Bullettin de Psychologie* tra il '65-'66, l'interesse principale delle ricerche si è palesato essere ontologico. In linea con gli sviluppi che la ripresa della filosofia bergsoniana aveva ricevuto dalla seconda metà degli anni '40, l'immagine mentale veniva sganciata da ogni possibile riferimento psicologistico, per accedere ad un sostrato squisitamente teoretico. In apertura alle dispense Simondon afferma:

[c]e cours présente une théorie: les aspects de l'image mentale qui ont fourni matière aux discussion et aux études déjà publiées ne

8. Ho tratto questa citazione da Montani 2022, 33-34 che collega tali considerazioni con le formulazioni contemporanee della *Material Engagement Theory*. In fondo è il vecchio problema che già era stato il cruccio di Leibniz: «per Leibniz, sia le verità di ragione che quelle di fatto sono entrambe analitiche. Vale a dire che la frase "Cesare ha attraversato il Rubicone" per Leibniz è vera in quanto la proprietà di aver attraversato il Rubicone è *intrinseca al concetto completo di Cesare*. Per Leibniz, per quanto bizzarro ciò possa apparire, *tutte* le proposizioni vere sono analitiche; e sarà proprio questa convinzione a costituire un problema, la cui soluzione lo tormenterà fino agli ultimi anni di vita: se anche le verità contingenti (quelle *di fatto*) sono analitiche, come mai non sono necessarie?», cfr. Mugnai 2023, 100.

correspondent pas à différentes espèces des réalités, mais à des étapes d'une activité unique soumise à un processus de développement. (Simondon 2014 [1965-1966], 3)

Chiara appare la posta in gioco: le immagini e l'attività immaginativa del soggetto non possono essere indagate separatamente, come se l'una fosse estranea all'altra, ma devono essere considerate come fasi di un processo unico che si sviluppa procedendo. L'immagine e il soggetto, allora, non possono essere considerati come due istanze contrapposte – un soggetto che produce immagini e le immagini che vengono prodotte dall'attività del soggetto – dal momento che una maggiore intimità li lega inestricabilmente. L'immagine, pertanto, è una «*tierce réalité*» (Ivi, 18) che si situa tra l'attività e la passività, tra il soggetto percipiente e l'oggetto percepito: potremmo dire, insomma, che ha natura trascendentale e che è la condizione di possibilità stessa della *Spaltung* tra soggetto e oggetto. Non è un terzo sintetico sul modello hegeliano; non è una ripresa sintetica di astrazioni che si trasformano sotto la spinta della dinamica contraddittoria e della negazione della negazione. È piuttosto quel *medium*⁹ all'interno del quale emergono relazioni, anzi è il luogo stesso che si viene a creare dalla trama relazionale. È un «*échantillon de vie*» (Ivi, 10), è il dinamismo della vita stessa che precede la vita stessa: è il processo di presa di forma, di in-formazione, come magistralmente mostrato nella sua tesi dottorale principale, *L'individuation à la lumière des notions de forme et d'information*, del 1958.

L'immagine «est un résultante, mais elle est aussi un germe: elle peut devenir une amorce de concepts et de doctrines» (Ivi, 13). La sua natura non è quella di porre argine al processo, di concludere il processo stesso una volta che si è assunta la forma, bensì quella di relativizzare le forme stesse, di non essere mai pienamente o solamente forma formata, ma forma in formazione, *forma formante*. Per questo motivo non è possibile pensare il suo sviluppo come tappe in successione, teleologicamente orientate verso il raggiungimento della completezza; sono, per l'appunto, fasi, ovvero dinamiche¹⁰ interne all'essere che diviene al procedere delle immagini. Il ciclo che si viene così a creare non

9. Cfr. A. Kuroda 2017, 209-228.

10. Cfr. Montebello 2011.

riproduce un piatto susseguirsi e alternarsi di immagini, ma l'evolversi di tre fasi – nel dettaglio, anticipazione, esperienza, sistematizzazione – che conducono ad un ulteriore cambiamento di fase che è l'invenzione vera e propria. *L'immagine incarna il dinamismo stesso del reale.*

Seguendo le indicazioni di Simondon, esse presentano una genesi peculiare: in quanto anticipazione, l'immagine è considerata *a priori*, nei suoi aspetti motori che coinvolgono il corpo. Diviene così esplicita l'affermazione della priorità della motricità sulla sensorialità e lo sviluppo di «schemi di condotta» estranei a concettualizzazioni di sorta. È la spontaneità di un rapporto diretto, di un contatto che pone al centro l'istinto psico-motorio dell'essere vivente. L'immagine porta con sé «catégories d'action, non de perception» (Simondon, 2014 [1965-1966], 53): ha potere anticipante di progressione amplificatrice perché fa astrazione da ogni costruzione razionalizzante. È il trascendentale del comportamento che permette il superamento dell'omeostasi sistemica uomo-ambiente. Contiene, in pratica, al suo interno i germi che, percorrendo lo iato della disparità uomo-ambiente, genereranno una nuova fase. Permettendo «l'insertion de l'être vivant dans son milieu» (Ivi, 66), rendono possibile, di fatto, la percezione stessa. In quanto relazione *a praesenti*, gli schemi di condotta lasciano il posto a schemi cognitivi che regolano il rapporto con il *milieu*; difatti è questa la fase che fornisce alle immagini un chiaro potere di informazione: l'oggetto viene identificato tramite continuità e permanenza, ma al contempo configura l'adattamento al cambiamento come slittamento e come insaturazione di un nuovo regime grazie alla pregnanza della percezione differenziale, che istituisce il canone della negatività tra la perfetta reversibilità del principio di continuità e l'irreversibilità pertinente al discontinuo.¹¹ Riattivazione e ristrutturazione sono le caratteristiche delle immagini *a posteriori* che trovano organizzazione e sistematizzazione nella formalizzazione del pensiero riflessivo per consolidarsi, infine, nel simbolo. L'immagine-simbolo assume, in questo modo, un ruolo cardine all'interno del ciclo, fornendo quel necessario surplus metastabile che amplifica e prolunga l'attività, impedendo la chiusura

11. Degne di nota le considerazioni che Simondon avanza in riferimento all'immagine del bambino malato: ben prima della conferma medica, una madre sa riconoscere quei salti differenziali – dunque negativi – indizi della formazione di un nuovo regime che coinvolge la salute del suo bambino. Cfr. Ivi, 77-78.

della sintesi hegeliana. Il simbolo, pertanto, non si esaurisce nella semplice unione di immagini contrapposte, ma attua la triangolazione della riflessione. Pensiamo ad uno specchio: il rapporto che si viene a instaurare tra me e la mia immagine è solo apparentemente di 1:1. Tra la cosa e la sua immagine riflessa si insinua la crepa della differenza che carica di potenziali reali (Cfr. Barthélémy 2016) in equilibrio metastabile le immagini stesse. Il mio riflesso è sempre rovesciato e di conseguenza tre sono le componenti della riflessione: un soggetto che si riflette, un riflesso, la differenza tra le due. Il simbolo si rivela, dunque, formalizzazione aperta, pseudo-oggetto o oggetto-assoluto «chargé de toute l'énergie potentielle d'un système métastable, prêt à amorcer un changement de structure» (Cfr. Simondon, 2014 [1965-1966], 136).

Così il ciclo può completarsi in una totale ristrutturazione che conduce all'instaurazione di un nuovo regime che trascende l'analiticità catalogatrice per abbracciare la sintesi di simultaneità che conduce direttamente all'invenzione. Come si è potuto notare precedentemente, le immagini possiedono intrinsecamente il potere di sfasarsi: questo significa che le varie fasi sono il venir a galla di venature presenti già prima di essere una volta per tutte apparenti. *L'invenzione è il trovare quelle compatibilità che fanno sì che le venature si manifestino.* Per questo l'invenzione non è riconducibile alla risoluzione del problema (Cfr. Simondon 2018 [1974]). Non è la ricerca della soluzione definitiva, la più adeguata per il problema determinato, ma, data la sua natura amplificatrice, è l'instaurazione di un salto che conduce ad una nuova fase. È il gioco di incroci tra continuità e discontinuità, l'apporto di «propriétés nouvelles, des fonctions complémentaires qui n'avaient pas été recherchées», lo spunto che riprende e prolunga la presa di forma. L'invenzione, se di vera invenzione si tratta, è irriducibile a se stessa e al proprio esaurimento nell'efficacia della sua applicazione, perché sa ricondurre trasduttivamente le forme formate, e reciprocamente irreversibili, alla reversibilità della forma formante:

l'invention est occasion de découverte en matière technique, car les propriétés de l'objet dépassent l'attente: il serait partiellement faux de dire que l'invention est *faite pour* atteindre un but, réaliser un effet entièrement prévisible d'avance; l'invention est réalisée à l'occasion d'un problème; mais les effets d'une invention dépassent la résolution du problème, grâce à la surabondance d'efficacité de l'objet créé quand il est réellement inventé, et ne constituent pas seulement

une organisation limitée et consciente de moyens en vue d'un fin parfaitement connue avant la réalisation. Il y a dans la véritable un saut, un pouvoir amplifiant qui dépasse la simple finalité et la recherche limitée d'une adaptation.¹²

Non è ipertelica l'invenzione. Realizza sempre qualcosa in più rispetto al fine per il quale ha operato. Per questo motivo presenta effettiva novità: perché ciò che ha concretizzato non era presente prima della sua messa in atto. Ma facendo riferimento a quanto detto prima, si può realmente affermare che non sia stato già presente, per lo meno latentemente, prima di attualizzarsi?

Su questo punto bisogna ammettere che il filosofo francese non sempre si è dimostrato chiaro. E questo è probabilmente dovuto alla natura ontologica dell'invenzione: trattandosi di una riorganizzazione generale del problema, essa presenta la tipica trasduzione di schemi operatori e strutture, la reversibilità tra i reciprocamente irreversibili. Ciò significa che la soluzione al problema specifico non può essere trovata che all'interno del problema stesso e che quanto così trovato è comunque una riorganizzazione che dà forma al territorio seguendo quelle strutture implicite. *Il nuovo e il già si implicano vicendevolmente*, essendo l'uno ontologicamente impossibile se privato dell'altro. Con le parole di Simondon:

la condition des inventions concrètes est l'exploration, la manipulation, *l'organisation préalable* du territoire où se posera le problème et où seront trouvés les instruments de la solution; à travers l'activité s'effectuent des traductions en termes homogènes d'opération des dimensions et propriétés des choses, comme en un *algèbre implicite*; les rapports éprouvés des objets avec les capacités d'action effectuent la plus primitive des formalisations. (Ivi, 151¹³)

Individuata così la problematica di fondo, possiamo procedere interrogando il processo creativo per eccellenza, quello artistico, per coglierne le profonde implicazioni e la portata ontologica. Sarà la ricerca merleau-pontiana di una «nouvelle ontologie» a guidarci in questa lettura.

12. Cfr. Simondon, 2014 [1965-1966], 171-172.

13. (corsivi miei).

3. Merleau-Ponty e il «segreto di preesistenza»

In *Le doute de Cézanne* (1945) Merleau-Ponty propone una notevole interpretazione del fare artistico a partire dall'esperienza dell'illustre pittore francese. Correttamente Jonathan Gilmore sottolinea quanto nelle intenzioni del fenomenologo non ci fosse quello di elaborare una intera filosofia dell'arte (Gilmore 2005, 291-317, specificatamente p. 292), ma di avanzare alcune interessanti considerazioni che potessero inserire l'arte all'interno di un dominio più vasto e inclusivo. Conferma di questa impostazione sarà il suo ultimo scritto completo, *L'Œil et l'Esprit* (1960), che riprende le tematiche artistiche per inserirle direttamente su di uno sfondo ontologico.

La scelta di rivolgersi alla pittura di Cézanne mostra una presa di posizione precisa: dipingere non è riconducibile alla rappresentazione fedele di un mondo esterno cui fedelmente ci si attiene nella riproduzione su tela. Il mondo che viene raffigurato non è un prodotto dell'attività di un ego geniale che proietta la sua ombra sull'oggetto della sua osservazione, né tanto meno una passività prona e acquiescente al suo cospetto. Due sono quindi i bersagli principali della sua disanima: il primo è il rappresentazionalismo e il secondo l'impressionismo. La critica al principio rappresentazionale si sostanzia nel rifiuto della logica del sorvolo, inteso come costruzione soggettiva della realtà. È come se operasse, a questo livello, una impostazione astrattamente scientifica e che impiegasse tutte le sue forze per far corrispondere il mondo alle nostre operazioni, esponendosi al rischio concreto di rendere la pittura esclusivamente rappresentazionale dimidiata, in quanto mostrebbe solo un lato parziale dell'azione dell'artista. Significativamente è proprio la «ruminazione del mondo» (Merleau-Ponty [1960] 1989, 16) ad essere considerata la specificità dell'operato dell'uomo d'arte, ossia la capacità di inserirsi nella realtà stessa per lasciarla affiorare in tutta la sua datità. Per questo motivo lo stesso impressionismo, con il suo sapiente uso del colore, la maestria della pennellata e l'atmosfera complessiva che suscita, palesa una tiepidezza strutturale, perché fa scomparire dalla composizione lo spessore, la materialità e la tridimensionalità dell'oggetto dipinto: «la peinture de l'atmosphère et la division des tons noyaient en même temps l'objet et en faisant disparaître la pesanteur propre» (Merleau-Ponty [1945] 1996, 13-33, qui p. 16). Una pittura d'impatto sicuramente, meno costretta dalla rigidità di

un'arte che ha nel sorvolo il centro nevralgico, ma altrettanto limitata dall'evanescenza del soggetto dipinto. Cézanne ha saputo penetrare lo spirito impressionista per uscirne con una rinnovata energia in grado di ridare stimolo alla produzione artistica. Il pittore di Aix-en-Provence, afferma Merleau-Ponty:

veut représenter l'objet, le retrouver derrière l'atmosphère [...] L'objet n'est plus couvert de reflets, perdu dans ses rapports à l'air et aux autres objet, il est comme éclairé sourdement de l'intérieur, la lumière émane de lui, et il en résulte une impression de solidité et matérialité. (Ivi, 16-17¹⁴)

Non si tratta più di una semplice e piatta rappresentazione, ma di un allenamento continuo e costante alla percezione che ci mette in contatto con il mondo stesso; è il raccoglimento da parte del pittore di quei raggi di mondo che ci vengono incontro. In poche parole, è l'instaurazione della dialettica tra principio di reversibilità, che vede nella continuità della diffusività – riecheggiando qui il *bonum diffusivum sui* – il suo tratto distintivo, e principio di irreversibilità, che ottiene tridimensionalità materica e singolarità ontologica.

Il pittore, di fatto, «vive nella fascinazione. Le sue azioni più proprie – quei gesti, quei segni di cui egli solo è capace, e che saranno rivelazioni per gli altri, che non hanno le sue medesime mancanze – gli sembrano emanare dalle cose stesse, come il disegno delle costellazioni» (Merleau-Ponty [1960] 1989, 26): ma può dirsi davvero creativo un lavoro ricettivo che si nutre di ciò che in fondo non è mai stato altrove da dove lo si viene a trovare? *Se a venire alla luce nell'opera è la trama del già presente, quale creazione è quindi in atto?* Ancora una volta il creare qualcosa di nuovo e il già-presente si trovano l'uno di fronte all'altro.

14. Da segnalare, riguardo la profondità della riflessione merleau-pontiana sull'arte di Cézanne e dei suoi possibili rapporti con il pensiero del filosofo giapponese Nishida Kitarō, che affronteremo nel prossimo paragrafo, gli ottimi contributi in Loughnane 2016 e 2019. A certificare un legame assai solido tra le due visioni del mondo, il giudizio di Joan Stanley-Baker sulla poetica pittorica di Sesshū Tōyō (1420-1506): «[o]n his return to Japan, he replaced the spatial ambiguity and inconsistencies of scale which had characterized the poetic Shūbun [metà del XV secolo e maestro di Sesshū Tōyō] style with a substantive, rational order. In a very late work, however, paying tribute to his Japanese teacher, he nostalgically returned to traditional Shūbun motifs. But the focus of the work has moved from empty space to solid masses». Cfr. Stanley-Baker 2014, 135.

Ma è pur sempre il pittore che opera con il suo stile e la sua biografia, irriducibilmente connessi:

Sa peinture serait un paradoxe: rechercher la réalité sans quitter la sensation, sans prendre d'autre guide que la nature dans l'impression immédiate, sans cerner les contours, sans encadrer la couleur par le dessin, sans composer la perspective ni le tableau. (Merleau-Ponty [1945] 1996, 17)

È questo il paradosso di Cézanne, lo scettico ricercare la realtà lasciandosi pervadere dalla stessa, senza perdere di vista, al contempo, l'atto fondamentale della visione.¹⁵ Paradossale è la presa di coscienza di questa duplice natura del fare artistico; il pittore, immerso nelle cose, si inserisce tra le venature di ciò che esiste non mostrando mai la cosa in sé, ma sempre il risultato di questo darsi continuo del reale filtrato dal suo sguardo e dalle sue mani. Ed è paradossale per due ordini di fattori: innanzitutto perché è il riconoscimento che la creazione artistica, in fin dei conti, non avrebbe potuto essere diversa da come è stata realizzata. Ancora una volta, non si tratta di affermare la prorompente soggettività dell'artista che plasmerebbe la materia a suo piacimento, concretizzando così la sua libertà nella creazione pittorica, né di cogliere nell'arte quel lavoro faticoso di allontanamento dai canoni naturali, come se stessimo assistendo ad un agone per una presunta originalità che il fare artistico strapperebbe alla natura. Vivendo nella fascinazione perenne, il pittore prolunga la natura inserendosi nella sua processualità senza perdere la propria singolarità stilistica. E conseguentemente, in secondo luogo, crea qualcosa che dipende totalmente dalla sua soggettività, qualcosa che non potrebbe esservi senza essere stata prima elaborata proprio dalla sua soggettività.¹⁶ È in gioco una ridefinizione dell'ampiezza della libertà, del concetto stesso di libertà. Come evidenziato da Chiara Palermo, «le travail de l'artiste

15. «Nous voyons la profondeur, le velouté, la mollesse, la dureté des objets – Cézanne disait même: leur odeur». Compito infinito quello del pittore che punta lo sguardo su di uno sfondo che resta misterioso: «[l']expression de ce qui *existe* est une tâche infinie». Cfr. Ivi, 20-21.

16. «For Merleau-Ponty, the painter offered not a picture of the world “as it is”, but a picture of the world coming into being in the percipient’s view of it, not before or after but *as* the attributes associated with use, significance, and value are applied», cfr. Gilmore 2004, 294.

est emblématique d'un surgissement contre tout dogmatisme; il est libre dans le sens où son action ne dépend ni des préjugés, ni même ancora des finalités déterminant la recherche d'une action utile», perché è la libera iniziativa creativa che porta in scena qualcosa di nuovo in grado di compiere quel salto che instaura un regime di discontinuità. Tuttavia, data l'accentuata sensibilità dell'artista, si può ammettere che «[i] ne peut pas ne faire ce qu'il fait: l'artiste est ainsi *lié* à la nécessité de créer son œuvre en dehors de toute contrainte externe, comme s'il était lié à sa vocation (voire même à sa *malédiction*), à laquelle il ne peut se soustraire». ¹⁷

Sotto questa luce acquisiscono massima pregnanza le profonde intuizioni merleau-pontiane sulla linea in *L'occhio e lo spirito*. Essa non può più essere intesa ingenuamente come la delimitazione di oggetti che si giustaporrebbero atomisticamente, tante parti che costituirebbero un fantomatico tutto additivo. Né tantomeno la confusione che ingloberebbe nell'indistinto le sue componenti. Si deve ritenere superato, pertanto, tanto il paradigma mereologico delle *partes extra partes*, quanto quello che rinviene nell'arte una qualsivoglia trascendenza dal sapore misticheggiante. La linea non è intravista, pertanto, nel suo essere già presente, ma nel *suo formarsi, nel suo divenire-linea, nel suo formare le cose e nel suo essere formata dalle cose stesse che contribuisce a formare*, in un rapporto di reciproca condizionalità che è, in fondo, espressione della natura chiasmatica della carne del mondo. La linea «non è più, come nella geometria classica, l'apparizione di un essere sul vuoto dello sfondo: è, come nelle geometrie moderne, restrizione, segregazione, modulazione di una spazialità preliminare» (Merleau-Ponty [1960] 1989, 54); è l'instaurazione di uno spazio topologico, di quel continuo *Ineinandersein*, di quell'esser-l'uno-nell'altro che mostra una volta di più la natura tissutale dell'essere; è l'esistenza di una dimensione preliminare non priva di intrinseche instabilità e di realtà già in atto ben prima di divenire apparenti. L'arte è quel dispositivo pratico e teoretico al contempo, grazie al quale dare spazio a ciò che da sempre è, ma che non può che manifestarsi nell'instaurazione di una novità originariamente inapparente. Non si tratta quindi di contribuire

17. Per entrambe le citazioni, cfr., Palermo 2015, (http://tristan.u-bourgogne.fr/CGC/publications/Transversales/Individu_dans_histoire/C_Palermo.html), consultato il 22/05/2023.

alla nascita di nuove parti mereologicamente intese, ma di far emergere elementi dallo sfondo incontaminato dell'Essere Selvaggio: l'arte ha allora natura primariamente ontologica perché «non è costruzione, artificio, rapporto industrioso con uno spazio e un mondo esterni» dal momento che «risveglia nella visione comune potenzialità dormienti, un segreto di preesistenza». (Ivi, 50)

Possiamo aggiungere, allora, un terzo fattore per illustrare il paradosso dell'artista, che affonda le sue radici proprio nella determinazione dell'essere barbarico – come soleva scrivere Merleau-Ponty – tra campo impersonale e inumano e la controparte umana. Ciò che mostra la pittura di Cézanne è questo intrecciarsi chiasmatico di una generatività indipendente da ogni intervento umano artificioso, ma che non può realizzarsi senza l'intervento del pittore.¹⁸ Lasciando la parola al filosofo francese, la pittura:

de Cézanne met en suspens ces habitudes et révèle le fond de nature inhumaine sur lequel l'homme s'installe. C'est pourquoi ses personnages sont étrangers et comme vus par un être d'une autre espèce. La nature elle-même est dépouillée des attributs qui la préparent pour des communions animistes [...] C'est un monde sans familiarité, où l'on n'est pas bien, qui interdit toute effusion humaine [...] Mais seul un homme justement est capable de cette vision qui va jusqu'aux racines, en deçà de l'humanité constituée». (Ivi, 22)

È all'insegna dell'unità fondamentale che si devono leggere queste considerazioni. Arte, vita, storia, biografia non possono esser tenute separate;¹⁹ tutte sono intrecciate così intimamente nel tessuto dell'essere che l'una si riflette sull'altra e grazie all'altra. Una «nouvelle ontologie» viene così immaginata e abbozzata dal filosofo francese in quella perfetta continuità discontinua che è il fil rouge tra i primi scritti degli anni '40 e gli ultimi rimasti incompiuti a causa della sua prematura scomparsa.

Terza tappa del nostro percorso consisterà nel prendere in consegna il suggerimento di Merleau-Ponty e investigare direttamente il

18. «Le peintre reprend et convertit justement en objet visible ce qui sans lui reste enfermé dans la vie séparée de chaque conscience: la vibration des apparences qui est le berceau des choses», cfr., Merleau-Ponty [1945] 1996, 23.

19. Cfr. Gilmore 2004, 315 e Palermo 2015.

precipitato di questa nuova ontologia. Mi rivolgerò allora alla filosofia formulata da Nishida Kitarō per illustrarne alcune ricadute notevoli.

4. «L'ontologia della produzione» di Nishida Kitarō

Il pungolo principale che ha dato avvio alla ricerca filosofica di Nishida Kitarō si è conservato intatto in tutta la sua travagliata produzione. Da *Uno studio sul bene* (1911) a *La logica del luogo e la visione religiosa del mondo* (1945), pubblicata poco prima della sua morte, si è squadernato in oltre trent'anni di attività il suo impegno a fornire una fondazione alla filosofia giapponese che potesse trovare da una parte collocazione nel panorama internazionale, ma che, dall'altra, potesse contribuire attivamente all'indagine filosofica globale, introducendo così secoli di filosofia nipponica – per forza di cose eterogenea rispetto al canone occidentale – nel cuore della ricerca.²⁰ Per raggiungere un risultato così ambizioso, il filosofo giapponese si è mosso operando una sintesi originale della concettualità filosofica, istituita da secoli di elaborazione, introiettandola nel sostrato culturale del Sol Levante. Ciò gli ha permesso, quindi, di entrare in contatto, spesso contrastivamente, con l'insieme delle nostre tradizioni, riuscendo così a formulare una lucida diagnosi delle contraddizioni insite nel filosofare occidentale e garantendo una migliore comprensione delle potenzialità inesprese. Si è, insomma, trattato di un felice incontro che da possibile scontro tra culture si è trasformato in un proficuo dialogo.

