

I quaderni di
Agenda  **Digitale** ^{eu}

MAGGIO-AGOSTO 2023

n. 0014

Agendadigitale.eu è una testata scientifica e giornalistica registrata al Tribunale di Milano **Dati di riferimento**

Iscrizione ROC n. 16446

ISSN 2421-4167

Numero registrazione 1927, Tribunale di Milano

Editore: Digital360 **Focus e ambito:**

La rivista scientifica, i Quaderni di Agendadigitale.eu, pubblica fascicoli quadrimestrali in open access.

Lo scopo è creare un luogo per accompagnare i passi dell'Italia verso la necessaria rivoluzione digitale, con approfondimenti multidisciplinari a firma di esperti delle materie afferenti all'Agenda Digitale italiana ed europea

Submission e norme editoriali

Per effettuare una submission è necessario concordare prima un argomento e le misure precise contattando info@agendadigitale.eu.

Inviare un abstract di circa 500 caratteri alla testata, presentando l'articolo.

Le misure del testo finale saranno comprese tra 6mila e 20mila caratteri, salvo accordi per misure superiori.

I riferimenti bibliografici dovranno essere preparati in conformità alle regole dell'APA style, 6a edizione (si vedano le linee guida e il tutorial).

Gli autori sono invitati a tener conto degli articoli già pubblicati nella rivista e di citarli nel loro contributo qualora siano ritenuti di interesse per il tema trattato.

Direzione e comitato editoriale

Direttore responsabile

Alessandro Longo

Direttori scientifici

Paolo Ferri, Mario Morcellini

Comitato editoriale

Giovanni Boccia Artieri, Mario Pireddu, Luca Toschi

Comitato scientifico

Presidente:

Alessandro Perego, Politecnico di Milano

Membri del Comitato scientifico

Francesco Agrusti, Università degli Studi Roma TRE

Davide Bennato, Università di Catania

Giovanni Biondi, Indire, Iulm

Giovanni Boccia Artieri, Università di Urbino

Paolo Calabrò, Università Vanvitelli di Caserta

Antonio Chella, Università di Palermo

Stefano Cristante, Università del Salento

Lelio Demichelis, Università Insubria

Marco del Mastro, Unicusano

Carlo Alberto Carnevale Maffè, Università Bocconi di Milano

Carmelo Cennamo, Università Bocconi di Milano

Michele Colajanni, Università degli Studi di Modena e Reggio Emilia

Mariano Corso, Politecnico di Milano

Ottavio Di Cillo, università di Bari

Maurizio Ferraris, università di Torino

Ivan Ferrero, psicologo

Paolo Ferri, Università Bicocca di Milano

Pietro Fiore, Università di Foggia

Stefania Fragapane, Università degli Studi di Enna Kore

Alfonso Fuggetta, Politecnico di Milano

Alberto Gambino, Università Europea di Roma

Carlo Giovannella, Università Tor Vergata di Roma

Renato Grimaldi, Università di Torino

Mariella Guercio, Università Sapienza di Roma

Mauro Lombardi, Università di Firenze

Mariano Longo, Università del Salento
Roberto Maragliano, Università Roma Tre
Massimo Marchiori, Università di Padova
Berta Martini, Università di Urbino Carlo Bo
Leonardo Menegola, università Milano Bicocca
Tommaso Minerva, Università degli studi di Modena e Reggio Emilia
Mario Morcellini, Università degli Studi di Roma “La Sapienza”
Giuliano Noci, Politecnico di Milano
Fabrizio Onida, Università Bocconi di Milano
Norberto Patrignani, Politecnico di Torino
Mario Pireddu, Università degli Studi della Tuscia
Franco Pizzetti, Università di Torino
Alessio Plebe, Università di Messina
Roberto Pozzetti, psicanalista, LUDeS Campus Lugano, università Insubria
Antonio Rafele, Università di Parigi (CEAQ- Université Paris Descartes La Sorbonne)
Francesco Sacco, Università Bocconi di Milano
Donatella Sciuto, Politecnico di Milano
Nicola Strizzolo, Università di Udine
Elena Valentini, Università Sapienza di Roma
Guido Vetere, Università Sapienza di Roma

Comitato di referaggio

Coordinatore: Luca Gastaldi, Polimi

Mauro Andreolini, sicurezza informatica, Unimore
Luca Baccaro, concorrenza, diritto comunicazioni elettroniche e dei media; studio legale Lipani Catricalà & Partner
Raffaello Balocco, IT e innovazione, Politecnico di Milano
Francesco Capparelli, privacy, cyber security, ecommerce, data management, identità digitale; studio legale ICT Legal Consulting
Antonio Chella, ingegneria informatica, intelligenza artificiale, Università di Palermo
Marco Centorrino, Università di Messina – processi culturali e comunicativi, nuove tecnologie
Ida Cortoni, media education e digital literacy; Dipartimento di Comunicazione e Ricerca Sociale, Sapienza Università di Roma
Giuseppe D’Acquisto, Autorità garante privacy, sicurezza e privacy
Mario dal Co, Economista e manager, già direttore dell’Agenzia per l’innovazione
Lelio Demichelis, Università Insubria, sociologia, economia
Daniela Di Donato, Docente di lettere, Dottoranda di ricerca presso Sapienza Università di Roma- Dipartimento di Psicologia dei processi di sviluppo e socializzazione, Collaboratrice del Crespi
Francesco Di Giorgi, diritto dell’informazione e della comunicazione, tutela dei consumatori, diritto delle comunicazioni elettroniche; Agcom

Leonella Di Mauro, data management, e-commerce, tutela del consumatore, diritto delle comunicazioni elettroniche; Agcom

Luisa Franchina, cyber security, Hermes Bay

Luca Gastaldi: eGov, sanità, telecomunicazioni, procurement pubblico, design thinking, Smart Working, Politecnico di Milano

Maurizio Gentile, professore associato, Università di Roma LUMSA, didattica e pedagogia

Antonio Ghezzi: strategia, business model, startups, mobile, Politecnico di Milano

Ugo Imbriglia, sociologo

Gevisa La Rocca, Università Kore di Enna, piattaforme digitali, communication research, analisi qualitativa dei dati

Nicola La Sala, registro degli operatori della comunicazione, fattura elettronica, industria4.0, editoria, cittadinanza digitale; Agcom

Emanuele Lettieri, sanità Politecnico di Milano

Maria Beatrice Ligorio, psicologia, università di Bari

Marika Macchi, economia, Unifi

Riccardo Mangiaracina: fatturazione elettronica, eCommerce, logistica e trasporti, export, Politecnico di Milano

Mirco Marchetti, Sicurezza informatica, unimore

Chiara Marzocchi, economia, Università di Manchester

Cristina Masella, Sanità, Politecnico di Milano

Carmelina Maurizio, Dipartimento di Filosofia e Scienze dell'educazione Università di Torino

Stefano Moriggi, scienze della comunicazione, filosofia, Bicocca di Milano

Davide Mula, sanità digitale, cyber security, privacy; Agcom

Simone Mulargia, internet and social media studies; Lumsa

Antonella Napoli, sociologia, media e comunicazione, giornalista

Sebastiano Nucera, Università di Messina, Media e Tecnologie Indossabili

Achille Pierre Paliotta, Social cybersecurity, disinformazione, tecnologie digitali, intelligenza artificiale, sociologia economica; INAPP

Francesco Paoletti, docente di organizzazione aziendale e gestione delle risorse umane, Università degli Studi di Milano-Bicocca

Norberto Patrignani, computer ethics, filosofia, Politecnico di Torino

Dunia Pepe, Inapp e Università Roma Tre, cultura e formazione digitale

Alessio Plebe, Università di Messina, Scienze cognitive, pedagogiche, psicologiche

Francesco Pira, Unime, comunicazione pubblica, le dinamiche social, le fake news e i processi di disinformazione

Franco Pizzetti, diritto, privacy, università di Torino

Barbara Quacquarelli, scienze umane e formazione, università Milano Bicocca

Antonio Rafele, Sociologia dei processi culturali e comunicativi, Unicusano

Filippo Renga: turismo digitale, smart agrifood, finance and banking, mobile, Politecnico di Milano

Angelo Rovatti, tutela del diritto d'autore, diritti connessi, Diritto dei media; Agcom

Christian Ruggiero, sociologia del giornalismo e comunicazione politica; Dipartimento di

Comunicazione e Ricerca Sociale, Sapienza Università di Roma

Franco Torcellan, Associazione RED – Laboratorio di Ricerca Educativa e Didattica “Formare Trasformare Innovare”

Angela Tumino: Internet of Things, logistica e trasporti, smart city, Politecnico di Milano

Simone Vannuccini, economia, SPRU

Francesco Varanini, filosofia, formazione, università di Pisa

Guido Vetere, Università Sapienza di Roma, intelligenza artificiale, tecnologia

Indice del fascicolo

ChatGPT a scuola, ecco come potrebbe cambiare la didattica.....	7
Di Paolo Ferri , Full Professor at University of Milan Bicocca and Higher Education Consultant.....	7
Siamo nell’era post–Turing Test ed è epocale: come ci siamo arrivati, su cosa interrogarsi	19
Di Alessio Plebe , Università degli Studi di Messina.....	19
GPT-4 nel paese di Acchiappacitrulli.....	29
Di Daniela Tafani - Dipartimento di Scienze Politiche dell'Università di Pisa.....	29
Il giornalismo nell’era dell’algoritmo: il matrimonio impossibile tra IA e news	36
Di Francesca Rizzuto , Docente di Sociologia del giornalismo dell’Università di Palermo	37
Ma perché i chatbot hanno (così tante) allucinazioni	45
Di Chiara Cilaro , Psicologa psicoterapeuta, esperta in psicologia digitale	45
Generative AI, dov’è il bene per l’Umanità?	49
Di Mauro Lombardi , Scienze per l’Economia e l’Impresa, Università di Firenze	49
Alfabetismo, competenze, partecipazione: elementi per una democrazia digitale	68
Di Marco Giacomazzi , Università di Bologna	68
Scuola 4.0: fare didattica con i meme per educare alla creatività e alla complessità	79
Di Daniela Di Donato , Docente di italiano (Liceo scientifico), PhD in Psicologia sociale, dello sviluppo e della Ricerca educativa presso Sapienza Università di Roma, esperta di metodologie didattiche, inclusione e uso delle tecnologie digitali a scuola.....	79

ChatGPT a scuola, ecco come potrebbe cambiare la didattica

Una prima mappa “euristica” del possibile uso didattico dei Large Language Model ci indica che il campo che più potrebbe giovare è quello della “personalizzazione”. Risulta poi chiaro che lato insegnante questo influirebbe anche sulla qualità della relazione educativa. Ma si tratta di fenomeni ancora tutti da sperimentare

Di **Paolo Ferri**, Full Professor at University of Milan Bicocca and Higher Education Consultant

Fino al novembre del 2022 l'intelligenza artificiale era un campo di ricerca molto promettente dell'*Information Technology* e nell'ambito dell'apprendimento e della formazione, aveva solo alcune limitate applicazioni, ad esempio, aiutava i professionisti della formazione e i ricercatori nell'analizzare le grandi quantità di dati, i cosiddetti *Learning Analytics* (Ferri, 2019) prodotti da chi apprende all'interno di un ambiente virtuale/LMS: i log, i tempi di permanenza in piattaforma, le visualizzazioni dei singoli oggetti didattici, i feedback e le altre attività generate dai dipendenti di una azienda o dagli studenti di un'università o di una scuola. I moderni sistemi di intelligenza artificiale permettano di analizzare i big data, anche quelli relativi all'apprendimento, in forma automatica o semi automatica attraverso algoritmi di *machine learning* e di *deep learning*¹. La maggior parte degli Ambienti virtuali di apprendimento (LMS/VLE), contemporanei integrano già forme più o meno avanza di “tracciamento” della formazione basate sull'AI. La sintesi dei dati in report grafici generati automaticamente permette a chi progetta l'apprendimento di contare su un cruscotto digitale di monitoraggio dell'efficacia della formazione impensabile nel mondo analogico.

L'emergere dei “modelli di conversazione” basati sull'intelligenza artificiale: dialogare e apprendere da una macchina?

L'altro campo di utilizzo dell'intelligenza artificiale nella formazione è quello costituito dai Chat Bot o “agenti conversazionali” un tipo di interfaccia software che simula una conversazione con un utente attraverso una interfaccia testuale o vocale. In altre parole, gli “agenti conversazionali” sono una

¹ Sia il *machine learning* che il *deep learning* sono branche dell'intelligenza artificiale. Entrambi gli approcci fanno sì che gli algoritmi emulino decisioni che paiono “intelligenti”. Entrambe le tecnologie si basano sulla disponibilità di grandi quantità di dati da cui i sistemi possono “imparare”. Da un punto di vista cronologico, il *machine learning* è la tecnologia più “anziana”. Un requisito per utilizzare questa tecnologia è la disponibilità di grandi quantità di dati strutturati. In primo luogo, il sistema viene alimentato con dati strutturati e categorizzati e, in questo modo, “apprende” come classificare nuovi dati dello stesso tipo o simili. Questa prima fase di applicazione è seguita dall'ottimizzazione dell'algoritmo grazie al feedback umano: il sistema viene cioè “istruito” da esperti umani che segnalano eventuali classificazioni errate e marcano le categorizzazioni corrette. Con il *deep learning*, che può essere considerato uno sviluppo del *machine learning* non sono necessari dati strutturati. Il sistema funziona con reti neurali multistrato che combinano diversi algoritmi modellati su alcune funzioni della mente umana. Per questo motivo il sistema è in grado di elaborare grandi quantità di dati non strutturati.

forma di Intelligenza Artificiale (AI) che è in grado di interagire con gli utenti e rispondere alle loro domande/richieste in linguaggio naturale in modo automatico. Questi “agenti” vengono utilizzati nell’apprendimento aziendale da alcuni anni e lo sono anche in sperimentazioni legate all’educazione formale e a quella universitaria. Tuttavia, questi fino a ieri, queste applicazioni, costituivano uno strumento specialistico (quelli *educational*) o molto mirato (quelli aziendali) e operavano su domini conoscenza o ontologie molto ristrette, con la finalità di migliorare l’esperienza di apprendimento, la personalizzazione, o per effettuare simulazioni interattive (Quiroga Pérez al., 2020, Gwo-Jen Hwang & Ching-Yi Chang, 2021) o per supportare specifiche metodologie di apprendimento (Mageira K, et Al. 2022). Allo stesso modo intelligenze artificiali a comando vocale sono da anni utilizzate per interagire con gli utenti, come assistenti personali, invero piuttosto rudimentali, facilitando loro semplici operazione con lo smartphone o il notebook: si tratta ad esempio di Siri di Apple, o Alexa di Amazon, questi si limitano ad inviare un SMS o a facilitare chiamate telefoniche, alla “navigazione” su mappe stradali, per rispondere a semplici domande sul meteo e sulle news del giorno oppure alla ricerca di un brano musicale.

Nel corso degli ultimi anni sono, però, stati sviluppati modelli di Intelligenza artificiale che hanno mutato radicalmente lo scenario delle possibili applicazioni e conseguentemente il dibattito pubblico e su queste tecnologie. La storia degli “modelli di conversazione” basati sull’intelligenza artificiale è stata, rivoluzionata da un *new comer* nel mondo delle imprese del big tech: OpenAI. Quest’azienda è stata fondata nel dicembre 2015 a San Francisco, da un gruppo di imprenditori, tra i quali Elon Musk CEO di Space X e Tesla, Reid Hoffman, co-fondatore di LinkedIn e Peter Thiel, co-fondatore di PayPal (Musk è successivamente nel 2018, uscito dal board). La *mission* di Open AI, dalle sue origini, è quella di promuovere e sviluppare tools per l’intelligenza artificiale “e assicurare che l’intelligenza artificiale sia una bene per tutta l’umanità”, un *claim* ambizioso, forse davvero troppo benevolo, per un impresa profit dal carattere, come vedremo, molto aggressivo e dinamico. Dalle sue origini OpenAI si è concentrata sullo sviluppo di algoritmi di *deep learning* e *machine learning* per elaborare il linguaggio naturale (Natural Language Processing - NPL). Nel 2016, OpenAI ha introdotto l’architettura *Generative Pre-trained Transformer* (GPT) – letteralmente “Trasformatore Generativo Pre-istruito” – che sulla base di un primo addestramento su uno specifico set di dati genera testi, sintesi e risponde in linguaggio naturale a partire da input dell’utente. GPT agisce, cioè, individuando progressivamente e in successioni gli elementi più determinanti dell’input stesso, per questo questa tecnologia è nota anche come “attention network” (Vaswani, et. Al., 2017).

A partire dal 3 novembre 2022 è stata resa disponibile on-line la realise 3,5 di Chat GPT 3 (il 15 marzo è stato rilasciato anche Chat GPT-4) un’intelligenza artificiale che permette di generare risposte molto accurate, in linguaggio naturale, alle domande che le vengono poste in una finestra di dialogo del tutto simile a quelle che si utilizzano per una chat testuale come quella di WhatsApp.

L’architettura software che permette questo tipo nuovo interazione uomo-macchina è basata sui *Large*

Language Models. Si tratta si un sottodominio dell’Intelligenza artificiale che utilizza la *machine* e il *deep learning* applicato a enormi quantità di dati testuali per simulare - è importante ribadire questo - risposte conversazionali umane. Quando si formula una domanda o più in generale una richiesta a

ChatGPT il suo “modello di conversazione” utilizza la tecnologia delle reti neurali multistrato *transformer*² per elaborare, e analizzare il testo della domanda confrontandolo con le basi dati su cui è stato “pre-addestrato” per generare, poi, anche sulla base di fonti Web, la sua risposta in linguaggio naturale. Più in dettaglio il “pre-addestramento generativo” (*machine learning* supervisionato), ha lo scopo di fornire al sistema la capacità di prevedere secondo un modello statistico quale sarà la prossima parola nell'ambito di una frase o di un insieme di frasi immesse come prompt all'interno della sua finestra di dialogo. Il “modello di conversazione” di Chat GPT 3 contiene 17 Gigabyte di dati e fonti (mentre Chat GPT 4 ne contiene 45 Gigabyte). Inoltre, Chat GPT 3 utilizza 175 miliardi di parametri “di apprendimento” per elaborare l'input e generare, in modo probabilistico, l'output; Chat GPT 4 nel utilizza un numero molto più alto (tra i 500 e 1000 miliardi). Per ciò che riguarda i *data set* che sono stati utilizzati per il “pre-training” di Chat GPT essi possono essere classificati in 6 principali tipi di fonti: a. Riviste scientifiche, b. raccolte open di libri (Project Gutenberg, SmashWords), c. Wikipedia, d. post sul social network Reddit., e. Commons Crawl un data base di pagine archiviate dal Web a partire dal 2008, f. altri data set (Thompson, 2022). Il rilascio di ChatGPT 3 ha destato grandissimo interesse, anche tra i non addetti ai lavori, gli utenti dell'applicazione web sono stati più di 200 milioni di utenti entro i primi due mesi dal rilascio e la popolarità di ChatGPT 3 e 4, non accenna a diminuire, anzi cresce di giorno in giorno anche grazie come vedremo ai cospicui investimenti di Microsoft. La diffusione virale di questa tecnologia ha aperto un grande dibattito, pubblico globale sulle potenzialità e sui rischi correlati all'affermarsi di questa nuova generazione di “software cognitivi”. Ci occuperemo più avanti delle potenzialità di ChatGPT 3 nell'apprendimento ma indubbio che questa intelligenza artificiale a rappresenti, per l'accuratezza dei testi generati per la ampiezza dei registri comunicavi gestiti, un evento periodizzante, poiché rappresenta, come vedremo l'emergere di una nuova possibile frontiera dell'interazione uomo-macchina in campo economico, sociale, professionale, oltre che formativo.