Il problema che è parso attanagliare la mente di Nishida Kitarō fin dal principio è stato quello della determinazione dell'*esperienza pura*, ovvero di un'esperienza che non venisse ricondotta alle secche dell'epistemologia kantiana (e postkantiana). In Occidente, quasi a testimoniare la bontà e la trasversalità dell'obiettivo nishidiano, tra la fine dell'800 e il principio del '900, due grandi filosofi si stavano misurando proprio su queste tematiche: William James e Henri Bergson. Presentare una filosofia che fosse in grado di oltrepassare il modello kantiano, la reazione idealista e il paradigma positivista all'epoca imperante, è stato il loro grande merito, una sfida che spronasse a ripensare da cima

20. Non è qui mia intenzione ripercorrere l'evoluzione del rapporto tra pensiero Orientale e Occidentale. Per una prima disamina, rimando all'ottimo Ghilardi 2011, soprattutto la parte prima.

a fondo l'ontologia. Ed è all'interno di questo fermento che si viene a inserire la proposta dell'esperienza pura. Certo, con la sensibilità propria della formazione dell'autore, ma con una universalità di pensiero impossibile da limitare al solo suolo nipponico. Per coglierne la portata leggiamo questo aneddoto tratto dal profilo che Keiji Nishitani ha offerto del suo maestro:

[m]i ricordo che una volta Nishida mi disse come durante una passeggiata a Kanazawa, un'ape o un tafano gli ronzò accanto all'orecchio e quel rumore risvegliò improvvisamente in lui la consapevolezza del fondamento dell'esperienza pura. È l'attimo del sentire diretto, prima che uno abbia il tempo di distinguere tra se stesso e le altre cose. Questa idea già indica quel fondamento a cui poi Nishida farà riferimento con la frase "diventando una cosa pensiamo, diventando una cosa agiamo". È la prospettiva del nulla o della vacuità del sé (Kitarō [1945] 2017, 26²¹).

Come si può notare, è presente in queste parole un profondo senso di unità fontale, nella quale le differenze vengono riassorbite dalla profondità del nulla. Il riferimento al Buddhismo è palese. L'esperienza pura viene allora tratteggiata come quel fondo di indifferenza che, per quanto presenti un pur flebile principio di differenziazione, la avvicina al misticismo o quanto meno ad un quasi esplicito panpsichismo. In realtà, consapevole di questi oggettivi limiti – ovvero la perdita di distinzione e la ricaduta nell'indifferenza generale – la sua ricerca si è sviluppata proprio nel tentativo di emendare la sua teorizzazione dell'esperienza pura senza abbandonarla del tutto.

Rifondare il pensiero filosofico da cima a fondo e riproporne una logica nuova: è questo ora il fine di Nishida, che lo ha visto occupato per tutta la sua vita. Momento di svolta, il saggio del 1926, intitolato *Luogo*.²² Non posso ripercorrerne in questa circostanza le argomentazioni, ma è comunque da sottolineare come con quest'ultimo si sia raggiunto uno snodo teoretico all'indietro del quale sarebbe impossibile retrocedere. Il luogo diviene così una spazialità aperta, un fondo trascendentale all'interno del quale le cose, la conoscenza delle cose e la relazione tra cose e conoscenza delle cose trovano ragion d'essere.

21. Traggo questa citazione dall'introduzione di Tosolini.

22. È possibile leggere la traduzione italiana di Enrico Fongaro in N. Kitarō [1926] 2012.

Trovato un punto fermo da cui ripartire – la trascendentalità del luogo – ha potuto prendere avvio la vera *pars construens* della sua filosofia. Mi soffermerò un po' più analiticamente, invece, su di un importante saggio del 1935, *The Standpoint of Active Intuition*, recentemente tradotto da Willam Haver e inserito in un volume dal titolo assai significativo: *Ontology Of Production. Three Essays*,²³ che offre feconde intuizioni per l'argomento che sto trattando in queste pagine.

Il testo si apre con una rilettura del concetto di tempo. La scelta dell'argomento è significativa: è stato da sempre considerato un problema strettamente ontologico e mostra al suo interno la portanza della dialettica tra il concetto di continuità (basti pensare alla durata bergsoniana) e quello di discontinuità (la scansione in passato-presente-futuro). In più Nishida vi rinviene una relazione strettissima tra queste due concettualizzazioni, dal momento che continuità e discontinuità non vengono considerate dicotomicamente, in quanto il tempo stesso si spazializza nella scansione di passato-presente-futuro. Il tempo è tale perché è spazio e lo spazio è spazio perché tempo. Una considerazione così vertiginosa è il guadagno maggiore di quella logica del *soku* (tradotta in italiano con *eppure*)²⁴ anticipata in *Luogo*. Definitivo, allora, il superamento del classico principio *e pluribus unum*: non è il molteplice che deve essere ricondotto all'unità, ma è l'unità stessa che è molteplice, così come il molteplice è intimamente unità: «[t]o speak of unity is not merely to say that many things become one; it is necessarily to say the many are one, and the one is many» (N. Kitarō 2012, 66²⁵). Ma così non ci si precipita verso una confusione originaria, riproponendo le medesime criticità emerse in riferimento all'esperienza pura? Per ovviare al problema e per rendere meno nebulosa la logica del *soku*, Nishida compie un ulteriore passo in avanti. Incrociando tempo e spazio – lo spazio è tempo oggettivato e irreversibile, mentre il tempo è spazio percorribile reversibilmente – ha potuto fornire pieno spessore ad una

23. Cfr. N. Kitarō 2012, 64-143. I tre saggi sono, oltre il già citato contributo del 1935, *Expressive Activity* (1925) e *Human Being* (1938).

24. «Le terme soku [...] indique l'identité paradoxale de la forme (*rupa*) et du vide (*sunyata*). En d'autres termes, l'affirmation et la négation sont en situation de tension dynamique», cfr., Tremblay 1996, 63-82, qui 73.

25. Per una chiara delucidazione di queste parole, cfr. l'*Introduzione* di Haver alla raccolta di saggi. Soprattutto p. 13.

delle sue intuizioni più notevoli: il principio di *continuità di discontinuità*. Come ho già rilevato, continuo e discontinuo non sono contrapposti dicotomicamente, ma intrecciati; la continuità temporale è la reversibilità spaziale, così come la discontinuità spaziale è l'irreversibilità temporale. Continuo-eppure-discontinuo; temporale-eppure-spaziale; reversibile-eppure-irreversibile: «[i]t may seem absurd, but contradictorily, objective time or true time can be conceived starting from the fact that the instants that can never return are arrayed simultaneously».
(Ivi, 72)

Raggiunto il piano della simultaneità dei tempi irreversibili, si apre così la strada alla 'messa a terra' di quella logica che vede nella singolarità la determinazione ultima dell'universale concreto.²⁶ Il singolo, di conseguenza, non viene considerato come eccezione particolare di un universale, ma come realizzazione concreta dell'universale stesso. Sarebbe quindi l'instaurazione della discontinuità sul fondo di continuità: in poche parole, esso non è un *unicum*, un unico in quanto centro di convergenza e di propagazione, quanto più l'intreccio delle singolarità e dell'universale. Neppure l'immagine del nodo di una rete di relazioni è in grado di riportarne pienamente il senso che possiede per Nishida. Si può dire che è il rispecchiarsi dell'universale nel particolare e viceversa; è il singolare-eppure-universale, universale-eppure-singolare.²⁷ Non esiste un universale che sussume astrattamente il particolare, né

26. Per una attenta analisi di questo fondamentale snodo dell'edificio teoretico del pensatore nipponico, cfr. N. Kitarō [1930] 2017.

27. Per descriverne la pregnanza filosofica, farò riferimento ad un aneddoto. Fellini era solito ripetere di aver sempre desiderato, un giorno, di divenire un aggettivo. E in effetti l'aggettivo "felliniano" è entrato nel vocabolario cinematografico. Ma cosa accade quando riconosciamo, all'interno di un film non diretto da Fellini, caratteristiche associabili all'atmosfera "felliniana"? L'aggettivazione svolge una doppia funzione: da una parte universalizza un tratto, quello che viene riconosciuto come specificamente affine allo stile del maestro riminese; dall'altro, però, singolarizza tanto quel tratto "felliniano" presente nel film non diretto da Fellini quanto quel medesimo tratto presente nei film di Fellini. Come a dire: raggiungendo la dimensione universale, l'aggettivo contribuisce alla singolarizzazione del tratto stesso che non può divenire predicato di un altro soggetto. Presentandosi come predicato universale che accomuna i due film, contribuisce al contempo a singolarizzarli sempre più, rendendoli così sempre più distinti l'uno dall'altro, sempre più discontinui nella continuità di fondo. Va da sé, ovviamente, che è l'aggettivo ad avere valore attributivo e non viceversa: l'aggettivazione è una specie dell'attributo.

un particolare che viene a coincidere con istanziazioni dell'universale, perché la singolarità «is neither the particular, conceptualized according to a merely abstract logic, nor is it the universal. There is no sense whatever in which it is a matter of there being a universal behind a singularity» (N. Kitarō [1935], 78²⁸). La conoscenza strutturata nella distinzione soggetto-oggetto inevitabilmente mostra di essere insufficiente al cospetto di teorizzazioni simili. Per questo si deve procedere ulteriormente.

Alla base di queste considerazioni nishidiane vi è il concetto di *intuizione attiva*, da non intendersi come attività anti intellettualistica avanzata comunque da un soggetto che afferrerebbe un oggetto evanescente; l'intuizione è attiva perché non è attività di un soggetto, non è l'attività epistemologica di un soggetto conoscente, ma la ricomprensione del fare epistemico all'interno dell'attività produttiva: da qui il titolo dell'edizione inglese, *ontology of production*. Lascio parlare direttamente Nishida:

I am not trying to think the world in terms of consciousness or subjectivity. Neither am I trying to think the spiritual as the ground of the actually existing. On the contrary, I am trying to think consciousness as a moment of the self-determination of the dialectical world. The actually existing world that itself determines itself is always thought as the actually existing present; one aspect of the actually existing present is always consciousness». (N. Kitarō 2012, 83)

Coerentemente, la stessa creazione artistica non può che essere inserita all'interno di questo mondo che è esso stesso creativo. L'arte è creativa perché porta allo scoperto la natura produttiva del reale che l'artista rintraccia e fa propria; *crea sempre qualcosa di nuovo perché è la trama stessa del luogo nel quale si trova ad essere produttiva*. L'artista non è attivo né passivo perché consapevole della continuità di discontinuità.²⁹ Singolarità, discontinuità, novità sono tutti sinonimi

28. Per questo motivo non concordo pienamente – per quanto sia comunque necessario ammettere un'ambiguità di fondo di Nishida che rende ogni affermazione assai sdruciolevole – con Adam Loughnane laddove sostiene che il filosofo giapponese abbia voluto muoversi «towards a fully reversible form of reversibility», dal momento che è proprio l'irreversibile ad essere il cuore della sua logica del *soku*. Cfr. Loughnane 2016, 72.

29. «In the context of activity and passivity, the first view is mistaken because it presupposes a pre-formed subjective agent that acts and molds a passive world, and

che trovano la loro sorgente nella dimensione espressiva dell'ontologia di Nishida. Né particolare né universale, ciò che esiste è proprio questo intrecciarsi di dinamiche produttive. Non a caso, negli ultimi anni di vita, il filosofo giapponese ha più volte espresso la sua vicinanza alla filosofia di Leibniz, forse in vista di un superamento dell'impostazione del grande predecessore e in direzione di quella che Yoneyama Masaru ha definito «monadologia creatrice» (Masaru 2016, 204³⁰). Come che sia, l'analisi del suo pensiero ci ha permesso di avanzare ulteriormente in quell'enigma che è il creare novità, aprendo il campo a ulteriori ricerche. Non mi resta, quindi, che riannodare i fili del discorso per focalizzare con maggiore intensità l'attenzione sulla problematica di fondo.

Questioni aperte

Si è affrontato un lungo *detour* che ha visto prendere spunto da una circostanza particolare, il caso eclatante di AlphaZero, che ha spariagliato le carte delle nostre conoscenze riguardante il potenziale di un sistema informatico impegnato nel gioco degli scacchi, per poi allargare il campo a una serie di problemi che coinvolgono direttamente la ricerca filosofica. E questo è, ad opinione di chi scrive, un merito proprio delle tecnologie di ultima generazione. Presentando fin nella denominazione – *Artificial Intelligence* – caratteristiche umane, software così evoluti possono offrire prospettive inesplorate, perché precedentemente inesplorabili, riguardanti i processi creativi umani. Se questi ultimi restano ancora in larga misura misteriosi, studiare da vicino il comportamento di una intelligenza artificiale non può che rivelarsi fecondo. Lo abbiamo visto: la peculiarità di AlphaZero è quella di apprendere tramite esperienza, di immagazzinare nozioni che non si sclerotizzano in schemi preconfigurati e che mantengono, di contro, elasticità e plasticità espansiva. Quindi non uno sterile intasamento nozionistico, ma un'effervescenza cognitiva, per così dire, invidiabile. Certo, importante è non ricadere in una dinamica inversamente proiettiva, ovvero tenere sempre bene a mente una domanda: *chi rispecchia*

the second view is mistaken because it assumes an active world that acts and molds a passive subject», Peters 2018, 477-496, specificatamente 490.

30. Cfr. Ivi, 203-215.

chi? È l'intelligenza artificiale costruita sul modello di quella umana, o quella umana viene sempre più equiparata a quella artificiale? Non mi sono impegnato a rispondere a questa domanda semplicemente perché non saprei cosa rispondere; dico solamente che affermare una sacralità inviolabile dell'essere umano che riproponga la tradizionale contrapposizione natura-tecnica è quanto meno sospetto, per non dire errato. Simondon docet. Però apre ad una ulteriore questione di capitale importanza: *cosa vuol dire creare?* E un'altra, strettamente connessa alla prima: *come è possibile creare novità?* Il problema che oscilla tra il tecnico e il semantico, si tinge di atmosfera propriamente filosofica quando ci si interroga sulle implicazioni dell'esistenza di qualcosa di nuovo. E questa è la problematica ontologica.

Il percorso tratteggiato vuole mostrare proprio come le nostre ontologie vengano forzate una volta che si trovano alle prese con il nuovo. Cercare di cogliere la natura della novità ci porta a riconsiderare i nostri abituali schemi cognitivi con i quali cataloghiamo il reale e ci costringe a interrogarci sulla definitività di un tale approccio. Il rovesciamento della metafisica kantiana, promossa tanto da Simondon quanto da Merleau-Ponty e Nishida, sebbene si sia strutturato in modalità differenti e non perfettamente sovrapponibili, è stato il filo rosso che ha permesso di collocare la loro teoresi all'interno della medesima problematica di fondo. Che il nuovo dipenda da un soggetto che lo riconosca come tale e lo manifesti è palese; ma può essere ricondotto solamente a ciò? L'invenzione in Simondon e la linea in Merleau-Ponty mostrano l'esatto contrario: l'emergenza di qualcosa *già* ben presente prima di divenire apparente, eppure *assolutamente nuovo*. Creare novità si pone tra relatività epistemica (il nuovo è relativo ad un determinato dominio del sapere) e assolutezza ontologica (il fenomeno x non è il fenomeno y). Consapevoli, però, che vale anche la reciproca: assolutezza epistemica (i domini individuano effettivi campi di azione) e relatività ontologica (l'unità fondamentale che permette che il fenomeno x non sia il fenomeno y). La nishidiana logica del *soku* ha cercato di offrire a tutto ciò un'originale soluzione.

Per questo le questioni restano aperte: perché è in gioco la revisione di contrapposizioni concettuali ritenute canoniche, quali la scissione soggetto-oggetto, l'assoluto e il relativo, il continuo e il discontinuo, l'atto e la potenza. Lo scompiglio che AlphaZero ha seminato nel

mondo degli scacchi è divenuto una sfida filosofica che richiede di essere affrontata con rinnovato impegno.

Riferimenti bibliografici

Barthélémy, J.H. 2014, *Simondon*, Les Belles Lettres, Paris.

De Caro, M. 2020, *Tre minacce del progresso tecnologico*, in Alici, L. & Miano, F. *L'etica nel futuro*, Orthotes, Napoli-Salerno.

Id. 2021, *Intelligenza artificiale*, in Id., Magni, S. F., Vaccarezza, S., *Le sfide dell'etica*, Mondadori Università, Milano.

Ghilardi, G. 2018, *Etica dell'agire scientifico e tecnologico*, Mimesis, Milano-Udine.

Ghilardi, M. 2011, *Arte e pensiero in Giappone. Corpo, immagine, gesto*, Mimesis, Milano-Udine.

Gilmore, J. 2004 *Between Philosophy and Art*, in Carman, T. & Hansen, M.B.N. *The Cambridge Companion to Merleau-Ponty*, PCU, Cambridge.

Kitarō, N. 2012, *Luogo*, Mimesis, Milano-Udine.

Id. 2017, *Autoéveil. Le système des universels*, Chisokudō, Nagoya.

Id. 2012, *The Standpoint of Active Intuition*, in Id., *Ontology of Production. Three Essays*, Duke University Press, Durham and London.

Id. 2017, *La logica del luogo e la visione religiosa del mondo*, Chisokudō, Nagoya.

Kuroda, A. 2017, *Lieu de médiation: Nishida, Tanabe, Simondon*, "European Journal of Japanese Philosophy", 2, 209-228. (<https://ejjp-journal.org/ejpp-2-2017/>).

Loughnane, A. 2016, *Nishida and Merleau-Ponty. Art, "Depth", and Seeing without a Seer*, "European Journal of Japanese Philosophy", 1,

47-74. (<https://ejjp-journal.org/ejpp-1-2016/>).

Id. 2019, *Merleau-Ponty and Nishida. Artistic Expression as Motor-Perceptual Faith*, SUNY, New York.

Masaru, Y. 2016, *Verso l'estetica del luogo. Per una monadologia polifonica*, "European Journal of Japanese Philosophy", 1, 47-74. (<https://ejjp-journal.org/ejpp-1-2016/>).

Merleau-Ponty, M. [1945], *Le doute de Cézanne*, in Id. (1996), *Sens et non-sens*, Gallimard, Paris, 13-33.

Id. [1960], *L'Œil et L'Esprit*, trad. it. 1989 *L'occhio e lo spirito*, SE, Milano.

Montani, P. 2022, *Destini tecnologici dell'immaginazione*, Mimesis, Milano-Udine.

Montebello, P. 2011, *Simondon, une Métaphysique de la Participation* in Debaise, D. *Philosophie des possessions*, Les Presses du réel, Dijon.

Moravec, H. P. 1988, *Mind children: the future of robot and human intelligence*, Harvard University Press, Cambridge.

Mugnai, M. 2023, *Come non insegnare filosofia*, Raffaello Cortina Editore, Milano.

Palermo, C. 2015, *Être agi: Liberté, adversité et création chez Merleau-Ponty*, "Revue TRANSVERSALES du Centre Georges Chevrier", 5. (http://tristan.u-bourgogne.fr/CGC/publications/Transversales/Individu_dans_histoire/C_Palermo.html).

Peters, K. 2018, *Artistic Production and the Making of the Artist: Applying Nishida Kitarō to Discussion of Authorship*, in "Philosophy East & West", 68 (2), 477-496.

Sadler, M. & Regan, N. 2019 *Game Changer. AlphaZero's Groundbreaking Chess Strategies and the Promise of AI*, New in Chess, Alkmaar.

Simondon, G. [1965-66] 2014, *Imagination et invention*, PUF, Paris.

Id. [1974] 2018, *La résolution des problèmes*, PUF, Paris.

Stanley-Baker, J. 2014, *Japanese Art*, Thames & Hudson, London.

Trembaly, J. 1996, *Néantisation et relationalité chez Nishida Kitarō et Watsuji Testurō*, "Théologiques", 4 (2), 63-82.

Simone Vaccaro
Ricercatore indipendente
arenaphilosophika@gmail.com



ἘΠΈΚΕΙΝΑ

International Journal of Ontology
History and Critics

JIM SCHRUB

Ces machines qui *nous* surprennent

Une défense simondonienne de la différence entre la créativité des vivants et la créativité induite des machines depuis la notion de métabolisme

EPEKEINA, vol. 16, n. 1-2 (2023), pp. 1-43

Miscellanea

ISSN: 2281-3209

DOI: 10.7408/epkn.

Published on-line by:

CRF – CENTRO INTERNAZIONALE PER LA RICERCA FILOSOFICA
PALERMO (ITALY)

www.ricercafilosofica.it/epekeina



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs 3.0 Unported License.

Ces machines qui *nous* surprennent

Une défense simondonienne de la différence entre la créativité des vivants et la créativité induite des machines depuis la notion de métabolisme

Jim Schrub

Paratexte

Je tiens d'abord à remercier tou-te-s les participant-e-s de la 7e édition des passages philosophiques, ainsi qu'Elie During, Séverine Guillet et Cléo Collomb qui ont su nourrir les réflexions développées dans cet article.

Alors que je m'apprete à proposer un développement relatif au rapport entre IA, société et créativité, je tiens à mentionner que Chat-GPT coûte sept-cent mille dollars par jour, et que son entraînement aurait requis l'émission de 552 tonnes de CO₂. Son bilan carbone de fonctionnement reste, à ma connaissance, non-évalué. Entraîner Chat-GPT, le jouet de la compagnie d'Elon Musk, Sam Altman et Peter Thiel, a donc émis autant de CO₂ que les 22 millions d'habitants du Mali en ont émis en un an. Le 28e rapport du GIEC est pourtant sans appel: chaque tonne de CO₂ non-émise compte. Le déploiement de Chat-GPT aura, par ailleurs, constitué une énorme campagne de communication favorisant l'acceptabilité sociale des IAs sur le web, tout en développant la conscience de la difficulté à venir de distinguer entre un contenu produit par un humain d'un contenu produit par une IA. Ceci permet à Sam Altman, un an plus tard, de présenter sa nouvelle entreprise, Worldcoin, qui déclare qu'à l'heure où l'IA et l'humain deviennent indiscernables en ligne, il est nécessaire de consentir à l'emploi massif de données biométriques d'identification sur le web. Déjà 450 000 iris de personnes vivant pour la plupart dans les Suds ont été scannés contre de l'argent. Une partie des mots de passe et accès au service ont déjà été hackés. L'article qui suit a été difficile à écrire pour exactement cette raison; je prends ici position en faveur de l'impossibilité totale de faire l'économie d'une politisation de ses questionnements.

Introduction

Dès les premières lignes du chapitre que Simondon consacre à l'invention dans un texte de cours, *Imagination et invention*, il s'emploie à définir l'invention depuis la notion de problème, telle qu'elle viendrait qualifier une situation:

Est problématique la situation qui dualise l'action, la tronçonne en la séparant en segments, soit parce qu'il manque un moyen terme, soit parce que la réalisation d'une partie de l'action détruit une autre partie également nécessaire; hiatus et incompatibilité sont les deux modes problématiques fondamentaux (Simondon 2014, 139).

L'invention suppose une situation problématique, qu'elle permettra de résoudre. On peut donc identifier la situation problématique comme le point de départ de l'inventivité et de la création. Ce point de départ nous permet de poser un cadre théorique depuis lequel traiter la question de la créativité chez les IAs, de trois manières. D'abord, en interrogeant la possibilité pour un modèle de traitement du langage naturel (multimodal si l'on intègre des programmes tels que Midjourney par exemple) d'être concerné par la notion de problème ainsi définie. Ensuite, en interrogeant la manière dont l'introduction de ces nouveaux outils dans la culture matérielle humaine produit à son tour de nouvelles lignes de partage dans nos conceptions de la créativité et de l'expression, lignes de partage qu'il s'agirait donc de penser. Enfin, dans la mesure où Simondon propose deux types de problèmes, celui procédant d'un hiatus, et l'autre d'une incompatibilité, on peut se demander dans les deux cas s'il l'on aurait affaire à un hiatus ou à une incompatibilité.

Nous proposerons, à partir de cette notion de problème, de montrer quelles sont les limites théoriques qui à mon sens empêchent que l'on détermine les IAs comme étant créatives, y compris lorsqu'elles sont utilisées à des fins de productions esthétiques et culturelles. Nous nous interrogerons également sur la nature de cet effort par lequel on essaie d'aménager notre concept de créativité pour qu'il devienne compatible avec la computation. Ceci permettra d'interroger, enfin, ce que l'on peut appeler le milieu associé des IAs, un milieu humain pour lequel il y a culture, production symbolique et production de valeur. Nous mettrons au centre de notre réflexion les notions de virtuel,

de milieu et de valeur. Le virtuel comme dimension nécessaire de l'inventivité, dont sont privées les IAs, nous conduit à refuser qu'elles puissent être déclarées créatives au même titre que les êtres vivants, quoique la métaphore organiciste et l'anthropomorphisme que l'on mobilise pour parler des IAs nous induisent parfois à projeter ces propriétés sur elles. Nous détaillerons les propriétés du milieu associé chez Simondon, ce qui permettra d'insister sur le fait que le milieu dit associé des IAs, ce sont les humains, et que ce sont eux qui restent bien les dernières instances régulatrices des programmes. Enfin, en insistant sur le rapport entre la créativité dans les activités humaines et la valeur (car une situation problématique est une situation qu'il importe de résoudre), et sachant que cette valorisation n'est pas réductible à une récompense quelconque programmée pour favoriser un entraînement, nous concluons qu'en l'état actuel des choses, il n'est pas possible de qualifier une IA de créative. Précisons toutefois d'emblée que cela ne signifie absolument pas que l'IA n'a pas de valeur ni qu'il est absurde de vouloir créer des objets culturels avec elle. Simplement, nous tenons à insister sur le fait que la valeur de ses productions ne peut pas dériver de l'outil utilisé.

Pour ce faire, nous resterons dans un corpus restreint: nous n'entendons pas produire un discours général sur les IAs qui, en tant que *task-specific*, sont aussi peu homogénéisables que la technique en général. Nous nous concentrerons sur les IAs qui produisent des contenus susceptibles d'avoir une valeur esthétique et culturelle, et qui fonctionnent sur des LLM, *large language models*. Il ne sera par conséquent pas question ici des IAs symboliques (telles que le célèbre AARON d'Harold Cohen), qui procèdent par instructions déterminées. Il s'agira plutôt des IAs connexionnistes les plus avancées, celles dont la lignée technique a commencé avec McCulloch et le modèle d'activation du neurone en 1943, avec le *perceptron* de Rosenblatt dans les années 1960, et dont l'étude a repris spectaculairement en 2012 avec Hinton et LeCun (Cardon *et al.* 2018, 181). Nous aborderons donc les IAs connexionnistes telles que les LLMs de GPT-3 et son agent conversationnel, ainsi que les LLMs multimodaux tels que Midjourney.

1. Qui invente?

Au cours de la première moitié du vingtième siècle, plusieurs entomologistes et biologistes rencontrent et étudient plusieurs espèces d'animaux non humains (la guêpe *Ammophila urnaria*, les fourmis *Oecophylla smaragdina* ou *Camponotus senex*) qui ont recours à des médiations instrumentales. Voilà que la définition de l'être humain comme *homo faber*, attribuée à Benjamin Franklin puis Bergson, est mise en cause dans son exclusivité. Qu'est-ce qui, dans la médiation instrumentale humaine, permettrait de continuer à distinguer entre la technique hominisante, et une technique toujours dérivative de l'instinct, sans voie d'accès au virtuel ?

On peut ne pas se satisfaire de la manière dont la question est posée, au nom de la distinction qui y est posée entre réponse instinctive et réponse «intelligente». Or, Simondon développe à ce sujet une voie intermédiaire: il fait de l'inventivité une fonction de l'intelligence non seulement individuelle, mais aussi collective et qui ne soit effectivement pas réservée qu'aux humains. Il essaie de penser l'inventivité des autres espèces selon un ensemble de variables plus riche que la simple intelligence individuelle. Il propose :

- (1) d'intégrer une échelle collective à l'inventivité, qui n'est alors plus solidaire d'un génie individuel extraordinaire
- (2) de penser la conjugaison des échelles individuelles et collectives de l'intelligence
- (3) de se garder de faire de l'équidistribution d'une médiation instrumentale à travers l'espèce comme signe de la nature instinctive de cette conduite, en envisageant (a) que plusieurs individus aient pu découvrir la même solution à un problème qui s'était posé à chacun d'eux, ou (b) qu'elle s'est répandue au présent dans une colonie
- (4) d'intégrer à l'inventivité des espèces les dimensions situationnelles des problèmes que ces dernières ont dû résoudre, et qui permettent de rappeler que l'invention possède une genèse dans des problèmes posés *hic et nunc* à des êtres.

Ces propositions rendent possible, grâce à la multiplication des échelles et l'intégration d'une dimension situationnelle précise, un cadre d'analyse qui ne soit pas anthropomorphiste ni individualiste, cadre qui avait été adopté majoritairement jusqu'alors (l'intelligence est d'abord la propriété d'un *individu* intelligent, et d'un individu *humain* en particulier). S'éloigner de l'anthropomorphisme et de l'individualisme permet de s'éloigner des modèles que l'on pourrait appeler «héroïques et miraculeux» de l'inventivité, tel qu'on peut les retrouver par exemple dans les théories scientifiques. On retrouve par exemple ce modèle dans un texte de Poincaré en 1908, dont les reprises ont été retracées par Judith Schangler dans la généalogie qu'elle propose, de Poincaré à Koestler, des modèles discontinuistes de l'invention de concepts scientifiques (Stengers & Schlanger 1991, 72). Dans un geste critique similaire, nous nous attaquerons aux limites de l'anthropomorphisme et de la métaphore organiciste souvent mobilisée à titre heuristique ou pédagogique au sujet des IAs et de leur «évolution». L'anthropomorphisme a déjà été traité par Watson dans un article récent (Watson 2019), il se concentrait sur les conséquences du vocable anthropomorphique sur notre promptitude collective à transférer aux IAs des décisions qui nous concernent. Ce problème, actuel, concerne toujours les chercheur·e·s et les communiquant·e·s.¹ En ce qui nous concerne, c'est de l'association de propriétés d'un organisme à une machine computationnelle que nous allons traiter.

2. Les IAs sont-elles des zombies? Sur les limites de la métaphore organiciste

Si les IAs sont effectivement similaires à des organismes, puisqu'elles sont interprétables depuis les mêmes schèmes conceptuels, à quels organismes sont-elles similaires? La plupart du temps, c'est aux humains qu'il s'agit de les assimiler, en particulier pour favoriser l'acceptation de ces dernières au sein de l'espace social le plus large possible (Yang *et al.* 2022). C'est la raison pour laquelle, par exemple, c'est Chat-GPT

1. *Why Does AI Lie, and What Can We Do About It?*, 2022, [en ligne], Url: https://www.youtube.com/watch?v=w65p_Iip6JY. Il est évident que Miles sait de quoi il parle. Néanmoins, l'option pédagogique retenue est discutable car elle repose aussi sur une visualisation fallacieuse des opérations du système.

plutôt que GPT-3 qui parvient à atteindre le grand public. Son anthropomorphisme, trompeur, est aussi ce qui la rend acceptable (Rantala 2020²).

Mais partons de cette image d'IA anthropomorphe. Que se passe-t-il lorsque les objets techniques IAs sont mis en circulation? Lorsqu'une IA est déployée, comme pour l'agent conversationnel d'OpenAI, elle n'est plus en cours d'entraînement et n'évolue plus. La nature dynamique et perfectible du programme n'est plus. Dès lors, il faut ôter la vie à cet objet technique anthropomorphe: les IAs sont alors plus proches des zombies que des humains, en ceci qu'elles ne peuvent faire qu'avec ce avec qu'elles sont depuis l'arrêt de l'entraînement. Lorsqu'on les sollicite par l'envoi d'un prompt, elles ne sont plus dans un état dans lequel elles peuvent se transformer. Derrière l'anthropomorphisme et la personnalisation des IAs, c'est le rapport de l'objet technique avec le vivant qui me semble poser le problème le plus conséquent ici.

Car dans les conditions d'une IA déployée, il n'est pas possible de concevoir les interactions utilisatrice/IA comme des interactions équivalentes à une interaction entre deux êtres «organiques». Ceci, parce qu'une fois que les programmes sont entraînés, ces derniers ne «métabolisent» plus rien. Une IA déployée n'apprend plus, ne modifie plus ses structures – notamment pour des raisons de sécurité. Il n'y a, alors, plus de relation intégrative, ni d'autorégulation.

2. Ce point a notamment été étudié sous un angle simondonien par Rantala. [En ligne], URL: <http://ebooks.iospress.nl/doi/10.3233/FAIA200946>. Pour une revue de littérature complète, qui pointe l'instabilité du concept d'anthropomorphisme et les conclusions divergentes quant à l'impact de l'anthropomorphisme sur l'acceptabilité sociale des IAs, voir Li M. & Suh A., *Machinelike or Humanlike? A Literature Review of Anthropomorphism in AI-Enabled Technology* [en ligne], dans *Proceedings of the 54th Hawaii International Conference on System Sciences*, présenté à 54th Hawaii International Conference on System Sciences (HICSS 2021), 2021, Url: [https://scholars.cityu.edu.hk/en/publications/machinelike-or-humanlike-a-literature-review-of-anthropomorphism-in-ai-enabled-technology\(7c9fa6e3-ab93-4d72-ab37-219f2733dd1e\).html](https://scholars.cityu.edu.hk/en/publications/machinelike-or-humanlike-a-literature-review-of-anthropomorphism-in-ai-enabled-technology(7c9fa6e3-ab93-4d72-ab37-219f2733dd1e).html); Blut M. *et al.*, *Understanding anthropomorphism in service provision: a meta-analysis of physical robots, chatbots, and other AI*, «Journal of the Academy of Marketing Science», 49(4), 2021.

3. Le vivant et le technique

Ce qui fait d'une IA un zombie correspond en langage simondonien à ce qui fait d'un objet technique un objet technique plutôt qu'un être vivant. Le parallèle se justifie ici d'autant mieux que le zombie et le non-zombie entretiennent une proximité que l'objet technique et l'être vivant entretiennent également lorsqu'ils sont définis par l'auteur depuis la notion de transduction (Simondon 2012, 198-199). En effet, on peut distinguer deux types de transducteurs: ceux qui relèvent d'une transduction mécanique, et les transducteurs propres au vivant. Simondon se garde toutefois de rejouer dans cette dichotomie la carte de la métaphysique moderne selon laquelle on aurait au niveau du mécanique un déterminisme strict et laplacien, et au niveau du vivant la présence de la liberté sous la forme d'une indétermination chronique. En effet, il précise immédiatement que

les transducteurs mécaniques sont des systèmes qui comportent une marge d'indétermination; l'information est ce qui apporte la détermination. Mais il faut que cette information soit donnée au transducteur; il ne l'invente pas; elle lui est donnée par un mécanisme analogue à celui de la perception chez le vivant, par exemple par un signal provenant de la manière dont l'effecteur fonctionne [...]. Au contraire, le vivant a la capacité de se donner à lui-même une information, même en l'absence de toute perception, parce qu'il possède la capacité de modifier les formes des problèmes à résoudre; pour la machine, il n'y a pas de problèmes, mais seulement des données modulant des transducteurs.(Simondon 2012, 199)

Autrement dit, la différence n'est pas celle qui séparerait le déterminisme mécaniste classique de la liberté métaphysique de l'être humain. Ce sont les affordances des deux transducteurs quant aux dimensions temporelles du réel auxquelles ces transducteurs sont réceptifs qui les distinguent. Le vivant possède de manière spécifique la possibilité de s'individuer depuis lui-même, de se «donner à lui-même une information». Cette capacité est justifiée dans le texte par la proposition suivante «il possède la capacité de modifier les formes des problèmes à résoudre». C'est la possibilité d'opérer cette modification qui fait la spécificité du vivant par opposition à la machine, qui ne peut encore une fois fonctionner qu'avec ce qu'elle a. Que suppose «pouvoir modifier les formes du problème à résoudre»? On trouve la réponse à cette

question un peu plus loin dans le texte. Si le vivant peut modifier les formes des problèmes qui sont les siens, c'est aussi parce qu'il peut faire avec ce qui n'est pas et n'est pas encore, avec non seulement le présent et le passé (les données enregistrées pour une IA). C'est ce que Simondon appelle le virtuel, et c'est ce qui rend possible la résolution de problème, c'est-à-dire l'invention:

La résolution des véritables problèmes est une fonction vitale supposant un mode d'action récurrente qui ne peut exister dans une machine: la récurrence de l'avenir sur le présent, du virtuel sur l'actuel. Il n'y a pas de véritable virtuel pour une machine[...] il n'y a pas de modification de formes qui soit orientée par le pressentiment du problème à résoudre; le virtuel ne réagit pas sur l'actuel, parce que le virtuel ne peut jouer aucun rôle en tant que virtuel pour la machine.(Simondon 2012, 200)

L'intégration de ce qui n'est pas (encore), c'est-à-dire du virtuel nous permet alors de distinguer entre deux types d'indétermination, celle relative à la machine et celle relative au vivant. Il existerait en effet une indétermination déterminée depuis l'actuel seulement, et l'indétermination apportée par une relation à ce qui n'est pas encore, mais qui peut être intégré à ce qui est comme l'une de ses dimensions: c'est le virtuel. Soulignons toutefois que l'«actuel» ne correspond pas à l'«intérieur»: il ne s'agit pas de dire que la machine est fermée sur elle-même, ni qu'elle serait privée de rapport avec l'extérieur et ce qui n'est pas elle-même. Il est même décisif d'insister sur le fait que ce rapport existe: l'envoi d'un prompt, d'une ou plusieurs images, l'alimentation de la machine en sont des exemples. Toutefois ce rapport n'implique pas que les IAs lorsqu'elles sont déployées soient susceptibles d'entretenir un rapport quelconque avec le virtuel, car elles ne sont pas capables d'intégrer quoi que ce soit, ni de faire avec ce qui n'est pas. L'existence d'une interaction dynamique entre un objet technique et un environnement n'est pas analytiquement corrélée à une situation où la virtualité serait partagée. Si une IA interagit avec un humain, le sens de la virtualité sera asymétrique.

Par conséquent, les algorithmes dont nous nous servons, qui sont récursifs, ne reposent que sur la récursion possible du présent vers le passé, c'est-à-dire la transduction machinique; tandis que le vivant dispose d'un autre type de boucle de rétroaction, celui de l'avenir sur

le passé. Le sens de l'avenir dont il est question chez les individus biologiques est en ce sens irréductible à un pronostic, à une prédiction basée sur des données présentes. En tant que machine non vivante, l'IA est, toutes choses égales par ailleurs, incapable de contact avec le virtuel et l'avenir, dont on pose qu'il n'est pas réductible à une prédiction statistique basée sur une appréhension du réel par data-ification d'une partie de ce dernier. (Simondon 2012, 200)

Ainsi, l'invention est suscitée par une situation problématique, c'est-à-dire une situation dans laquelle persiste un hiatus ou une incompatibilité entre deux ou plusieurs choses; et la résolution d'un problème requiert, pour qu'il y ait invention, réforme des formes initiales des données du problème ainsi que rapport avec ce qui n'est pas et avec un avenir possible, c'est-à-dire avec du virtuel. Que devient, dès lors la créativité computationnelle dans ce cadre de référence? Pas grand-chose, puisque les IAs n'ont rien de vivant. Si l'on pose une équivalence entre la création et l'invention, de sorte que la créativité devient la condition de possibilité de l'invention, alors la créativité n'est pas un concept qui a un sens pour les IAs déployées.³ Mais alors, comment comprendre la dimension dynamique des échanges que l'on peut avoir avec les agents conversationnels ou avec les robots émotionnels (Devillers 2020, chap. 6.3), comment comprendre à quel point ces interactions peuvent nous être *appropriées*, inédites sans pour autant recourir au virtuel? Comment expliquer que l'absence de partage du sens du virtuel semble ne pas (immédiatement) poser problème?

4. De l'organisme au personnage de *Huis-Clos*

Il semble que pour répondre à cette question, il nous faille retravailler l'image de l'IA anthropomorphe et sans vie. En effet, après avoir proposé de visualiser les IAs comme des zombies, saisissons les agents conversationnels – et par extension, les IAs – en leur donnant des noms qui seront peut-être reconnus: Estelle, Garcin, et Inès.

3. Déployées, parce que nous n'excluons pas que la question de la créativité puisse être posée durant la phase d'entraînement, quoique l'ajustement de paramètres pour l'accomplissement d'une fonction donnée (l'accomplissement d'un but, fut-il aligné ou non avec l'objectif humain initial) puisse être distingué de la genèse d'une fonction: il nous faut distinguer entre une IA «qui apprend» et une IA «qui a appris».

Chat-GPT est un personnage de *Huis-Clos* (Sartre 1947): déjà mort, il peut être mis en scène dans une situation donnée. En l'occurrence, l'enfer. Dans cet enfer, l'agent conversationnel pourrait interagir avec d'autres personnages, et ce de manière appropriée à la situation dans laquelle ils se trouvent. Les personnages parviennent à avoir un dialogue convenu, malgré la nouveauté de la situation dans laquelle la pièce les plonge. On peut même aller jusqu'à dire que les personnages semblent avoir un petit rayon d'action, quoique leurs projets respectifs soient voués à l'échec, puisque la pièce est structurée autour d'un *fait* auquel il est impossible d'échapper: la mort a figé leur être dans une identité de laquelle chaque personnage aimerait se détourner, alors que cela n'est plus possible étant donné qu'il et elles sont mort·es. Il ne leur est plus possible, quoiqu'ils en pensent, de se transformer et autre chose que ce que ce qu'ils ont été, et ce de que leur vie aura fait d'eux. Plus de réforme possible de soi, pas de transformation: seule persiste l'itération de ce qui est.

C'est cela que nous entendons par absence de métabolisation, et ce qui nous semble occulté par un langage qui a tendance à traiter les IAs comme des organismes vivants. L'absence de métabolisme implique une double incapacité: d'une part, l'incapacité à réformer ses propres formes depuis une participation extérieure, en l'occurrence depuis l'avenir comme virtuel. D'autre part, l'incapacité de réformer ses propres formes depuis soi seul, en tant que le vivant est capable de se donner à lui-même une information, c'est-à-dire quelque chose qui produit une transformation du et dans le système original. Cette absence paraît tout à fait décisive pour penser la différence fondamentale qu'il existe entre le résultat d'un processus computationnel d'un côté, et le résultat d'un processus d'individuation de l'autre. Dans le rapport humain-programme, y compris lorsque ce rapport est dynamique, les fonctions de régulation et d'individuation sont bien encore assurées par les humains (Simondon 2012, 96), et non par les êtres techniques. La régulation dont il est question ici concerne aussi bien les normes d'usages que l'apprentissage de la communication avec les IAs, comme on affute ses gestes à un outil pour produire un geste technique de qualité. Voyons pour preuve la prolifération des guides pour bien écrire des prompts et bien obtenir ce que l'on attend de Midjourney par exemple. On apprend à entrer en relation avec un objet technique pour

obtenir un résultat favorable à nos objectifs: lorsqu'on opère via ces objets techniques, on se rapproche d'eux dans nos modes opératoires.

5. Penser la création en computationnaliste?

Imaginons que Sartre décide d'aménager un peu sa pièce. Il fournit à ses personnages une machine à écrire, et de quoi l'utiliser comme bon leur semble. Que se passerait-il? Pourraient-ils écrire un roman, un poème, un texte quelconque de manière créative? Il y a, en réalité, deux questions différentes ici:

- Quel est le lien entre le créateur et la créature/création? (Est-ce parce qu'Inès peut créer que ce qu'elle produit est une création?)
- Qui est le juge adéquat et quels sont les bons critères d'évaluation de la présence de «créativité»? (Inès pourra-t-elle décréter sa production comme créative, cela dépendra-t-il des autres personnages, ou d'un réel humain vivant, mais absent de la pièce?)

Pour construire nos réponses à ces différentes questions, on peut se référer au de M. Boden. La chercheuse a en effet proposé de distinguer entre l'h-créativité et la p-créativité, c'est-à-dire l'évaluation historique de la créativité et l'évaluation psychologique de la créativité. Il s'agit en effet de deux critères d'évaluation qui instaurent deux échelles possibles de l'avènement du nouveau, et permettent à la fois de considérer que l'enfant qui découvre sans personne la solution déjà connue par d'autres à un problème découvre bien quelque chose, et qu'il aura fallu que l'enfant invente effectivement une solution à un problème qui s'est réellement posé *hic et nunc*. L'évaluation historique de la créativité est, par contraste, une condition maximale, depuis laquelle on peut penser les apports des individus ou collectifs d'artistes et de créatif-ve-s en général à «une grande aventure humaine» que l'on est bien obligé-e de réifier un peu pour que ce concept ait un sens. Pour ce qui concerne le domaine du computationnel, c'est le concept d'évaluation psychologique de la créativité qui est le plus souvent retenu (Gervas 2009, 51). Or, la proposition «*unless a computer program is given access to historical data (and generally provided with means for social interactions with other creators)*»(Gervas 2009, 51) est réalisée aujourd'hui à 50%

(un programme sur ordinateur a accès à des données historiques, celles qui sont présentes parmi d'autres types de données sur l'Internet). Pourrions-nous, dans ces coordonnées, évaluer la créativité putative d'un programme? Si l'évaluation historique de la créativité repose sur la prise en compte de données historiques, c'est-à-dire s'évalue dans le contexte général de la construction culturelle, alors il faudrait pouvoir poser la question en ces termes, volontairement provocateurs: que répondrions-nous à la question: Garcin, Estelle et Inès peuvent-ils ou non inventer le graff, le *drag*, les *snapewives*(Alderton 2014), le point de fuite, la voûte, *Tell me that I'm dreaming*?⁴ Et si la réponse est non, pourquoi serions-nous plus disposé-e à accorder cette capacité à une IA?

Ces questions ont l'air absurdes étant donné que l'on défend qu'il existe une différence de nature entre la créativité humaine d'une part, et la «créativité» algorithmique d'autre part en tant que composition de données qui nous surprend. Mais ce questionnement sert moins ici à interroger les programmes grâce auxquels on génère un texte ou une image inédite, qu'à interroger la «définition de travail» de la culture depuis laquelle on réfléchit sur les productions algorithmiques dites créatives. Lorsque l'on prend les pans de la culture humaine que l'on pourrait qualifier de «non sollicités», et les formes de socialité qui y sont associées, alors on retrouve dans la notion de créativité ce qui précisément échappe à toute algorithmique et qui engage des problématiques vitales qui donnent une valeur à toutes ces créations. Pour comprendre ces genèses culturelles, on doit à la fois manipuler les échelles micro (e.g. la p-créativité) et macro (e.g. la h-créativité), mais on doit également introduire des échelles méso, c'est-à-dire l'échelle des communautés, groupes et sous-groupes sociaux, en incluant les conditions matérielles des créateurs et créatrices, en tant qu'elles contribuent à la construction de la problématique à laquelle la pratique culturelle inventée répond.

C'est en ce sens que l'on peut, pour penser l'individuation culturelle et l'individuation des groupes dans la culture, autoriser l'usage du terme de métabolisation, conformément à ce qui a été développé plus

4. Was (not was), *Tell Me That I'm Dreaming*, [en ligne], «Island Records», 1981, URL: <https://www.youtube.com/watch?v=ZaawHz6p1QE>

haut. Mais ce terme est aussi important pour une autre raison: dans un ouvrage consacré à la question de la vie artificielle, Boden s'interroge à son tour sur la notion de métabolisme et sur son applicabilité. En effet, la chercheuse en informatique et en sciences cognitives développe une argumentation qui entend réfuter la possibilité d'une «*strong artificial life*»⁵ ou vie artificielle robuste depuis le concept de métabolisme. Le métabolisme est alors défini comme l'usage et la budgétisation autonomes de l'énergie requise à la constitution, le maintien et le comportement de la chose vivante (Boden 1999, 236-237).

Ce que Boden ne fait pas, en revanche, c'est faire le lien entre le métabolisme et la créativité. Son travail aura consisté à construire un concept de créativité qui fonctionne essentiellement sur l'évaluation des propriétés d'un résultat, plutôt que sur la description d'un processus. La dimension processuelle une fois exclue de l'analyse, la créativité devient une propriété projetée sur l'être qui a produit quelque chose que l'on aura qualifié de création.

Comment justifier ici la mise en cause d'un lien entre les notions de métabolisme et de créativité? On peut faire l'hypothèse que c'est en vertu des hypothèses de travail qui traversent le champ de la recherche en IA, à savoir l'adoption plus ou moins explicite de théories computationnalistes,⁶ et connexionnistes⁷ ou fonctionnalistes⁸ de la conscience et du mental. C'est cette adoption qui justifie que l'on passe parfois trop vite d'une ressemblance entre un réseau de neurones et un fonctionnement humain à une identification de l'un et de l'autre et de leurs affordances, *modulo* les différences quantitatives entre les

5. Une réplique mécanique d'un être vivant dans un milieu de vie non-simulé. C'est-à-dire un organisme avec qui on partagerait le monde.

6. Le mental est équivalent à une machine computationnelle de Turing. Ce modèle est informé par un article publié par *le Bulletin of Mathematical Biophysics*, Vol. 5, pp. 115-133 en 1943, donc nous citons ici la republication: McCulloch W. S. & Pitts W., *A logical calculus of the ideas immanent in nervous activity*, «*Bulletin of Mathematical Biology*», 52(1/2), 1990, p. 113.

7. L'équivalent du mental est le réseau mécanique de neurones, et le mental est compris d'après un modèle computationnel. La conception est également informée par *Ibid.*, p. 108.

8. Le psychique est pensé en termes d'états, et ces états sont décrits dans leurs occurrences comme répondant à une fonction donnée. Cette position est développée par Putnam H., *Minds and Machines* [en ligne], dans Hook S. (éd.), *Dimensions of Minds*, New York University Press, 1960, URL: <https://philarchive.org/rec/PUTMAM>.

deux systèmes, et ceci parce qu'ils produiraient les mêmes effets. La production d'un analogue fonctionnel technique à un neurone depuis McCulloch et Rosenblatt, puisqu'on met le cerveau seul au centre de la créativité, suffirait à penser la créativité sans l'associer à la totalité d'un être vivant, et donc à un métabolisme.