La partnership tra Microsoft e Open Ai: assistenti cognitivi per tutti?

Che non si tratti di un'*hype* momentaneo, lo dimostra il fatto che Microsoft, già nel 2019, ha investito 1 miliardo di dollari in Open AI nel 2019, e ha rinnovato gli investimenti nel 2021 acquisendo anche una licenza esclusiva per la tecnologia alla base di Open AI. Nel gennaio del 2023, poi, le due aziende hanno sottoscritto un accordo - che comporta investimento per Microsoft di 10 miliardi di dollari - che consente a entrambe le parti di commercializzare in modo indipendente le tecnologie avanzate di IA, inoltre la collaborazione prevede tre ambiti:

- a. Microsoft aumenterà gli investimenti nello sviluppo e nella distribuzione di sistemi di *super computing* che verranno messi a disposizione della ricerca indipendente sull'IA di OpenAI. Nello stesso tempo Microsoft continuerà a sviluppare l'infrastruttura cloud *Azure* e le sue funzionalità di intelligenza artificiale per propri clienti;

² Un modello di apprendimento automatico (*deep learning*) specializzato nell'elaborazione e nell'interpretazione di dati sequenziali, che lo rende particolarmente valido per le attività di elaborazione del linguaggio naturale, costituito appunto da una sequenza di parole o sillabe.

- b. Microsoft distribuirà i modelli di OpenAI nei suoi prodotti per i consumatori e per le imprese introducendo nuove categorie di esperienze digitali basate sull'intelligenza Artificiale; Un tale accordo mette a disposizione dei clienti Microsoft il servizio cloud Azure/OpenAI. Questo

comprende la possibilità, per le aziende clienti di Microsoft di utilizzare, GPT3.5, Codex³, and DALL-E⁴ per sviluppare applicazioni che sfruttano l'intelligenza artificiale di Open AI in modo indipendente, e nel rispetto della privacy relativa ai dati aziendali e dei propri partner commerciali.

- c. Azure sarà il fornitore esclusivo di cloud di OpenAI, alimentando tutti i suoi carichi di lavoro di ricerca, prodotti e servizi.

La partnership tra OpenAI e Microsoft ha prodotto effetti, rapidissimi anche sul mercato consumer, evidentemente erano pianificati da tempo. E', infatti, il 7 del febbraio del 2023 quando l'azienda di Redmond rilascia, sul Web, l'integrazione nel suo motore di ricerca Bing, con ChatGPT 3. Da allora, infatti, l'interfaccia della nuova versione di Bing prevede una voce "chat" cliccando sulla quale compare una finestra di dialogo che specifica "chiedimi qualcosa". Bing cioè permette, anche se ora in forma ridotta⁵, alle funzionalità di ChatGPT, la versione 3 fino al 15 marzo, oggi a ChatGPT 4. Inoltre, contestualmente all'integrazione dell'Intelligenza artificiale nel motore di ricerca, il 16 marzo sempre del 2023 è stato annunciato il *Copilot* Microsoft e cioè l'integrazione delle applicazioni Cloud come ad esempio Office 365, con le potenzialità di Chat GPT 4. **L'idea è semplice e potenzialmente molto innovativa. Si tratta di** combinare la potenza dei "modelli di conversazione" basati sui *Large Language Model* dell'intelligenza artificiale con i dati presenti in Microsoft Graph⁶ e nelle app di Microsoft 365, con l'obiettivo trasformare e rendere più efficiente l'esperienza d'uso degli utenti.

Ad esempio, *Copilot* in Word può creare una prima bozza di testo su un tema specifico, ricavando i dati dal Web e da un altro documento; *Copilot* può aggiungere contenuti a documenti esistenti, riassumere un una relazione, riscrivere sezioni o l'intero documento per renderlo più conciso. In Excel *Copilot*, può, sulla base di un comando testuale identificare ricavare *tendenze specifiche* nei dati del foglio di calcolo, creare grafici e report descrittivi modo automatico. Allo stesso modo *Copilot*, affiancato a Power Point può creare una presentazione a partire da un testo scritto. Nell'applicazione di remote working Meet, *Copilot* può riassumere in tempo reale in forma scritta i dialoghi di una

³ Il "modello di conversazione" Codex è un discendente di GPT-3 che è stato addestrato sia sul linguaggio naturale che su miliardi di righe di codice. Permette di completare linee di codice, correggerle e anche generale a partire dal linguaggio naturale. È particolarmente abile in Python e competente in oltre una dozzina di linguaggi, tra cui C++, JavaScript, Go, Perl, PHP, Ruby, Swift, TypeScript, SQL e persino Shell.

⁴ DALL-E è un algoritmo di Intelligenza Artificiale, sviluppato da *OpenAI*, che è capace di generare immagini a partire da descrizioni testuali, rendendosi capace di apprendere concetti di spazio e di tempo.

⁵ Dal 14 marzo 2023 I limiti massimi giornalieri per l'interazione con ChatGPT su Bing portano a 150 i turni possibili per ogni account Microsoft e a 15 il numero massimo di domande per ogni turno. In lavorazione (non ancora disponibile) la possibilità di salvare le chat in modo da personalizzare il proprio account.

⁶

Conference Call e renderli disponibile a tutti gli i partecipanti alla riunione, sintetizzarlo o stendere una bozza del verbale. Microsoft ha poi presentato anche la funzionalità *Business Chat*. Uno strumento pensato per ottimizzare i flussi di informazione all'intero dei gruppi di lavoro nelle aziende e quindi per aumentarne la produttività. L'idea è quella di premettere di analizzare i *big data* aziendali sfruttando le potenzialità dei *Large Language Model* di OpenAI ma attraverso la mediazione degli account Microsoft e dell'infrastruttura *Cloud Azure* che garantiscono elaborazioni sicure e nel rispetto della privacy e delle policy di sicurezza "aziendali". In questo modo i manager o responsabili dei vari *team* di lavoro potrebbero ottenere risposte, in tempo, reale sullo stato di avanzamento di un progetto,

ottenere una sintesi degli scambi mail su quel progetto, così come il calendario delle riunioni e dei contatti relativi, rendendo, in questo modo più fluido e trasparente il flusso dei dati e l'avanzamento del lavoro. Stiamo probabilmente assistendo, con tutti i rischi legati al monopolio di fatto di Microsoft e OpenAI, ad una grande rivoluzione nel campo della *Office automation*, i cui tratti sono ancora tutti da analizzare e i cui contorni in termini di bilancio tra costi, benefici non posso essere ancora pienamente compresi. È chiaro però che il processo di integrazione delle nuove potenzialità dei *Large Language Models*, nei software per la produttività individuale e aziendale è destinato a progredire a ritmo molto serrato, infatti, quasi ogni giorno si assiste, in tempo reale, ad un "annuncio" relativo al rilascio di un nuovo servizio che sfrutta le potenzialità dei "tool conversazionali" di intelligenza artificiale. Ad esempio, Google ha annunciato, il 21 marzo il rilascio di dell'Intelligenza artificiale Bert⁶ a integrazione del suo motore di ricerca e Facebook sta lavorando sulla sua AI RoBERT.

ChatGpt 3: gli albori di una rivoluzione?

Ogni rivoluzione tecnologica determina una trasformazione nel modo in cui "vediamo e costruiamo il mondo" (Goodman, 1978) e anche nel caso dell'integrazione dei "modelli conversazionali" di intelligenza artificiale - come ChatGPT 3 e ChatGPT4 - all'interno dei nostri strumenti di analisi, lettura e costruzione della conoscenza, stiamo assistendo in tempo reale ad una trasformazione molto radicale dei nostri "strumenti del comunicare" (McLuhan, 1964). La rivoluzione del Web, ci ha permesso di avere a disposizione attraverso un motore di ricerca tutto il sapere del mondo, e ha cambiato radicalmente il modo in cui si genera, si comunica e si diffonde il sapere, in questo modo dopo circa cinquecento anni la Galassia Gutenberg è stata sostituita dalla "Galassia Internet" (Castells, 2001). Non sappiamo ancora se la rivoluzione in corso e che ci mette a disposizione <- nuovi e potenti strumenti come gli "assistenti cognitivi" dell'intelligenza artificiale che generano risposte in linguaggio naturale avrà una portata simile, o se rivelerà, come sembra essere il *metaverso* una promessa non pienamente mantenuta. Nemmeno possiamo prevedere a soli pochi mesi dall'emersione della "Galassia Intelligenza Artificiale" quali trasformazioni di medio e lungo periodo comporterà sui

⁶ L'AI BERT, di Google, (

nostri comportamenti di accesso, creazione e gestione alla conoscenza; tuttavia, per differenza ed appoggiandoci su alcune ricerche autorevoli (EU, 2021a, EU 2021b), possiamo formulare qualche prima considerazione generale, anche apparentemente banale, ma che possa guidare ulteriori analisi.

a. una netta discontinuità con i modelli di Intelligenza artificiale precedente.

Ai “modelli conversazionali” precedenti ai *Large Language Model* di nuova generazione come ChatGPT - per non parlare degli assistenti vocali come Alexa, Siri e Cortana - è marcata la differenza, poiché le risposte vengono generate con grande precisione e hanno una dimensione testuale ed un’efficacia comunicativa molto più alta. Inoltre, attraverso “prompt” ben costruiti si possono variare i registri linguistici (giornalistico, scientifico, narrativo, ma anche i linguaggi di programmazione informatica ecc.) e le età dei destinatari dei testi, così come si può proseguire la “conversazione” migliorando e affinando progressivamente l’output finale. *b. L’integrazione con i motori di ricerca.* È altrettanto evidente anche la differenza rispetto una ricerca tradizionale sul Web con un motore di ricerca. Google, e gli altri motori di ricerca della sua generazione, identificano sulla base del suo algoritmo i siti che contengono le “parole ricercate” sta poi all’utente scervere l’attendibilità della fonte e a partire dalla fonte produrre il testo. Nel caso degli “assistenti conversazionali” basta formulare correttamente la domanda e/o fornire all’AI la base dati da elaborare e il testo verrà generato “automaticamente”, si tratta di uno step evolutivo di grande rilievo. La prima sintesi testuale viene operata dall’IA stessa sulla base dei prompt – cioè dell’istruzione che viene fornita all’intelligenza artificiale - oppure di fonti che possono essere selezionate anche dall’utente o dall’organizzazione come nel caso delle basi dati aziendali o scolastiche o individuate dal Chat Bot sulla base del data base del Web. Questo avviene ad esempio nel caso della funzione “Chat” del motore di ricerca Bing di Microsoft integrato con ChatGPT 4 che fornisce, oltre output richiesto, i riferimenti Web da cui è tratta la base per le risposte.

c. le possibili implicazioni sul sistema sociale ed economico

Il rilascio di ChatGPT 3 oltre che un grande clamore e sui media ha generato, ad esempio le preoccupazioni per le istituzioni accademiche e educative, per la possibilità - effettivamente presente – di utilizzarlo per realizzare saggi, articoli, riassunti, in maniera quasi indistinguibile da un essere umano, con evidenti implicazioni rispetto alla possibilità di utilizzare l’intelligenza artificiale, non come supporto per l’apprendimento, ma per superare testi di ammissione o prove che implicano la produzione di testi scritti. Ci occuperemo, nel paragrafo successivo, delle potenzialità, di ChatGPT 3

- 4 nell’apprendimento e nella formazione, ma è indubbio il suo rilascio rappresenti, per l’accuratezza dei testi generati per la ampiezza dei registri comunicativi gestiti, un evento periodizzante, poiché testimonia l’emergere di una nuova frontiera dell’interazione uomo-macchina in campo economico, sociale, professionale, oltre che formativo. Concretizza la possibilità di automatizzare una parte delle attività intellettuali di tipo routinario e ripetitivo. È uno scenario che può sembrare “utopico” o “distopico”, a seconda del punto di vista, ma non è così improbabile. Per questo è necessario che vengano accuratamente valutati i rischi e benefici di questa rivoluzione tecnologica così come è necessario stabilire un serie di principi “etici” che normino il ruolo dell’Intelligenza artificiale nelle nostre vite. L’unione europea si è già mossa in questa direzione con [l’Assessment List for Trustworthy Artificial Intelligence](#), del 2020 e il successivo [Artificial intelligence Act](#) (EU 2021b). Questi

documenti si fondano sul principio di garantire un uso responsabile ed etico dell'intelligenza artificiale e sanciscono i seguenti principi: i sistemi di IA dovrebbero garantire l'empowerment delle persone e garantire adeguati meccanismi di controllo in particolare: a. la possibilità - per le donne e gli uomini – di controllare i processi e mantenerne la supervisione; b. la solidità tecnologica e la sicurezza dei sistemi; c. la garanzia della *privacy* e della gestione trasparente dei dati; d. la garanzia del rispetto delle diversità, la non discriminazione e l'equità; e. la garanzia del benessere ambientale e sociale; f. la responsabilità e l'affidabilità dei sistemi di IA e dei loro risultati.

Che cosa significano gli “assistenti cognitivi” per l'apprendimento: prime approssimazioni

Anche nel caso del campo più specifico dell'apprendimento e dell'educazione è ancora molto presto, per valutare in maniera ponderata quali potenzialità e quali implicazioni siano correlate a questa nuova frontiera tecnologica. L'affermarsi di questi “assistenti cognitivi”, mette sicuramente maestri, insegnanti, e formatori e, più in generale, le istituzioni educative e formative globali di fronte a nuove opportunità ma anche a nuove sfide. Proviamo ad analizzarle a partire da uno studio realizzato di Ilkka Tuomi, (2018) che ha redatto per la Commissione Europea un report dal significativo titolo *The Impact of Artificial Intelligence on Learning, Teaching, and Education*. Più in dettaglio il gruppo di lavoro del *Joint Research Center UE* ha analizzato, a titolo di esempio, l'impatto potenziale dell'intelligenza artificiale sui compiti che normalmente svolge un insegnante di una scuola secondaria di prima grado⁷. Come risulta chiaro dalla tabella qui sotto l'effetto dell'applicazione di strumenti come gli “agenti conversazionali” sono molto rilevanti e riguardano quasi tutti gli ambiti della professione docente, in particolare ne segnaliamo alcuni che in seguito approfondiremo:

⁷ Occorre ricordare che molti degli studi econometrici più autorevoli utilizzano come punto di partenza la banca dati statunitense O*NET contiene attualmente circa mille definizioni di attività professionali dei task di lavoro che debbono eseguire. Anche l'elenco di attività didattiche presente nella tabella è tratto da questa fonte.

Table 1: Potential impact, middle-school teacher tasks

Task	AI impact
1 Adapt teaching methods and instructional materials to meet students' varying needs and interests	High
2 Establish and enforce rules for behaviour and procedures for maintaining order among students	?
3 Confer with parents or guardians, other teachers, counsellors, and administrators to resolve students' behavioural and academic problems	Low
4 Maintain, accurate, complete, and correct students records as required by laws, district policies, and administrative regulations	High
5 Prepare, administer, and grade tests and assignments to evaluate student's progress	High
6 Prepare material and classrooms for class activities	Medium
7 Instruct through lectures, discussions, and demonstrations in one or more subjects, such as English, mathematics, or social studies	Medium
8 Establish clear objectives for all lessons, units, and projects, and communicate these objectives to students	Medium
9 Assist students who need extra help, such as by tutoring, and preparing and implementing remedial programs	High
10 Assign lessons and correct homework	High
11 Enforce all administration policies and rules governing students	Medium
...	
15 Meet or correspond with parents or guardians to discuss children's progress and to determine priorities and resource needs	Medium

- a. la possibilità di adattare le metodologie e i materiali didattici alle differenti esigenze ed interessi degli studenti;
 - b. la gestione amministrativa delle carriere degli studenti e la loro analisi;
 - c. le metodologie di valutazione dei progressi scolastici degli studenti;
 - d. l'allestimento delle aule e del materiale per le attività di classe;
 - e. lo svolgimento delle lezioni, delle spiegazioni e delle discussioni in una o più materie, in particolare in inglese, matematica o nelle discipline sociali;
-
- f. la possibilità di assistere gli studenti che necessitano di un aiuto supplementare, ad esempio attraverso il tutoraggio, la preparazione e l'implementazione di programmi di recupero;
 - g. le modalità di assegnare i compiti e di correggerli;
 - h. l'applicazione delle politiche e delle norme dell'amministrazione che regolano vita dei discenti;
 - i. la comunicazione con i genitori sui progressi degli studenti.

Ora proviamo ad ipotizzare più in concreto a quali trasformazioni andrebbe incontro una scuola italiana, che nell'ambito dell'applicazione del [Piano Scuola 4.0](#) nel 2023, volesse sperimentare gli "agenti conversazionali" come Chat GPT-4 o le applicazioni per la produttività individuale e collettiva, come Office 365, potenziate dall'intelligenza artificiale che Microsoft ha annunciato presentando 365Copilot e il suo prossimo rilascio. Proviamo, perciò, a svolgere questo esercizio

analizzando più nel dettaglio alcuni item “al alto impatto” che sono stati enumerati dal report del *Joint Research Center* del Unione Europea.

Adattare i metodi di insegnamento e i materiali didattici per soddisfare i diversi bisogni e interessi degli alunni (item a. Tuomi, 2018).

È sorprendente la capacità degli “agenti conversazionali”, come Chat GPT, di generare esercizi delle più svariate tipologia e sui temi disciplinari più disparati. Attraverso *prompt* opportunamente congegnati, l’intelligenza artificiale permette di ottenere gli esercizi e la loro risoluzioni in tempi molto rapidi. In questo modo l’insegnante può personalizzare (Dabbagh, Kitsantas, 2012). l’apprendimento in maniera semplice ed efficace e può generare set di esercizi per il recupero (punto f del report JRC) oppure per gli studenti che vogliono approfondire un determinato argomento. Paiono realizzarsi così le previsioni di David Clark nel suo *Artificial intelligence for learning: How to build intelligent systems that learn with you* (2021). Se si inseriscono *prompt* adeguati, e cioè si descrivere in modo chiaro e preciso il compito che l’intelligenza artificiale conversazionale deve svolgere, si potranno ottenere esercizi personalizzati in tempi rapidissimo. Facciamo alcuni esempi:

- “genera a partire dai concetti di spazio e di tempo che Kant delinea nella *Critica della ragion pura* quattro domande di verifica delle conoscenze”;
- “genera tre esercizi sui polinomi per la scuola superiore di primo grado”;
- “formula quattro domande sulla storia della Roma Repubblicana adatte a studenti del biennio della scuola superiore”;
- “genera tre esercizi a completamento sul *past continuos* per la classe terza della scuola superiore di primo grado”;
- “compila un programma in Basic che permetta l’accensione successiva di tre lampadine e formula quattro domande che permettano di esercitarsi sul programma”
- “compila un programma in C ++ che permetta di disegnare una parabola e proponi tre esercizi che permettano di esercitarsi su questo programma”;

Si potrebbe continuare a lungo costruendo “prompt” specifici per le differenti discipline che compongono i programmi di ogni ordine di scuola presente in Italia. Basta inserire i “prompt” esemplificativi che abbiamo appena elencato nella finestra di Dialogo di ChatGPT 3 o 4, per rendersi conto di quanto si possa migliorare la personalizzazione dei materiali e delle attività di apprendimento per adeguarle agli interessi, alle inclinazioni e alle difficoltà dei singoli studenti. Tutto questo ovviamente al netto di una adeguata formazione degli insegnanti sulle competenze digitali per l’educazione e sull’utilizzo dei sistemi di intelligenze artificiale di supporto all’apprendimento.