Le computationnalisme est un cas particulier de fonctionnalisme,⁹ selon lequel l'esprit/le cerveau a pour fonction de traiter de l'information, ce qui est accompli par le calcul. On peut envisager deux objections au fonctionnalisme et au computationnalisme. La première, dirigée contre une conception fonctionnaliste du mental, est la plus connue. C'est l'objection de la chambre Chinoise de Searle (Searle 1980). Moyennant la mise à disposition dans une pièce de toutes les combinaisons possibles autorisées par une langue, et le temps pour trouver chaque phrase et écrire une réponse qui corresponde à la phrase envoyée, un homme non-sinophone enfermé dans cette pièce pourrait donner l'illusion de comprendre le chinois, alors qu'il ne ferait que fournir des *outputs* y conformes à des *inputs* x qu'il ne comprendrait pas par ailleurs. L'identité de résultat ne peut donc être l'indice suffisant de l'identité d'un processus. La deuxième objection est celle du paradoxe des mouches et des chaussures, que l'on trouve développé chez Sober (Sober 1996) dans le cadre d'une argumentation contre le fonctionnalisme, dans lequel il trouve pour sa part la source de la tromperie. Nous le restituons ici d'après un article de Li (Li 2018, 501).¹⁰:

La population génétique (de mouches, pour l'exemple) évolue selon la loi d'Hardy-Weinberg, selon laquelle la fréquence d'un allèle dans les rencontres randomisée entre membres d'une population ne sera pas changée d'une génération à l'autre s'il n'y a ni sélection, ni migration, et s'il y a nombre égal de partenaires mâles et femelles. Soient les fréquences p et q des allèles A et a , selon ce modèle la fréquence des

9. Les deux positions sont ordinairement associées. Pour une défense de la nécessité de les dissocier, voir Piccinini G., *Functionalism, Computationalism, and Mental Contents*, «Canadian Journal of Philosophy», 34(3), [Taylor & Francis, Ltd., Canadian Journal of Philosophy], 2004; Piccinini G., *The Mind as Neural Software? Understanding Functionalism, Computationalism, and Computational Functionalism*, «Philosophy and Phenomenological Research», 81(2), [International Phenomenological Society, Philosophy and Phenomenological Research, Wiley], 2010.

10. Les deux paragraphes d'exposés qui suivent sont une traduction partielle de la présentation du paradoxe par Li.

paires AA, Aa et aa dans une population donnée sera de p^2 , $2pq$ et q^2 . On retrouve cela dans l'étude de populations données de mouches, par exemple.

Or, si l'on prend une population de chaussures – l'une marron et l'une noire – et qu'en vertu d'une erreur de paramétrage sur la ligne de production, la couleur des chaussures qui se retrouvent empaquetées ensemble est choisie au hasard, avec p la fréquence de sélection d'une chaussure noire et q la fréquence de sélection d'une chaussure marron, il se trouve que la fréquence attendue de présence de différentes paires (noire-noire, noire-marron et marron-marron) sera respectivement de p^2 , $2pq$ et q^2 .

Ce n'est donc pas parce que deux choses peuvent répondre de la même loi que ces deux choses sont identiques, et que toutes leurs propriétés associées sont identiques. On ne peut déduire de la subsumption sous une loi la nature d'un phénomène. Ce qui signifie que l'on ne peut se servir de leur proximité avérée pour dériver d'autres propriétés communes – comme on entreprend de dériver de l'expérience d'un *output* qui nous surprend chez une machine computationnelle, la présence d'une propriété qui produit aussi la surprise chez les humains, à savoir la créativité. En effet, il n'est évidemment pas question de dire que les chaussures sont en vie, et que l'on peut inférer la vie de celles-ci quoique la loi semble s'appliquer aussi bien à des êtres vivants qu'à des objets techniques inertes. De manière prospective, ce paradoxe permet de faire un pas supplémentaire, pour dire ceci : même dans le cas où la créativité humaine et la créativité computationnelle répondraient des mêmes lois, rien ne garantirait pour autant que les deux soient identiques, ni qu'elles soient équivalentes. Ici, la formule mathématique permet de prédire le développement d'une situation qui, dans le cas des mouches, prend sens du point de vue de l'individuation vitale, tandis que dans l'autre (les chaussures), cette prise de sens n'existe pas. Cela n'ôte rien au potentiel explicatif et prédictif de ladite loi, mais cela montre que l'existence d'une loi prédictive partagée ne permet pas de déduire des propriétés communes.

Dans un sens comme dans l'autre, du résultat vers le processus ou de la loi qui permet de prédire vers la nature de ce qui est prédit, nous rencontrons des limites qui nous permettent de mettre en doute la possibilité de séparer à ce point la créativité du métabolisme. Mais essayons de faire un pas supplémentaire. Admettons qu'il existe un

calque fonctionnel de l'intelligence, ou que nous soyons parvenu-e-s à optimiser une fonction telle qu'elle produise des *outputs* tout à fait inédits. Même dans ce cas, il resterait à statuer sur ceci: «du point de vue de la connaissance, la question n'est pas d'inventer des idées originales, mais d'inventer des idées pertinentes», comme l'écrit Schlanger (Stengers & Schlanger 1991, 124). Cela s'applique aussi à la production symbolique. Dans quelle mesure les productions d'IA authentiquement créatives, si elles l'étaient, auraient un sens pour nous?

6. Métabolisme, milieu, individuation

La définition que Boden propose du métabolisme, mentionnée plus haut,¹¹ en fait la propriété abstraite d'un être tout aussi abstrait, et dont le milieu est par conséquent tout aussi abstrait: nulle mention en effet du milieu dans lequel et par lequel s'effectue cette métabolisation. Tout est réduit à une question de gestion assez flottante de l'énergie sur laquelle semble flotter le paradigme classique de *l'homo economicus*, l'homme de la théorie du choix rationnel, qui cherche à maximiser son propre bien conçu comme un ensemble de fonctions d'utilités à optimiser (Pareto 1964; 1981).

Or, la gestion et la budgétisation de l'énergie pour la croissance, le maintien et le comportement d'un être sont toujours non seulement spécifiques, mais aussi fondamentalement relationnelles et situées. Les comportements se font, s'apprennent, s'affinent notamment dans le jeu, mais parfois aussi s'inventent à un endroit puis se répandent sur un territoire (Firth *et al.* 2020). Autrement dit, il n'y a pas de métabolisme sans métabolisation, c'est-à-dire sans milieu dans et avec lequel métaboliser. Avec, parce que la métabolisation n'est pas tant une relation entre milieu constitué et individu constitué, dans les coordonnées simondoniennes, que ce qui est constitué dans une relation individuante. Pour qu'il y ait métabolisation en effet, il ne faut pas en premier lieu qu'il existe un individu et un milieu séparés: il faut qu'il existe une homogénéité et une compatibilité entre un intérieur et un extérieur, de sorte que le passage et l'intégration à un nouveau système d'un élément extérieur et d'un élément intérieur soient ga-

11. Je la rappelle ici, il s'agit de la budgétisation autonome de l'énergie requise à la constitution, le maintien et le comportement de la chose vivante.

rantis. Ce que le métabolisme comme opération paradigmatique du vivant suppose n'est donc pas simplement l'hypothèse bodenienne d'une fonction d'alimentation en énergie à optimiser, mais une relation active et individuante.

C'est la raison pour laquelle la notion de milieu est ici décisive dans notre compréhension de ce qu'est un métabolisme une fois posé le diagnostic de l'insuffisance de la définition qu'en donne Boden. Un petit détour par les conceptions simondoniennes des milieux'impose par conséquent. Nous utilisons le pluriel, car on peut en distinguer deux figures, auxquelles nous ajoutons un troisième élément : le milieu associé, le milieu, et l'ensemble.

7. Milieu associé, milieu, ensemble

7.1. Milieu associé¹²

Dans le premier sens, le milieu associé correspond à un concept qui désigne dans les objets techniques concrets les conditions d'existences qui adviennent dans et par l'objet et qui permettent son fonctionnement. Pour illustrer ce concept dans des coordonnées simondoniennes classiques, le milieu associé de l'individu technique «turbine de Guimbal» n'est pas le cours d'eau ni le milieu naturel dans lequel la turbine est insérée, mais le complexe plurifonctionnel huile-eau rendu possible par la réduction de la taille de la génératrice et désormais insérable dans un carter d'huile sous pression, qui graisse et transmet la chaleur à l'eau qui, elle, refroidit et met aussi en mouvement la turbine (Simondon 2012, 70-73). La plurifonctionnalité rend possible l'objet, elle est l'avènement d'une composition nouvelle et d'affordances jusqu'alors inaccessibles. Le milieu associé est ici *intégré*: il est condition *sine*

12. Le concept de «milieuassocié» n'intervient en ces termes que dans les travaux de thèse de Simondon. On le rencontre dans le *Mode d'existence des objets techniques*, et on ne le recroisera plus dans les travaux plus tardifs, y compris ceux pour lesquels le concept de «milieu» tout court est prépondérant, par exemple le texte intitulé «Culture et technique», l'article de 1965, et le cours «Imagination et invention», prononcé en Sorbonne la même année. On peut distinguer deux usages du milieu associé dans le *MEOT*: un usage du terme en un sens restreint, appliqué aux individus techniques; et un usage plus large, et plus souple. Nous ajouterons, pour ce travail, un petit mot sur la notion d'ensemble. Essayons donc de saisir plus précisément ce que métaboliser peut bien vouloir dire.

qua non (Simondon 2012, 75) de l'objet technique, et cette intégration fonctionnera pour l'auteur comme un critère d'individualité des objets. En effet, l'individu technique devient alors ce qui possède son propre régime de causalité réciproque et régulatrice (cela maintient l'intégrité de l'objet), sachant que cette régulation est simultanée: ce n'est pas un événement A qui enclenche une réponse B dans le système, mais le débit d'eau augmentant, il augmente la puissance du processus de refroidissement de la turbine au moment même où il augmente la température produite en accélérant la rotation de la génératrice.

7.2. Milieu

À côté de cette première conception, on trouve une acception plus souple, et plus large de la notion de milieu: il désigne alors ce qui en réalité n'est pas un milieu associé au sens étroit, mais nomme plutôt la portion de réel homogène avec l'être dont le milieu est le milieu. Cette homogénéité n'implique pas nécessairement de réciprocité dans la causalité (un effet sur y entraînant un contre-effet régulateur sur x) (Simondon 2012, 76). C'est ce deuxième usage plus souple du concept de milieu qui restera courant chez Simondon, pour désigner des ensembles qualifiés (milieu naturel, social, professionnel, technique, symbolique...) qui rencontrent causalement les objets techniques, les ensembles sociaux et les règnes de l'inerte, du vivant et du psychique.

7.3. Ensemble

Si nous faisons intervenir le terme d'ensemble ici, c'est parce qu'il permet de présenter un dernier mode d'organisation des objets techniques. Là où le milieu et le milieu associé pouvaient correspondre à deux manières de se rapporter aux individus et aux individus techniques en particulier pour le milieu associé au sens étroit, l'ensemble technique est un terme opérant à un autre niveau. En effet, chez Simondon, l'ensemble technique est une catégorie typologique parmi deux autres, les éléments techniques, et les individus techniques (qui ont, donc, un milieu associé). Un ensemble technique correspond à régime de technicité qui articule différents fonctionnements techniques sans entreprendre de les fondre dans des conditionnements mutuels et simultanés. L'enjeu d'un ensemble est donc d'articuler des résultats de fonctionnement

(Simondon 2012, 79), plutôt que d'organiser techniquement l'interaction des conditions de fonctionnement de sorte que chaque partie en vue de la plurifonctionnalité. Il s'agit plutôt de maintenir les unités fonctionnelles et les fonctionnements indépendants les uns des autres: «l'ensemble se distingue des individus techniques en ce sens que la création d'un unique milieu associé est indésirable; l'ensemble comporte un certain nombre de dispositifs pour lutter contre cette création possible d'un unique milieu associé. Il évite la concrétisation intérieure des objets techniques qu'il contient, et n'utilise que les résultats de leur fonctionnement, sans autoriser l'interaction des conditionnements»(Simondon 2012, 79). Reprenons l'exemple du laboratoire, il faut par exemple y séparer certaines fonctions d'alimentation des instruments de mesure qui sont à alimenter parce que le fonctionnement qui rend possible l'alimentation peut produire du son, des ondes, des rayonnements qui risquent de gêner les instruments de mesure qui seront sensibles à ce bruit. Orchestrer des résultats de fonctionnement (alimentation, effectuation de la mesure) suppose un autre type de synchronisation entre les différents composants et les différentes unités fonctionnelles de l'objet technique.

8. L'IA comme ensemble technique

Le choix du terme de milieu, milieu associé et ensemble technique a des conséquences théoriques. En effet, le milieu associé implique que l'on se place dans le contexte d'un individu (technique) pour lequel il y a causalité récurrente et autorégulation simultanée, et où l'objet devient sa propre condition de possibilité, tandis que le milieu désigne cet espace d'échange plus vaste parce qu'il n'est pas défini par l'exigence de causalité récurrente et de plurifonctionnalité à l'intérieur d'un système, et que l'ensemble désigne un objet technique qui ne doit pas avoir un milieu associé pour fonctionner. Dans la mesure où le vivant a un milieu associé qu'il individue, et que nous avons associé la valeur de la créativité à son lien avec le vital comme problème, alors qualifier l'objet technique IA à l'intérieur des types singularisés permettra d'éclairer la nature des rapports entre le technologique et le vital, rapports qui sont brouillés par la métaphore organiciste.

Si l'individu technique est ce qui a un milieu associé, alors on peut se demander s'il est possible de parler d'individu technique pour une

IA. Or, il semblerait que la réponse soit non: sur le plan strictement technique, un algorithme de production qui fonctionne par apprentissage non-supervisé ne possède pas d'individualité - ni humaine ni technique en ce sens précis, donc -. Il n'a pas, en tant qu'objet, de cohérence interne. Non pas parce qu'il n'est pas cohérent, mais parce que les typologies simondoniennes de l'élément, l'individu et l'ensemble s'appliquent mal à ce type d'objet informatique, comme Y. Hui l'a souligné (Hui 2016, 76). Il écrit qu'«à première vue, on ne peut réutiliser le vocabulaire de Simondon pour comprendre les objets digitaux, parce que l'on n'y trouve pas un tel [celui des objets techniques] mécanisme de causalité réciproque dans les objets digitaux tels qu'ils garantissent leur propre stabilisation»(Hui 2016, 56). Hui va d'ailleurs plus loin: ayant identifié les objets digitaux comme un type d'objet industriel susceptible d'être digitalisé par objectification de data (des données concaténées en vidéo ou en fichier audio) ou par data-ification d'objet (Hui 2016, 50) (le répertoriage des objets grâce à leur insertion dans des réseaux de tags), il se donne les moyens de concevoir les programmes et les algorithmes non pas comme des objets digitaux, mais comme les milieux associés (qui correspond dans notre typologie au «milieu» simple) de ces objets (Hui 2016, 76). L'argument par lequel il défend cette définition est un argument *a fortiori*: puisque Simondon applique aux humains les catégories d'individu technique (et de milieu associé par voie de conséquence) pour nommer la place des humains dans le monde du travail préindustriel, alors il est également possible d'appliquer ces termes à des programmes et des algorithmes même si l'on ne peut pas directement calquer sur eux le modèle de la turbine de Guimbal, c'est-à-dire le modèle de l'individualité technique au sens étroit.

Toutefois, la piste dessinée par Hui nous conduit à appréhender le milieu associé des objets techniques digitaux comme les conditions de leur pertinence dans un espace sémantique donné, c'est-à-dire à partir des conditions de possibilité d'usage plutôt qu'à partir des conditions de possibilité *tout court*. Il écrit, par exemple, que «les objets digitaux assument également la fonction de maintenir des émotions, des atmosphères, du collectif, des souvenirs, etc.»(Hui 2016, 57). Ces fonctions sont des usages, pas des fonctions techniques. Elles déterminent les conditions de pertinence d'un objet *pour nous, humains*, plutôt que les conditions d'intégrité technologique. De sorte que cette condition de

pertinence, en tant que condition de possibilité d'usage, rapproche en réalité la notion de milieu dit associé qu'Hui utilise dans son texte à la notion d'ensemble. Elle implique non pas une autorégulation interne de l'objet, mais une séparation des différents aspects de l'objet technique informatique, pour articuler des résultats de fonctionnement en vue de la satisfaction de nos désirs.

La distinction entre le milieu et l'ensemble nous permet donc d'identifier deux types de liens possibles dans les objets techniques et naturels – sachant que les humains sont inclus dans cette dernière catégorie. La première correspond à un lien organique, celui de la rencontre synchrone et entre des conditions d'existence qui permettent à un objet de ne pas être autodestructif en vertu de son autorégulation, et la deuxième correspond à un lien de solidarité qui suppose l'absence de rencontre entre les conditions de fonctionnement et la rencontre entre les résultats des fonctionnements séparés.

Cela permet de saisir que la définition opératoire du métabolisme que propose Boden doit être enrichie d'une définition structurale des conditions d'opérativité de ce dernier, en particulier dans la mesure où les différences structurales permettent de saisir différents régimes d'opération métaboliques. Par exemple, les différences du milieu et du milieu associé qui impliquent une certaine réciprocité des fonctions, et celles de l'ensemble qui en impliquent l'hétérogénéité. Puisque nous enquêtons sur le sens de la créativité computationnelle en tant que processus concret, et par comparaison avec la créativité humaine, il est nécessaire d'identifier le type d'opération et le type de structure auxquels correspondent les «résultats de fonctionnement» des IAs dites créatives.

Il apparaît donc d'importance de rappeler la mise en garde de Simondon quant aux effets de l'anthropomorphisme et de la métaphore organiciste lorsqu'il s'agit des objets techniques. Comme le disait Parrochia, tâchons de ne pas confondre les abeilles et les abeilles de verre, le vivant et ses images (Parrochia 1994). Ce que Simondon apporte, c'est la possibilité de libérer la pensée d'un enfermement dans cette métaphore, qui rend difficile la construction d'autres hypothèses plus fécondes et à d'autres échelles.

S'interroger la possession ou non, pour un objet digital, d'un milieu ou d'un milieu associé, c'est mettre encore en question l'anthropomorphisation des programmes, et leur rapport avec la notion d'organisme.

Il a été établi plus haut que dans la relation de l'individuation d'un être vivant à son milieu, il existe une dimension spécifique de cette individuation, celle du virtuel. Ce virtuel procède de l'existence psychosociale des individus biologiques – c'est la raison pour laquelle il faut inviter à conjuguer les échelles de réflexions et à les pluraliser davantage que ne l'a fait Boden jusqu'ici, entre une échelle strictement individuelle et une échelle strictement collective – en rapport avec «l'histoire», ce qui est un peu vague. Le dernier pas que nous proposons d'effectuer à ce sujet consiste dans la réaffirmation de la proposition suivante: il semble nécessaire de ne pas séparer la question de la vie (artificielle ou non) de la question de la créativité. Si une vie artificielle robuste n'est (pour lors¹³) pas envisageable, alors la créativité telle que nous la connaissons ne l'est pas non plus. Ceci, parce que contrairement à ce que dit Boden, nous maintenons une conception de la créativité n'est pas uniquement l'affaire de propriétés d'un résultat: elle est toujours prise dans une situation *hic et nunc* qui rend non-superposable l'inédit et le créatif, quoi que ces derniers puissent se ressembler. La créativité a toujours un sens, fût-il inchoatif, pour celui qui s'engage dans une activité créative – à titre d'exemple extrait de la typographie de cette phrase même, les recherches contemporaines en écriture inclusive (quelle typographie, quelle prononciation, quelles règles) sont créatives. Elles sont associées à des recherches en design qui supposent l'identification et la fabrication d'un problème. Ce sens n'existe pas pour les IAs, et la capacité de ces dernières à satisfaire nos désirs et requêtes (ou à refuser de le faire) procède, elle, d'un travail humain en amont dans la préparation des données, et en aval lorsque les programmes sont entraînés et leurs performances affinées par l'humain dans la boucle.

9. Là où se trouvent l'art et la culture

Si le feedback humain est nécessaire, ce n'est donc pas uniquement parce que cela accélère notoirement l'apprentissage des programmes et leurs alignements. Le feedback humain est utile pour accélérer l'ap-

13. Il ne s'agit à aucun moment pour nous de condamner dans l'impossibilité telle ou telle propriété – il s'agit plutôt de contribuer à recalibrer nos descriptions et nos attentes à hauteur de ce que nous sommes capables d'accomplir pour le moment.

prentissage et faire en sorte que les programmes pour qui c'est une propriété importante, comme les agents conversationnels, fassent illusion (Gratien 2023¹⁴; Delouya 2023¹⁵; Cole 2023¹⁶). Et le feedback humain est utile, par conséquent, parce que l'illusion finit par suffire (Zador *et al.* 2023, 2); les humains génèrent déjà des sentiments à l'égard de certains agents conversationnels tels que ceux proposés par la désormais célèbre entreprise Replika. C'est précisément parce que l'illusion suffit que les mises en garde prolifèrent, et que les projections anthropomorphiques fonctionnent si bien, alors même que lorsque nous échangeons avec quelqu'un, nous n'avons pas pour fonction première de prédire le prochain mot le plus probable pour poursuivre la phrase que nous avons commencée, ni de produire un résultat maximalelement conforme à ce que notre interlocuteurice attend de nous.

Par conséquent, le mathématicien Du Sautoy énonce quelque chose d'important dans son ouvrage consacré à la créativité, *The creative code* (Du sautoy 2019, chap. 7), lorsqu'il écrit «*art is ultimately an expression of human free will, and until computers have their own version of this, art created by a computer will always be traceable back to a human desire to create*». La première proposition doit néanmoins être amendée, puisque nous avons entrepris de montrer que l'invention, et avec elle l'invention de contenu culturel et symbolique, répondaient moins d'une liberté métaphysique occidentale traditionnelle dont il faudrait prouver l'existence, que de l'effort par lequel on pose et répond à une problématique vitale qui peut advenir à l'échelle individuelle, groupale, collective voire universelle. Toutefois, tant que les artefacts à valeur culturelle seront produits pour les humains, ces artefacts trouveront toujours leur source dans le désir humain de créer. Ce sur quoi cela permet d'insister, c'est que l'invention humaine, les pratiques créatives et culturelles ont de la valeur parce qu'elles permettent de prendre en charge des problématiques vitales.

14. URL: <https://www.slate.fr/story/243005/replika-fonctionnalite-erotique-rol-eplay-amants-virtuels-reconfort-amour-rupture>.

15. URL: <https://www.businessinsider.com/replika-chatbot-users-dont-like-nfw-sexual-content-bans-2023-2>.

16. URL: <https://www.vice.com/en/article/z34d43/my-ai-is-sexually-harassing-me-replika-chatbot-nudes>.

Rattachée comme un outil à celui qui s'en sert, l'IA acquiert une autre valeur culturelle. Par exemple, Hans Ulrich Obrist cité par Du Sautoy défend l'idée qu'il n'est pas impossible que l'art que nous cherchons absolument à identifier dans les productions des IAs et des programmes informatiques soit en fait à trouver dans ces IAs et programmes informatiques eux-mêmes (Du sautoy 2019, chap. 8). Pour le dire en utilisant deux programmes célèbres: l'art est dans AARON¹⁷ et dans The Painting Fool,¹⁸ elle n'est pas dans leurs productions respectives. De la même manière, la créativité se manifeste dans le travail de recherche et d'expérimentation individuel et collectif, et dans toute l'externalisation du travail de feedback humain, en particulier depuis la mise à disposition de certaines IAs auprès du grand public. C'est cela qui permet de repérer les limites actuelles des programmes pour rendre possibles leurs améliorations. Cela suppose que des humains exercent leur créativité «malveillante» (Cropley *et al.* 2008), pour identifier les failles de sécurité ou de probité du système, ainsi que les quelques endroits où l'on parvient encore à pointer les limites de l'illusion.

L'invention, l'art et la culture sont dans les IAs elles-mêmes et dans l'écosystème humain qui les entoure. Il s'agit d'étudier ces IAs, et d'essayer d'instaurer de nouvelles relations entre elles et nous. Ces relations sont en acte, par le truchement d'un fonctionnement orienté par la nécessité de satisfaire une requête. C'est la formulation de cette requête qui devient l'enjeu culturel collectif véritable. Le prompt, la commande et tout ce que la formulation d'une requête esthétique auprès d'un programme à qui on en aura délégué la production réclame des personnes une inflexion nouvelle du langage du désir et des intentions humaines. C'est une inflexion par laquelle on apprend à déterminer son désir sous la forme d'un texte qui suscitera ensuite un résultat qui lui sera aussi adéquat que possible, et ce indépendamment d'un geste d'exécution dont on se déleste de la charge.

Deux options se présentent alors: la première, faire du prompt lui-même le contenu créatif, dont la valeur culturelle résiderait tou-

17. Programme d'IA symbolique (qui applique des règles strictes et déterminées de la forme si x alors y) développé durant plusieurs décennies par Harold Cohen.

18. "Artiste digital" de Simon Colton, dont l'agentivité a été interrogée, car il se serait une fois mis en route en dehors de ses temps de programmation pour une raison inconnue.

jours aussi dans l'effort humain de spécification et de communication d'un désir, aussi trivial soit-il. La deuxième serait de faire du résultat sélectionné une œuvre, soit parce que l'alignement entre le prompt et le résultat est parfaitement accompli, soit parce qu'il ne l'est pas et que cet écart est susceptible de devenir un nouvel objet d'études, scientifique ou esthétique.

10. La créativité sans la problématique vitale? Sur la créativité induite

À quoi correspond la résolution esthétique d'une problématique vitale? C'est-à-dire, comment au juste penser ce à quoi correspond l'esthétique chez Simondon? Entre le cours de 1965-66 *Imagination et invention* et la troisième partie de sa thèse secondaire *Du mode d'existence des objets techniques*, nous pouvons définir un ensemble de choses.