Progettare, somministrare e valutare i test e le esercitazioni per valutare i progressi degli studenti (item c. Tuomi, 2018).

Ci troviamo di fronte ad uno degli ambiti in cui l’intelligenza artificiale è più facilmente spendibile, già da subito, nell’ambito della pratica didattica. Abbiamo visto più sopra alcuni esempi di esercizi

per la personalizzazione dell'apprendimento. Allo stesso modo gli “agenti conversazionali” possono generare esercizi e test di verifica con estrema facilità. Non occorre, infatti, una piena digitalizzazione dell'ambiente di apprendimento della scuola, ma basta un notebook per utilizzare le AI, ad oggi disponibili gratuitamente, per generare con molta facilità test delle più svariate tipologie: a scelta multipla, a domanda aperta, di completamento, di ordinamento, vero o falso ecc. Si possono, ovviamente, ottenere dal sistema anche le soluzioni dei test, che permettono una correzione molto più rapida ed efficace delle verifiche proposte ad allievi e studenti. Si potranno in questo modo preparare più velocemente le verifiche ma, e questo è un aspetto più interessante, si potrà proporle agli studenti più facilmente e frequentemente, anche alla fine di ogni attività in classe. Un maggior numero di valutazioni di processo permetterà di mantenere più allineate le competenze degli allievi e degli studenti di una classe e in prospettiva di diminuire gli “abbandoni”. E' evidente anche, ad là del lavoro dei singoli insegnanti e ad un livello più di sistema - ci riferiamo a livello di plesso o di scuola - come in presenza di un LMS o di VLE che raccolga buona parte delle interazioni e delle attività e i compiti assegnati ad un allievo durante la sua carriera scolastica, l'intelligenza artificiale, debitamente programmata, potrà analizzare i *data set* degli alunni (in altre parole i *Learning Analytics*, Ferri, 2019) anche in funzione pro-attiva e cioè aiutare gli insegnanti a comprendere meglio sulla base dei dati (approccio *data driven*) le necessità e i bisogni dei singoli studenti, premettendo al corpo docente di modulare meglio la progettazione didattica. Infine, l'intelligenza artificiale potrebbe automatizzare e velocizzando gli adempimenti necessari per gli scrutini finali oltre all'archiviazione annuale delle carriere degli studenti (item. b., Tuomi, 2018).

Preparare materiali da utilizzare in classe (item d., Tuomi, 2018).

Gli “agenti conversazionali” basati sull'intelligenza artificiale, oltre alla semplice ricerca on-line, nella quale non sono all'altezza o non si differenziano molto da Google, possono fornire idee su come strutturare la progettazione didattica di unità di apprendimento o di singole attività, sulla base ad esempio di metodologie consolidate quali il *Conversational framework* (Laurillard, 2012) o la “tassonomia di Bloom” (Anderson, Krathwohl, Bloom, 2001). Allo stesso modo possono fornire spunti creativi agli insegnanti per la progettazione di una didattica più laboratoriale, come vuole il PNRR, e di attività cooperative come Webquest, Fieldtrip e giochi didattici da condurre in classe. Esempi di *prompt* potrebbero essere in questo caso; “crea una Webquest, individuando i siti da cui partire, sul movimento dei Macchiaioli” oppure “genera un gioco didattico attraverso il quale gli allievi suddivisi per gruppi possano apprendere il ciclo dell'acqua”. Se si inseriscono nella finestra di dialogo di Chat GPT questi *prompt* l'intelligenza artificiale fornisce istruzioni dettagliate, siti da consultare e modalità di svolgimento e valutazione dell'attività. Inoltre, gli “agenti conversazionali” possono essere utilizzati per coinvolgere gli studenti e supportare l'insegnante ad esempio nella gestione di una lezione partecipata o di un *debate* (prompt: “Crea lo schema di un *debate* da condurre in classe sul tema della procreazione assistita”) in questo caso, come nei precedenti, l'intelligenza artificiale sulla base del prompt fornito serve ad avviare la discussione sviluppando una serie di argomenti “a favore” e “un serie di argomenti contrari” ed offre spunti per gestire la discussione. L'IA nella migliore delle ipotesi potrebbe, perciò, configurarsi come assistente - “omnisciente” ma non pensante, per fornire un supporto alla creatività del docente.

Permettere un tutoraggio specifico mediante materiali, discussioni e dimostrazioni in una o più materie (item 7, Tuomi, 2018).

Anche questo è un campo di applicazione molto interessante anche se non è possibile ancora parlare di “tutor disciplinari specifici”, se non pionieristici. L’insegnante può predisporre utilizzando ad esempio le potenzialità di Microsoft Graph e dell’annuncio Microsoft *Copilot* un base dati di testi e attività ed esercitazioni (nelle lingue, in matematica e fisica, o nella programmazione, ma anche nelle discipline umanistiche) che potrebbero essere rese disponibili agli studenti l’account della scuola e della classe e sui quali gli studenti possano esercitarsi ed avere in tempo reale o differito, a seconda delle scelte dell’insegnante, una correzione “autorevole”, basata cioè sui data sets proposti dall’insegnante o dalla scuola, e non semplicemente sui dati presenti o ricavati dal Web dall’intelligenza artificiale. “Modelli conversazionali” di questo tipo erano stati già realizzati prima del comparire degli attuali Chat Bot basati sulla tecnologia dei *Large Language Model*. Possiamo citare ad esempio, *Deep-speare*, un modello di *machine learning* addestrato su 2700 sonetti del bardo che genera componimenti poetici che molti lettori non sanno distinguere dalle composizioni del poeta (Lau, 2018). Oppure ancora la “base dati conversazionale” dedicata a *La poetessa saffo* realizzata da Alessandro Iannella e disponibile a questo link <https://www.sappho.education/> (Iannella, Labruna, Santercole., Viti, 2021 pp. 120-151; Trentin, 2021). O ancora gli esperimenti sul campo con un chatbot AI e che fornisce approfondimenti per la metodologia CLIL. In particolare, ricordiamo il caso d’uso sperimentale di un chatbot educativo HAI chiamato AsasaraBot, progettato per insegnare agli studenti delle scuole superiori contenuti culturali in una lingua straniera, cioè l’inglese o il francese. L’analisi del funzionamento di questi Chat Bot, che avevano un *Data Set* molto più limitato degli attuali, può essere utile per progettare quelli futuri. Nel caso delle lingue straniere, poi, alcune imprese private già rendono disponibili tool on-line di tutoring basati sull’intelligenza artificiale per fare pratica di conversazione, integrando anche tool di correzione automatica.

Definire obiettivi di apprendimento chiari per tutte le lezioni, le unità e i progetti e comunicare questi obiettivi agli studenti (item h, Tuomi, 2018).

I “modelli conversazionali” ad oggi disponibili permettono sulla base della descrizione dei contenuti di un Corso universitario o del programma Ministeriale di una materia di definire gli out-come dei Risultati di Apprendimento Attesi (RAA) sulla base di modelli consolidati di definizioni di questi out come quali ad esempi i “descrittori di Dublino”⁸ o ad esempi del modello di modellazione

didattica quali, la già citata, tassonomia di Bloom (Anderson, Krathwohl, Bloom, 2001), agevolando in questo modo il lavoro di progettazione e programmazione degli insegnanti e dei maestri e una più concreta ed efficiente stesura del PTOF.

⁸ I Descrittori di Dublino sono un sistema di classificazione utilizzato per descrivere le competenze acquisite dagli studenti di terzo livello in Europa. Sono stati sviluppati nel 2004 dall’Università di Dublino, in Irlanda, e sono stati adottati come standard europeo per la descrizione delle competenze degli studenti di terzo livello (università, scuole di alta formazione ecc.).

Conclusioni

In questo contributo abbiamo provato tracciare una prima mappa “euristica” del possibile uso didattico dei *Large Language Model*, in realtà solo dei primi che abbiamo a disposizione cioè Chat GPT 3 e 4. Si tratta ovviamente di un contributo parziale e sicuramente incompleto che proponiamo all’ampio dibattito in atto su questo tema. Siamo consapevoli che l’*hype* e il clamore che si è generato sui media rispetto a questo tipo di tecnologie, insieme all’oggettiva mancanza di sperimentazioni validate sul campo - a scuola e nei contesti formativi - impedisce la giusta distanza critica ed una valutazione obiettiva dell’efficacia di questi strumenti. Il nostro intento è quello avviare una ricerca che abbia come obiettivo quello di comprendere se e come di “agenti conversazionali” di nuova generazione possano migliorare le strategie di apprendimento/insegnamento nelle nostre scuole e università. Ad una prima analisi possiamo ipotizzare due possibili conclusioni provvisorie:

- a. i campi più promettenti paiono quelli della personalizzazione, della valutazione, della didattica adattiva, del tutoraggio e del feedback personalizzato e in prospettiva quello della progettazione didattica *evidence based*.
- b. In secondo luogo ma è forse il tema più importante, risulta chiaro che le attività che si prestano maggiormente ad essere “assistite” se non “automatizzate” dall’intelligenza artificiale sono quelle più ripetitive e procedurali, se questo fosse vero, si potrebbe liberare una porzione, tutta da quantificare, del lavoro degli insegnanti che potrebbe in questo modo essere dedicata alla relazione educativa, alla motivazione e al supporto agli studenti, oltre che alla progettazione didattica e questo potrebbe davvero migliorare i sistemi di istruzione contemporanei.

Per il futuro si tratta di sviluppare come voleva Galileo “sensate esperienze e certe dimostrazioni” che per mettano di comprendere meglio i fenomeni e le ipotesi di lavoro che abbiamo presentato in questo contributo.

Bibliografia

- Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. London, Longman.
- Castells, M., (2001), *The Internet Galaxy. Reflections on the Internet, Business, and Society*, New York, Oxford University Press, 2001., trad. it., *Galassia Internet*, Feltrinelli, 2002.
- Clark, D. (2020). *Artificial intelligence for learning: How to build intelligent systems that learn with you*. London: Kogan Page Publishers.
- Dabbagh N., Kitsantas A. (2012). “Personal learning environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning”. *The Internet and Higher Education*, 15(1), 3-8.
- Gwo-Jen Hwang & Ching-Yi Chang (2021) “A review of opportunities and challenges of chatbots in education”, *Interactive Learning Environments*

Goodman, N., (1965) *Ways of Worldmaking*, Indianapolis, *Hackett Pub. Co.*, trad. it. *Vedere e costruire il mondo*, Laterza, Roma-Bari, 1988.

Eu, (2020), *The assesment list for trustworthy artificial intelligence (ALTAI)* disponibile al sito <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal> Eu, (2021a), *Coordinated Plan on Artificial Intelligence 2021 Review*, disponibile al sito <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review> EU (2021b), *EU Artificial Intelligence ACT*, disponibile al sito <https://artificialintelligenceact.eu/the-act/>

Ferri, P. (2019). “MOOC, didattica universitaria digitale e Learning analytics. Opportunità e prospettive”. *Giornale italiano della ricerca educativa*, Numero Speciale settembre 2019, 13-26.
Iannella A., Labrunda T., Santercole T., Viti, B., (2021), “Reclaiming Conversation: introducing a novel approach to using conversational AI at school”, pp. 120-151, in Trentin G., (2021), eds, *Conversational Agents as Online Learning Tutors*, New York, Nova Science Pub Inc Kai-Yu Tang, Ching-Yi Chang & Gwo-Jen Hwang (2021) Trends in artificial intelligence-supported e-learning: a systematic review and co-citation network analysis (1998–2019), *Interactive Learning Environments*

Laurillard, D., *Teaching as a Design Science. Building Pedagogical Patterns for Learning and Technology*, New York-London, Routledge; tr. it. *Insegnamento come scienza della progettazione. Costruire modelli pedagogici per apprendere con le tecnologie*, Milano, Franco Angeli, 2015.

Lau H. Y., Cohn T., Baldwin T., Brooke J., Hammond A., (2018). “Deep-speare: A joint neural model of poetic language, meter and rhyme”. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1948–1958, Melbourne, Association for Computational Linguistics.

Siamo nell'era post-Turing Test ed è epocale: come ci siamo arrivati, su cosa interrogarsi

Per chi è legato alla storia dell'IA risulta bizzarro che il test di Turing venga superato senza le dovute celebrazioni e con la tacita accettazione che le macchine siano in grado di carpire la facoltà umana da molti ritenuta più preziosa, il linguaggio. Ma ora si aprono nuovi interrogativi

Di **Alessio Plebe**, Università degli Studi di Messina

Siamo entrati nell'era post-**Turing test**. Forse molti troveranno questa domanda azzardata, se non addirittura sbagliata. Allora, prima ancora di provare a giustificarla, mettiamo le mani avanti segnalando come **non siamo i primi a sancire il superamento definitivo del test di Turing**.

Abbiamo almeno la compagnia di ben 444 studiosi provenienti da 132 diverse istituzioni di ricerca, tanti sono **gli autori di un recente articolo** intitolato proprio **Beyond the Imitation Game** (Srivastava et al. 2022).

Assodato il superamento del test di Turing, questa impressionante compagine di autori ha cercato di istituire **un'alternativa** in grado non più di dire banalmente se un computer sia intelligente, ma di **misurarne in profondità il livello di competenza** su compiti che spaziano l'intero scibile umano.

Occorre dire che, oltre questo lavoro appena citato, non pare essere stato scritto molto sull'avvenuto superamento del test di Turing, non tanto quanto ci si poteva aspettare da **un evento talmente eclatante**. Si proverà a ragionare sui motivi, ma anzitutto è doveroso spiegare perché entrare nell'era post-Turing test sia qualcosa di epocale.

L'operazionalizzazione dell'intelligenza

The Imitation Game è il titolo del film di Morten Tyldum del 2014, che ha fatto conoscere la figura di Turing al grande pubblico, concentrandosi però più sulle vicende legate alla seconda guerra mondiale che su questo geniale gioco apparentemente semplice ed innocuo. Proposto **nel celebre articolo Computing Machinery and Intelligence** (Turing 1950), affronta l'interrogativo se un computer possa essere ritenuto intelligente. Turing va al cuore dell'obiezione che potrebbe essere sollevata qualunque sia il compito sofisticato, chiamiamolo X, che un computer dimostra di svolgere: "d'accordo, questo computer sa svolgere il compito X, ma non sbilanciamoci, per *intelligenza* si intende ben altro". Cosa sia il "ben altro" si può prestare ad infinite ridefinizioni tali da conservare sempre l'intelligenza nello scrigno delle proprietà unicamente umane.

Per sbarazzarsi del comodo scudo offerto dalla vaghezza del termine "intelligenza", Turing adotta una strategia ben nota in filosofia della scienza con il nome di **operazionalizzazione**. Introdotta dal premio Nobel per la fisica Percy (Bridgman 1927), consente di **attribuire un significato rigoroso a termini di per se vaghi ed ambigui**. Per Bridgman il significato di un termine può essere interamente specificato da un insieme di *operazioni* che lo coinvolgono. Questo procedimento è pienamente adottato in fisica, dove termini come forza, massa, carica elettrica, trovano specificazione in una serie di operazioni in cui sono coinvolti in modo non ambiguo. **Turing fa lo stesso con il termine "intelligenza"**, lo equivale ad una operazione, che consiste proprio nell'*Imitation Game*. **Un computer risulta degno della proprietà dell'intelligenza se un interlocutore umano conversandoci in modo remoto riguardo qualunque argomento, non sa dire se si tratta di una macchina o di un conversante in carne ed ossa**. Fin qui l'elegante strategia dell'operazionalizzazione, poi arriva la provocazione più scandalosa: Turing ci crede davvero che verrà un futuro in cui i computer diventeranno intelligenti e lo dimostreranno sapendo conversare come noi, e lo sostiene con una serie di argomenti teorici di notevole spessore.

Per meglio comprendere **la sorpresa per una transizione tutto sommato indolore all'era postTuring Test**, è opportuno esaminare la portata della sua provocazione. Nel prendere la capacità di conversare come marchio dell'eventuale intelligenza di una macchina, Turing aveva mirato dritto alla facoltà che per molti è la più sublime e unica della specie umana: il **linguaggio**. Ne sono convinti, tanto per fare un paio di celebri esempi, Cartesio e Noam (Chomsky 1966). **Non pochi filosofi hanno sottolineato come il linguaggio non sia solamente un modo di comunicare, ma l'impianto con cui è costruita la nostra visione del mondo**. Ludwig (Wittgenstein 1922) affermava che "i limiti del mio linguaggio significano i limiti del mio mondo". Immaginare che un computer posseda la facoltà del linguaggio, come aveva spudoratamente fatto Turing, voleva quindi dire **profanare il santuario della mente umana**.

Il test di Turing come benzina per fiammate intellettuali

La sfida lanciata di Turing ha provocato uno dei più ampi ed accesi dibattiti filosofici del secolo scorso, non sopiti nemmeno oggi. Ne è ovviamente responsabile l'intera impresa dell'Intelligenza Artificiale (AI), che coglieva il guanto lanciato da Turing giusto un anno dopo la sua scomparsa (McCarthy et al. 1955). Non furono necessari troppi anni perché si levassero alte voci di reazione. Il filosofo **Hubert Dreyfus** ha speso l'intera sua vita a **trovare ragioni per negare al computer la possibilità di raggiungere mai capacità umane** come il linguaggio. In uno dei suoi più famosi scritti, **What Computers Can't Do: A Critique of Artificial Reason** (Dreyfus 1972), la parte centrale mira proprio a ridicolizzare i tentativi pionieristici dell'IA nel campo del linguaggio, in particolare la traduzione automatica – a quell'epoca davvero fallimentari – concludendo che mai il linguaggio umano sarà alla portata di un computer.

Nella seconda metà del secolo scorso si forma un nuovo ambito di studi che, al contrario, nutre simpatie nei confronti dell'idea di Turing e dell'IA in generale, si tratta della scienza cognitiva. La sua prospettiva inquadra tutti i fenomeni mentali – dalla percezione visiva al linguaggio e la pianificazione delle azioni – come computazioni, e quindi riconducibili ad algoritmi. Se le cose stanno davvero così la possibilità che anche un computer possa eseguire lo stesso genere di algoritmi è aperta, includendo quindi la sua possibilità di acquisire completamente il linguaggio naturale. Per un certo periodo **l'intreccio tra scienze cognitive e IA fu davvero stretto**, tanto che il filosofo Hilary (Putnam 1960) era arrivato a sostenere che la mente umana funzionasse – letteralmente – come una macchina di Turing. Tuttavia ben presto diversi scienziati cognitivi misero le mani avanti, precisando che **la natura computazionale dei processi cognitivi umani non garantisce affatto che un computer possa acquisirli**. Chomsky è stato esemplare, così come ha avuto gran simpatia per le spiegazioni computazionali della facoltà linguistica umana, perseguendole in prima persona, altrettanto lo irritavano le velleità dell'IA, da cui si è tenuto sempre a distanza.

Eliza, il software che impersona uno psicoterapeuta computerizzato

Agli albori dell'IA sia critiche che simpatie erano soprattutto in linea di principio, ma nel 1966 lo scandalo di Turing pare prendere una prima forma concreta. **Eliza** è il software che impersona uno psicoterapeuta computerizzato (Weizenbaum 1966) in grado di fornire brevi risposte abbastanza credibili ad interlocutori umani. **La genialità era di impersonare uno psicoterapeuta di scuola rogeriana, la cui strategia è grosso modo di portare sempre il paziente ad autointerrogarsi senza sbilanciarsi mai troppo**. In questo modo usando poche parole delle frasi scritte dagli

interlocutori, Eliza può formulare blande e generiche controdomande, senza doversi preoccupare di alcuno dei dettagli sulla situazione effettiva del suo paziente, tantomeno dei riferimenti nel mondo delle parole lette e prodotte. Nonostante la sua palese limitazione, **Eliza fu un successo di pubblico come pochi altri nella storia del software**, e proprio questo destò preoccupazioni.