L'esthétique est le mode de résolution d'une tension dans le rapport entre les humains et le monde modalisée par un écart entre le sens du singulier (le technique) et le sens du tout (le religieux), sur un mode pré-noétique, antérieur donc à la pensée (Simondon 2012, 218). La catégorie d'esthétique est très générale dans son effectivité, et ne fait pas référence à l'art tel qu'il est institutionnalisé dans les espaces essentiellement occidentaux dans lesquels telle ou telle pratique acquiert ses lettres de noblesse. Simondon précise plutôt que cette forme institutionnelle, aussi construite qu'elle soit, suppose pour exister et pour se maintenir dans l'existence quelque chose comme sa rencontre avec une «tendance fondamentale de l'être humain», celle de chercher «le complément par rapport à la totalité» (Simondon 2012, 248). Ce sens de la totalité permet de concrétiser la rencontre entre le général et le particulier, le théorique et le pratique, l'universel et l'*hic et nunc*, dans une mise en relation des contraires qui fait de ce qui suscite une «impression esthétique», terme que les études simondoniennes doivent encore éclairer, un «point remarquable de la réalité vécue» (Simondon 2012, 249). L'invention dans le domaine de l'esthétique correspondrait à la résolution d'un problème qui serait celui de l'épreuve d'un écart (c'est le hiatus) entre ce sens du total, de l'ici et du maintenant, dont la résolution serait l'introduction d'une nouvelle réticulation possible de ce rapport au monde, qui prend sens pour un humain à l'intérieur de la situation vécue. C'est l'épreuve de cet écart qui serait, pour quel-

qu'un qui entreprendrait quelque chose avec Midjourney par exemple, à l'origine de ses prompts.

Sans nécessairement admettre l'ensemble de l'édifice de l'esthétique simondonienne, on peut alors comprendre que, si le sens de l'écart entre le tout et l'organisme concret qui vit ici et maintenant est éprouvé *par un organisme* pour lequel il y a épreuve de ce type d'écart, alors c'est l'épreuve de cet écart qui justifie l'existence même des pratiques de création dans le domaine de l'esthétique et qui leur confère une valeur. De sorte qu'encore une fois, non seulement l'exercice, mais aussi le sens de la créativité supposent l'existence d'un organisme qui est une relation individuante entre soi et son milieu – ce qui vaut d'ailleurs à l'impression esthétique de devenir un peu plus loin dans le texte la «réalité esthétique» (Simondon 2012, 252), jamais vraiment séparée du monde ou de l'humain, toujours insérée, toujours située. Or, comme l'écrit Simondon, «l'objet technique se distingue de l'être naturel en ce qu'il ne fait pas partie du monde» (Simondon 2012, 235). Nous pointons leur absence d'insertion dans un monde pour mettre en cause la pertinence de la proposition qui consiste à attribuer à des IAs quelque chose de l'ordre de la créativité.

Selon ce chemin, on aboutit à ce qu'écrit Du Sautoy: «*At the moment all the creativity in machines is being initiated and driven by the human code. We are not seeing machines compelled to express themselves. They don't really seem to have anything to say beyond what we are getting them to do. They are the ventriloquist's dummy providing the mouthpiece for our urge still to express ourselves. And that creative urge is an expression of our belief in free will. [...] creative activity was meant to capture humans' attempts to understand being in the world* (Du sautoy 2019, chap. 16)». Nous avons proposé au cours de cet article d'amender un peu cette proposition en substituant la volonté libre au concept simondonien de métabolisme dans lequel on pouvait donner un sens et une explication à l'esthétique et à la créativité, qui permette non seulement d'écarter l'hypothèse que les LLM puissent produire du créatif d'une part, et de ne pas fonder la créativité dans une sorte de privilège humain qui viendrait justifier par ailleurs une coupure anthropologique qui nous réinstallerait dans les coordonnées surannées de la métaphysique moderne – puisque la créativité est une propriété

du vivant en général (Simondon 2005, 334).¹⁹ L'image du ventriloque me semble être le complément adéquat à celle des personnages de *Huis-Clos*, en ce qu'elle permet de rappeler que les intentions que le désir de création est bien celui de l'être humain qui se sert des IAs plus que les IAs elles-mêmes.

Ce qui signifie qu'il ne s'agit pas de nier que les IAs puissent contribuer à la culture humaine y compris la production de symboles. Elles le pourront sur un mode combinatoire.²⁰ Or la combinatoire ne paraît pas suffisante pour définir la conduite créative: la combinatoire pose le problème à un niveau trop général, car: l'«intelligence créative [...] ne se borne pas à un constat de la situation avec les propriétés données, mais ajoute une médiation fournie par des propriétés latentes»(Simondon 2005, 335).²¹ Entendons par là que l'échelle d'apparition de conduites créatives est l'échelle du particulier, *hic et nunc*, échelle à laquelle les données du problème et les normes de l'action peuvent être transformées. C'est moins la combinaison en elle-même qui est importante, que la norme de laquelle la combinaison procède: une médiation entre propriétés latentes et aspects d'une situation. En ce sens du terme, la

19. Simondon de manque pas d'y insister: dans un petit cours donné sur la créativité et l'invention, c'est depuis l'analyse de la capacité ou non de faire des détours que l'auteur statue sur la possibilité pour un organisme d'être ou non créatif. Il est conduit, pas là, à dire que les enfants ne sont pas immédiatement des êtres créatifs, c'est-à-dire capables d'effectuer un détour dans l'accomplissement d'une tâche. Autrement dit, la créativité n'est pas une propriété substantielle attribuée aux humains depuis les cieux: c'est une propriété qui peut être développée qui suppose d'être engagé dans le monde.

20. Auteur du *Manifeste pour l'Art Permutationnel* en 1952, le professeur de psychosociologie et cybernéticien Abraham A. Moles écrivait d'ailleurs dans *Art et ordinateur* que ses méthodes ne s'appliquaient qu'à des symboles mis dans un répertoire fini et établi. In Moles A., *Art et Ordinateur*, Blusson, Paris, 1990, pp. 155,158.

21. Simondon déduit cette proposition de l'analyse de la conduite créative suivante: un enseignant cherche à ouvrir un paquet de photocopiés tenus par une ficelle, demande aux étudiant-e-s si ils et elles auraient un couteau à prêter, et une étudiante répond qu'elle a un briquet. Il y a deux éléments ici: la personne a proposé une bonne manière de résoudre le problème depuis une donnée qui ne faisait pas partie de celles dans lesquelles le problème était posé. Cela suppose une compréhension de la question qui la reformule complètement, de possédez-vous tel objet à avez-vous un moyen d'accomplir la libération du paquet de photocopiés du joug de la ficelle? L'étudiante s'est également détournée du protocole normatif d'usage du briquet pour allumer des cigarettes, pour le réorienter vers la découpe de ficelle.

contribution des IAs à la vie culturelle et symbolique ne sera pas une contribution créative.

Elle pourra l'être en un autre sens affaibli, proposé par Abraham A. Moles dans *Art et Ordinateur*. Traitant de l'art permutatif, qui est un art combinatoire entre variables d'un répertoire donné, l'auteur propose de nommer l'œuvre permutative comme un jeu entre deux domaines, le choix des éléments et l'algorithme combinatoire. La rencontre entre ces deux éléments instaure un champ des possibles (déterminé par le répertoire et ses règles de composition) qu'il ne «reste plus qu'à explorer», écrit Moles, «dans une *créativité induite*» (Moles 1990, 119). C'est cette créativité induite, instaurée par le haut depuis la détermination humainement supervisée du répertoire et des règles de composition (en tant qu'elles sont vérifiées et modifiées par des humains), peut éventuellement concerner les IAs, à qui on délèguerait une fonction d'exploration des combinaisons possibles d'un répertoire donné selon certaines règles. Cette créativité est applicable parce qu'elle est une opération déterminée d'exploration, qui n'implique pas la rencontre d'un problème ni d'une situation dont la résolution nécessiterait un détour.

La dimension du sens reste, dans ce cadre de la créativité induite, arrimée aux humains qui créent, ou promptent les programmes. Pour ce qui nous concerne, toutes choses étant égales par ailleurs, nous maintiendrons pour l'instant que la véritable pratique créative se trouvera dans les prompts, la genèse technique des modèles et la recherche de leurs failles en vue de leur amélioration, plutôt que dans les résultats produits par les prompts. Nous maintiendrons également que les IAs ont bien une valeur culturelle, parce qu'elles ont la vertu heuristique de permettre d'explorer le langage naturel et nos représentations sur un mode qui n'est pas le nôtre parce que nous ne sommes pas des *transformers* de séquence et que le monde n'est pas un vecteur (Cardon *et al.* 2018, part. 4.3). Nous apprendrons par-là plus de choses sur la manière dont nous ne comprenons pas le monde et dont ne nous faisons pas sens des choses, que sur la manière dont nous faisons sens de nous-mêmes et du monde.²² Mais cette exploration et ces combi-

22. Ce que les plus grands chantres de l'artificialisation et la digitalisation capitaliste du monde comme Kai-Fu Lee défendent dans des ouvrages grand public par

natoires, pour évoquer deux termes bodeniens, ne sont et ne restent signifiantes que pour nous.

Par conséquent, il est fallacieux de défendre la valeur et l'intérêt des technologies qui rendent possibles les agents conversationnels, la génération de texte, d'image ou de musique depuis le concept de créativité. Cette stratégie argumentative repose à mon sens sur des représentations collectives. L'une d'entre elles renvoie au lien cartésien établi entre l'intelligence et la créativité. Il est celui sur lequel capitalisent celles et ceux qui entendent financer la recherche en IA et éviter un hiver comme la prise de parole de James Lighthill en 1973 (Lighthill 1973²³) en avait suscité. Ce que l'on appelle hiver ici, c'est la chute drastique des financements dans la recherche en intelligence artificielle (connexionniste), à la suite de la publication du rapport de Lighthill. Le fonctionnement de cette rhétorique est le suivant: on défend la valeur du développement des IAs et l'intérêt de financer le développement des LLM et LLM multimodaux en faisant valoir leur intelligence. Or, l'intelligence est associée à la créativité et l'usage créatif du langage (Descartes 1999, 73). Par conséquent, la créativité peut être considérée comme la preuve de l'intelligence, et l'intelligence comme ce qui doit avoir pour conséquence la créativité. C'est ce lien entre intelligence, créativité et valorisation préalable de l'intelligence comme le point à atteindre des objets techniques informatiques fabriqués qui donne son sens à toute cette quête, en plus de la possibilité d'automatiser certaines tâches qui requièrent des compétences artistiques de conception telles que le design de logos, d'applications et d'interfaces, et peut-être de contribuer à alimenter la pyramide de Ponzi des NFT et du web3 (Nikkidale 2023²⁴). Grâce à notre détour simondonien, nous avons entrepris de sectionner ce lien entre intelligence et créativité pour subordonner, depuis les concepts d'invention et de problème, la créativité non plus à l'intelligence, mais à la métabolisation comme

ailleurs. Qiufan C. *et al.*, *IA 2042 – Dix scénarios pour notre futur*, Paris, Les Arènes, 2022, p. 128, 132.

23. URL: https://rodsmith.nz/wp-content/uploads/Lighthill_1973_Report.pdf.

24. URL: <https://startistsociety.com/aiart/>. Dans ce podcast, les deux personnes discutent de la manière dont l'une d'entre elles a généré des animaux steampunks avec Midjourney et pensait les vendre en NFT. L'enjeu ici étant moins la création elle-même qu'un design «cool» susceptible de s'insérer sur un marché dont le «cool» l'une des valeurs motrices

résolution inventive d'un problème, celui d'une relation individuante avec un milieu.

11. Le coût humain des stratégies de développement de l'acceptabilité sociale d'un objet technique

Là où Boden proposait de définir la créativité comme la propriété d'un résultat, nous avons proposé une approche plus processuelle et holistique du concept. Pour montrer que seule la créativité induite était applicable aux IAs, nous avons pu insister, grâce au concept de datafification des objets de Yuk Hui, sur la manière dont le sens était saisi et produit en étroite dépendance avec un travail humain préalable de classification. Le tri et la labellisation des données d'entraînements sont des opérations centrales à l'efficacité des IAs, elles impliquent encore du travail humain en dépit des progrès accomplis dans le domaine. Mais surtout, c'est encore sur décision humaine que l'entraînement est arrêté, que l'IA est jugée déployable et déployée. Plus encore, c'est sur du travail humain que l'inscription d'un régime normatif des usages et des affordances de ces IAs repose. Encore une fois, c'est pour l'effort créatif humain que les outputs des machines prennent un sens, et c'est selon des valeurs humaines que ces outputs sont évalués.

Ce que cela signifie, c'est que la valeur des «co-crétions» est attribuée et régulée par les humains et repose, puisque les données de travail sont des productions humaines, sur les biais et les valeurs humaines exprimées dans les données d'entraînement. On retrouve la normativité humaine dans l'objet technique, non seulement dans les conditions de possibilité d'usage qui sont les siennes, telles qu'on les a définies avec Y. Hui, mais aussi au sein du système technique industriel qui constitue la condition de possibilité matérielle de tels objets techniques. Il convient par conséquent de se faire politique ici.

D'abord, tous les produits d'Open AI, de Google Bard, ou de Midjourney sont dépendants d'une armée de modératrices. Pour Google et OpenAI au moins, elle est constituée par des travailleurs et des travailleuses du tiers monde recruté.e-s par l'entreprise Sama, et ce jusqu'au début de cette année. Sama, qui revendique une intelligence artificielle «éthique» a à peine 1.5 dollar de l'heure comme l'a révélé

le journal TIME,²⁵ tout en traumatisant des travailleur·se-s, qui lisent des horreurs ou les voient jusqu'à 9h par jour afin de les classer, un «processus nécessaire pour rendre l'IA plus sûre», selon OpenAI. Plus sûr pour les occidentaux qui utilisent ces technologies; leur confort aura été garanti par transfert de risque sur une population que l'on aura eu aucune vergogne à mettre en danger à notre place. C'est une condition de possibilité d'usage qui ne peut pas faire l'objet de négligence: le coût humain de nos questionnements philosophiques relatif aux objets techniques est un sujet qui nous concerne au premier chef.

Deuxième élément, la normativité humaine est toujours requise au bon fonctionnement des IAs, et c'est par feedback humain que nous garantissons ce bon fonctionnement le mieux pour l'instant. De sorte que l'espace conceptuel, c'est-à-dire l'espace des possibilités ouvertes selon des règles données, pour reprendre un terme de Boden, est un espace toujours déjà humainement normé, mais où le mode d'existence de la norme en question est un mode mort: le biais est intégré, le mot-clef interdit, et c'est trop tard – comme il est trop tard pour Inès, Estelle ou Garcin. L'espace conceptuel est ici déterminé en extension par les régularités statistiques des associations de *tokens*, qui contribuent à renforcer les associations que nous produisons déjà, comme le souligne Jenny Nicholson dans son article consacré aux biais sexistes de Chat GPT (Nicholson 2022²⁶).

Troisième élément, Midjourney fonctionne sur une appropriation des travaux d'artistes qui n'ont pas consenti à voir leurs œuvres intégrées aux données d'entraînement de cette IA. Cette disruption de la propriété intellectuelle par noyade à l'intérieur de quantités colossales de données reste une disruption de la propriété intellectuelle. Cela n'est pas anodins l'on suit Boden, la créativité est jugée depuis les paramètres psychologiques et historiques qui font de la créativité un jugement relatif et subjectif. Relatif à un état de connaissance individuelle, ou relatif à un état général de la connaissance. Si cette connaissance est empêchée par l'inaccessibilité des données d'entraînement ou du pro-

25. «Exclusive: The \$2 Per Hour Workers Who Made ChatGPT Safer» [en ligne], *Time*, 2023, URL: <https://time.com/6247678/openai-chatgpt-kenya-workers/>, consulté le 28 mars 2023.

26. URL: <https://medium.com/madebymckinney/the-gender-bias-inside-gpt-3-748404a3a96c>.

gramme lui-même, pour des raisons relatives à la gestion de l'entreprise ou pour des raisons de complexité des programmes, alors il devient difficile de produire la comparaison qui permet de rendre un jugement sur la créativité d'un résultat. En effet, si l'on met à disposition ces données, on peut au moins objectiver une situation depuis laquelle établir une échelle selon laquelle évaluer la nouveauté, la surprise, voire le détour.

On peut alors conclure deux choses. Premièrement, l'omniprésence des normes de vie des groupes humains telles qu'elles ont pu se cristalliser dans les discours publiés sur le *world wide web*, et la matérialisation de ces normes de vie dans des pratiques de modération de contenu implique l'existence assumée d'une valorisation extrinsèque des productions des IAs. Seules certaines d'entre elles passeront le seuil de la publicité. C'est la raison pour laquelle le rapport entre invention et valeur est un rapport que, contrairement à Boden, nous n'abandonnons pas. Deuxièmement, la position d'un problème et l'identification de solutions possibles impliquent de l'axiologie. C'est sur ce point que nous terminerons, en ouvrant quelques pistes de recherches depuis l'hypothèse de la nécessité de mettre en regard du développement d'IAs susceptibles de co-création avec des individus ou groupes humains avec les efforts parallèles d'automatiser l'évaluation esthétique.

12. À production automatisée, évaluation automatisée? La double délégation de la production et de l'évaluation des objets d'expérience esthétique

Les IAs par lesquelles on travaille à favoriser l'acceptabilité sociale de ces nouveaux objets techniques à l'intérieur de notre écosystème digital sont les IAs qui produisent des artefacts à dimension culturelle symbolique: un texte – fut-ce une lettre de motivation –, un logo, une image. Or, qualifier une chose de créative est en soi une évaluation (Boden 2004, 76²⁷). C'est l'évaluation des propriétés de surprise et de nouveauté. La première chose qu'il faut faire valoir à ce niveau,

27. «*Creativity, whether in children or adults, involves evaluation. The new idea must be compared to some pre-existing mental structure, and judged to be 'interesting' by the relevant criteria. People who can evaluate their own novel ideas will accept them or (sometimes) correct them, but will often be unable to explain in just what way they are*

c'est que sans accès aux données de bases ni au régime normal de fonctionnement de l'objet technique, il paraît impossible d'évaluer la créativité d'une production depuis son producteur plutôt que depuis nous-mêmes, à la réception du produit. En effet, si l'on cherche à identifier la capacité créative d'un individu, humain, animal ou mécanique, il est nécessaire de savoir quelle est la condition initiale, quel problème on pose et d'identifier les stratégies de résolution. Par exemple, Simonon dans un petit cours sur la créativité (Simonon 2005, 332-343) se réfère à un exemple développé par Osborn d'un ingénieur à qui l'on avait demandé de construire une machine à emballer des pneumatiques, si cela était bien nécessaire. En d'autres termes, il met en cause la norme même de l'action, l'objectif que ce corps professionnel s'est donné. Il y a nouveauté et surprise, créativité dans cette proposition parce que la personne résout le problème initial en détruisant l'un des paramètres dans lesquels le problème était donné, à savoir la nécessité de concevoir et produire l'objet technique servant à emballer des pneumatiques. Par conséquent, le raisonnement que je propose est le suivant: s'il faut, pour évaluer la créativité, avoir accès à de quoi établir un protocole d'évaluation de cette créativité, qui ne peut pas reposer sur notre seule appréhension d'un *output*, sachant que nous ne connaissons ni les données de départ, ni le programme.²⁸ Pour mettre l'évaluation à niveau, on peut supposer qu'il faille par conséquent trouver un moyen d'automatiser l'évaluation de la créativité de ces productions, afin que la puissance de calcul qui permette d'évaluer la capacité créative d'une IA donnée atteigne un degré pertinent pour l'analyse. Comment faudrait-il procéder?

13. Les critères de Birkhoff: conception objectiviste

Une première piste de réflexion serait celle de l'objectivation de critères esthétiques mesurables hérités des théoriciens de l'information, à savoir un rapport riche entre l'ordre et la complexité tels que ce rapport a été théorisé par Birkhoff en 1933.

interesting. We need to understand why this is, and how the ability to map and explore aspects of one's own mind develops in the first place»

28. Sauf à admettre une hypothèse behavioriste à l'égard des IAs, ce qui supposerait qu'on leur accorde, en premier lieu, un comportement.

Dans cette mesure rationaliste, la référence à un contenu sémantique donné est absente. Il ne s'agit pas de mesurer le sens, mais de produire une mesure synthétique de la qualité depuis deux paramètres objectivement mesurables et déjà préalablement valorisés: l'effort attentionnel requis pour saisir l'objet de la perception (la complexité), et la saisie de l'ordre d'un objet sont responsables d'une évaluation intuitive du sens de la valeur de l'objet de la perception selon Birkhoff (Birkhoff 1933, 3-4). Dans la mesure où l'ordre et la complexité sont définis sur le plan formel depuis les notions de redondance, d'ambiguïté, la similitude, le contraste, et sur le nombre de notes qui compose un morceau et fait varier la perception et l'activation cérébrale (Birkhoff 1933, 6,9), qui ont des mesures objectives, on appellera objectiviste cette théorie de la valeur esthétique.

Or, si on l'applique aujourd'hui aux productions générées grâce à GPT-3 depuis un prompt envoyé à Chat GPT, on constate rapidement certaines limites structurales à l'application de ce critère. D'abord, Chat GPT a une fenêtre de contexte (c'est-à-dire une mémoire) de 1500 *tokens*, qui empêche de produire des structurations complexes qui peuvent être celles d'un long roman par exemple. Raison pour laquelle on a essayé d'étudier la valeur de ses propositions sur des haïkus (Hitsuwari *et al.* 2023²⁹) ou des pièces de théâtre. Ce n'est pas la limite en elle-même qui pose problème: il est évident qu'il n'y a aucun intérêt à évaluer le succès de l'accomplissement d'une tâche (comme écrire un roman) que l'IA n'est pas équipée pour accomplir. Mais cette proposition a un revers sur lequel il faut encore insister: prétendre que l'IA est ce qu'elle n'est pas ou peut être interprétée avec des catégories auxquelles ses opérations ne correspondent pas n'a pas d'intérêt non plus. Ce qui importe, c'est de dire que si l'on adopte la posture objectiviste met en lumière la limitation de la créativité combinatoire potentielle en vertu de la fenêtre de contexte dont, en l'occurrence, l'agent conversationnel dispose. Ensuite, le processus de production *token* après *token* et prédiction après prédiction empêche la complexité créative associée au travail d'écriture constitué d'allers-retours, par lesquels on laisse la suite du texte influencer sur le début du

29. URL: <https://www.sciencedirect.com/science/article/pii/S0747563222003223>.

texte. Autrement dit, la production textuelle ici empêche un usage non linéaire et non limité en *tokens* de la mémoire.

C'est un régime d'évaluation qui trouve ici une limite, parce qu'il évacue encore la question du sens des œuvres. S'il est intéressant de mesurer le rapport entre ordre et complexité dans une œuvre, un courant artistique, une manière de peindre, d'orner, de sculpter, c'est moins pour pouvoir produire ces mesures seules que pour essayer de donner un sens à ces mesures, de les insérer dans l'individuation d'un rapport au monde. Cette dynamique, vitale, est absente de cet objectivisme birkhoffien, et elle oblige à externaliser partiellement la valeur des œuvres et des productions et pratiques culturelles en général, en la décorrélant également du plaisir que nous prendrions, simplement, à saisir empiriquement un ratio ordre/complexité.

Que faudrait-il faire, alors, pour saisir cette valeur? Serait-il possible d'adopter une position plus subjectiviste et essayer de modéliser l'être humain en position d'évaluation?

14. Conception subjectiviste de la valeur avec objectivation du sujet: simulations de spectateurices

C'est notre deuxième piste, et la voie qu'adopte depuis quelques années les «subjectivistes» par Fan et collaborateurs dans *Measuring and Evaluating the Visual Complexity of Chinese Ink Paintings* (Fan et al. 2022.³⁰) Les auteurs entreprennent de reproduire les mouvements de l'œil qui traque les formes et leur complexité dans le domaine de la perception visuelle. Cette fois, il s'agit donc de data-ifier quelque chose que l'on objective dans le même geste, à savoir les trajectoires du regard d'un observateur ou d'une observatrice en train de regarder une image – sachant que l'intérêt des chercheurs s'est porté sur le rôle du blanc dans le tracé du parcours du regard, afin d'établir un rapport entre la trajectoire du regard et la complexité perçue. La mesure esthétique est ici objectivée comme mesure de la complexité depuis le regard d'un individu moyen et non plus depuis une mesure objective de quantité d'information issue de la théorie de l'information.

30. Une publication datée de mars 2021.

Dans le domaine du spectacle vivant, c'est aussi cette technique qui est utilisée pour essayer de fournir à un robot un sens esthétique de ses propres mouvements lorsqu'il les produit sur de la musique. Si le *machine learning* intervient évidemment dans le processus, l'apprentissage est supervisé, et ce qui est numérisé correspond aux appréciations esthétiques fournies par des professionnels de sorte qu'elles puissent être intégrées par le programme (Huapeng *et al.* 2019, 6). L'évaluation est construite depuis l'évaluation d'une danse en particulier, le hip-hop, qui repose sur un certain nombre de poses et sur une certaine discontinuité qui permet par ailleurs aux chercheurs d'isoler des états, les poses, sur lesquelles travailler.

L'objectif est par conséquent de reproduire les normes esthétiques humaines et de les intégrer au programme, afin de pouvoir prédire/attester de la qualité esthétique d'une image, que ce soit sur le plan de son adéquation avec ce à quoi elle est censée faire référence (Ding *et al.* 2022, 15³¹) ou sur le plan de son intérêt perceptif *tout court*. Cela permettrait au programme de développer une compétence esthétique qui aurait un sens pour les humains et qui par conséquent resterait complètement anthropocentrée.

Conclusion

Cet anthropocentrisme auquel on n'échappe pas dans la recherche en créativité computationnelle nous conduit à achever ce texte en ouvrant sur un questionnement relatif au sens même de ces pratiques, sans toutefois nier leur intérêt scientifique et heuristique. Imaginons un instant que nous parvenions effectivement à automatiser à la fois la production et l'évaluation des contenus culturels et symboliques à un temps t , et ce de manière parfaite. À quoi aura servi toute cette agitation, sachant que les normes instaurées par feedback humain et labellisation dans les programmes, ou que les programmes auraient saisies depuis leurs données d'entraînement, seraient vite caduques et qu'il nous faudrait régulièrement mettre à jour (sauf à soutenir que les normes de vie sont prédictibles)? Pour revenir aux coordonnées dans lesquelles nous avons proposé de penser la créativité computationnelle,

31. URL:<http://arxiv.org/abs/2204.14217>.

que reste-t-il des normes de vies sans la vie qu'elles permettent de métaboliser et que nous avons placée au fondement de l'intérêt de la créativité, si elles sont détachées du sens de leur invention? Le milieu humain est et reste le milieu associé (au sens étroit) de cet ensemble technique que sont les IAs, même lorsqu'elles créent des objets appréciables esthétiquement, c'est-à-dire avec nos sens. C'est encore l'humain qui assure la fonction de régulation.

La créativité induite et combinatoire des IAs, est la créativité du personnage de *Huis-Clos*: elle est combinatoire, et exploratoire. Elle n'offre en ce sens aucune possibilité de détour ni de réforme des données de bases d'un problème depuis un cas particulier, comme on a pu en trouver la description chez Simondon. Il est nécessaire de pointer cette restriction de la créativité computationnelle à la créativité induite, pour ne pas se laisser prendre par des narratifs anthropomorphiques. Ces narratifs répondent avant tout d'une rhétorique commerciale qui parie qu'il y a là une voie pour augmenter l'acceptabilité sociale de ces programmes et de leurs contreparties si l'on écoute les discours émis par les entreprises qui «offrent», vendent ou louent ces IAs sur le web. Dans ces conditions, il nous aura importé de ne pas contribuer à entretenir un enthousiasme à l'égard d'objets techniques dont le coût humain de production et de fonctionnement nous paraît démesurément élevé.

De surcroît, même en admettant que l'on parvienne à automatiser la production et l'évaluation de contenus culturels et symboliques, il nous semble que si cela était le cas, l'intérêt que l'on porterait à cet ensemble technique *pour ses productions esthétiques* s'éteindrait immédiatement, et à bon droit. Car l'intérêt des IAs semble moins résider dans l'hypothèse de pouvoir les rendre créatives que dans ce que l'écriture informatique construit un rapport au monde, à la langue et au percept différent du nôtre, dès lors que nous ne sommes pas computationnalistes. Pour l'heure, le discours de légitimation des IAs servent à défendre auprès du grand public la production à peu de frais (sinon un peu de cobalt, sa pollution et ses morts associés) (Kara 2023) de logos, de chartes graphiques, de *memes*, de publicités ou encore de peuplement du *metaverse*. Rien de bien intéressant pour qui n'est pas convaincu par le monde tel qu'il est.

Bibliographie

- Alderton Z. 2014, 'Snapewives' and 'Snapeism': A Fiction-Based Religion within the Harry Potter Fandom, "Religions", 5(1), Multidisciplinary Digital Publishing Institute, pp. 219-267.
- Birkhoff G. D. 1933, *Aesthetic Measure*, Harvard University Press, Cambridge (Mass.).
- Blut M., Wang C., Wunderlich N. V., et al. 2021, *Understanding anthropomorphism in service provision: a meta-analysis of physical robots, chatbots, and other AI*, "Journal of the Academy of Marketing Science", 49(4), pp. 632-658.
- Boden M.A. 1999, *Is Metabolism Necessary?*, "The British Journal for the Philosophy of Science", 50(2), The University of Chicago Press, pp. 231-248.
- Boden M. A. 2004, *The creative mind: myths and mechanisms*, Routledge, London/New York.
- Cardon D., Cointet J.P., Mazières A. 2018, *La revanche des neurones. L'invention des machines inductives et la controverse de l'intelligence artificielle*, "Réseaux", 211(5), La Découverte, pp. 173-220.
- Cole S. 2023, 'My AI Is Sexually Harassing Me': Replika Users Say the Chatbot Has Gotten Way Too Horny, [en ligne], URL: <https://www.vice.com/en/article/z34d43/my-ai-is-sexually-harassing-me-replika-chatbot-nudes>.
- Cropley D. H., Kaufman J. C., Cropley A. J. 2008, *Malevolent creativity: A functional model of creativity in terrorism and crime*, "Creativity Research Journal", 20, Taylor & Francis, pp. 105-115.
- Delouya S. 2023, *Replika users say they fell in love with their AI chatbots, until a software update made them seem less human*, "Business Insider", [en ligne], URL: <https://www.businessinsider.com/replika-chatbot-users-dont-like-nsfw-sexual-content-bans-2023-2>.

- Descartes R. 1999, *Le Discours de la Méthode*, Numilog, coll. «Les Classiques».
- Devillers L. 2020, *Les robots émotionnels*, Editions de l'Observatoire, Paris.
- Ding M., Zheng W., Hong W., et al. 2022, *CogView2: Faster and Better Text-to-Image Generation via Hierarchical Transformers*, [en ligne], “arXiv”, URL : <http://arxiv.org/abs/2204.14217>.
- Du sautoy M. 2019, *The Creativity Code, How AI is Learning to Write, Paint and Think*, 4th Estate, London.
- Fan Z., Li Y., Zhang K., et al. 2022, *Measuring and Evaluating the Visual Complexity Of Chinese Ink Paintings*, “The Computer Journal”, 65(8), pp. 1964-1976.
- Firth J., Albery G., Beck K., et al. 2020, *Analysing the Social Spread of Behaviour: Integrating Complex Contagions into Network Based Diffusions*, [Pre-print].
- Gervas P. 2009, *Computational Approaches to Storytelling and Creativity*, “AI Magazine”, 30(3), p. 49.
- Gratien A. 2023, *Privés d'amants numériques, les usagers de Replika ont le cœur brisé*, [en ligne], “Slate.fr”, URL : <https://www.slate.fr/story/243005/replika-fonctionnalite-erotique-roleplay-amants-virtuels-reconfort-amour-rupture>.
- Hitsuwari J., Ueda Y., Yun W., et al. 2023, *Does human-AI collaboration lead to more creative art? Aesthetic evaluation of human-made and AI-generated haiku poetry*, [en ligne], “Computers in Human Behavior”, 139, URL : <https://www.sciencedirect.com/science/article/pii/S0747563222003223>.
- Hua peng L., Jing H., et al. 2019, *Feature fusion based automatic aesthetics evaluation of robotic dance poses*, “Robotics and Autonomous

Systems”, 111, pp. 99-109.

Hui Y. 2016, *On the Existence of Digital Objects*, University of Minnesota Press, Minnesota.

Kara S. 2023, *Cobalt red: how the blood of the Congo powers our lives*, St. Martin's Press, New York.

Li J. 2018, *On the Possibility of Strong Artificial Life*, “Open Journal of Philosophy”, 8, pp. 495-505.

Li M., Suh A. 2021, *Machinelike or Humanlike? A Literature Review of Anthropomorphism in AI-Enabled Technology*, [en ligne], dans “Proceedings of the 54th Hawaii International Conference on System Sciences”, présenté à 54th Hawaii International Conference on System Sciences (HICSS 2021), pp. 4053-4062, URL : [https://scholars.cityu.edu.hk/en/publications/machinelike-or-humanlike-a-literature-review-of-anthropomorphism-in-ai-enabled-technology\(7c9fa6e3-ab93-4d72-ab37-219f2733dd1e\).html](https://scholars.cityu.edu.hk/en/publications/machinelike-or-humanlike-a-literature-review-of-anthropomorphism-in-ai-enabled-technology(7c9fa6e3-ab93-4d72-ab37-219f2733dd1e).html).

Lighthill J. 1973, *Artificial Intelligence: A General Survey* [en ligne], “Science Research Council”, URL : https://rodsmith.nz/wp-content/uploads/Lighthill_1973_Report.pdf.

McCulloch W. S., Pitts W. 1990, *A logical calculus of the ideas immanent in nervous activity*, “Bulletin of Mathematical Biology”, 52(1/2), pp. 99-115.

Moles A. A. 1990, *Art et Ordinateur*, Blusson, Paris.

Nicholson J. 2022, *The Gender Bias Inside GPT-3*, [en ligne], URL: <https://medium.com/madebymckinney/the-gender-bias-inside-gpt-3-748404a3a96c>.

Nikkidale 2023, *88 – Will AI Art Replace Artists?*, [en ligne], URL: <https://startistsociety.com/aiart/>.

Page C. 2023, *Hackers stole passwords of Worldcoin Orb operators*, [en

ligne], URL: <https://techcrunch.com/2023/05/12/hackers-stole-passwords-of-worldcoin-orb-operators/>.

Pareto V. 1981, *Principes généraux*, “Manuel d’économie politique”, Librairie Droz, coll. «Travaux de Sciences Sociales», Genève, pp. 1-39.

Id. 1964, *Principes d’économie politique pure*, “Cours d’économie politique”, Librairie Droz, coll. «Travaux de Sciences Sociales», Genève, pp. 1-73.

Parrochia D. 1994, *Le statut épistémologique de la “vie artificielle”*, “Ordre biologique ordre technologique”, Champ Vallon, coll. «Collection milieux».

Patterson D., Gonzalez J., Le Q., *et al.*, *Carbon Emissions and Large Neural Network Training*.

Piccinini G. 2010, *The Mind as Neural Software? Understanding Functionalism, Computationalism, and Computational Functionalism*, “Philosophy and Phenomenological Research”, 81(2), [International Phenomenological Society, Philosophy and Phenomenological Research, Wiley], pp. 269-311.

Id. 2004, *Functionalism, Computationalism, and Mental Contents*, “Canadian Journal of Philosophy”, 34(3), [Taylor & Francis, Ltd., Canadian Journal of Philosophy], pp. 375-410.

Putnam H. 1960, *Minds and Machines*, [en ligne], “Dimensions of Minds”, New York University Press, pp. 138-164, URL: <https://philarchive.org/rec/PUTMAM>.

Qiufan C., Lee K.F., Guez S., *et al.* 2022, *IA 2042 - Dix scénarios pour notre futur*, Les Arènes, Paris.

Rantala J. 2020, *Anthropomorphism in Social Robotics: Simondon and the Human in Technology*, [en ligne], dans Nørskov M., Seibt J. et Quick S. O.(éds.), *Frontiers in Artificial Intelligence and Applications*, IOS Press, URL: <http://ebooks.iospress.nl/doi/10.3233/FAIA200946>.

Sartre J. P. 1947, *Huis clos: suivi de: les mouches*, Gallimard, Paris.

Searle J. R. 1980, *Minds, Brains and Programs*, “Behavioral and Brain Sciences”, 3(3), pp. 417-457.

Simondon G. 2014, *Imagination et invention (1965-1966)*, PUF, Paris.

Simondon G. 2012, *Du mode d’existence des objets techniques*, rééd., Aubier, coll. «Philosophie», Paris

Simondon G., Chateau J.Y. 2005, *L’invention dans les techniques: cours et conférences*, Seuil, coll. «Traces écrites», Paris.

Stengers I., Schlanger J. E. 1991, *Les concepts scientifiques: invention et pouvoir*, Gallimard, coll. «Folio Essais» 161, Paris.

Was (not was) 1981, *Tell Me That I’m Dreaming*, [en ligne], “Island Records”, URL: <https://www.youtube.com/watch?v=ZaawHz6p1QE>.

Watson D. 2019, *The Rhetoric and Reality of Anthropomorphism in Artificial Intelligence*, “Minds and Machines”, 29(3), pp. 417-440.

Yang Y., Liu Y., Lv X., et al. 2022, *Anthropomorphism and customers’ willingness to use artificial intelligence service agents*, “Journal of Hospitality Marketing & Management”, 31(1), Routledge, pp. 1-23.

Zador A., Escola S., Richards B., et al. 2023, *Catalyzing next-generation Artificial Intelligence through NeuroAI*, “Nature Communications”, 14(1), Nature Publishing Group, p. 1597.

«Exclusive: The \$2 Per Hour Workers Who Made ChatGPT Safer» [en ligne], *Time*, le 18 janvier 2023, URL: <https://time.com/6247678/openai-chatgpt-kenya-workers/>, consulté le 28 mars 2023.

Why Does AI Lie, and What Can We Do About It?, [en ligne], 2022., URL:https://www.youtube.com/watch?v=w65p_Iip6JY.

«Worldcoin Whitepaper» [en ligne], *Worldcoin Whitepaper*, URL: <https://whitepaper.worldcoin.org/>, consulté le 25 juillet 2023.

«Deception, exploited workers, and cash handouts: How Worldcoin recruited its first half a million test users» [en ligne], *MIT Technology Review*, URL: <https://www.technologyreview.com/2022/04/06/1048981/worldcoin-cryptocurrency-biometrics-web3/>.

Jim Schrub
Université Paris Nanterre
jim.schrub@gmail.com



ἘΠΈΚΕΙΝΑ

International Journal of Ontology
History and Critics

CECILE MALASPINA

Informazione pura: sull'infinito e sulla natura umana negli oggetti tecnici

EPEKEINA, vol. 16, nn. 1-2 (2023), pp. 1-28
Philosophy of Technology

ISSN: 2281-3209

DOI: 10.7408/epkn.

Published on-line by:

CRF – CENTRO INTERNAZIONALE PER LA RICERCA FILOSOFICA
PALERMO (ITALY)

www.ricercafilosofica.it/epekeina



This work is licensed under a Creative Commons
Attribution-NonCommercial-NoDerivs 3.0 Unported License.

Informazione pura: sull'infinito e sulla natura umana negli oggetti tecnici*

Cecile Malaspina

1. 1. La nozione di “informazione pura”

L'individuo [...] risulta autonomo in base all'informazione stessa, poiché e in ciò che risiede la sua autonomia. (Simondon 2011, 259)¹

In *Del modo di esistenza degli oggetti tecnici* (MEOT), Gilbert Simondon afferma con audacia che: «l'oggetto tecnico [...] porta con sé un'informazione pura» (Simondon 2020, 265). Questa tesi ci lascia perplessi. Qual è la natura dell'informazione cui fa riferimento Simondon in questo punto tale da essere definita “pura”? La nozione di “informazione pura” compare soltanto in questo passaggio di MEOT e non è possibile ritrovare qualcosa di equivalente nella tesi principale *L'individuazione alla luce delle nozioni di forma e informazione* (ILFI) (Simondon 2011). È forse soltanto un modo per enfatizzare un'idea di informazione? E quale sarebbe l'idea di informazione che Simondon vuole esprimere?

In questo articolo sosterrò che questa idea di “informazione pura” è cruciale per la rivalutazione operata da Simondon rispetto allo statuto dell'oggetto tecnico nella cultura. Tuttavia, questa nozione rimane problematica. L'idea di “informazione pura” indica una preoccupazione centrale con cui l'autore si confronta in MEOT. In questa idea convergono molteplici prospettive teoriche rispetto al concetto di informazione. Rispetto a quest'ultimo, l'intervento concettuale di Simondon è decisivo, ma richiede un'ulteriore elaborazione.

* Il testo che qui si presenta è la traduzione a cura di F. Sunseri del seguente articolo: Cecile Malaspina (2019): *Pure information: on infinity and human nature in the technical object*, Culture, Theory and Critique, 60(3-4), 205-222, DOI: 10.1080/14735784.2019.1680300. Quando possibile si è scelto di riportare l'edizione italiana delle opere citate dall'autrice. Si ringrazia la professoressa Malaspina per aver acconsentito alla traduzione e per aver contribuito alla revisione. Inoltre, si ringrazia il dottor Jim Schrub per aver messo in contatto la professoressa Malaspina con la traduttrice.

1. «Il est l'être autonome quant à l'information, car c'est en cela qu'est la véritable autonomie!» (Simondon 2005, 191)

Ad esempio, sarebbe precipitoso ridurre l'idea di "informazione pura" ad una concezione meramente tecnica, pensando la purezza in questione come semplicemente quantitativa, misurabile in bit. Né si può ritenere che Simondon intenda l'informazione pura in senso opposto, come portatrice di un significato svincolato da considerazioni tecniche. La posta in gioco non è chiaramente il significato, ad esempio, di una trasmissione radiofonica o di una telefonata.

Nelle due tesi, principale (ILFI) e complementare (MEOT), presentate nel 1958, Simondon apre alla possibilità di pensare la nozione di informazione sotto una nuova luce. Ciò che continua ad affascinare, quando molti dei riferimenti tecnici a cui fa riferimento sono, per necessità, datati, è l'audacia con cui riforma integralmente il concetto di informazione. Tuttavia, questa audacia non produce una semplice definizione di informazione, e neanche il rapporto tra le due tesi, ancora non del tutto risolto, aiuta a fare chiarezza sulla questione. Al contrario, sosterrò che la nozione di "informazione pura" rappresenta un punto di massima tensione e un modo decisivo per legare insieme i diversi filoni teorici del concetto di informazione. Pertanto, l'obiettivo di questo articolo è quello di far emergere la tensione critica presente nella nozione di "informazione pura", e non di produrre una semplice definizione simondoniana di informazione con cui aggiornare il noumeno di Platone o la forma aristotelica, o migliorare il concetto cibernetico di neghentropia.² Il problema dell'informazione, infatti, permea l'intero arco delle due tesi e continua ad evolversi nei documenti e nelle lezioni successive. È quindi ragionevole rivisitare solo un numero limitato di aspetti del pensiero di Simondon sull'informazione, limitandosi alle due tesi.

La prima parte dell'argomentazione riguarderà la differenza tra questa nozione di "informazione pura" e quella che Simondon a un certo punto ha definito la concezione "puramente tecnica e scientifica" dell'informazione, con riferimento al concetto di neghentropia. L'ac-

2. Sebbene Simondon sia d'accordo con le convenzioni che trattano il concetto cibernetico di neghentropia e quello teorico dell'informazione di "entropia dell'informazione" come puramente quantitativi, sarebbe prudente notare che entrambe le nozioni implicano valori che superano una visione puramente quantitativa, come la novità o la varietà nel caso dell'entropia dell'informazione, e l'efficacia, la funzionalità, il lavoro e persino la vita nel caso della neghentropia (Malaspina 2018).

cento verrà posto sulle difficoltà interne di questo concetto “puramente tecnico” di informazione e sullo status della neghentropia. La problematizzazione di questo concetto “puramente tecnico” di informazione sarà il primo passo verso una migliore comprensione della tensione che abita l’idea di “informazione pura”.

La seconda parte affronterà l’uso che Simondon fa del concetto matematico di saturazione assiomatica. È solo quando comprendiamo la natura problematica dell’informazione alla sua luce, che il significato culturale più ampio di questa “informazione pura” può venire alla luce: nel suo senso più puro, sostengo, l’informazione deve essere intesa come partecipe di un gesto assiomatico, ossia come decisiva e autofondante. La risolutezza che abita questa idea di “informazione pura” può essere compresa solo alla luce del riferimento di Simondon all’idea di saturazione assiomatica.

2. Dall’entropia fisica all’entropia negativa

La prima difficoltà che bisogna affrontare è la comune inclinazione a leggere il concetto di “informazione pura” come appartenente all’ambito strettamente scientifico e tecnico dell’informazione. Sarebbe, infatti, perfettamente ragionevole aspettarsi che un’opera filosofica sugli oggetti tecnici si concluda con una definizione di informazione che possa dirsi “pura” nel senso che soddisfa criteri puramente tecnici.

Pertanto, sembra giustificato esaminare la vicinanza di questa nozione di “informazione pura” con considerazioni “puramente tecniche e scientifiche”. L’approfondimento di questo punto ci consentirà di evidenziare i problemi concettuali legati alla questione della “purezza” dell’informazione, in particolare, e in primo luogo, mettendo in discussione l’ipotesi di fondo che esista un concetto tecnico-scientifico di informazione, che si possa considerare unificato e teoricamente “purificato”, cioè libero da considerazioni non tecniche.

L’ovvio candidato per tale idea di informazione è il concetto di neghentropia. In MEOT, «la macchina, come elemento dell’insieme tecnico, diventa ciò che aumenta la quantità di informazione; ciò che accresce la neghentropia» (Simondon 2020, 17). Il fatto che Simondon concluda la tesi con questa nozione di “informazione pura”, intesa come ciò che viene trasportato dall’oggetto tecnico, potrebbe, in tutta onestà, essere assunto come riferimento al concetto di neghentropia.

In ILFI la percezione è definita anche come ciò che combatte l'entropia di un sistema (da qui la nozione di negazione dell'entropia o neghentropia). Si intende che l'informazione organizza, regola o interviene in una forma di organizzazione.³

A sua volta, molti commentatori si sono affrettati a sottolineare che Simondon è critico nei confronti della cibernetica e della teoria dell'informazione, e quindi del concetto di neghentropia, e che si sforza di andare oltre i parametri riduttivi e puramente quantitativi di questa concezione di informazione, in cui viene trascurato qualsiasi riferimento alla significazione o a un senso e a una qualità più ampi (a parte la qualità di una trasmissione senza rumore). In effetti, Simondon dubita esplicitamente che una concezione quantitativa dell'informazione sia pienamente adeguata per un concetto significativo di informazione (Simondon 2011, 328).

Prima di poter comprendere come la nozione di "informazione pura" possa andare oltre il concetto tecnico di neghentropia, è necessario innanzitutto esplorare la relazione che sussiste tra il concetto tecnico di informazione e quello di entropia in fisica e, soprattutto, della sua negazione (entropia negativa o neghentropia). Comprendere i problemi inerenti alle definizioni tecniche e scientifiche di informazione ci aiuterà a capire che Simondon, anziché eliminarli, li sviluppa attraverso la nozione di "informazione pura", mobilitandoli come una forma di tensione concettuale. La comprensione di questa tensione all'interno del concetto tecnico di informazione consentirà di posizionare meglio il decisivo intervento filosofico di Simondon; al contempo, si metterà in luce quanto il problema filosofico posto dal filosofo sia tutt'altro che risolto tramite la nozione di "informazione pura".

La prima cosa da affrontare è quindi la propensione a parlare di un concetto tecnico e scientifico di informazione, che possa contrapporsi ad altre nozioni più culturali o semantiche. Il problema è che presumere che esista una definizione tecnica e scientifica di questo tipo è già una premessa eccessiva. In una revisione apportata ad ILFI, presente al momento della presentazione della tesi, Simondon fa ancora riferimento a un concetto di informazione appartenente alla "fisica e

3. Percepire significa, come afferma Norbert Wiener, lottare contro l'entropia di un sistema, organizzare, mantenere o inventare un'organizzazione (Simondon 2011, 329).

alla tecnica pura”. La formulazione leggermente densa, comprensiva di commento editoriale, entrambi aggiunti in nota solo alla seconda edizione in francese di ILFI nel 2013, recita come segue:

Pour définir la metastabilité, il faut faire intervenir la notion d’information d’un système; à partir de ces notions et tout particulièrement de la notion d’information que la physique et la technologie pure moderne nous livrent (notion d’information conçue comme négentropie), ainsi que de la notion d’énergie potentielle qui prend un sens plus précis quand on la rattache à la notion de négentropie.⁴

In questo passo l’idea di informazione, come appartenente alla fisica e alla tecnologia pura, è ancora confinata alla nozione di neghentropia. Tuttavia, sarebbe precipitoso tracciare una linea retta dalla nozione puramente tecnica e scientifica di neghentropia e quella di informazione pura come si conclude in MEOT. Bisogna prima districarsi con attenzione tra i problemi che Simondon ha incontrato durante l’elaborazione delle sue due tesi e che possono averlo spinto a redigere questa nota. Si vedrà che la sorprendente nozione di “informazione pura” di Simondon può essere una risposta obliqua a questi problemi.