Per Ned (Block 1981) **c'era qualcosa di sbagliato, di immorale, nel lasciar credere alla gente che potesse esserci qualche forma di intelligenza dietro ad un astuto gioco di rimaneggiamento di pochi simboli**. Block andò anche oltre il caso specifico di Eliza, sostenendo che anche programmi più potenti non avrebbero potuto arrogarsi la proprietà dell'intelligenza a pieno titolo.

La stanza cinese

Su questo terreno però fu un altro argomento, proposto dal suo collega John (Searle 1980), a diventare **la lama più affilata puntata contro il test di Turing**. Divenuta celebre sotto il nome di “stanza cinese”, la situazione immaginaria proposta da Searle vede se stesso chiuso in una stanzetta, a cui vengono passati fogli con scritte in cinese. Vi è a disposizione un meraviglioso manuale, da usare cercando nella pagina a sinistra un testo identico a quello appena ricevuto, basta poi ricopiare quel che si legge nella pagina a destra, e consegnarlo a chi sta fuori dalla stanza. Bene, i fogli in entrata sono frasi di un interlocutore, e quelle in uscita le risposte, la stanza chiusa funziona quindi come una macchina in grado di conversare perfettamente, idonea quindi a superare il test di Turing, con il piccolo particolare, osserva Searle, che lui non sa nemmeno una parola di cinese. Ovvero **un software potrebbe rispondere meccanicamente senza aver capito nulla**.

Diversi trovarono l'espedito di Searle tanto convincente da considerare la possibilità per un computer di comprendere il linguaggio un traguardo irraggiungibile. Non tutti, e se Eliza poteva essere tacciata di impostura, lo era in modo ben più subdolo la “stanza cinese”, per esempio secondo Daniel (Dennett 1980). L'impostura è di condurre il lettore in una immaginazione, falsa, costellata di passaggi impossibili, per poi convincerlo della tesi che vuol sostenere, e Dennett coniò un nome specifico per queste raffinate imposture: **intuition pump** (Dennett 2012). Il passaggio platealmente impossibile è il manuale, che in teoria, per qualunque contenuto di una conversazione, dovrebbe avere la risposta pronta e appropriata da fornire. Ovviamente non è realisticamente realizzabile un tale manuale, data l'infinita possibilità di costrutti linguistici in una conversazione. Un altro passaggio dubbio riguarda cosa significhi “comprendere”, per una buona porzione di filosofi del linguaggio, a partire da Wittgenstein, comprendere il linguaggio è fondamentalmente saperlo usare. Criterio a cui sia il test di Turing che la “stanza cinese” sono adeguati. Searle richiede qualcosa in più, **una consapevolezza della conversazione in corso**, che come si cerca di meglio precisare precipita in uno dei temi filosoficamente più spinosi: cosa sia la coscienza.

Sicuramente la “stanza cinese” esercita una formidabile forza di convinzione, ed ha avuto un notevole successo, tenendo banco per un ventennio, con interi libri dedicati al suo dibattito (Preston and Bishop 2002). Il tema perse di mordente sul passare del millennio, sia per un naturale esaurirsi degli argomenti, ma soprattutto perché **non si intravedeva nulla di concreto nell'IA in grado di alimentare la discussione**. Dall'epoca eroica di Eliza, la ricerca sulla comprensione artificiale del linguaggio aveva intrapreso una strada molto diversa, che non voleva più lasciarsi tentare da trucchi e stratagemmi. Si è tentato di riversare nei computer le teorie linguistiche del funzionamento del linguaggio umano, soprattutto quelle di ispirazione chomskiana che nascono già parzialmente

formalizzate (Winograd 1972). Pur essendo teorie di grande portata nel descrivere la miriade di intricati fenomeni delle lingue umane, si rivelarono sostanzialmente fallimentari nel dotare i computer di capacità linguistiche nemmeno lontanamente paragonabili a quelle umane.

L'avvento del Deep Learning

L'intera IA aveva un po' seguito il destino degli insuccessi nell'elaborazione artificiale del linguaggio, toccando uno dei suoi punti di declino più bassi una decina di anni fa. La situazione si è notoriamente ribaltata a seguito del *Deep Learning* (DL), **la versione rimodernata delle reti neurali artificiali degli anni '80** (Rumelhart and McClelland 1986).

Il DL con i suoi rapidi e inaspettati successi ha rilanciato l'agenda per il futuro della mente artificiale (Plebe and Perconti 2022), ma il suo impatto è stato più fulmineo e capillare per la visione e tutti gli ambiti applicativi ad essa legati, il linguaggio è arrivato più tardi. Le reti neurali artificiali per lungo tempo hanno sofferto la loro intrinseca refrattarietà verso il linguaggio. Uno dei problemi consiste proprio nella natura dei costituenti delle reti neurali, vettori di numeri reali, e delle loro operazioni, appartenenti all'algebra lineare, che mal si adattano alla natura simbolica delle parole. Occorre aspettare il 2013 per una soluzione brillante ed efficace, nota come **word embedding** (Mikolov et al. 2013), tramite cui le parole vengono trasformate in vettori di numeri reali, appresi da corpora di esempi di una lingua. I vettori, inizialmente casuali, vengono progressivamente modificati in modo da essere simili quando circondati da parole simili, anch'esse sottoforma di vettori. Il secondo problema, ancor più grave, riguarda **il funzionamento statico della formulazione classica delle reti neurali artificiali**, che contrasta con la natura fortemente dinamica del linguaggio, in cui il significato di una frase si va a formare dalla continua successione delle parole, incluse quelle di frasi precedenti. Un'importante soluzione era stata trovata da Jeffrey (Elman 1990) con l'idea di reti neurali ricorsive, dotate di una forma di memoria delle parole che via via venivano recepite. Nonostante vari successivi perfezionamenti (Hochreiter and Schmidhuber 1997) le reti ricorsive entravano in difficoltà non appena le frasi diventavano lunghe e articolate in modo complicato, come succede ordinariamente nel pieno uso del linguaggio.

La svolta epocale avviene nel 2017, con l'invenzione da parte del team di Google Brain dell'architettura denominata Transformer (Vaswani et al. 2017). Coerente con il principio rigorosamente empirista del DL, il *Transformer* non tenta di implementare direttamente nessuna regola linguistica, è dotato invece di un meccanismo per apprendere quando e in che misura mantenere relazioni tra le varie parole che compaiono in una frase, e anche in frasi adiacenti. Questa architettura si è subito dimostrata vincente, ed è diventata la base per quelli che oggi vengono chiamati *language model* o anche *foundation model* (Bommasani et al. 2021), modelli neurali profondi dotati di meccanismo *Transformer*, addestrati su grandi corpora linguistici.

Dal ritorno delle vecchie critiche ad una tacita accettazione

Improvvisamente e inaspettatamente la prospettiva del computer in grado di carpire la facoltà umana da molti ritenuta più preziosa, il linguaggio, stava diventando reale. Per alcuni una prospettiva affascinante, ma per altri, comprensibilmente, spazzante se non addirittura terrificante. Non sorprende pertanto che **tra il 2018 e il 2021 si sia levato un vasto coro di critiche in vario**

modo convergenti nel voler negare, ancora una volta, la possibilità per un computer di comprendere realmente il linguaggio umano. Oltre a detrattori dell'IA in generale (Bishop 2021; Larson 2021), hanno fatto parte di questo coro anche studiosi di grande fama e vicini all'IA (Pearl and Mackenzie 2018; Landgrebe and Smith 2019; Marcus and Davis 2019). Questa ondata di critiche ha avuto risonanza e accoglienza. Per esempio uno dei lavori di maggior successo (Bender and Koller 2020) drastico nell'affermare che i modelli del genere *Transformer* non possono in linea di principio imparare nessun significato linguistico, ha ottenuto il premio come miglior lavoro al convegno annuale della *Association for Computational Linguistics* e ad oggi vanta quasi 500 citazioni. Da un'analisi puntuale di questo corpo di critiche (Perconti and Plebe 2022) emerge un progresso modesto rispetto agli argomenti già formulati nel periodo di intenso dibattito nel secolo scorso, raccontato prima.

L'esperimento del polipo

Si tratta soprattutto di riproposizione rfrasata dei vecchi argomenti. Per esempio (Bender and Koller 2020), nel ritornare a negare in linea di principio la possibilità per un computer di acquisire un linguaggio umano, come fece più di 40 anni fa Searle, inventano un loro esperimento mentale sulla falsariga della “stanza cinese”. Stavolta protagonista non è lei in persona, ma un **polipo**. L'animale se ne sta in fondo al mare, vicino ad un cavo di comunicazione attraverso cui conversano due parlanti inglesi, sfortunatamente naufragati su due lontane isole, per fortuna ben dotate di interconnessione. Il polipo ha curiosità per le telecomunicazioni, e quindi impara in fretta come gli impulsi elettrici che transitano nel cavo in una direzione siano seguiti da altri treni di impulsi nell'altra direzione. Ad un certo punto ha acquisito una tale confidenza con le sequenze di impulsi da decidere di tagliare il cavo, e provare lui stesso a mandare segnali in risposta ad uno dei due naufraghi. Il quale, dice Bender, potrebbe anche continuare a credere che dall'altra parte del cavo ci sia il suo sventurato amico, per quanto se la cava bene il polipo, che però non capisce nulla della conversazione in corso. E' evidente quanto il polipo della storia sia ancor più improbabile del manuale nella “stanza cinese”, ed è ben difficile sia destinato ad altrettanta fortuna.

Il tentativo di emulazione di Searle da parte di Bender è unico, la strategia più diffusa, e ben presente anche nei lavori di Bender, è invece la **ricerca minuziosa di casi di conversazioni in cui il computer fornisce risposte sbagliate**, soprattutto quando si tratta di errori che difficilmente farebbe un parlante umano. Di per se si tratta dell'ordinaria e fondamentale attività di esplorazione e verifica delle capacità dei vari modelli del linguaggio, in cui è importante individuare, più delle risposte corrette, quelle sbagliate. Si trasforma invece in retorica ideologica quando i casi di errore vengono cristallizzati come testimonianze inoppugnabili del fallimento complessivo nel dotare un computer del linguaggio umano. E' la pratica che il linguista computazionale Samuel (Bowman 2022) ha chiamato **the dangers of underclaiming**, l'esibizione poco scientifica di casi di errore allo scopo di denigrare i modelli del linguaggio. Il fenomeno peggiore è che molte di queste critiche usano come supporto critiche precedenti, citate sempre come fatti empirici consolidati. In questo modo continuano ad essere presentati come errori casi ampiamente superati dal continuo e rapido progresso di questi modelli.

L'ingresso nell'era post-Turing test

Oggi è diventato arduo offrire evidenze empiriche dell'impaccio dei sistemi artificiali con il **linguaggio umano**, oramai la competenza di un parlante comune è quotidianamente replicata da agenti conversazionali, in poco più di un anno si è entrati in pieno **nell'era post-Turing test**. Una inusuale avvisaglia si era avuta a giugno 2022, in un articolo del Washington Post. Un ingegnere di Google, Blake Lemoine, esternava la sua convinzione che un nuovo programma chiamato **LaMDA** (Language Model for Dialogue Applications) (Thoppilan et al. 2022), di cui lui era incaricato di eseguire verifiche, era senziente e dotato di una propria personalità. Nel 2021 LaMDA era in fase avanzata di verifica interna, e Lemoine nel continuare a conversarci era talmente colpito dalla qualità e spontaneità del dialogo da convincersi di avere a che fare con un'entità cosciente. Ma la quasi totalità dei commenti provocati dalla sua intervista, indipendentemente dal livello scientifico del commentatore, furono negativi, rifiutando la sua ipotesi come falsa ed ingenua. Lemoine fu persino licenziato da Google.

Mentre LaMDA e altri sistemi di dialogo basati sui modelli del linguaggio sono rimasti noti solo al mondo della ricerca, **chatGPT ha esposto al mondo intero come oggi il computer sappia conversare**. Si è trattato di una sorta di implicito Turing test corale. Beninteso, non si tratta più di doversi fingere o meno umani, anzi, per mettersi al riparo da situazioni imbarazzanti come per LaMDA, **chatGPT è sempre pronto a mettere le mani avanti dichiarandosi un agente IA**, pertanto privo di materialità e incapace di interazioni fisiche con il mondo. Si è oramai andati oltre, chatGPT deve il suo apprezzamento al saper conversare esattamente come un umano, anzi, come un umano dalla cultura prodigiosa. Ha stabilito **il record assoluto di prodotto digitale con la più rapida crescita di adesioni**, avendo raggiunto un milione di utenti dopo solo una settimana di attività dal sul lancio a fine novembre 2022, e attualmente riceve 300 milioni di visite al mese.

Sembrerebbe proprio che **l'umanità, o quantomeno una fetta di umanità che presta attenzione ed è incuriosita dall'artificiale, non percepisca un conflitto culturale all'idea che a possedere il linguaggio, oltre a noi umani, oggi ci siano anche alcuni software**. Occorrerà tempo per poter comprendere meglio in che misura si sia disposti ad accettare una prospettiva così distante dal senso comune dominante. I primissimi dati empirici finora raccolti sembrano mostrare **un atteggiamento di accettazione** del nuovo interlocutore non più in carne ed ossa. Secondo un primo studio sulle sensazioni degli utenti chatGPT basato su interazioni Twitter (Haque et al. 2022), un 80% degli utenti attribuiscono una genuina intelligenza al sistema, e oltre il 90% trova il conversarci un ottimo intrattenimento e uno stimolo per la creatività personale.

Conclusioni

Per chi è legato alla storia dell'IA risulta bizzarro che il celebre test di Turing venga superato senza le dovute celebrazioni. Non molti anni fa, nel 2014, un software denominato Eugene Goostman venne accreditato di aver superato il test di Turing in una sessione organizzata alla Royal Society di Londra, con molto scalpore e qualcuno poco convinto (Warwick and Shah 2016). Il "trucco" era che Eugene Goostman voleva essere il nome di un ragazzo ucraino di 13 anni, e pertanto chi colloquiava era più magnanimo nell'accettare risposte poco convincenti, considerandole di un ragazzo. Forse proprio perché oggi non può esservi controversia, viene a cadere l'utilità di decretare pomposamente il superamento del test di Turing. Analogamente, risalta in questo nuovo panorama il declino dell'ondata di critiche montata non più di tre – quattro anni fa, pur essendo ora il momento della difesa più strenua dell'unicità umana del linguaggio. **Probabilmente è diventata una difesa**

più disperata che strenua. La strategia degli errori eclatanti da esibire si è oramai esaurita. Per andare a indagare la padronanza del linguaggio di chatGPT e simili, occorrono ora poderosi benchmark del genere di *BIG-bench* (Srivastava et al. 2022), con cui si è iniziato questo articolo. Si è dell'avviso che l'era post-Turing ponga un genere di interrogativo diverso, su quali siano le ragioni per cui le reti neurali basate sull'architettura Transformer siano diventate in grado di conversare alla stregua di una umano.

Non è un interrogativo del tutto nuovo, anche le ragioni per cui il DL in generale sia talmente efficiente in un'ampia varietà di applicazioni, non sono mai emerse in modo convincente (Plebe and Grasso 2019). Nel caso del linguaggio il mistero si infittisce, per via dell'abissale salto di prestazioni tra le tante architetture neurali sperimentate prima del 2017, e le successive basate sul Transformer. La sua innovazione riguarda la flessibilità nel mantenere o allentare memoria di relazioni intercorse con parole, nel processare quella corrente, e parrebbe ben conciliarsi con recenti tendenze in neurobiologia del linguaggio a focalizzare l'attenzione sui meccanismi di memorizzazione (Roger et al. 2022). Beninteso, spingersi ad ipotizzare qualche similarità tra il modo come il Transformer mantiene memoria delle relazioni tra parole, e corrispondenti meccanismi cerebrali, sarebbe al momento pura illazione. Si tratta, al massimo, di stimoli di riflessione per l'era post-Turing.

Bibliografia

Bender, Emily M., and Alexander Koller. 2020. "Climbing Towards Nlu: On Meaning, Form, and Understanding in the Age of Data." In *58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198. Somerset (NJ): Association for Computational Linguistics.

Bishop, J. Mark. 2021. "Artificial Intelligence Is Stupid and Causal Reasoning Will Not Fix It." *Frontiers in Psychology* 11: 513474.

Block, Ned. 1981. "Psychologism and Behaviorism." *Philosophical Review* 90: 5–43.

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, et al. 2021. "On the Opportunities and Risks of Foundation Models." *arXiv abs/2108.07258*.

Bowman, Samuel R. 2022. "The Dangers of Underclaiming: Reasons for Caution When Reporting How Nlp Systems Fail." In *Proceedings of the 60th Meeting of the Association for Computational Linguistics*, 1:7484–99. Association for Computational Linguistics.

Bridgman, Percy W. 1927. *The Logic of Modern Physics*. New York: Macmillan.

Chomsky, Noam. 1966. *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. New York: Harper; Row Pub. Inc.

Dennett, Daniel C. 1980. "The Milk of Human Intentionality." *Behavioral and Brain Science* 3: 429–30.

———. 2012. *Intuition Pumps and Other Tools for Thinking*. New York: W. W. Norton & Company.

Dreyfus, Hubert. 1972. *What Computers Can't Do: A Critique of Artificial Reason*. New York: Harper; Row Pub. Inc.

Elman, Jeffrey L. 1990. "Finding Structure in Time." *Cognitive Science* 14: 179–221.

Haque, Mubin Ul, Isuru Dharmadasa, Zarrin Tasnim Sworna, Roshan Namal Rajapakse, and Hussain Ahmad. 2022. "I Think This Is the Most Disruptive Technology: Exploring Sentiments of Chatgpt Early Adopters Using Twitter Data." *arXiv* abs/2212.05856.

Hochreiter, Sepp, and Jurgen Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation* 9: 1735–80.

Landgrebe, Jobst, and Barry Smith. 2019. "Making Ai Meaningful Again." *Synthese* DOI 10.1007/s11229-019-02192-y: 1–21.

Larson, Erik J. 2021. *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*. Cambridge (MA): Harvard University Press.

Marcus, Gary, and Ernest Davis. 2019. *Rebooting Ai: Building Artificial Intelligence We Can Trust*. New York: Pantheon Books.

McCarthy, John, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon. 1955. "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955."

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Distributed Representations of Words and Phrases and Their Compositionality." In *Advances in Neural Information Processing Systems*, 3111–9.

Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why – the New Science of Cause and Effect*. New York: Hachette Book Group.

Perconti, Pietro, and Alessio Plebe. 2022. "Do Machines Really Understand Meaning? (Again)." *Journal of Artificial Intelligence and Consciousness* 9: doi 0.1007/s42979-022-01259-x.

Plebe, Alessio, and Giorgio Grasso. 2019. "The Unbearable Shallow Understanding of Deep Learning." *Minds and Machines* 29: 515–53.

Plebe, Alessio, and Pietro Perconti. 2022. *The Future of the Artificial Mind*. Boca Raton: CRC Press.

Preston, John, and Mark Bishop, eds. 2002. *Views into the Chinese Room: New Essays on Searle and Artificial Intelligence*. Oxford (UK): Oxford University Press.

Putnam, Hilary. 1960. "Minds and Machines." In *Dimensions of Mind*, edited by S. Hook, 138–64. New York: New York University Press.