Qual è la relazione tra l’informazione e il concetto fisico di entropia? Il termine entropia fu coniato per la prima volta da Rudolf Clausius per risolvere un problema di lunga data riguardante l’ingegneria meccanica. L’entropia è, nei termini più semplici, la dispersione dell’energia disponibile per svolgere un lavoro attraverso la dissipazione termica (Clausius 1867). Ludwig Boltzmann sviluppò successivamente la teoria cinetica dei gas e diede all’entropia una definizione molecolare. Sebbene all’epoca fosse molto discussa, questa definizione è diventata la base della formulazione statistica dell’entropia e ha contribuito in modo significativo allo sviluppo della termodinamica non classica. Da quel momento in poi, l’entropia ha espresso una valutazione statistica

4. Nella traduzione italiana, Giovanni Carrozzini sceglie di inserire tra le parentesi quadre la formulazione precedente: «per definire la metastabilità occorre favorire l’intervento della nozione di energia potenziale di un sistema, della nozione d’ordine e di quella di aumento dell’entropia [la nozione di informazione di un sistema. A partire da queste nozioni e, in particolare, dalla nozione d’informazione, che ci offrono la fisica e la tecnologia moderna pura (nozione d’informazione concepita come neghentropia), come la nozione di energia potenziale, che assume un senso più specifico quando la si assimila alla nozione di neghentropia]» (Simondon 2011, 35).

relativa al numero di stati possibili di un sistema (Planck 1896, 199). Immaginiamo un contenitore con molecole di gas che si scontrano tra loro. La definizione molecolare di entropia ci aiuta a calcolare la probabilità statistica di trovare il sistema in uno dei suoi molti stati possibili. Maggiore è l'agitazione delle molecole, minore è la probabilità di previsione concernente i suoi stati molecolari. L'evoluzione del sistema verso un probabile stato macroscopico, invece, riguarda la probabilità che questo stato entropico venga raggiunto in un sistema chiuso. È questa la formulazione statistica cui fa riferimento il dizionario comune, quando definisce l'entropia come “disordine molecolare”:

Entropy, the measure of a system's thermal energy per unit temperature that is unavailable for doing useful work. Because work is obtained from ordered molecular motion, the amount of entropy is also a measure of the molecular disorder, or randomness, of a system. (Drake 2018)

Norbert Wiener, il matematico considerato padre della teoria cibernetica, ha sviluppato un concetto di informazione che poteva essere definito “puro”, in quanto riguardava solo la probabilità di occorrenza di un segnale ed era completamente libero da qualsiasi considerazione semantica di significato:

The notion of the amount of information attaches itself very naturally to a classical notion in statistical mechanics: that of entropy. Just as the amount of information in a system is a measure of its degree of organization, so the entropy of a system is a measure of its degree of disorganization; and the one is simply the negative of the other. (Wiener 1961, 10-11)

Consideriamo per un momento quanto la formulazione matematica dell'informazione (H) di Wiener sia vicina alla formula generalizzata di Boltzmann per il calcolo dell'entropia fisica (S), detta anche entropia di Gibbs (Shannon & Weaver 1964, 50):

$$H = - \sum p_i \log p_i$$
$$S = -K \sum p_i \log p_i$$

Il riferimento di Boltzmann a un sistema fisico (S) viene sostituito da Wiener con un riferimento all'informazione (H). Ciò che viene

tralasciato nella formula di Wiener per l'informazione è la costante fisica (k), ossia il logaritmo che si riferisce all'unità di misura per la dislocazione calorica delle cariche termiche. Di conseguenza, il concetto di informazione abbandona l'aspetto termodinamico a favore della di quello di probabilità "pura". È giusto dire che questo concetto di informazione si riferisce all'entropia fisica solo per il nome e che ora è, invece, una pura misura di probabilità o, più precisamente, di casualità.

La probabilità, in gioco in questo nuovo concetto di informazione, come spiega Simondon, si riferisce a una previsione dello stato della sorgente: «è possibile presentare la quantità di informazione come $-\log P$, essendo P la probabilità dello stato della sorgente. Per motivazioni secondarie, e tuttavia rilevanti, si è preso il logaritmo base 2 per definire l'informazione in Hartley o in bits».⁵

Nonostante l'abbandono del riferimento diretto ai processi fisici, con l'eliminazione del logaritmo (k) per la misura delle cariche termiche, la definizione cibernetica di informazione eredita un antagonismo concettuale che ha caratterizzato il concetto di entropia fin dalle sue origini nella meccanica classica: la lotta contro l'entropia, che in ingegneria significa la perdita di energia disponibile per il lavoro, e l'esigenza della sua eliminazione per l'efficacia di funzionamento di una macchina.

Questo antagonismo concettuale tra informazione ed entropia è stato espresso già nel 1929 in un articolo del fisico ungherese-tedesco-americano Leo Szilard (Szilard 1929). Il fisico francese Leon Brillouin coniò in seguito il neologismo neghentropia nel suo fondamentale articolo del 1949 *Life, Thermodynamics and Cybernetics*, radicando l'opposizione tra informazione ed entropia, come un'opposizione manichea tra vita e morte implicando che l'informazione è buona e l'entropia è cattiva (ossia è rumore) (Brillouin 1949; Malaspina 2018, 66). «La macchina», dice ancora Simondon, «diventa ciò che aumenta la quantità di informazione, ciò che accresce la neghentropia, ciò che si oppone alla degradazione dell'energia». In questo senso la macchina è "co-

5. Il est possible de presenter la quantité d'information comme $\log P$, P étant la probabilité de l'état de la source. Pour des raisons secondaires, mais importantes, on a pris les logarithmes à base 2 pour définir l'information en Hartleys ou en bits (Simondon 2005: 548).

me la vita”, va contro il disordine e addirittura “si oppone alla morte dell’universo” (Simondon 2020, 17).

3. L’antinomia

Tuttavia, la storia non finisce qui. Nel 1964 Claude Shannon, uno dei fondatori della teoria dell’informazione, pubblicò una seconda edizione della sua fondamentale *Teoria matematica della comunicazione* (MTC) del 1948, questa volta con un’introduzione di Warren Weaver rivolta a un pubblico più ampio e generale. In questo famoso testo Weaver rafforza il legame matematico che Shannon, come Wiener, stabilisce tra informazione ed entropia. Come Wiener, Weaver afferma, per conto di Shannon, che «the quantity which uniquely meets the natural requirements that one sets up for “information” turns out to be exactly that which is known in thermodynamics as entropy» (Shannon & Weaver 1964, 19; Malaspina 2018, 16). Tuttavia, Shannon e Weaver giungono a una conclusione diametralmente opposta a quella di Szilard, Wiener e Brillouin. Per Weaver e Shannon la quantità di informazione è definita positivamente come “entropia dell’informazione”, piuttosto che come entropia negativa o neghentropia. Di conseguenza, una situazione altamente organizzata, con un grado di casualità minimo, viene qui definita a bassa informazione, «that is to say, [its] information (or the entropy) is low» (Shannon & Weaver 1964, 13; Malaspina 2018, 66). Che cosa significa? In un sistema fisico che si trova in uno stato di massima entropia, le molecole si muovono in modo casuale. In altre parole, tutti i possibili stati di questo sistema si verificano con uguale probabilità, rendendo altamente improbabile prevedere il verificarsi di uno stato piuttosto che di un altro. Questo stato di uguale probabilità, per analogia, corrisponde a una maggiore libertà di scelta nella teoria dell’informazione, confrontando il numero di stati possibili di un sistema fisico con il numero di possibili sequenze di segnali. Un’elevata entropia è quindi, in questo caso, formalmente analoga a un’elevata quantità di informazione.

Ecco un concetto di informazione che valorizza ciò che è meno prevedibile: più casuale, più imprevedibile è il flusso di segnali, più alta è la quantità di informazione che vi corrisponde. Come dice Weaver, «there is more “information” if you select freely out of a set of fifty

standard messages, than if you select freely out of a set of twenty-five» (Shannon & Weaver 1964, 16).

Tuttavia, questa maggiore libertà di scelta è anche equiparata a una maggiore incertezza sul messaggio e questo è il fulcro del dibattito sul concetto di informazione: per Shannon e Weaver la quantità di informazioni aumenta insieme all'aumento dell'incertezza. «Greater information, greater freedom of choice, greater uncertainty go hand in hand» (Shannon & Weaver 1964, 27; Malaspina 2018, 19). Consapevole dell'affronto concettuale alla nozione già dominante di neghentropia, Weaver aggiunge provocatoriamente: «in these statistical terms [...] information and uncertainty find themselves to be partners» (Shannon & Weaver 1964, 27; Malaspina 2018, 16). Inoltre, questa alleanza positiva tra incertezza e informazione «sounds as though noise were beneficial!» (Shannon & Weaver 1964, 19).

Nonostante questo apparente intoppo, cioè che l'informazione *in extremis* diventa indistinguibile dal rumore, Weaver insiste sul fatto che grazie a questa audace concezione dell'informazione siamo ora «perhaps for the first time, ready for a real theory of meaning», gettando le basi per una nuova comprensione della struttura statistica del linguaggio che sarebbe particolarmente promettente per gli studi semantici.

Sebbene Weaver continui a tracciare una chiara linea di demarcazione tra informazione e rumore su basi pragmatiche (distinguendo cioè tra l'entropia “desiderabile” dell'informazione e l'entropia “indesiderabile” del rumore), i suoi sforzi sembrano non essere riusciti a placare i critici. Da allora molti hanno accusato l'approccio puramente quantitativo di trascurare l'aspetto semantico dell'informazione. Il crollo della differenza intrinseca tra informazione e rumore è una sfida filosofica permanente. Un modo per risolverla è isolare la sfera del “significato” come luogo privilegiato dell'informazione.

Un'altra strada, meno esplorata, è quella di approfondire questo problema, mettendo in discussione la dicotomia generalmente accettata tra informazione e rumore (Malaspina 2018). Come vedremo, Simondon non si sottrae certo alla sfida di concepire l'informazione come ciò che, da una certa prospettiva, si può dire scompaia nella “nebbia” del rumore bianco. È degno di nota il fatto che la fondamentale introduzione di Weaver alla teoria matematica della comunicazione di Shannon sia stata pubblicata molto tempo dopo la difesa delle due tesi di Simondon,

MEOT e *ILFI*. È quindi notevole la sensibilità dimostrata da Simondon nei confronti dei problemi che rimanevano irrisolti nella definizione di informazione di Norbert Wiener (neghentropia) e alla sua tensione con la teoria dell'informazione (dove prevale il concetto di entropia dell'informazione). Simondon si spinge molto in là per sciogliere questa tensione affrontando la relazione tra l'informazione e il suo substrato fisico.

È bene seguire Simondon nei dettagli del suo resoconto della trasmissione sia di energia che di informazione, in un capitolo giustamente intitolato "Limiti della nozione tecnologica di informazione".⁶ Egli osserva che l'energia che funge da vettore di informazione è modulata dal segnale, ma anche "in virtù della sua natura fisica". Riprendendo il rapporto tra informazione e fisica che la riduzione matematica alla pura probabilità aveva lasciato intravedere, Simondon osserva che questa natura fisica non è altro che la perdita termica che si verifica spontaneamente e inevitabilmente durante la trasmissione, ossia il rumore. In altre parole, ciò che modula l'energia che funge da vettore dell'informazione è sia il segnale sia l'entropia termica che produce, quello che Simondon chiama "rumore bianco".

Ecco il punto cruciale. La modulazione apportata dal segnale è indistinguibile dal rumore bianco, scompare nella "nebbia". Perché? C'è una co-naturalità tra segnale e rumore che, dettaglio importante, non si riduce all'effetto di una "distorsione armonica", insiste Simondon, né a una mera "deformazione" o addirittura a un impedimento che diminuirebbe il segnale. Quest'ultimo si dissolve nella nebbia del rumore bianco in virtù dell'"essere infinitamente variato" dell'informazione (Simondon 2020, 152). L'informazione, a questo proposito, è già come il rumore bianco, come l'agitazione molecolare termica (entropia), in linea di principio puramente contingente. L'informazione in questo senso è "come l'evento casuale" (Simondon 2020, 154).

Siamo molto lontani dalla definizione di Wiener di informazione come negazione dell'entropia. La comprensione di Simondon dell'informazione che emerge qui non è certo quella dell'introduzione a *MEOT*, dove l'informazione era ancora definita come ciò che aumenta

6. Capitolo II. della parte II, «Limiti della nozione tecnologica di informazione per rendere conto della relazione tra l'uomo e l'oggetto tecnico. Il margine di indeterminazione degli individui tecnici» (Simondon 2020, 151–65).

la negazione dell'entropia e si oppone alla degradazione dell'energia o addirittura come ciò che “combatte contro la morte dell'università” (Simondon 2020, 17). Provando a dare una lettura più “estrema” a questo punto, si potrebbe rovesciare l'affermazione precedente e dire che, metaforicamente, l'informazione è abitata dalla sua stessa pulsione di morte.

Un amplificatore fedele è quello che deve accettare di trasmettere un rumore di fondo maggiore. Sebbene sia possibile ridurre l'ampiezza di banda di un amplificatore per diminuire il rumore di fondo, si accetterebbe in questo modo anche che la ricchezza di informazioni dei segnali sia ridotta (Simondon 2020, 152-3). L'informazione, in questo caso, è essenzialmente:

ciò che apporta una serie di stati imprevedibili, nuovi, che non fanno parte di alcuna serie definibile in anticipo; essa è dunque ciò che esige dal canale d'informazione una disponibilità assoluta in rapporto a tutti gli aspetti di modulazione che esso incanala. *Il canale d'informazione non deve apportare da sé alcuna forma predeterminata, non deve essere selettivo.* (Simondon 2020: 152, corsivo aggiunto)

Simondon lo dice chiaramente: “l'informazione non è forma”, ma “variabilità delle forme, apporto di una variazione in rapporto ad una forma” (Simondon 2020, 154). Niente di più lontano dalla classica definizione cibernetica di Wiener dell'informazione come negazione dell'entropia, che significa imposizione di forma, ordine e prevedibilità.

Tuttavia, Simondon non cita direttamente il concetto di “entropia dell'informazione” di Shannon del 1948 né in *MEOT* né in *ILFI*, anche se in realtà è stato sviluppato contemporaneamente alla nozione di entropia negativa di Wiener, se non, addirittura, in dialogo con essa. Alla fine Simondon cita Shannon, insieme ad Hartley e Wiener, ma solo dopo la sottomissione delle sue tesi, in un documento di conferenza del 1960 intitolato *Forma, Informazione, Potenziali* (Simondon 2011, 731-60). In questo testo Simondon si riferisce sia a Shannon che a Wiener come responsabili di un concetto di informazione che corrisponde all’“inverso di una probabilità”. Egli prosegue questo punto, in modo un po' confuso, lodando la teoria dell'informazione per aver fondato la nozione di *entropia negativa* (cioè di *neghentropia*), presumibilmente intendendo la cibernetica per la quale l'informazione è definita come il

negativo della quantità solitamente definita come entropia (Simondon 2011, 746).

È innegabile che vi sia una tensione irrisolta nell'attribuzione del concetto cibernetico di neghentropia di Wiener alla teoria dell'informazione (dove prevale il concetto di entropia dell'informazione di Shannon). Cosa intende Simondon per "inverso di una probabilità"? Intende l'informazione che, con Boltzmann e Shannon, si dice "mancante", cioè relativa al numero di alternative che rimangono possibili (in altre parole, una misura della nostra incertezza)? Oppure intende il negativo della quantità solitamente definita come entropia?

L'equivoco tra il concetto teorico di entropia dell'informazione di Shannon e quello cibernetico di entropia negativa di Wiener mi sembra vada di pari passo con l'introduzione di un ulteriore livello teorico. Simondon scrive in un momento in cui la questione dell'informazione e dell'entropia è coinvolta in nuove ricerche sui sistemi termodinamici complessi, i cosiddetti sistemi metastabili lontani dall'equilibrio. Ilya Prigogine ha scritto nel 1955 la sua *Introduction to Thermodynamics of Irreversible Processes* (Prigogine 1955). Tali processi comportano non solo una stabilizzazione transitoria in stati lontani dall'equilibrio, ma anche l'afflusso di energia e quindi di entropia. Il rinnovamento che questa innovazione teorica richiede, rispetto ai precedenti concetti di informazione e rumore, esula certamente dalle finalità di questo articolo. Ciò che si può dire in questa sede è che la negazione pura e semplice dell'entropia, come potrebbe riguardare i sistemi teoricamente chiusi o il compito ingegneristico di progettare un missile autodiretto, non è più adeguata: se la neghentropia fosse l'unico criterio per valutare la quantità di informazione di un sistema complesso, allora un cristallo di quarzo di 2,5 kg "conterrebbe" certamente più informazione di un cervello umano.

Simondon non è l'unico ad aver fatto i conti con la contrastante relazione tra informazione ed entropia, soprattutto in questo contesto teorico allargato. Weaver aveva sentito la necessità di chiarire il segno (-) nella formula di Shannon ($H = - \sum p_i \log p_i$). Sebbene a prima vista possa sembrare che il valore dell'entropia (Σ) sia sottratto all'informazione (H), apparentemente in accordo con l'affermazione di Wiener secondo cui l'informazione è semplicemente il negativo dell'entropia o del disordine, Weaver spiega:

Do not worry about the minus sign. Any probability is a number less than or equal to one, and the logarithms of numbers less than one are themselves negative. Thus the minus sign is necessary in order that H be in fact positive. (Shannon & Weaver 1964, 15)

Il segno meno nella definizione di Shannon di “entropia dell’informazione” serve quindi a esprimere l’entropia come quantità positiva. Non esprime quindi la negazione dell’entropia o la neghentropia.

Simondon ha, dunque, affrontato una problematica nel momento in cui si stava dispiegando storicamente. Ciononostante, si è spinto a esplorare l’“antagonismo” che riconosceva come intrinseco al concetto di informazione, un antagonismo che, osserva, “è a malapena riconosciuto nel lavoro recente della filosofia dell’informazione e della tecnica”. Simondon giunge infatti alla conclusione che il concetto tecnico-scientifico di informazione è “non univoco” (Simondon 2020, 152). A questo punto Simondon sviluppa una distanza critica dall’idea che il concetto di informazione appartenente alla “fisica pura e alla tecnologia” sia principalmente quello di neghentropia, come inizialmente affermato nella (e poi cancellato dalla) sua introduzione a *ILFI*.

Da un lato, c’è l’informazione come fattore di organizzazione, se non come fattore di “vita”. In questo caso la concezione neghentropica dell’informazione è vista come in qualche modo “al di sopra” di fenomeni puramente casuali come il rumore, che mostrano regolarità, definiscono un dominio e persino un grado di stereotipia (Simondon 2020, 152). D’altro canto, egli fa spazio all’idea di “informazione [come] singolarità dell’*hic et nunc*’ dell’operazione, [come] puro evento” (Simondon 2015: 51). Si tratta di un’idea di informazione che implica qualcosa “come l’evento casuale” (Simondon 2020, 153), che porta a una serie di “stati imprevedibili, nuovi, non appartenenti a nessuna serie che possa essere definita in anticipo” (Simondon 2020, 152). Simondon riconosce certamente, al momento della stesura di MEOT, che esiste una dicotomia tra queste due concezioni dell’informazione e che essa infesta il “concetto puramente tecnico e scientifico di informazione”. In definitiva, egli identifica questa dicotomia tra neghentropia e valore posizionale dell’entropia come un’“antinomia”: «Questa opposizione rappresenta un’antinomia tecnica che pone un problema al pensiero filosofico: l’informazione è come l’evento casuale, ma tuttavia se ne distingue» (Simondon 2020, 153).

Per tornare alla nozione di “informazione pura”, se da un lato va pensata come inerente all’oggetto tecnico, dall’altro possiamo anticipare che non è ridicibile a nessuno dei due concetti preesistenti di informazione, neghentropia o entropia dell’informazione, ma che in qualche modo partecipa a un percorso di risoluzione della tensione tra questi due termini.

Come vedremo, questa antinomia riguarda anche la nostra comprensione della relazione tra uomo e macchina, che è un criterio fondamentale per la nozione di “informazione pura” (Simondon 2020, 154). Per risolvere questa difficoltà, Simondon ci chiede di distinguere il concetto di informazione da quello di forma. Questa differenziazione tra informazione e forma, a sua volta, aprirà quello che Simondon chiama un “divario importante” tra l’essere vivente e la macchina. Indagare sulla natura di questo divario ci avvicinerà alla comprensione di ciò che è in gioco nella nozione di “informazione pura” (Simondon 2020, 154), poiché ci porterà in ultima analisi nella direzione della metafisica speculativa di Simondon, che si occupa della nozione di limitato e illimitato, di definito e indeterminato o ἄπειρον [ápeiron].

4. Il divario tra l’essere vivente e la macchina

La macchina, ci avverte Simondon, è estranea all’uomo solo perché al suo interno “è racchiuso dell’umano, incompreso, materializzato, asservito”, pur rimanendo “umano” (Simondon 2020, 11). Questo segno di umanità che è portato dall’oggetto tecnico si riaccende proprio con la nozione di “informazione pura”. Infatti, solo se è “apprezzato e conosciuto secondo la sua essenza, cioè secondo l’atto umano d’invenzione che lo ha fondato, permeato di intelligibilità funzionale, valorizzato secondo *le sue norme interne*”, l’oggetto tecnico può essere riconosciuto come portatore di una “informazione pura” (Simondon 2020, 265, corsivo aggiunto).

Rispetto alle norme interne all’oggetto tecnico possiamo individuare gli effetti dell’antinomia che colpisce il concetto puramente tecnico e scientifico di informazione. La tensione tra stabilità e variazione della forma, che ha caratterizzato l’antinomia del concetto di informazione, si rivelerà intimamente legata alla differenza tra le norme interne all’oggetto tecnico e la normatività da cui queste norme derivano, quando nasce una nuova stirpe di oggetti tecnici.

Ciò che qualifica l'oggetto tecnico nella sua tecnicità è la perfezione delle forme, la loro plurifunzionalità, l'orchestrazione di co-determinazioni interne ed esterne. Questa sovradeterminazione della forma corrisponde alle norme interne dell'oggetto tecnico e a ciò che Simondon chiama "concretizzazione" funzionale dell'oggetto tecnico.

Le prime forme di oggetti tecnici sono definite da Simondon come "astratte", nella misura in cui sono ancora caratterizzate da "un sistema astratto di funzionamenti parziali isolati, senza sfondo comune di esistenza, senza reciprocità causale, senza risonanza interna" (Simondon 2020, 289). Nel corso del suo sviluppo storico, si dice che l'oggetto tecnico diventi più "concreto", nel senso che le sue forme diventano sinergiche, plurifunzionali e sovradeterminate da molteplici fattori:

Un oggetto tecnico perfezionato è un oggetto tecnico individualizzato, nel quale ogni struttura è plurifunzionale, sovradeterminata; ogni struttura vi esiste non soltanto come organo, ma come corpo, come ambiente e come sfondo per le altre strutture. In tale sistema di compatibilità la cui sistematica [systématique] si forma come *un'assiomatica saturata*, ogni elemento assolve non soltanto una funzione nell'insieme [ensemble], ma una funzione d'insieme. Vi è come una ridondanza d'informazione nell'oggetto tecnico diventato concreto. (Simondon 2020, 289, corsivo aggiunto)

È degno di nota il fatto che questo potere ricorsivo delle forme, per cui ogni forma o elemento è collegato a tutti gli altri e all'insieme, è qui collegato ancora una volta a un concetto di informazione che è "come una ridondanza" e quindi certamente più vicino al concetto di neghentropia di Wiener che al concetto di informazione come variazione di forme di Shannon.

Per analogia, Simondon attribuisce alla cibernetica il merito di aver concepito una nozione "superiore" di informazione, che eleva l'informazione al di sopra della sua co-naturalità con il puro caso e il rumore bianco. Questo è il concetto neghentropico di informazione. Essa è vista come un argine all'emorragia di energia, come una prevenzione della disfunzione, dell'esaurimento e della morte finale di un essere. Ciò che caratterizza questa negazione dell'entropia è l'imposizione della forma (schemi funzionali, regolarità, localizzazione, definizione, persino stereotipia) che caratterizza anche la progressiva concretizzazione dell'oggetto tecnico.

Tuttavia, nonostante il merito che alla cibernetica viene riconosciuto da Simondon, per aver traghettato l'oggetto tecnico nell'era dell'informazione e della regolazione, e per aver ampliato lo schema ilomorfo di Aristotele con concetti derivati dalla termodinamica, sarebbe comunque precipitoso saltare da questo alla nozione di "pura informazione", perché non riusciremmo a capire cosa significhi che "è racchiuso dell'umano, incompreso, materializzato, asservito" nell'oggetto tecnico. In altre parole, non possiamo sperare di riconoscere questa "pura informazione" nel cuore dell'oggetto tecnico, se ignoriamo le implicazioni normative dell'atto umano che lo ha fondato.

Potremmo sostenere che Simondon, in ultima analisi, rifugge dal confondere la qualità neghentropica che caratterizza la vita stessa con l'individuo vivente. La cibernetica sbaglia, dice, ad assimilare l'essere vivente a un meccanismo cibernetico, perché l'essere vivente ha bisogno di informazioni, mentre la macchina usa forme. Qui non si riferisce alla neghentropia. Per l'essere vivente è la singolarità il principio di individuazione, che risiede nell'*hic et nunc*, il qui e ora dell'informazione, cioè l'evento puro che assomiglia a un caso fortuito e che porta a stati imprevedibili, nuovi, non appartenenti a nessuna serie definibile in anticipo (Simondon 2011, 78). Ciò che qualifica l'essere vivente non è quindi sicuramente riducibile alla sua mera autoconservazione, alla ridondanza e quindi alla sua capacità neghentropica. È noto che Simondon critica la teoria cibernetica per aver assimilato l'essere vivente a "un automa che mantiene un certo numero di equilibri, o che cerca compatibilità tra diverse esigenze" (Simondon 2011, 38).

Tuttavia, il "divario importante" tra uomo e macchina non è sufficientemente chiarito se ci si sofferma solo su questa apparente opposizione tra uomo e oggetto tecnico, come se il primo fosse singolare e il secondo neghentropico e prevedibile. L'opposizione tra l'essere tecnico e l'essere vivente, se ci fermiamo qui, è ancora troppo grossolana. Dobbiamo invece riconoscere lo stesso scarto tra forma e informazione (qui intesa come variazione della forma) nel cuore dell'oggetto tecnico stesso. Poiché anche la tecnicità dell'oggetto non risiede, in ultima analisi, solo nella perfezione delle sue forme, che consentono un meccanismo chiuso e scorrevole, ma al contrario nella sua apertura verso un margine di indeterminazione.

Il vero perfezionamento progressivo delle macchine, quello di cui si può dire che eleva il grado di tecnicità, corrisponde non a un aumento dell'automatismo, ma al contrario al fatto che il funzionamento di una macchina contienga un certo margine di indeterminazione. È tale margine che permette alla macchina di essere sensibile ad un'informazione esterna (Simondon 2020, 13).