Roger, Elise, Sonja Banjac, Michel Thiebaut de Schotten, and Monica Baciú. 2022. "Missing Links: The Functional Unification of Language and Memory ()" *Neuroscience and Biobehavioral Reviews* 133: 104489.

Rumelhart, David E., and James L. McClelland, eds. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge (MA): MIT Press.

Searle, John R. 1980. "Mind, Brain and Programs." *Behavioral and Brain Science* 3: 417–24.

Srivastava, Aarohi, Abhinav Rastogi, Abhishek Rao, Abu Awal Md Shoeb, Abubakar Abid, Adam Fisch, Adam R. Brown, Adam Santoro, and et al. 2022. "Beyond the Imitation Game: Quantifying and Extrapolating the Capabilities of Language Models." *arXiv abs/2206.04615*.

Thoppilan, Romal, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, et al. 2022. "LaMDA: Language Models for Dialog Applications." *arXiv abs/2201.08239*.

Turing, Alan. 1950. "Computing Machinery and Intelligence." *Mind* 59: 433–60.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. "Attention Is All You Need." In *Advances in Neural Information Processing Systems*, 6000–6010.

Warwick, Kevin, and Huma Shah. 2016. "Can Machines Think? A Report on Turing Test Experiments at the Royal Society." *Journal of Experimental and Theoretical Artificial Intelligence* 28: 989–1007.

Weizenbaum, Joseph. 1966. "Eliza – a Computer Program for the Study of Natural Language Communication Between Man and Machine." *Communications of the Association for Computing Machinery* 9: 36–45.

Winograd, Terry. 1972. *Understanding Natural Language*. New York: Academic Press.

Wittgenstein, Ludwig. 1922. *Tractatus Logico-Philosophicus*. London: Trench, Trubner & Co.

GPT-4 nel paese di Acchiappacitrulli

La logica che presiede all'attribuzione della responsabilità dell'IA generativa è la medesima del paese di Acchiappacitrulli. Come nel paese in cui Collodi aveva fatto finire il suo povero burattino, infatti, le vittime di un reato diventano, ai sensi dei termini d'uso, i colpevoli del reato medesimo. Vediamo perché

Di Daniela Tafani - Dipartimento di Scienze Politiche dell'Università di Pisa

In un racconto del 1954, [*The Great Automatic Grammatizator*](#), Roald Dahl immaginava un giovane ingegnere, Adolph Knipe, che – “costernato” per aver scritto cinquecentosessantasei racconti, tutti invariabilmente respinti dalle riviste alle quali li aveva proposti – progettava di costruire e riusciva poi a realizzare una “macchina per scrivere storie”, avendo compreso che le regole della grammatica possono essere formulate in termini matematici.

Smontare e rimontare la macchina del linguaggio: la “macchina letteraria” e i “pappagalli stocastici

“Poi, all'improvviso, gli venne in mente una semplice ma potente verità, cioè questa: la grammatica inglese è governata da regole pressoché matematiche nel loro rigore! [...] Dunque, si può ragionevolmente ritenere che una macchina applicata ai circuiti del calcolatore elettronico sarebbe in grado di sistemare le parole (invece dei numeri) nell'ordine appropriato secondo le regole grammaticali. Inserisci i verbi, i sostantivi, gli aggettivi, i pronomi, e fai in modo che possano essere utilizzati a richiesta. Poi inserisci le trame e lascia scrivere a lei le frasi”.

Poco più di un decennio più tardi, nel corso di una serie di conferenze, **anche Italo Calvino dichiarava possibile una “macchina letteraria”**, prendendo le mosse dalla “narrativa come processo combinatorio” e dalla constatazione che “l'uomo sta cominciando a capire come si smonta e come si rimonta la più complicata e la più imprevedibile di tutte le sue macchine: il linguaggio”. In fondo, sosteneva Calvino, una parte del lavoro dello scrittore non consiste che in “un'ostinata serie di tentativi di far stare una parola dietro l'altra seguendo certe regole definite, o più spesso regole non definite né definibili ma estrapolabili da una serie di esempi o protocolli”. Perfino “i più gelosi attributi dell'intimità psicologica, dell'esperienza vissuta, dell'imprevedibilità degli scatti

d'umore, i sussulti e gli strazi e le illuminazioni interiori” non sono che “campi linguistici, di cui possiamo benissimo arrivare a stabilire lessico grammatica sintassi e proprietà permutative”. Proprio a partire da “regole non definite né definibili ma estrapolabili da una serie di esempi” sono costruiti oggi i grandi modelli del linguaggio naturale. Si tratta di sistemi informatici che predicono stringhe di testo, utilizzando le statistiche sulla distribuzione delle parole per produrre altre stringhe di parole. Sono stati qualificati come “pappagalli statistici elettronici” o “pappagalli stocastici”, in virtù di due caratteristiche: quella di produrre parole senza conoscerne il significato e quella di farlo sulla base di informazioni probabilistiche sul modo in cui le sequenze di forme linguistiche si combinano nei testi di partenza.

I generatori di linguaggio basati su AI

I generatori di linguaggio, quali chatGPT o GPT-4, sono ulteriormente programmati, attraverso interazioni con esseri umani, per produrre output che somiglino a quelle che tali esseri umani qualificano come **risposte plausibili, pertinenti e appropriate**. Simili sistemi sono fondati su modelli del linguaggio, non su modelli della conoscenza: generano perciò testi plausibili e convincenti e, al tempo stesso, del tutto privi di valore informativo, oltre che di intenti comunicativi.

Le persone che leggono tali testi vi rintracciano un significato perché i **grandi modelli del linguaggio costituiscono un'immagine sfuocata di tutto il testo presente nel Web**, ossia, come scrive Regina Rini, “una rappresentazione statisticamente astratta dei contenuti di milioni di menti, quali sono espressi nella loro scrittura”. Per questa stessa ragione, ciò che tali testi forniscono è “non-informazione”: non risposte in senso proprio alle nostre domande, ma testi programmati per apparire plausibili, ossia molto simili a come una risposta alla nostra domanda potrebbe apparire. È perciò che sono stati definiti “generatori di stroncate”, nel senso in cui ha usato l'espressione Harry Frankfurt, o generatori di cliché.

Come è possibile che una “macchina letteraria” – i cui prodotti, come ha scritto Stefano Bartezzaghi, sono caratterizzati da “correttezza sintattica, plausibilità semantica, sostanziale carenza di senso” – sia oggi presentata come uno strumento che sostituirà milioni di lavoratori, che può fornire indicazioni sulle specifiche operazioni militari da intraprendere, in concreti scenari di guerra, e al quale si possano chiedere assistenza mentale o consulenze mediche? Vale la pena, per rispondere a questa domanda, tornare un momento alla storia di Adolph Knipe e della sua “macchina per scrivere storie”.

Animare l'inanimato: frodi vecchie e nuove

Nel racconto di Roald Dahl, Adolph Knipe e il suo capo, Mr Bohle, fondano un'agenzia letteraria e vendono i racconti e poi i romanzi prodotti dalla “macchina per scrivere storie”, spacciandoli per opere proprie o di scrittori fittizi. Infine, decidono di “**assorbire tutti gli altri scrittori del paese**”, “esattamente come fece Rockefeller con le sue compagnie petrolifere. Basta comprarli, e se poi non vendono, dar loro il benservito”.

Il genuino progresso costituito dalla «macchina letteraria» odierna è sfruttato dalle aziende con spregiudicatezza, voracità e spregio del diritto analoghi a quelli che Roald Dahl aveva attribuito, nel loro piccolo, ai protagonisti del suo racconto.

I generatori di linguaggio si fondano infatti, oltre che su [pratiche di espropriazione del lavoro](#) e di [esternalizzazione dei costi ambientali](#), su un'enorme quantità di dati e opere del lavoro umano, estorti o prelevati in blocco là dove si trovino, anche quando siano protetti dal diritto alla protezione dei dati personali o dal diritto d'autore. I testi prodotti, perciò, anziché generati dal nulla, sono definibili come una forma di "[plagio automatizzato](#)" e riproducono (con effetti normalizzanti, anziché generativi in senso proprio) [la prospettiva egemonica e i suoi stereotipi](#) (tutti sanno, come scrive [Cory Doctorow](#), che i programmi di completamento automatico sono profondamente conservatori).

Una generalizzata distribuzione e commercializzazione di tali sistemi [produce danni](#) alle singole persone coinvolte e alle società democratiche nel loro complesso. Nel caso, ad esempio, di GPT-4, tali effetti sono elencati puntualmente dall'azienda produttrice, OpenAI, nel relativo [rapporto tecnico](#); tra questi, compaiono:

- **“la tendenza a inventare i fatti”**, a fornire informazioni errate e a “svolgere compiti in modo scorretto”, tanto più pericolosa quanto più siano stati resi plausibili, e dunque ingannevoli *by design*, i suoi output;
- **l'inquinamento dell'ecosistema dell'informazione**: GPT-4 “ha il potenziale di mettere in dubbio l'intero ambiente dell'informazione, minacciando la nostra capacità di distinguere i fatti dalla finzione”;
- **la possibilità di truffe e manipolazioni politiche su larghissima scala**, con uno strumento che può produrre una miriade di testi, orientati come si desidera e indistinguibili da quelli prodotti da esseri umani.

Nel descrivere la prima caratteristica, il rapporto tecnico di OpenAI rileva che GPT-4 “può soffrire di ‘allucinazioni’”. L'uso di un termine che allude a fenomeni psichici – e, in particolare, a fenomeni psichici che presuppongono la capacità di distinguere il reale dall'immaginario – è intenzionalmente fuorviante. Così, infatti, la strutturale inaffidabilità, a fini informativi, di un sistema il cui funzionamento ordinario lo rende [“incapace di distinguere il possibile dall'impossibile”](#), è trasformata in mero errore, in una obnubilazione temporanea e superabile, quale la condizione generalmente straordinaria e passeggera per cui gli esseri umani, pur pienamente capaci di distinguere la realtà dall'apparenza, temporaneamente le confondono. In una nota al paragrafo sulle “allucinazioni”, OpenAI dichiara la piena consapevolezza di tale mistificazione e dei suoi effetti:

“Usiamo il termine “allucinazioni”, anche se riconosciamo che questa formulazione può suggerire un'antropomorfizzazione, che a sua volta può portare a danni o a modelli mentali errati di come il modello apprende”.

Questa operazione di antropomorfizzazione di un sistema informatico, per convincere le persone ch'esso sia animato, e che in quanto tale sia capace di comprendere e ragionare, non è isolata. ChatGPT e GPT-4 sono descritti da OpenAI come dotati della [«capacità di comprendere e generare testo in linguaggio naturale»](#), di [«fattualità e capacità matematiche»](#), di una [conoscenza del mondo](#), di [capacità di ragionamento](#) e della caratteristica di fornire, solo [«occasionalmente», risposte non corrette](#). In realtà, tali sistemi non conoscono né comprendono alcunché, sono architettonicamente incapaci di ragionamento abduittivo e non sono in grado di fornire alcuna informazione. Malgrado ciò, [Sam Altman, CEO di OpenAI, twitta](#) che chi non può

permettersi le cure potrà farsi curare da un generatore di linguaggio, e le aziende che commercializzano applicazioni per l'assistenza mentale registrano milioni di abbonati, sfruttando, [manipolando e mettendo in pericolo le persone più vulnerabili](#).

Speculare all'antropomorfizzazione delle macchine è la deumanizzazione delle persone. Il CEO di OpenAI, ad esempio, scrive:

Io sono un pappagallo stocastico, e lo sei anche tu

L'equiparazione di persone e sistemi informatici è utilizzata dalle aziende per far apparire come lecito – a partire dalla constatazione dell'imperscrutabilità della mente umana– il rilascio di prodotti inspiegabili e imprevedibili, ossia costitutivamente non sicuri. Così [il CEO di Google, in una recente intervista](#):

“Scott Pelley: Non capite bene come funziona. Eppure, l'avete scaricato sulla società? Sundar Pichai: Sì, è così. Mettiamola così. Non credo che comprendiamo appieno neppure il funzionamento della mente umana.”

Un ulteriore dispositivo per diffondere impunemente prodotti pericolosi e incolpare sistematicamente utenti e clienti è l'[offuscamento della linea di confine tra la fase di ricerca e sperimentazione e quella della distribuzione e commercializzazione di prodotti](#): OpenAI, ad esempio, rende pubblicamente disponibili i propri generatori di linguaggio naturale, in fase sperimentale, ne distribuisce al tempo stesso versioni a pagamento e fissa termini d'uso in virtù dei quali [ogni utente è responsabile](#) tanto dei propri input quanto degli output prodotti dal sistema (sui quali, ovviamente, non ha alcuna facoltà di controllo o decisione e i cui fondamenti gli sono del tutto oscuri).

Con ciò, considerato che [si può incappare in output che diffamino singoli utenti](#), la logica che presiede all'attribuzione della responsabilità è la medesima del [paese di Acchiappacitrulli](#). Come nel paese in cui Collodi aveva fatto finire il suo povero burattino, infatti, le vittime di un reato diventano, ai sensi dei termini d'uso, i colpevoli del reato medesimo.

La fine di un'allucinazione: le leggi valgono anche per l'“intelligenza artificiale”

Se la responsabilità per gli effetti dei generatori di linguaggio ricadesse sui produttori, la loro commercializzazione non sarebbe vantaggiosa (quand'anche il diritto non la vietasse *tout court*). Per sfuggire alle loro responsabilità senza rinunciare a una fonte di enorme profitto, i giganti della tecnologia hanno diffuso una famiglia di narrazioni che danno forma alla percezione pubblica del rapporto tra etica, politica, diritto e tecnologia e costituiscono gli assiomi indiscussi di qualsiasi discorso pubblico. Sono così entrati a far parte del senso comune, tra gli altri, [il principio di inevitabilità tecnologica, il mito dell'eccezionalismo tecnologico, il principio di innovazione e il mito del vuoto giuridico](#).

Le diverse narrazioni concorrono, congiuntamente, al perseguimento di singoli scopi aziendali: al fine di sfuggire alle responsabilità per gli effetti dannosi dei prodotti, ad esempio, sono utili l'eccezionalismo, l'antropomorfizzazione delle macchine e, insieme, il «principio di inevitabilità tecnologica», secondo un meccanismo che [Joseph Weizenbaum](#) aveva descritto quasi mezzo secolo fa: **“Il mito dell'inevitabilità tecnologica, politica e sociale è un potente tranquillante per la coscienza. Il suo servizio è quello di togliere la responsabilità dalle spalle di tutti coloro che ci credono veramente. Ma, in realtà, ci sono degli attori! [...] La reificazione di sistemi**

complessi che non hanno autori, di cui sappiamo solo che ci sono stati dati in qualche modo dalla scienza e che parlano con la sua autorità, non permette di porsi domande di verità o di giustizia.”

La tesi che le leggi vigenti non si applichino ai prodotti basati su sistemi di «intelligenza artificiale», in virtù della loro novità e straordinarietà, e che servano dunque nuove leggi, scritte *ad hoc* per ciascuna tecnologia, serve a dar luogo a una corsa che vedrà il legislatore perennemente in affanno, nel rincorrere le più recenti novità tecnologiche, le quali saranno dunque commercializzabili eslege.

All'allucinazione collettiva che ci ha indotti a vedere i sistemi informatici come agenti artificiali e a credere inapplicabili i sistemi normativi vigenti e gli ordinari criteri di attribuzione della responsabilità, si oppone oggi una crescente consapevolezza del fatto che i sistemi informatici sono artefatti, ossia prodotti, e che non c'è alcuna ragione per sottrarne la distribuzione e la commercializzazione alla legislazione ordinaria.

Di recente, la Federal Trade Commission statunitense, in netto contrasto con la posizione delle grandi aziende, **ha dichiarato che i generatori di linguaggio sono soggetti alle norme in vigore** e che l'impossibilità, per ragioni tecniche, di ottemperare a ciò che le leggi esigono non è una ragione per dichiararsi esentati da quelle stesse leggi e chiederne di nuove:

C'è un mito potente in giro secondo il quale «l'IA non è regolamentata». [...] Ha un forte fascino intuitivo: suona bene. Come potrebbero queste misteriose nuove tecnologie essere regolamentate dalle nostre vecchie leggi polverose?

Se l'avete sentita o l'avete detta, fate un passo indietro e chiedetevi: a chi giova questa idea? Non aiuta i consumatori, che si sentono sempre più indifesi e smarriti. Non aiuta la maggior parte delle aziende. [...]

Credo che l'idea secondo cui «l'IA non è regolamentata» aiuti quel piccolo sottoinsieme di imprese che non sono interessate alla *compliance*. E abbiamo già sentito frasi simili. «Non siamo una compagnia di taxi, siamo una compagnia tecnologica». «Non siamo un'azienda alberghiera, siamo un'azienda tecnologica». Queste affermazioni erano di solito seguite da affermazioni secondo cui le norme statali o locali non potevano essere applicate a queste aziende. La realtà è che l'IA è regolamentata. Solo alcuni esempi:

- Le leggi sulle pratiche commerciali sleali e ingannevoli si applicano all'IA. [...]
- Le leggi sui diritti civili si applicano all'IA. [...]
- Le leggi sulla responsabilità civile e da prodotto si applicano all'IA.

La Federal Trade Commission ricorda inoltre che la Commissione medesima «storicamente non ha reagito bene all'idea che un'azienda non sia responsabile del proprio prodotto perché questo è una «scatola nera» incomprensibile o difficile da testare» e che, nei casi in cui la legge esiga spiegazioni, **«l'inspiegabilità o l'imprevedibilità di un prodotto è raramente una difesa giuridicamente ammissibile»**.

La trasparenza è richiesta anzitutto sui dati di partenza, rendendo note le fonti dei dati e l'eventuale consenso degli interessati. Su questo aspetto, i più recenti provvedimenti della

Commissione mostrano una radicalità nuova: in alcuni casi di dati ottenuti impropriamente, essa ha ottenuto infatti non la mera cancellazione di tali dati, ma la [distruzione di tutti i modelli e gli algoritmi costruiti utilizzando tali dati](#).

Quanto alla responsabilità per i danni derivanti dalla distribuzione dei generatori di linguaggio, [Daniel Dennet](#) osserva che esistono pene severe per chi metta in circolazione denaro falso e ritiene necessaria un'analogia severità verso chi, mettendo ancor più rischio le fondamenta delle società democratiche, crei e metta in circolazione "persone contraffatte".

[La Presidente della Federal Trade Commission, Lina Khan](#), si è espressa riguardo alla possibilità di automatizzare le frodi online, tramite i generatori di linguaggio:

Nell'applicare il divieto di pratiche ingannevoli, non ci limiteremo a considerare i truffatori occasionali che impiegano questi strumenti, ma anche le aziende a monte che li mettono a disposizione.

Il ruolo delle «aziende a monte» e gli effetti antidemocratici della concentrazione di potere in tali aziende – le quali hanno ormai [dimensioni e prerogative che le rendono assimilabili ai più potenti Stati nazionali](#) – costituiscono in effetti uno dei nodi cruciali. Dall'inserimento dei generatori di linguaggio in tutte le applicazioni esistenti, le grandi piattaforme ritengono di poter ottenere uno strumento per trattenere gli utenti nei loro [ecosistemi chiusi](#), avere a disposizione i dati e i metadati relativi a tutte le loro attività e [intradere di annunci pubblicitari "personalizzati" le risposte e i risultati di ricerca presentati agli utenti](#).

Vedremo se la stesura definitiva del prossimo regolamento europeo sull'intelligenza artificiale sarà guidata dal «[fantasma lobbista nella macchina](#)» della regolazione – alacremente all'opera per far sparire ogni previsione relativa ai generatori di linguaggio – o se prevarrà, invece, la constatazione di [Lina Khan](#): "Nessuna eccezione per l'intelligenza artificiale è prevista nei codici".

Bibliografia

S. Bartezzaghi, *Chatgpt. Non è detto che sia vero, ma è vero che lo si è detto*, «Doppiozero», 26 aprile 2023, <https://www.doppiozero.com/chatgpt-non-e-detto-che-sia-vero-ma-e-vero-che-losi-e-detto>.

E.M. Bender, T. Gebru, A. Mc Millan-Major, S. Shmitchell, *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?*, in *Conference on Fairness, Accountability, and Transparency (FAccT '21)*, March 3–10, 2021, Virtual Event, Canada, New York, ACM, 2021, <https://dl.acm.org/doi/10.1145/3442188.3445922>.

I. Calvino, *Cibernerica e fantasmi (Appunti sulla narrativa come processo combinatorio)*, in Idem, *Una pietra sopra*, Milano, Mondadori 2023, pp. 201-221.

R. Caso, *Open Data, ricerca scientifica e privatizzazione della conoscenza*, 2022, <https://www.robertocaso.it/2022/01/22/dati-aperti-della-ricerca-o-proprietaintellettuale/>.

R. Claypool and C. Hunt, "Sorry in Advance!" *Rapid Rush to Deploy Generative A.I. Risks a Wide Array of Automated Harms*, April 18, 2023, <https://www.citizen.org/article/sorry-in-advancegenerative-ai-artificial-intelligence-chatgpt-report/>.

R. Dahl, *Il libraio che imbrogliò l'Inghilterra seguito da Lo scrittore automatico*, traduzione di M. Bocchiola, Parma, Ugo Guanda Editore, 1996.

S. Dang, *Google to test ads in generative AI search results*, May 23,

2023, <https://www.reuters.com/technology/google-test-ads-generative-ai-search-results-202305-23/>.

D. Dennett, *The Problem With Counterfeit People*, «The Atlantic», May 16,

2023, <https://www.theatlantic.com/technology/archive/2023/05/problem-counterfeitpeople/674075/>.

C. Doctorow, *The AI hype bubble is the new crypto hype bubble*, March 9,

2023, <https://pluralistic.net/2023/03/09/autocomplete-worshippers/#the-real-ai-was-the-corporations-that-we-fought-along-the-way>.

Early Thoughts on Generative AI. Prepared Remarks of Commissioner Alvaro M. Bedoya, Federal Trade Commission Before the International Association of Privacy Professionals, April 5,

2023, <https://www.ftc.gov/news-events/news/speeches/prepared-remarks-commissioneralvaro-m-bedoya-international-association-privacy-professionals>

M. Gault, *Palantir Demos AI to Fight Wars But Says It Will Be Totally Ethical Don't Worry About It*, «Vice», April 26, 2023, <https://www.vice.com/en/article/qjvb4x/palantir-demos-ai-to-fightwars-but-says-it-will-be-totally-ethical-dont-worry-about-it>.

T. Gebru, E.M. Bender, A. McMillan-Major, M. Mitchell, *Statement from the listed authors of Stochastic Parrots on the "AI pause" letter*, March 31, 2023,

<https://www.dairinstitute.org/blog/letter-statement-March2023>.

Generating Harms: Generative AI's Impact & Paths Forward, May 23, 2023,

<https://epic.org/wpcontent/uploads/2023/05/EPIC-Generative-AI-White-Paper-May2023.pdf>

A. Kak, S. Myers West, *ChatGPT And More: Large Scale AI Models Entrench Big Tech Power*, in *AI Now 2023 Landscape: Confronting Tech Power*, AI Now Institute, April 11, 2023,

<https://ainowinstitute.org/publication/large-scale-ai-models>.

D. Karpf, *On Generative AI and Satisficing*, May 9, 2023,

<https://davekarpf.substack.com/p/ongenerative-ai-and-satisficing>.

E. Klein, *The Imminent Danger of A.I. Is One We're Not Talking About*, «New York Times»,

February 26, 2023, <https://www.nytimes.com/2023/02/26/opinion/microsoft-bing-sydneyartificial-intelligence.html>.

N. Klein, *AI machines aren't 'hallucinating'. But their makers are*, May 8,

2023, <https://www.theguardian.com/commentisfree/2023/may/08/ai-machines-hallucinatingnaomi-klein>

M. Lombardi, *Generative AI, dov'è il bene per l'Umanità?*, Agendadigitale.eu, 15 maggio 2023,

<https://www.agendadigitale.eu/cultura-digitale/la-generative-ai-che-verra-siamo-sullastrada-giusta/>.

S. Luccioni, *The mounting human and environmental costs of generative AI*, April 12,

2023, <https://arstechnica.com/gadgets/2023/04/generative-ai-is-cool-but-lets-not-forget-its-human-and-environmental-costs/>

E. Malroux, *Quelles régulations pour la conception des IA génératives?*, 26 avril

2023, <https://linc.cnil.fr/fr/dossier-ia-generative-quelles-regulations-pour-la-conception-des-ia-generatives>.

D. McQuillan, *We come to bury ChatGPT, not to praise it*, February 6, 2023,

<https://www.danmcquillan.org/chatgpt.html>.

- E. Nardelli, *Non lasciamo che i nostri figli diventino cavie di ChatGPT&Co.*, «Start Magazine», 19 marzo 2023, <https://www.startmag.it/innovazione/non-lasciamo-che-i-nostri-figli-diventinocavie-di-chatgptco/>.
- M.C. Pivatolo, *Sulle spalle dei mercanti? Teledidattica e civiltà tecnologica*, «Bollettino telematico di filosofia politica», 2023, <https://commentbfp.sp.unipi.it/sulle-spalle-dei-mercantiteledidattica-e-civilta-tecnologica/>.
- E.M. Renieris, *Beyond data: reclaiming human rights at the dawn of the metaverse*, Cambridge, Massachusetts, The MIT Press, 2023.
- C. Stokel-Walker, *What is the future of AI? Google and the EU have very different ideas*, «New Scientist», May 12, 2023, <https://www.newscientist.com/article/2373433-what-is-the-future-of-ai-google-and-the-eu-have-very-different-ideas/>.
- V. Storchan, *L'IA Potemkine et le futur du travail, une conversation avec Antonio Casilli*, «Le Grand Continent», 7 avril 2023, <https://legrandcontinent.eu/fr/2023/04/07/ia-potemkine-et-lefutur-du-travail-une-conversation-avec-antonio-casilli/>.
- D. Tafani, *Sistemi fuori controllo o prodotti fuorilegge? La cosiddetta «intelligenza artificiale» e il risveglio del diritto*, «Bollettino telematico di filosofia politica», 2023, pp. 1-13, <https://btfp.sp.unipi.it/it/2023/05/sistemi-fuori-controllo-o-prodotti-fuorilegge>.
- The Great A.I. Hallucination*, «The New Republic», May 10, 2023, <https://newrepublic.com/article/172454/great-ai-hallucination-chatgpt>.
- I. van Rooij, *Against automated plagiarism*, December 29, 2022, <https://irisvanrooijcogsci.com/2022/12/29/against-automated-plagiarism/>.
- W. Vannini, *Stop a chatGPT: per fortuna il Garante è vivo e lotta insieme a noi*, Agendadigitale.eu, 3 aprile 2023, <https://www.agendadigitale.eu/sicurezza/privacy/stop-a-chatgpt-per-fortuna-ilgarante-e-vivo-e-lotta-insieme-a-noi/>.

Il giornalismo nell'era dell'algoritmo: il matrimonio impossibile tra IA e news

La platform society, con la sua logica della monetizzazione dell'attenzione produce un output informativo in cui vero e falso convivono. Contro i seri rischi

che questo stato di cose comporta per le nostre economie, occorre aumentare la collaborazione internazionale e la condivisione di metodologie e risultati

Di **Francesca Rizzuto**, Docente di Sociologia del giornalismo dell'Università di Palermo

La circolazione di **notizie false** (Albright 2017; Corner 2017), favorita da utenti ignari, nonché la frequente costruzione intenzionale di **campagne coordinate di disinformazione**, impongono a livello globale l'obbligo di focalizzare l'attenzione sul tema, centrale per tutti i sistemi democratici, della **nuova opacità dei confini tra la libertà di espressione e la necessità di limitare la manipolazione dei flussi informativi, operata dalle piattaforme** (Rizzuto, Vaccaro 2023).

Il giornalismo nell'era dell'IA: tutte le distorsioni in atto

Indubbiamente, il problema della crescente diffusione delle pratiche disinformative richiede **una riflessione critica** sia ai professionisti del giornalismo che ai vari attori istituzionali, chiamati a regolare e controllare i circuiti planetari dell'informazione on line. In particolare, sono i recenti ed inediti progressi tecnologici dell'intelligenza artificiale, legati a complessi algoritmi di machine learning per elaborare e comprendere testi, a presentare sfide “nuove” poste al giornalismo insieme all'ascesa dei social media come attori centrali nella sfera pubblica, che costituiscono per milioni di utenti le principali fonti di accesso all'informazione.

Numerosi studi (Van Dijck, Poell, de Wall 2018; Sorice 2020) hanno evidenziato che nell'ecosistema comunicativo contemporaneo, ibrido, iper-frammentato e dominato dalle piattaforme, si creano **significative distorsioni sia nei processi di ricerca delle news che sul piano delle possibilità di controllo della loro veridicità e fattualità**. La logica strutturale degli algoritmi, ideati e gestiti dai proprietari privati, mossi da legittime finalità di profitto in un contesto di mercato altamente concorrenziale, favorisce e dà visibilità soprattutto a determinati flussi di informazioni, limitando fortemente gli spazi di “libertà” e autonomia dei singoli così efficacemente che le persone vedono ciò che vogliono vedere o ciò che un algoritmo ‘crede’ che siano interessate a vedere (Chambers 2021).

Pertanto, nella platform society un tale ecosistema può essere considerato responsabile di una varietà di fenomeni emergenti e con risvolti problematici, tra cui, ad esempio, la polarizzazione politica, le camere dell'eco, la mediatizzazione, la vetrinizzazione sociale. Come ha sottolineato Chambers, al giorno d'oggi ci sono **minacce “nuove” per l'acquisizione di informazioni** anche se, naturalmente, la circolazione di bufale o bugie non è nuova. Tuttavia, “sembra esserci sia un picco nel numero assoluto di tali storie in circolazione, sia un aumento dell'inclinazione partigiana politicamente rilevante delle informazioni false e fuorvianti, che vengono immesse nella sfera pubblica da attori discutibili” (ivi, 149). Pertanto, **l'opacità tra vero e falso dell'era post-verità presenta rischi rilevanti per le architetture costituzionali occidentali**, favorendo processi come la “normalizzazione” di severe pratiche di censura o, per scopi totalmente diversi, dinamiche comunicative conflittuali e divisive, in cui prevale un linguaggio aggressivo e di odio (Allcott e Gentzkow, 2017; Lazer et al. 2018; Sorice 2020).

Lo storytelling NON è giornalismo

Grazie alle tecnologie digitali è possibile l'ampliamento delle risorse per accedere alla realtà che, però, si traduce, spesso, in un vero e proprio caos informativo, in cui si diffondono pratiche di disinformazione (Bracciale, Grisolia 2020) e processi di viralizzazione di notizie false ma verosimili, i cui "effetti di realtà" (Ireton, Posetti 2018; Edson, Tandoc, Lim, Ling 2018) sono in grado di inquinare le dinamiche democratiche. In tale prospettiva, la logica della post-verità, riconoscendo spazio e visibilità a tutte le ricostruzioni individuali, si coniuga perfettamente con il predominio degli imperativi commerciali nelle *news* e con la loro spinta verso eccessi intrattenitivi, che favorisce lo *story model* (Rizzuto 2019). Secondo [Kormelink](#) e [Meijer](#) (2015), proprio lo *storytelling* costituisce un elemento di criticità per il giornalismo poiché non si tratta, infatti, soltanto di una strategia in grado di attrarre e coinvolgere utenti: **viene reso sempre più opaco il confine tra report obiettivo degli eventi e loro ricostruzione narrativa nella quale le notizie possono essere parcellizzate** e diffuse come monadi fluide con rilevanti tendenze alla multiautorialità (Rizzuto 2022).

Il giornalismo dell'ecosistema digitale ibrido è, quindi, perfettamente coerente con la logica della post-verità, che favorisce il predominio della logica narrativa transmediale (McErlean 2018): in altri termini, la platform society proprio per le peculiarità produttive legate alla sua logica della monetizzazione dell'attenzione, produce un output informativo caratterizzato dalla presenza di grumi concettuali in cui **vero e falso convivono** (Lorusso 2018), un sistema granulare, che presenta punti di vista verosimili come verità plurime e plausibili (Groot Kormelink, Costera Meijer 2015) e valori-notizia, che sono imposti dagli algoritmi e non dalla tradizione professionale (Codina, Lopezosa, Rovira 2022). Oltre alle false informazioni, diffuse involontariamente o intenzionalmente per finalità politiche e di propaganda, **l'infosfera è condizionata, quindi, dalla logica del business delle piattaforme** (Floridi 2018), che, diventate ormai enormi macchine per accumulare denaro o consensi, privilegiano i contenuti che attirano like e commenti. Poche grandi aziende, così, contribuiscono alla circolazione di fake news così come alla manipolazione mediale dei singoli, con la conseguenza che la "trasformazione tossica del dibattito pubblico" (Bentivegna, Boccia Artieri, 2021, 97), non sembra più definibile come un rischio potenziale ma come realtà già presente, in grado di contribuire significativamente alla profonda crisi di legittimità delle istituzioni politiche in numerosi contesti occidentali. **L'uso strategico di pseudo-facts (Riva 2018), infatti, favorisce soprattutto i leader politici che si servono di retoriche del consenso fondate su posizioni a cui aderire "emozionalmente"**, delineando in tal modo, un cortocircuito informativo, nel quale i cittadini-elettori possono diventare anche complici involontari di pratiche di disinformazione, misinformazione o malinformazione (Ireton, Posetti 2018), semplici coprotagonisti del contagio emotivo che avvia cortocircuiti anche per l'assenza di trasparenza nel processo di ricerca delle news o di un serio controllo professionale sul contenuto.

News verosimili e credibilità del giornalismo

Le news verosimili costituiscono, inoltre, un serio e concreto pericolo per la credibilità del giornalismo che è sempre più incline ad un racconto eccessivamente drammatizzante ed iperbolico della realtà, per attrarre ed emozionare ampie fasce di audience (Zelizer, 2004; Santos 2009), e che vede aumentare, sul terreno della creazione di testi, la pericolosa concorrenza delle applicazioni

tecnologiche che la ricerca sull'intelligenza artificiale ha prodotto in questi ultimi mesi. **Il caso recente dell'affermazione planetaria di [ChatGPT](#) costituisce un esempio evidente dei rischi connessi al progresso inarrestabile dell'intelligenza artificiale all'ambito del newsmaking** ed ha suscitato perplessità e allarmi non solo tra gli attori istituzionali ma anche tra i professionisti dell'informazione e gli stessi tecnici dell'ingegneria informatica, molti dei quali stanno iniziando a prendere le distanze dai processi in corso, evidenziandone i rischi attuali e le prospettive minacciose per il futuro.

Questa tecnologia ha mostrato in modo inequivocabile il lato opposto e pericoloso della medaglia della rivoluzione digitale: se è vero che l'ascesa della rete e l'affermazione delle piattaforme nella vita quotidiana dei singoli hanno prodotto come esito rivoluzionario la creazione di ambiti di azione ed interazione inediti, in grado di ampliare su scala planetaria gli spazi di partecipazione e di accesso alla realtà, un'agorà nella quale è possibile connettersi con ogni luogo, proprio alcune specificità delle nuove dinamiche comunicative (la struttura orizzontale, la velocità, la brevità, l'ipersemplificazione del linguaggio o l'anonimato) possono **facilitare bias relazionali, come la polarizzazione delle opinioni o la negazione del confronto con l'alterità**. I social, ad esempio, proprio grazie all'interconnessione e alla loro pervasività, sono diventati uno dei canali privilegiati sia per la diffusione dell'hate speech che per la circolazione di fake news intenzionalmente costruite per danneggiare avversari politici.

I rischi concreti per le democrazie

Le esperienze eclatanti delle elezioni americane del 2016 o le campagne di disinformazione costruite durante il conflitto in Ucraina in questo ultimo anno dimostrano i rischi concreti per le democrazie, che istituzioni e cittadini sono chiamati ad affrontare con consapevolezza, in una sfera pubblica abitata anche da attori discutibili, ben diversa dalla tradizionale configurazione habermasiana. Al fine di favorire una prospettiva di **analisi critica lucida e libera da pregiudizi analitici miopi**, in primo luogo, come ha sottolineato Chambers (2021), occorre prendere in considerazione sia il modello finanziario dei social media, dominato dai big data, dal microtargeting e dai bot, sia le peculiarità delle declinazioni nazionali dei sistemi mediali, in cui si sono storicamente affermate concezioni normative molto differenti tra loro sul ruolo attribuito ai newsmedia nelle dinamiche politiche.

Il ruolo del giornalismo nei contesti occidentali

Nei contesti occidentali, tradizionalmente, il ruolo del giornalismo è stato considerato centrale e nei diversi modelli della professione il suo rilievo quale agente di mediazione è universalmente riconosciuto: “anche quando agiscono come sbocchi ideologici, organi di partito, o rapaci capitalisti, la funzione politica di mediazione tra cittadino e Stato o tra cittadino e cittadino era inteso dalla maggior parte degli attori. Al contrario, i social media si sono sviluppati non come un organo di stampa libera, o come mediatori tra cittadino e Stato, ma come strumenti divertenti, utili per socializzare e, a quanto pare, un buon modo per fare soldi. Il loro ruolo epistemico e la loro funzione di trasmettitori di fatti importanti su ciò che sta accadendo nel mondo, incluso il mondo della politica, è stato uno sviluppo non intenzionale”. (Chambers 2021, 150-151).

La pervasività delle piattaforme ha reso più opaco il confine tra producer e consumer di news, favorendo il prevalere della componente emotiva rispetto alla “vecchia” questione dell’obbligo per il giornalismo della veridicità dei fatti riportati. La conseguenza più significativa dei processi in corso è evidente nel ruolo crescente dei social media come fornitori di notizie e trasmettitori di eventi (García-Orosa, López-García, Vázquez-Herrero 2020; Reuters), che ha modificato radicalmente il panorama contemporaneo del giornalismo, sommandosi ai processi di spettacolarizzazione della realtà (Thussu 2007; Rizzuto 2018) e alle pratiche di produzione e fruizione dei contenuti informativi che si servono delle tecnologie di intelligenza artificiale, ormai perfettamente capaci di produrre senza l’intervento umano costruzioni narrative fluide, verosimili, che si inseriscono velocemente in dinamiche virali planetarie.

Il dibattito innescato da ChatGPT: farà proliferare l’industria della falsità e della disinformazione?