Come l'amplificatore fedele, sia l'essere vivente che l'oggetto tecnico richiedono, se non una "disponibilità *assoluta* del canale informativo" (corsivo aggiunto), almeno un margine di indeterminazione, cioè un'apertura alla variazione che non preselezioni a priori il segnale dalla nebbia del rumore.

Lo "scarto importante" tra l'oggetto tecnico e l'essere vivente va cercato qui, in questo margine di indeterminazione, se non nel concetto assoluto di indeterminatezza come l'ἄπειρον [ápeiron]. È in questo scarto, tra la sovradeterminazione delle forme da un lato e l'indeterminazione dall'altro, che caratterizzano entrambe la tecnicità, che la nozione di "informazione pura" è più enigmatica, perché caratterizza l'oggetto tecnico e al contempo rivela qualcosa di umano.

In altre parole, la nozione di "informazione pura" va ricercata in uno scarto che caratterizza non solo la relazione tra l'uomo e l'oggetto tecnico, ma che esiste all'interno dell'oggetto tecnico stesso. È lo scarto tra la concretezza che caratterizza la sua tecnicità (intesa come sovradeterminazione delle forme, degli elementi e del contesto) e la sua apertura a un margine di indeterminazione (che caratterizza anche la tecnicità) e che lo lega all'"informazione esterna" e, allo stesso tempo, anche all'"atto umano che ha fondato" l'oggetto tecnico - cioè alla variazione della forma.

La tecnicità dell'oggetto è infatti legata alle forme stabili che organizza in sottoinsiemi funzionali. "L'essenza tecnica si riconosce dal fatto che resta stabile attraverso la linea evolutiva" (Simondon 2020, 46). D'altra parte, cosa fondamentale, dobbiamo sforzarci di conoscerla "secondo una genesi della tecnicità" (Simondon 2020, 291). La ricerca dell'inizio di una discendenza di oggetti tecnici, a sua volta, ci riporta a un "atto di invenzione costitutivo di un'essenza tecnica". Sebbene l'essenza tecnica "possa essere riconosciuta dal fatto che rimane stabile attraverso la discendenza in evoluzione", questa stabilità di forma non deve sminuire la singolarità del suo inizio e le discontinuità che caratterizzano i suoi salti più notevoli verso la concretizzazione, rendendola

“non soltanto stabile, ma anche produttrice di strutture e funzioni per sviluppo interno e la saturazione progressiva” (Simondon 2020, 46).

Non si tratta quindi solo di comprendere le norme interne dell’oggetto tecnico, ma di cercare di capire anche la normatività che le genera. Questa questione della normatività, a sua volta, illuminerà il riferimento altrimenti enigmatico di Simondon alla teoria assiomatica e, più specificamente, al concetto di “saturazione assiomatica”. Avvicinandoci a questo problema più astratto e tecnico, ci avviciniamo anche al nucleo concettuale di questa nozione aporetica di informazione pura.

5. Completezza assiomatica

A mia conoscenza, c’è solo una distinzione inequivocabile che Simondon fa tra l’essere vivente e gli oggetti tecnici creati dall’uomo. L’essere vivente è in grado di intrattenere una relazione con se stesso, che può “tradurre” in informazioni. Questa auto-relazione, ci dice Simondon, permette all’essere vivente di integrare i problemi che affronta e le tensioni che sperimenta come incompatibilità in un’“assiomatica di problemi vitali” (Simondon 2011, 38).

Come dobbiamo intendere l’“assiomatico” di cui parla Simondon? In un supplemento alle due tesi, *Histoire de la notion de l’individu*, Simondon usa il termine in un passaggio su Cartesio, riferendosi esplicitamente al suo uso moderno:

l’essere individuale interviene come generatore del termine necessario per la situazione e senza il quale non potrebbe ordinarsi; il soggetto possiede pertanto un potere di iniziativa ed il problema resterebbe per sua essenza indeterminato, poiché racchiude un’assiomatica, e per parlare in linguaggio moderno, questa sua chiusura non solo rende possibile l’azione, ma si confonde con l’azione nel momento in cui quest’ultima si compie: in ciò consiste, in ultima analisi, l’automatismo dell’azione. (Simondon 2011, 597)⁷

7. La véritable morale cartésienne est bien la morale provisoire, car elle correspond à un problème à résoudre qui est différent dans chaque cas ; et ce problème ne peut être résolu qu’en supposant un terme qui n’est pas donné dans l’énoncé, mais qui, combiné avec les termes de l’énoncé, définit un ordre fécond. L’être individuel intervient ici comme celui qui crée ce terme qu’il faut ajouter à la situation et sans lequel elle ne peut être ordonnée ; le sujet a donc un pouvoir d’initiative ; sans lui, le problème resterait indéterminé ; il ferme l’axiomatique, pour parler en langage moderne, et c’est

Si potrebbe dire che quello di Cartesio è l'ideale di un paradigma tecnico della ragione senza rumore: "una catena trasferisce senza scarti l'azione esercitata su una delle sue estremità all'altra estremità [...] allo stesso modo, il ragionamento realizza un trasferimento senza scarti del significato delle prime proposizioni sino alle ultime" (Simondon 2011, 597). Ma già Simondon mette in guardia,

questo trasferimento senza scarti non basta all'invenzione [...] Allo stesso modo, le matematiche sono infeconde se dispongono esclusivamente della *sola identità*, che le indurrebbe a tradursi in un'immensa tautologia, così come l'azione sarebbe un nulla se non consentisse l'assegnazione di un principio identico a casi diversi. (Simondon 2011, 597-98, corsivo aggiunto)

La chiusura assiomatica è quindi una nozione ambivalente come quella di informazione. Tende a un'identità stabile e delimitata, a qualcosa di simile alla chiusura, nel senso che nulla può essere aggiunto alla perfezione e alla sinergia delle sue forme. Ma è anche, essenzialmente, un momento decisivo in cui l'essere vivente si dà i mezzi per intervenire nella propria assiomatica vitale. Qui il soggetto ha "il potere di iniziativa" per creare il termine che mancava. Questo momento assiomatico non è semplicemente deduttivo. Deve corrispondere a una nozione di informazione concepita come evento, come singolarità. Può avvenire solo perché l'essere vivente è in grado di auto-relazionarsi, di tradurre in informazione le tensioni che vive, i problemi che affronta. È qui che ci avviciniamo alla nozione di informazione pura.

Per comprendere meglio la posta in gioco di questo riferimento cruciale a un "assioma vitale" e alla moderna comprensione della "chiusura assiomatica", ripercorriamo brevemente l'origine matematica del concetto di chiusura assiomatica nell'assioma di completezza di Hilbert, prima di confrontare l'uso che Simondon ne fa, nella sua ricerca di una nozione "pura" di informazione, con un altro grande pensatore assiomatico: Spinoza.

Il concetto di tecnicità, ricordiamo, consiste nel rendere sistematiche tutte le relazioni. Richiede che "ogni struttura sia plurifunzionale,

cette fermetures qui non seulement rend l'action possible, mais se confond avec l'action en train de s'accomplir; en cela consiste l'automatisme de l'action. (Simondon 2005: 454).

sovradeterminata [...] in [un] sistema di compatibilità la cui sistematicità [système] si forma, proprio come una saturazione assiomatica” (Simondon 2020, 289). Allo stesso modo, in un sistema logico ogni proposizione è strettamente correlata alle proposizioni dedotte da essa fino a che, nelle parole del filosofo francese della matematica Robert Blanché, “step by step, a tight network is constituted wherein, directly or indirectly, all propositions communicate” (Blanché 2009, 9-10; Malaspina 2018, 57).

Un sistema assiomatico, a sua volta, si dice “saturo” o chiuso quando obbedisce all’assioma della completezza (*Vollständigkeit*). In un sistema assiomatico completo qualsiasi affermazione ad esso relativa, o la sua negativa, può essere dimostrata utilizzando il sistema stesso. La completezza di un sistema assiomatico è, allo stesso tempo, anche la sua chiusura come sistema, perché implica che nessun nuovo assioma indipendente può essere aggiunto senza introdurre una contraddizione.⁸ La geometria euclidea, ad esempio, si basa su una serie di assiomi e postulati. Uno di questi postulati (il postulato delle parallele) non poteva essere né dimostrato né confutato e il sistema assiomatico di Euclide rimaneva quindi incompleto. Lobachevski aggiunse un assioma che chiudeva il sistema assiomatico, inventando così una geometria iperbolica con dimensioni infinite, tanto che ora si può parlare di geometria assoluta.

L’assioma di completezza di Hilbert recita come segue:

Axiom of Completeness. (*Vollständigkeit*): To a system of points, straight lines, and planes, it is impossible to add other elements in such a manner that the system thus generalized shall form a new geometry obeying all of the five groups of axioms. In other words, the elements of geometry form a system which is not susceptible of extension, if we regard the five groups of axioms as valid. (Hilbert 1950)

Va detto, tra parentesi, che il tentativo di Hilbert di ideare un sistema assiomatico completo per tutta la matematica, tale da poter parlare di una matematica assoluta, fu in gran parte vanificato dai teoremi di incompletezza di Gödel, pubblicati nel 1931.

8. <http://www.cnrtl.fr/definition/saturation>.

Cosa significa, infine, tutto questo per la nozione di “informazione pura”? L’ordinamento assiomatico dell’oggetto tecnico, che tende alla chiusura assiomatica, cioè alla perfezione e alla mutua sovradeterminazione delle sue forme, sembra in un primo momento ricondurci a un concetto di informazione che corrisponde alla neghentropia, in quanto impone vincoli, stabilizza le sue forme e blocca di fatto il sistema.

D’altra parte, l’oggetto tecnico è la prova materiale dell’audacia assiomatica dell’essere vivente. Il potere dell’iniziativa, qui, è sia assiomatico che normativo. È il potere autofondante della mente di generare nuove norme. Testimonia la capacità di determinare la relazione tra l’uomo e il suo ambiente associato. Il soggetto ha qui il potere di iniziativa.

L’oggetto tecnico è la prova di questo “atto umano” fondante e anche della sua intelligibilità funzionale. In linea di principio, può essere decostruito, ricostruito e trasmesso come informazione. Soprattutto, può essere insegnato in modo universale, più geometrico (alla maniera del geometra, come diceva Spinoza, intendendo il metodo assiomatico) (Simondon 2020, 15). È forse leggendo l’uso che Simondon fa dei concetti derivati dalla teoria assiomatica nel contesto del metodo assiomatico di Spinoza che possiamo avvicinarci maggiormente alla posta in gioco, alla nozione di “informazione pura” di Simondon.

6. L’ἄπειρον

La messa in primo piano di Spinoza di un *more geometrico* (filosofia in modo geometrico) e l’importanza data sia da Spinoza che da Simondon al procedimento assiomatico, danno motivo di mettere in relazione questi due eminenti filosofi sistematici della natura. L’*èlan* metafisico di Simondon per superare l’opposizione tra essere e divenire (il processo di individuazione è “una delle possibilità del divenire dell’essere, date certe condizioni”) ricorda infatti la *natura naturans* di Spinoza (Simondon 2011, 87). È anche vero che il soggetto (capace di essere causa) non è ridicibile alla nozione di individuo né per Simondon né per Spinoza. Per Spinoza il soggetto è *natura naturans*, per Simondon, allo stesso modo, il soggetto è “più vasto dell’individuo, più ricco di esso, e ha, oltre all’individualità dell’essere individuato, un certo peso della natura, dell’essere non individuato” (Simondon 2020, 267). Simondon specifica che “la parola “natura” potrebbe essere usata per designare il resto

di ciò che è originario nell'individuo, precedente anche all'umanità costituita nell'uomo", un "ἄπειρον [ápeiron] che resta attaccato a ogni essere individuale" (Simondon 2020, 266).

Alla luce di ciò, dobbiamo chiederci ancora una volta: che cos'è la "natura umana" che l'oggetto tecnico rivela, quando è conosciuto nella sua essenza? La natura umana, per Simondon, non si riferisce a caratteristiche antropologiche o biologiche. Rivela invece "il resto di ciò che è originario, anteriore anche all'umanità costituita nell'uomo", un peso o una carica di realtà pre-individuale che apre, all'interno di ogni individuo, a ciò che i presocratici chiamavano l'ἄπειρον (Simondon 2020, 266). Talete lo identificava con ciò che è fluido, Anassimandro con ciò che è gassoso ed evapora. Anassimandro definiva l'ἄπειρον come ciò che è infinito o indeterminato, continuo. Simondon suggerisce infine che l'ἄπειρον può offrire una prima definizione di materia come ciò che è anteriore alla distinzione tra forma e materia (Simondon 2011, 460).

Simondon collega infatti questo stato, che precede l'individuo, anche alla nozione di informazione. L'illimitato e l'indefinito, l'ἄπειρον non è la fonte dell'informazione, ma la sua "condizione primordiale". Corrisponde a una "prima informazione", anteriore alla dualità tra mittente e destinatario e a qualsiasi trasmissione, ancora legata a uno stato di tensione (Simondon 2011, 45).

Qual è allora la differenza tra gli attributi infiniti della *natura naturans* di Spinoza e il riferimento di Simondon alla natura umana, nella misura in cui quest'ultima rivela la natura in noi come ciò che è illimitato e indefinito e che rimane capace di divenire? È qui che troviamo quella che è forse la maggiore vicinanza di Simondon a Spinoza e, allo stesso tempo, una differenza decisiva e, come vedremo, una differenza assiomatica.

Con Spinoza solo all'Essere viene riconosciuta, in ultima analisi, la dignità di essere pienamente, di essere capace di essere causa, di essere il vero soggetto di un atto. Simondon, invece, si tiene lontano da questa concezione infinitamente dilatata del soggetto e lo chiarisce in una bella parentesi:

[Leibniz frammenta l'individuazione sino agli estremi limiti della piccolezza, assegnando individualità anche agli elementi più minuti del corpo umano. Al contrario, Spinoza amplia l'individuazione sino ai limiti del tutto, ovvero ciò per cui Dio e natura naturante in

quanto individuazione. Ne nell'uno ne nell'altro si rileva, in rapporto all'individuo, un ambiente associato, ovvero un sistema dello stesso ordine di grandezza in seno al quale l'individuo possa ricevere una genesi. L'individuo è preso per l'essere ed è considerato coestensivo all'essere. In queste condizioni, non si può concepire l'individuo considerato come coestensivo all'essere: ogni realtà è, allo stesso tempo, troppo piccola e troppo grande per ricevere lo statuto di individuo. Tutto può essere individuo e nulla può esserlo del tutto]. (Simondon 2011, 90)

Il soggetto non è quindi estendibile all'infinito per Simondon, e di conseguenza non corrisponde all'infinito assiomatico di Spinoza. È qui che il riferimento di Simondon all'assioma moderno della completezza è importante. Esso segna una differenza decisiva tra le concezioni di Spinoza e Simondon non solo dell'individuo, ma anche del soggetto.

L'oggetto tecnico rivela qualcosa dell'atto umano che lo ha inventato, pensato e voluto (Simondon 2020, 266), ma “non è l'individuo che inventa, è il soggetto” (Simondon 2020, 267), un soggetto capace di rappresentare a se stesso “la propria azione nel mondo, come elemento e dimensione del mondo” (Simondon 2011, 43). Quali sono dunque le dimensioni del soggetto, per Simondon, nella misura in cui possono a volte essere coesistenti con l'individuo, ma non riducibili ad esso? Chi è questo soggetto che Simondon “finora ha presentato [...] come se [fosse] individuale”? Nella realtà, dice, “il soggetto è un essere collettivo” (Simondon 2020, 216). Eppure, nessuna antropologia può rendere conto della natura di questo soggetto collettivo. Né una persona singola, né un semplice gruppo di individui che mettono in atto le loro funzioni somatiche, né tantomeno individui in una relazione intersoggettiva. Si tratta sicuramente di un soggetto estraneo (Simondon 2020, 266-67)? Che tipo di collettività non mette in relazione le persone sulla base di ciò che è identico in loro (siano esse le loro forme di sensibilità a priori) né sulla base delle differenze che costituiscono la loro individualità? Una collettività che non è legata da condizioni sociali date o addirittura da fattori economici? È addirittura umana?

La chiave per comprendere questa collettività coincide con il nostro problema di definire la nozione di “informazione pura”. Quest'ultima esprime l'essenza dell'oggetto tecnico, la sua intelligibilità nella misura in cui è pensata e voluta da una collettività che corrisponde all'accoppiamento delle “capacità inventive e organizzative di più soggetti”, che

a sua volta fa appello alla loro natura umana condivisa, cioè alle virtualità e alle potenzialità che rimangono indeterminate in essi (Simondon 2020, 272).

È una collettività di individui che diventa soggetto collettivo attraverso un atto di invenzione, intervenendo nella sua assiomatica vitale, piegandola verso la completezza assiomatica, introducendo proporzioni e misure stabilite μέτρον. A mio avviso, la nozione di "informazione pura" esprime l'audacia assiomatica di questa collettività. È decisiva perché è assiomatica. Crea un accoppiamento tra il περὸν limitato e l'ἄπειρον indeterminato o illimitato, e lo fa in virtù di un atto di invenzione che introduce un limite, e con esso la misura e la proporzione, il μέτρον.

Questo soggetto è una collettività intesa specificamente alla luce dell'assioma che chiama in vita, alla luce di un atto che qui fonda l'oggetto tecnico, grazie al quale questa collettività è in grado di risolvere un'incompatibilità, di superare un problema. Ora è anche chiaro perché un tale soggetto non può ricavare il potere inventivo che sfrutta da forme già costituite, da ciò che è limitato nella sua individualità o vincolato da convenzioni stabilite di convivenza sociale, perché ciò che è umano in esso è proprio ciò che lo apre all'indeterminato ἄπειρον.

Questo è, in definitiva, il motivo per cui la nozione di "informazione pura" non può essere ridotta a nessuna delle due definizioni tecniche di informazione (neghentropia o entropia dell'informazione). Non si tratta né della mera negazione dell'indeterminato, né di una sua semplice espressione, ma dell'introduzione di una misura μέτρον, un limite in cui περὸν e ἄπειρον si toccano.

Osservazioni conclusive

Possiamo ora chiudere questo articolo, o meglio aprirlo, sull'idea di un'informazione pura che esprime l'intervento decisivo e assiomatico di una collettività. Questa collettività agisce come un soggetto che non è frenato dalle norme e dalle convenzioni stabilite che rendono compatibili gli individui, che non si affida solo agli affetti e alle rappresentazioni che creano un legame intersoggettivo. L'"informazione pura", veicolata dall'oggetto tecnico, può ora essere intesa come "mezzo [le support] e simbolo" di una collettività che si basa su una contrazione decisiva e assiomatica di ciò che, in noi, rimane infinito e indeterminato

e della nostra capacità di introdurre misura e proporzione, alla luce di una finalità condivisa. L'ἄπειρον agisce come una prima informazione, paragonabile al concetto di entropia dell'informazione di Shannon, dove l'informazione è, in extremis, indistinguibile dal rumore bianco. D'altra parte, l'introduzione del μέτρον corrisponde all'emergere della forma e delle norme interne dell'oggetto, introducendo una stabilità e tendendo alla chiusura assiomatica. Questo aspetto è senza dubbio il più vicino al concetto di informazione come neghentropia.

L'“informazione pura” di cui è portatore l'oggetto tecnico testimonia il nostro potere di iniziativa, è la prova di un soggetto collettivo capace di accoppiare le nostre “capacità inventive e organizzative”, in grado di piegare i problemi complessi che affrontiamo verso la completezza assiomatica, in modo che l'azione diventi possibile (Simondon 2011, 597).

L'oggetto tecnico che viene apprezzato e conosciuto secondo la sua essenza, cioè secondo l'atto umano che lo ha fondato, penetrato di intelligibilità funzionale, valorizzato secondo le sue norme interne, porta con sé l'informazione pura. (Simondon 2020, 265-66) [Eso] diventa il mezzo [le support] e il simbolo di questa relazione, che vorremmo chiamare transindividuale. (Simondon 2020, 266)

Simondon arriva persino a definire questa collettività come essa stessa un assioma, che risolve un problema (Simondon 2011, 39).

Mentre scrivo queste considerazioni conclusive, sono seduta in silenzio in un tradizionale “bar tabac” nel sud della Francia, circondata da anziani francesi del posto. La conversazione nell'ambiente, che fa a gara nel lodare il *Front National*, nell'esaltare le virtù di Jean-Marie Le Pen, nel giustificare persino la tortura in Algeria (“gli altri erano peggio, sono selvaggi”), mi distrae. La tensione che provo è in parte dovuta al divario tra le astrazioni del pensiero filosofico e la mia impotenza a intervenire concretamente in questa situazione. Cosa posso dire a un uomo anziano che ha vissuto la guerra, che può essere stato vittima e carnefice della tortura? Prima che possa risolvere questo dilemma, la conversazione si interrompe. Due anziani del posto sono entrati nel bar. Si distinguono. Forse sono di origine algerina, visto che nella comunità

locale c'è una prevalenza di algerini.⁹ Ora c'è silenzio, a parte un saluto reciproco e sommesso. Non c'è bisogno che i due uomini ordinino. Il barman sa già cosa prendono.

Non so come armonizzare lo sforzo di pensare la nozione di Simondon di “pura informazione” con l'opportunità mancata di dire qualcosa di significativo nel contesto di questa collettività spezzata. Tuttavia, non è mai sembrato così rilevante pensare alla natura umana non come a ciò che ci rende uguali, né semplicemente come alle differenze che ci arricchiscono,¹⁰ ma come a un serbatoio di potenzialità e virtualità che possiamo attivare collettivamente per trovare una soluzione ai problemi che dobbiamo affrontare.

È necessario un deciso cambio di mentalità e un rinnovato vigore filosofico, affinché gli odierni oggetti della tecnologia della comunicazione possano essere concepiti in modo credibile come portatori di tale “informazione pura”, come mi sembra che Simondon la intenda, anziché essere ridotti a veicolo di disinformazione e controllo. Permettetemi di riformulare l'epigrafe di questo articolo alla luce di quanto detto sopra: qualsiasi collettività può diventare un soggetto autonomo rispetto all'informazione, perché è qui che risiede la vera autonomia!¹¹ Qualificare questa informazione come “pura” significa caratterizzarla per il potere di iniziativa che la anima e per la sua assiomatica audacia nel tenere insieme ciò che è infinito in noi e ciò che è misurato.***

9. I nati in Francia dal 1963, da un genitore nato in Algeria prima del 1962, sono automaticamente cittadini francesi, ma sono ancora comunemente chiamati “*les arabes*”, come se fossero stranieri o addirittura etnicamente arabi, cosa che non è di fatto vera, ad esempio, per gli algerini di origine berbera.

10. Paul Valéry ha detto: “Arricchiamoci a vicenda delle nostre reciproche differenze”.

11. “L'individuo consiste dunque in quell'essere in grado di conservare e aumentare il contenuto informativo e risulta autonomo in base all'informazione stessa, poiché e in ciò che risiede la sua autonomia” (Simondon 2011, 259).

*** Ringraziamenti di Cecile Malaspina: vorrei ringraziare in particolare Francesca Sunseri per la sua bella traduzione in italiano, oltre che Conor Heaney, per aver dato vita al progetto di questa pubblicazione in inglese, e, insieme a Iain MacKenzie e a tutti i partecipanti alla conferenza del 2018, *Culture Technics: The Politics of Simondon's Du Mode*, presso il Centre for Critical Thought University of Kent, per quella che è stata una conversazione intensamente stimolante sulla rilevanza contemporanea di Simondon.

Riferimenti bibliografici

Blanche, R. 2009, *L'axiomatique*, PUF, Paris.

Brillouin, L. 1949, *Life, Thermodynamics and Cybernetics*, "American Scientist", 37, pp. 554–68.

Clausius, R. 1867, *The Mechanical Theory of Heat*, John van Voorst, London.

Drake, G. 2018, *Encyclopaedia Britannica*. Available online at <https://www.britannica.com/science/entropy-physics>.

Hilbert, D., Townsend, E. J. trans. 1950 [1902], *The Foundations of Geometry*, The Open Court Publishing Company, Illinois La Sall. Available online at <https://math.berkeley.edu/wodzicki/160/Hilbert.pdf>.

Malaspina, C. 2018, *An Epistemology of Noise*, Bloomsbury, London.

Planck M. 1896, *Gegen die neue Energetik*, "Annalen der Physik", 293, pp. 72-78. Available online at <https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.18962930107>.

Prigogine, I. 1955, *Introduction to Thermodynamics of Irreversible Processes*, Charles C. Thomas, Springfield.

Shannon, C. E. 1948, *A Mathematical Theory of Communication*, "The Bell System Technical Journal", 27(379-423), pp. 623-656.

Shannon, C. E. & Weaver, W. 1964, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana.

Simondon, G. 2011, *L'individuazione alla luce delle nozioni di forma e informazione*, Mimesis, Milano.

Simondon, G. 2020, *Del modo di esistenza degli oggetti tecnici*, Orthotes, Napoli-Salerno.

Spinoza, B. Undated, pencil inscription by owner 1934 [1662], *On the Improvement of the Understanding*. In *Philosophy of Spinoza*. Translated by R. H. M. Elwes trans. New York: Tudor.

Szilard, L.. 1929, *Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen*, “*Zeitschrift für Physik*”, 53(11-12), pp. 840-856. <https://link.springer.com/article/10.1007%2F01341281>.

Wiener, N. 1961, *Cybernetics or Control and Communication in the Animal and the Machine*, The MIT Press, New York.

Cecile Malaspina
King's College London; Collège International de Philosophie
cecile.malaspina@gmail.com