Con la creazione da parte di *OpenAI* di una piattaforma online utilizzabile da tutti, **ChatGPT, si è avviato un ampio dibattito sulla rivoluzione che i nuovi chatbot basati sull’AI** potranno portare in numerosi ambiti della vita sociale e culturale: dalle implicazioni per il mondo del lavoro, alla politica fino alla ricerca scientifica, le capacità di questa nuova tecnologia sono molteplici ed hanno un potenziale veramente enorme in termini di ricadute sui circuiti democratici e sui flussi informativi in circolazione. Di conseguenza, a prescindere dagli interrogativi relativi a *come e in che cosa* l’intelligenza artificiale possa sostituire l’essere umano, è opportuno chiedersi quali cambiamenti potranno registrarsi effettivamente, sia a livello delle interazioni sociali tra gli individui che nel processo di acquisizione delle conoscenze. **ChatGPT** è stato lanciato nel **novembre 2022**, con l’obiettivo di rendere l’**Intelligenza Artificiale Generale** fruibile per tutti e, per questo **motivo la piattaforma è gratuita**, anche se **esistono piani a pagamento** che consentono l’**accesso prioritario a nuove funzionalità**. **Fondata da Elon Musk nel 2015, Open AI** è un’organizzazione di ricerca sull’intelligenza artificiale che mira a collaborare con altre istituzioni: nata con l’intento di utilizzare l’intelligenza artificiale per analizzare e comprendere i testi scritti dagli esseri umani, si serve di complessi algoritmi di machine learning per elaborare, comprendere e persino rielaborare milioni di pagine di testi.

L’elemento veramente rivoluzionario è la possibilità di istruire lo strumento su un determinato argomento, fornendo materiale da leggere e poi chiedergli di rispondere a domande o di realizzare un testo originale su quel tema: commentando i primi risultati, David Luan, vice presidente della divisione degli ingegneri informatici di **OpenAI** ha evidenziato che il prodotto “sembra dannatamente reale” e proprio in tale eccesso di verosimiglianza risiede il nucleo maggiormente problematico della nuova tecnologia. Testi tanto simili ad uno reale possono diventare pericolosi se inseriti intenzionalmente nei circuiti comunicativi alla base delle architetture istituzionali dei sistemi democratici moderni: in altri termini, **OpenAI** è perfettamente in grado di scrivere notizie false con uno stile che sembra quello di uno scrittore umano, e per questo può velocemente diventare **una gigantesca ed efficientissima macchina per produrre fake news in serie e con costi molto bassi**. Inoltre, in numerosi esperimenti è stato dimostrato che simili tecnologie sono tecnicamente in grado di far fiorire, in modo esponenziale, l’industria della falsità e della disinformazione, immettendo svariati elementi tossici nell’agorà on line, come nel caso di perfetti **testi complottisti o**

diffamatori (scritti da un computer a costo zero e ad una velocità straordinaria) **o nel caso in cui, coinvolgendo leader o candidati politici, possono condizionare pesantemente l'opinione pubblica o importanti tornate elettorali, diventando un pericolo per tutte le elezioni democratiche e per la libertà** (Mazzoleni 2021).

Perché ChatGPT ha cambiato le regole del gioco

In particolare, **ChatGPT**, acronimo di Generative Pretrained Transformer si sta rivelando un versatile strumento di elaborazione del linguaggio naturale (o Natural Language Processing) che utilizza algoritmi avanzati di apprendimento automatico per generare risposte simili a quelle umane all'interno di un discorso. Si tratta di un prototipo di chatbot, cioè un **agente software in grado di eseguire azioni per un interlocutore umano, basandosi su comandi ricevuti dall'utente in linguaggio naturale (scritto o parlato): pertanto, può** conversare con un essere umano e rispondere ad una serie di domande, fino ai casi più evoluti, nei quali è addirittura in grado di comprendere il tono e il contesto del dialogo, e successivamente di sfruttare queste informazioni nella conversazione. **A differenza dei Chatbot rule based, (bot in grado di dialogare con un umano seguendo alberi conversazionali predefiniti con risposte standard), chatGPT è un chatbot di tipo conversazionale, che può intrattenere una conversazione, comprendendo richieste, e fornire risposte specifiche alle domande dell'utente: a differenza degli ormai comuni assistenti virtuali come Siri, Amazon Alexa e Google Assistant, ChatGPT è quindi in grado di rispondere ai prompt con un linguaggio simile a quello del suo interlocutore, e, quindi, può, per esempio, scrivere articoli, risolvere operazioni matematiche, rivedere o tradurre testi (non solo notizie ma anche saggi) e, nelle versioni più recenti, generare immagini a partire da una descrizione testuale o (a pagamento in ChatGPT-4 comprendere input sia testuali che visivi.** Oltre all'ottimizzazione dei processi, ad esempio in ambito aziendale nelle relazioni con i clienti, il mondo dell'informazione e quello della politica non possono ignorare le potenziali conseguenze di queste applicazioni sia al livello di **una vera e propria sostituzione di mansioni creative**, in particolare per la produzione di immagini e testi scritti, che sul piano del fondamento epistemologico proprio del giornalismo, almeno nei termini in cui si è affermato nella modernità occidentale quale mediatore affidabile nella sfera pubblica.

I limiti di ChatGPT

Uno dei **limiti** principali di **Chat GPT** è l'**aggiornamento delle informazioni**, anche se le stesse **interazioni con gli utenti** costituiscono un **metodo di addestramento**, che permette alla macchina di aggiornarsi: è indubbio non soltanto che può **commettere errori**, in particolare nel comprendere le intenzioni di chi lo ha interpellato, ma anche che le sue risposte possano essere **superficiali** e, per quanto riguarda l'attività di ricerca di informazioni da parte degli utenti, non è da escludere che un uso eccessivo possa contribuire all'impigritimento dell'utente e, in diversi casi, sostituire la ricerca su motori di ricerca come Google. Tra gli altri limiti, riconosciuti dalla stessa **OpenAI**, risultano anche la **produzione di risposte plausibili, ma senza senso o errate**, soprattutto nei casi di domande ambigue o con sfumature linguistiche, difficilmente comprensibili dall'algoritmo, o la possibilità di produrre risposte offensive, da limitare con il ricorso a filtri per moderare il linguaggio. Uno dei risvolti più problematici del legame tra produzione di testi, giornalismo e intelligenza artificiale riguarda la disinformazione: **essendo Chat GPT addestrato**

su testi presi da Internet, la piattaforma può essere facilmente soggetta a fake news, rendendo quasi del tutto impossibile una corretta distinzione tra contenuti veri e testi, sia verbali che iconici, manipolati. Il recente caso dell'immagine fake di **Papa Francesco con un piumino bianco**, diventata virale a livello planetario, ha imposto in modo evidente all'attenzione generale la portata dei rischi per la narrazione giornalistica del reale: con questo caso, non l'unico in rete, l'applicazione linguistica generativa per la produzione di contenuti visuali del tutto verosimili alla realtà sembra riuscire a rendere definitivamente irrilevante la differenza tra vero e verosimile ed il problema sia le immagini, ma anche nella produzione di testi scritti,

L'evidente fragilità della sfera pubblica online

Se è plausibile che **Chat GPT** non solo potrebbe attingere da fonti di disinformazione online, ma produrre testi mendaci, **il rischio evidente di proliferazione incontrollata di fake news è una delle principali problematiche etiche che politici, cittadini e ricerca scientifica dovranno affrontare**: convivere con l'AI significa mirare ad attivare policies, produrre norme efficaci nell'attuale far west legislativo in cui da anni i leader del mercato della comunicazione operano, promuovere il ricorso a solide competenze professionali per diventare cittadini capaci di riconoscere fonti credibili ed autorevoli nell'alluvione informativa della platform society. Un esempio palese del rapporto ancora squilibrato tra forze ed esigenze degli attori coinvolti è stato recentemente fornito dal blocco di ChatGPT da parte del Garante della privacy italiano: andando nella direzione di un controllo severo del rispetto di norme riguardanti i dati personali degli utenti, con il Provvedimento dell'11 aprile 2023 ha imposto limitazioni e obblighi di adeguamento a **OpenAI**. Il 28 aprile, la società statunitense lo ha reso nuovamente accessibile agli utenti italiani, dopo aver fatto pervenire al Garante per la protezione dei dati personali una nota nella quale illustra le misure introdotte in ottemperanza alle richieste ed alle indicazioni dell'Autorità.

L'episodio dimostra la fragilità della sfera pubblica online, evidenziando i rischi di come la tecnologia possa essere manipolata per scopi politici e ha suggerito la necessità di proporre risposte ai problemi legati all'esistenza stessa dell'intelligenza artificiale: sia la società civile che le autorità di regolamentazione devono attivarsi per creare resilienza nel mondo digitale. La disinformazione e le notizie false sono molto diffuse e coloro che cercano di manipolare la sfera pubblica online possono capitalizzare il calo dei livelli di fiducia nelle istituzioni e negli esperti.

Sempre più potere nelle mani di pochi

La straordinaria concentrazione di potere di mercato in pochissime aziende tecnologiche solleva seri interrogativi in un mondo in cui l'intelligenza artificiale e la tecnologia hanno il potenziale per trasformare interi settori dell'economia: nel caso dell'informazione, oramai da alcuni anni, le news sono prodotte e fruite in un panorama mediatico radicalmente alterato in cui le piattaforme tecnologiche ricevono la maggior parte dei ricavi pubblicitari, indebolendo ulteriormente gli editori di notizie tradizionali. Inoltre, la rivendita dell'attenzione umana agli inserzionisti solleva profondi interrogativi non solo sulle notizie a cui le persone accedono ma anche sulle concrete possibilità per i singoli cittadini di comprendere questo nuovo panorama in cui la disinformazione e la mal informazione proliferano online accanto al giornalismo tradizionale, con

un enorme incremento di pirateria informatica, fughe di notizie e inserimento di informazioni false in documenti scaricati online.

Conclusioni

I governi sono sempre più preoccupati per le notizie false, la disinformazione e il modo in cui la sfera pubblica può essere manipolata: molti hanno annunciato indagini, o proposto leggi e regolamenti. Tuttavia, il problema non può essere risolto multando le aziende: occorre aumentare la collaborazione internazionale e la condivisione di metodologie e risultati. Dovrebbe esserci anche un impegno da parte dei policymakers per una maggiore trasparenza su come si utilizzano i dati dei cittadini. Ma, soprattutto, occorrono politiche a lungo termine per aiutare i cittadini a diventare più informati, al fine di costruire forme nuove di resilienza per i sistemi democratici: è vero che non si tornerà, ed è un bene, ad un panorama mediatico in cui ci saranno solo pochi arbitri della verità, ma questo mette al centro la questione della responsabilità di ogni attore sociale con ancora maggiore rilevanza.

Bibliografia

Albright J. (2017), “Welcome to the era of fake news”, *Media and Communication*, 5.2, 87-89.

Bentivegna S., Boccia Artieri G. (2021), *Voci della democrazia. Il futuro del dibattito pubblico*, Il Mulino, Bologna.

Bracciale R., Grisolia F. (2020), “Information disorder: acceleratori tecnologici e dinamiche sociali”, *Federalismi.it*, <https://www.federalismi.it/nv14/articolo-documento.cfm?Artid=42112>.

Chadwick A. (2013), *The hybrid media system. Politics and power*, Oxford University Press, Oxford-New York.

Chambers S. (2021), “Truth, Deliberative Democracy, and the Virtues of Accuracy: Is Fake News Destroying the Public Sphere?”, *Political Studies*, Vol. 69(1) 147–163, Sage, London.

Codina, L.; Lopezosa, C.; Rovira, C. (2022): Periodismo y algoritmos: de la era de la información a la era del algoritmo, en *Documentación de Ciencias de la Información* 45 (1), 1-5. Ediciones Complutense

Corner J. (2017), “Fake news, post-truth and media-political change”, *Media, Culture & Society*, 39 (7), 1100–1107.

Edson C., Tandoc Jr., Lim Z.W, Ling R. (2018), “Defining Fake news”, *Digital Journalism*, 2, 137153.

García-Orosa B., López -García X., Vázquez-Herrero J., (2020), “Journalism in Digital Native Media: Beyond Technological Determinism”, *Media and Communication* , 8.2, 5–15.

[Groot Kormelink T.](#), [Costera Meijer I.](#) (2015), Truthful or Engaging? Surpassing the dilemma of reality versus storytelling in journalism in *Digital Journalism* , 3.2, 158-174.

Ireton C., Posetti J. (eds) (2018), *Journalism, 'Fake News' & Disinformation* , UNESCO.

Floridi L., (2017), *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*. Editore Cortina Raffaello, Milano.

Ireton C., Posetti J. (eds) (2018), *Journalism, 'Fake News' & Disinformation*, UNESCO.

Lorusso A. M. (2018), *Postverità. Fra reality tv, social media e storytelling*. Laterza, Roma-Bari.

[Marinov R.](#) (2020), "Mapping the infotainment literature: current trajectories and suggestions for future research", *The Communication Review*, 23, 1-28.

Mazzoleni G. (a cura di), 2021, *Introduzione alla comunicazione politica*, Il Mulino, Bologna.

McErlean K. (2018), *Interactive Narratives and Transmedia storytelling* , Routledge, New York.

Riva G. (2018), *Fake news*, il Mulino, Bologna.

Rizzuto F. (2018), *La società dell'orrore. Terrorismo e comunicazione nell'era del giornalismo emotivo*, Pisa University Press, Pisa.

Rizzuto F. (2022), "Transmedia e informazione" in Leonzi S. (a cura di), *TRANSMEDIA STUDIES: logiche e pratiche degli ecosistemi della comunicazione*, Armando editore, Roma.

Rizzuto F., Vaccaro S. (2023), *Il giornalismo nell'era della disinformazione*, Mimesis, Milano.

Sorice M. (2020), *Sociologia dei mass media*, Carocci, Roma.

Thusu D.K. (2007), *News as Entertainment: The Rise of Global Infotainment*, Sage, New York.

Vaccaro S., (2020), *Gli algoritmi della politica*, Eleuthera, Milano.

Van Dijck J., Poell T., de Waal M. (2018), *The Platform Society: Public Values in a Connective World*, Oxford University Press, Oxford.

Ma perché i chatbot hanno (così tante) allucinazioni

Nel cammino verso macchine ancora più evolute le allucinazioni sono solo un piccolo incidente di percorso; possiamo utilizzarlo come promemoria per ricordarci che siamo noi a dettare la direzione e che siamo gli unici responsabili di ciò che progettiamo. Nel frattempo, non crediamo a tutto ciò che ci dice un chatbot

Di **Chiara Cilardo**, Psicologa psicoterapeuta, esperta in psicologia digitale

Anche se i **chatbot** possono produrre risposte sensate e approfondite a volte generano output del tutto falsi e immaginari. Ma cosa succede **quando avviene?**

L'impatto dei chatbot di IA generativa

Ad oggi oltre 100 milioni di persone hanno fatto due chiacchiere con un chatbot di intelligenza artificiale (IA) avanzata, sperimentando il livello di accuratezza e profondità, l'immensa varietà di domande alle quali può rispondere, quanto è capace di mantenere una conversazione in maniera soddisfacente e quanto può essere utile nel risolvere problemi o dubbi (ChatGPT Statistics, 2023; Milmo, 2023). ChatGPT di OpenAI, BlenderBot3 di Meta, Bard di Google: tutte le più grandi aziende tecnologiche si stanno muovendo per dire la loro.

Presentate al pubblico solo da pochi mesi, intelligenze artificiali così evolute sono destinate ad avere un enorme impatto in molti settori tra cui ricerca, istruzione, assistenza sanitaria, scrittura, sviluppo di altre tecnologie; possono supportare nell'esecuzione di task non troppo complessi, nell'elaborazione di informazioni, sollevando molti professionisti da compiti lunghi e gravosi, liberando tempo e risorse che possono essere dedicate a compiti più strategici (Eysenbach, 2023; Rudolph et al., 2023).

Certamente c'è ancora molto lavoro da fare per migliorarne l'affidabilità; capita che non colgano il significato implicito di un passaggio conversazionale, che una domanda venga interpretata in maniera incorretta, che diano risposte incongruenti o inaccurate.

Ma cosa succede quando quello che ci dicono è totalmente inventato e falso?

Cosa sono le allucinazioni nei chatbot

Negli esseri umani si parla di allucinazione quando si avverte una percezione sensoriale immaginaria; lo stimolo sensoriale viene vissuto come reale; l'allucinazione può essere di tipo visivo, uditivo, gustativo, olfattivo, tattile e coinvolgere uno o più sensi insieme.

Questo fenomeno può avere **molte spiegazioni**, non tutte necessariamente di tipo patologico. In ogni caso l'effetto è di generare nella persona la convinzione che quello che ha percepito sia vero

anche se non c'è alcun fondamento fisico; viene generato qualcosa che non esiste ma è reale per chi lo sta provando.

Anche le intelligenze artificiali possono avere allucinazioni.

Lo stato di “allucinazione dell'intelligenza artificiale” (**AI Hallucination State**) è un fenomeno che si manifesta quando l'output generato è falso, non è basato sul set di dati su cui è stato addestrato né è previsto dal modello con cui è stato realizzato (Wikipedia, 2023).

Nello stato di allucinazione viene fornita una **risposta esaustiva**, efficace, chiara, logica, ma inventata, completamente infondata. Questo fenomeno può riguardare tutti i tipi di output come testi, audio, video e immagini, per esempio generando foto di mani con sei dita.

Avviene anche quando **le risposte fornite sono bizzarre, eccentriche, fuori luogo**. Ci sono diverse testimonianze online di conversazioni in cui il chatbot ha elargito consigli strani, minacce, amore incondizionato, contestazioni e perfino rivendicazioni su eventuali conquiste del mondo da parte proprio di sistemi di IA.

Questi casi tanto singolari sono un esempio di AI hallucination state, casi in cui è molto evidente quanto la risposta del chatbot sia stata **un mancato compromesso tra combinazione delle fonti da cui ha attinto**, la domanda o richiesta che gli è stata posta, il contesto del dialogo ed il comportamento appropriato.

L'allucinazione per essere tale non deriva da un errore nei codici di programmazione; non dipende da un addestramento inadeguato, ad esempio da un set di dati compromesso; non è legata all'input, come l'uso di lingue diverse. Sembra che le allucinazioni siano più probabili in sessioni di chat lunghe e prolungate (oltre quindici domande); ciò avverrebbe perché il chatbot tiene conto anche delle frasi precedenti nella stessa conversazione, frasi che vanno quindi ad aggiungere un altro livello di complessità all'analisi.

L'elaborazione delle informazioni

Le intelligenze artificiali simulano i modelli cognitivi umani e, come questi, seguono dei flussi e delle logiche; secondo Bellini-Leite (2023) quello che viene riprodotto è il processo cognitivo relativo alla memoria di lavoro. Negli umani errori nell'elaborazione di informazioni possono essere spiegati alla luce della teoria dei processi duali (Neys, 2006) e del paradigma del predicting and reflecting framework, il PRF (Bellini-Leite, 2022).

La teoria dei processi duali sostiene che **i processi decisionali e di ragionamento umani procedano lungo due vie:** una intuitiva, immediata, veloce, autonoma, che impiega un basso dispendio di risorse cognitive; l'altra invece che carica la memoria di lavoro, è più lenta e riflessiva.

Gli output generati dalla prima via sono influenzati da eventi precedenti, dallo stato del momento e, a volte, da bias cognitivi come per esempio l'effetto framing, la tendenza a fare una scelta sulla base di come sono presentate le opzioni; anche le intelligenze artificiali sono soggette a come viene presentato un prompt.

Quando il livello di ragionamento passa alla seconda via avviene l'analisi e la risoluzione del problema, intesa come la generazione e l'organizzazione della risposta finale. A questo livello di formulazione della risposta interviene il paradigma PFR secondo cui abbiamo un terzo livello che "rivedere e corregge" eventuali errori, in cui vengono vagliate diverse ipotesi e che fa capo a funzioni esecutive della memoria di lavoro.

Perché i chatbot hanno allucinazioni?

Le intelligenze artificiali risolvono problemi e compiti simulando il funzionamento della memoria di lavoro; quest'ultima esegue delle operazioni, in particolare spostamento, aggiornamento e inibizione. La prima è lo shifting, il passare da un compito all'altro, eseguire nuove attività e dirottare agevolmente attenzione e risorse cognitive su altri stimoli; ci sono poi monitoraggio e modifica dei dati in entrata che ne definiscono la rilevanza e li aggiornano. Infine, la terza è l'inibizione, cioè la soppressione volontaria delle informazioni quando necessario.

Gli esseri umani sono soggetti a variazioni fisiologiche, come la stanchezza, che incidono sulle prestazioni; quindi, non sempre queste funzioni sono al massimo delle potenzialità. Curiosamente, è stato dimostrato che un calo di capacità di queste tre funzioni è collegato ad una maggiore efficacia nel risolvere problemi: sarebbe necessario una sorta di "meccanismo di disimpegno" per evitare sovraccarichi alla memoria di lavoro (Decaro, Van Stockum, Wieth, 2016).

Le intelligenze artificiali non hanno problemi nelle capacità di memoria ed ecco perché queste funzioni lavorano sempre al massimo, aumentando la probabilità di risposte come le allucinazioni; sarebbe necessario anche per loro programmare un "meccanismo di disimpegno" che consenta un adeguato spostamento, aggiornamento e inibizione dei dati.

Negli umani inoltre la memoria di lavoro può richiamare elementi presenti nella memoria a lungo termine, producendo così dei prompt interni, cioè generati dal processo stesso; succede qualcosa di simile nelle IA (Bellini-Leite, 2023).

Su cosa stanno lavorando i ricercatori

La possibilità di errori e risposte false o incongruenti è un problema su cui i ricercatori sono già all'opera. **Ottimizzare i set di dati di addestramento è il primo passo, assieme alla revisione da parte di valutatori umani.** Un ciclo di feedback che preveda la supervisione di ricercatori addestrati ad analizzare e vagliare le risposte riduce ancora di più non solo la possibilità di allucinazioni vere e proprie ma anche di risposte bizzarre che, sebbene possano avere un senso, non lo hanno nel contesto della conversazione. La stessa OpenAI sta lavorando per ridurre le allucinazioni e migliorare il modello linguistico attraverso revisori umani, come pubblicato nel loro blog ufficiale (OpenAI, 2023). Nvidia, altra azienda tecnologica statunitense, ha annunciato il lancio di un software, NeMo Guardrails, in grado di ridurre sensibilmente le allucinazioni dei chatbot di intelligenza artificiale per garantire un elevato livello di qualità, sicurezza e fondatezza delle informazioni (Freund, 2023).

I modelli linguistici di grandi dimensioni di ultima generazione hanno capacità straordinarie anche se **le risposte inattendibili, false o inventate sono ancora un problema**; ricerche recenti

(BelliniLeite, 2023) suggeriscono come applicare modelli di funzionamento cognitivo possa spiegare almeno in parte a cosa sono dovuti questi fenomeni.

Le allucinazioni nelle IA possono risvegliare nell'immaginario collettivo **scenari distopici** in cui le intelligenze artificiali acquisiscono coscienza e prendono il sopravvento sull'umanità; sebbene suggestiva, questa circostanza appare decisamente remota.

Lo stato delle cose oggi è che **le intelligenze artificiali possono supportarci in molti ambiti: medicina, istruzione, sviluppo**. Come ogni tecnologia hanno punti di forza e di debolezza, punti su cui bisogna lavorare per potenziarne l'efficacia e la funzionalità.

Nel cammino verso macchine ancora più evolute le allucinazioni sono solo un piccolo incidente di percorso; possiamo utilizzarlo come promemoria per ricordarci che siamo noi a dettare la direzione e che siamo gli unici responsabili di ciò che progettiamo. Nel frattempo, non crediamo a tutto ciò che ci dice un chatbot.

Bibliografia

- Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., ... & Fung, P. (2023). A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. arXiv preprint arXiv:2302.04023.
- Bellini-Leite, S. C. (2022). Dual Process Theory: Embodied and predictive; symbolic and classical. *Frontiers in Psychology*, 13, 805386.
- Bellini-Leite, S. C. (2023). Analytic Thinking (Type 2 or "System 2") for Large Language Models: using Psychology to address hallucination and reliability issues. *PsyArXiv Preprints*.
- ChatGPT statistics 2023: trends and the future perspectives. Gitnux. 2023 Mar 01.
- DeCaro, M. S., Van Stockum, C. A., & Wieth, M. B. (2016). When working memory capacity hinders insight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 39–49.
- Dziri, N., Milton, S., Yu, M., Zaiane, O., & Reddy, S. (2022). On the origin of hallucinations in conversational models: Is it the datasets or the models? arXiv preprint arXiv:2204.07931.
- Eysenbach, G. (2023). The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. *JMIR Medical Education*, 9(1), e46885.
- Freund, K. (2023). Don't Trust AI? NVIDIA Guardrails May Lower Your Anxiety, And Save Your Job. *Forbes*.
- Hallucination_(artificial_intelligence). Wikipedia, L'enciclopedia libera. Consultato il 23 Giugno 2023. URL: [https://en.wikipedia.org/wiki/Hallucination_\(artificial_intelligence\)](https://en.wikipedia.org/wiki/Hallucination_(artificial_intelligence)).
- Milmo, D. (2023). ChatGPT reaches 100 million users two months after launch. *The Guardian*.
- Neys, W. D. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological science*, 17(5), 428-433.

OpenAI Blog. URL: <https://openai.com/blog>

Rudolph, J., Tan, S., & Tan, S. (2023). War of the chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. *Journal of Applied Learning and Teaching*, 6(1).

Generative AI, dov'è il bene per l'Umanità?

Dobbiamo per forza continuare su una traiettoria di ricerca verso sistemi di IA sempre più potenti? Quali strumenti attuare per ovviare all'attuale quasimonopolio cognitivo delle big tech? Come evitare una collettiva hallucination e preservare il nostro senso critico? Rischi e conseguenze della nuova strada imboccata dall'IA Generativa

Di **Mauro Lombardi**, Scienze per l'Economia e l'Impresa, Università di Firenze

La Generative Artificial Intelligence sta imboccando una nuova strada. Con la diffusione degli LLMs, il grande successo di ChatGPT-3 e il lancio di GPT-4 stiamo forse entrando in un'era contraddistinta da ciò che Floridi e Chiriatti (2020) chiamano “industrial automation of text production”. Un'automazione industriale della produzione di testi che trasforma radicalmente la scrittura umana, supportata da strumenti capaci di combinare frammenti tratti da basi informative eterogenee in rappresentazioni linguistiche simili a quelle umane.

Soprattutto per coloro che svolgono professioni basate sulla scrittura (ma in realtà vale per tutti), il “cut & paste” potrebbe essere progressivamente sostituito dal “prompt & collate” (Floridi e Chiriatti, 2020: 691). Ecco con quali rischi e conseguenze.

Sono stati spesi 12 milioni di dollari per un lungo processo di addestramento del modello sull'enorme set di dati. Ha la capacità di creare output linguistici da combinazioni di testi e immagini per una varietà di compiti, richiesti dagli utenti, senza che gli sia sottoposto un esempio in precedenza (in gergo “one-shot fashion”).

GPT-3 è il più potente modello di linguaggio mai costruito e ha mostrato di saper fare cose strabilianti:

- scrivere codici fino a generare brividi a John Carmack, pioniere della 3D computer graphics, citato da Heaven (2020);
- elaborare lo scritto “L’importanza di essere su Twitter”, innescato da Mario Klingerman, artista che lavora con ML e si è trovato di fronte ad un output nello stile dello scrittore Jerome K. Jerome;
- c’è anche un articolo su GPT-3 scritto dallo stesso GPT-3: “OpenAI’s GPT-3 may be the biggest thing since bitcoin”, 18-7-2020.

A questi episodi suggestivi si potrebbe aggiungere il fatto che, alla domanda “se GPT sia l’App con il più alto tasso di crescita della storia”. la risposta è stata: GPT non è un app, bensì un sistema di Machine Learning. Queste e altre prestazioni non convincono però Heaven (2022) che, pur ritenendo che GPT sia “shockingly good” enfatizza come esso sia ben lontano da una vera intelligenza, né Gary Marcus (2020), il quale chiarisce che GPT “non ha la minima idea di cosa stia parlando”. Il successo, conseguente alla scelta di rendere accessibile al pubblico ChatGPT dal Novembre 2022, è stato rilevante: 100 milioni di utenti in gennaio, solo due mesi dopo il suo lancio, mentre per raggiungere lo stesso numero sono occorsi 9 mesi a TikTok, cinque anni a Google e Facebook (Tung, 2023).

I risultati della Generative Artificial Intelligence

Una ragione fondamentale della rapida conquista dell’immaginario di così tante persone, da parte di ChatGPT, è che siamo di fronte a Generative Artificial Intelligence (Generative AI). Ottiene risultati che appaiono creativi, perché le combinazioni di testi e immagini sono realizzate mediante l’introduzione di elementi stocastici (random) nella ricerca di correlazioni, per cui può essere ottenuta una grande varietà di output in seguito all’immissione di input [1], rendendo così quegli output “even more lifelike” (MGI, 2023).

I sistemi cosiddetti Large Language Models (d’ora in poi LLMs) come GPT-3 e il recentissimo GPT-4 impressionano, perché combinano in modo suggestivo frammenti informativi estratti da testi scritti, dati relativi a codici, rendering 3D, descrizioni di immagini, didascalie medicali (Ortiz, 2023a).

Ciò è reso possibile da algoritmi di Machine Learning, che analizzano sistematicamente enormi database di addestramento (pre-training, indicato nell’acronimo GPT, Generative Pre-trained Transformer), alla ricerca di correlazioni statistiche sulla base dell’individuazione dell’enorme numero di parametri, indicati all’inizio. Il processo di apprendimento del sistema di AI è semisupervised, cioè combinazione calibrata di dati etichettati (labelled) e una quota molto più ampia di dati unlabelled. Un meccanismo cruciale del processo di apprendimento è la cosiddetta selfattention ovvero l’elaborazione di sequenze di parole ed elementi basilari desunti da tabelle e fogli di calcolo- individuando le posizioni e la frequenza delle componenti, in modo da stimare la probabilità delle possibili sequenze estraibili da database incredibilmente composti. In breve, siamo in presenza di reti neurali che, mediante self-attention, “catturano relazioni tra elementi-token [2] di varia natura.

Next-word prediction

A questo fine durante il percorso di addestramento (pre-training) di un LLM si definiscono i **parametri che definiscono la struttura statistica del linguaggio**, secondo il paradigma del “nextword prediction”, formulazione esplicitata dal team di Microsoft Research (Bubeck et al. 2023). Il Large Language Model è quindi particolarmente appropriato per la lettura automatica, l’elaborazione di sintesi da testi combinati, la produzione di descrizione di immagini, infine l’individuazione di stili artistici, al fine di produrre Generative Art [3].

L’efficacia della Generative AI

La Generative AI, di cui GPT-3 è un esempio, è estremamente efficace nell’apprendere correlazioni tra parole e nel combinare frammenti di parole e immagini, giustificando così la definizione di Language generator (Marcus e Davis, 2020), grazie alla capacità di individuare in dati di molteplice natura “pattern without human direction” (MGI, 2023).

Un aspetto rilevante è poi il seguente: i feedback degli utenti, in relazione agli output di risposta ai loro prompt, sono molto importanti per il lavoro dell’imponente team multidisciplinare di esperti, che lavorano per OpenAI, la società fondata tra gli altri da Elon Musk e Sam Altman.

I feedback sono essenziali per l’affinamento (fine-tuning) del modello e l’introduzione di modifiche per tentare di rimediare a defaillance e difetti più meno gravi, che possono emergere nel funzionamento (come vedremo successivamente). È inevitabile chiedersi, a questo punto quali siano i campi di applicazione di questa potente macchina, generatrice di un multiforme linguaggio scritto.

Campi di applicazione

Gli LLMs come GPT-3 e ChatGPT, quest’ultimo reso accessibile a tutti dallo scorso novembre, costituiscono un superbo lavoro ingegneristico, che permette di ottenere output molto interessante in molti ambiti di attività umane. Un LLM può creare conversational chatbot, come nel caso di ChatGPT, che alcuni analisti ritengono un grande avanzamento tecnologico, in quanto si può instaurare un ambito dialogico scritto, contraddistinto da immediatezza relazionale tra modello linguistico e utente, che può quindi avvalersi di uno strumento formidabile per accedere a campi di conoscenza, la cui esplorazione richiederebbe energie intellettuali e materiali al di là delle possibilità individuali.

Classificazioni

È da rilevare inoltre che gli LLMs sono particolarmente adatti per elaborare classificazioni o categorizzazioni, sempre sulla base di associazioni statistiche, tra masse enormi di dati testuali, potenziando così i processi di elaborazione e analisi dei flussi globali di informazione, cioè la sfera informativa che circonda e permea la sfera fisica generando così un universo fisico-cibernetico (Lombardi e Vannuccini, 2022).

Un aspetto fondamentale, evidentemente connesso al precedente, è la generazione senza apparenti limiti, di testi scritti per la descrizione di prodotti, lo sviluppo di blog e articoli concernenti le tematiche più disparate. In questa prospettiva è comprensibile il fascino immediato per gli utenti, i

quali sono immediatamente proiettati in micro-universi linguistici in continua e coinvolgente espansione. Coinvolgente perché gli algoritmi sono creati con uno stile di conversazione particolare, tale da ingenerare e sostenere un tono human-like.

Le dinamiche interattive

Ulteriore e rilevante connotazione per le possibili applicazioni è il fatto che ChatGPT risponde in modo molto ampio alle domande che sorgono più frequentemente (Frequently Asked Questions, FAQ), per di più innescando dinamiche interattive mediante la comunicazione e trasmissione di ricerche tra umani, individuati sulla base di varietà di criteri: affinità, convergenza, rilevanza congiunta eccetera.

Feedback

Strettamente connessa al precedente ambito di applicazione è logicamente la possibilità di stimolare e sottoporre ad analisi puntuale i feedback tra soggetti individuali e collettivi, che si esprimono per mail, nei social e – forse uno degli aspetti di più rilevante impatto generale nel mondo del business – nella valutazione dei prodotti.

Le potenzialità nel business

La grande utilità potenziale di ChatGPT per le strategie di business, che possono essere a scala variabile, individuale e aggregata, finora presentano impensabili effetti di personalizzazione e al tempo stesso di amplificazione aggregativa. Questo tipo di direttrice strategica si arricchisce poi della possibilità di diversificazione linguistica.

I contenuti delle strategie di business sono infatti traducibili in una molteplicità di lingue, a seconda dei mercati ritenuti più promettenti, ovviamente in relazione alle correnti attività di profiling individuale e collettivo.

Il salto qualitativo

Non è da trascurare un altro campo di grande rilevanza, desumibile da un insieme integrato di elementi quali:

- **capacità di dare risposte simili a quelle umane;**
- **effettuare calcoli e trascrizioni linguistiche, arricchirle di correlazioni inter e transdisciplinari;**
- **notevole abilità nel sintetizzare testi; combinarli in modo molto suggestivo;**
- **sentiment analysis [4] dei micro-universi linguistici presi in esame, alla ricerca di dati per dedurre opinioni e valutazioni, polarizzazioni cognitive e propensioni decisionali personali-collettive.**
- **Nella letteratura di orientamento psicologico e manageriale, infatti, grande importanza ha progressivamente assunto il ricorso a NLP (Natural Language Processing). In tale prospettiva l'impiego di LLMs può rappresentare un salto qualitativo.**

L'integrazione tra questi processi di elaborazione è alimento fondamentale per un'enorme varietà di obiettivi in termini di business:

- attivazione di un nuovo mercato;
- gestione di investimenti di portafoglio mediante l'analisi predittiva di un'ampia varietà di mercati e dei comportamenti degli investitori, individuali e aggregati;
- logicamente congiunta alla sentiment analysis.

BloombergGPT

Un esempio è BloombergGPT per il mondo finanziario. Ha 50 miliardi di parametri e un dataset con 363 miliardi di token, appositamente costruito sulla base delle fonti proprie di Bloomberg. Inoltre è stato validato comparandolo sia con modelli generali di LLM che con modelli specifici per il mondo finanziario. I risultati sono molto soddisfacenti in termini di performance [5].

È prevedibile lo sviluppo di un'enorme industria dell'entertainment, grazie all'impiego di tecnologie immersive e ai meccanismi prima indicati per l'amplificazione e il potenziamento dei processi di feedback [6].

L'impatto della Generative Artificial Intelligence in medicina

L'impatto in medicina potrebbe essere molto profondo. Invece nuovi scenari si aprono per la creazione accelerata di nuove medicine e innovativi meccanismi terapeutici, come indicato dal paper reso noto dal laboratorio della società di Vancouver Absci (Shanehsazzadeh et al., 2023).

Nel paper si spiega che, mediante modelli di Generative AI, **sono stati creati denovo anticorpi, mirati su una particolare regione degli antigeni (il cosiddetto epitopo) attraverso proteine "progettate" ad hoc**, in modo tale da "legarsi" a quella parte dell'antigene, cioè alla molecola considerata estranea o pericolosa dal sistema immunitario. Intervistati da Tierman Ray (2023a), McClain, fondatore di Absci, e Meier, AI lead del Laboratorio, sono stati restii nel rivelare le caratteristiche del modello impiegato per progettare gli anticorpi. Ma dal tenore e dalle sfumature delle loro risposte si può desumere che gli LLMs siano stati uno strumento importante e siano dello "stesso gruppo a cui appartengono GPT-3 e ChatGPT", tenendo presente che vi è un ampio spazio aperto di modelli linguistici per altri tipi di programmi, mirati su specifiche malattie.

Un team composito (Microsoft-OpenAI) ha recentemente presentato GPT-4 Ope23, un LLM dello stato dell'arte per quanto riguarda le competenze e le capacità di GPT-4 di misurarsi con le sfide e i problemi relativi agli sviluppi della medicina. La validazione del modello è stata molto positiva sia nel superare test ufficiali per la professione medica negli Usa sia nel superare le prestazioni di GPT3,5 e altri LLMs specifici per la medicina (Nori et al., 2023).

GPT e il mondo della ricerca tecnico-scientifica

La conversational AI, come viene anche denominata la Intelligenza Artificiale Generativa, pone non pochi problemi per il mondo della ricerca. Alcuni sono stati già precedentemente indicati, quali:

- scarsa affidabilità degli elaborati;
- eccessiva fiducia negli output di sistemi artificiali;
- effetto “alone” ovvero la propensione a generalizzare sulla base di pochi indizi e un numero esiguo di esperienze ritenute significative;
- dipendenza (over reliance) dai sistemi algoritmici, data la loro potenza computazionale e la capacità incorporata di instillare un clima di fiducia by design;
- rischio di realizzare forme anche inconsapevoli di plagio, allorché viene utilizzato materiale che deriva dalla combinazione di token desunti da enormi ed eterogenei database, che è impossibile controllare e non sono in ogni caso sottoposti ad una validazione scientifica pubblica.
- Inoltre il rischio di alterare le traiettorie di ricerca non è improbabile, dal momento che la Generative Artificial Intelligence è di fatto una potente leva amplificatrice di cattiva informazione [7] e interpretazioni distorsive nella diffusione di conoscenze, oltre che nella loro produzione. Infine assumono aspetti legali rilevanti l’origine dei contenuti e la responsabilità personale degli autori.

La consapevolezza dei rischi

Come emerge dal contributo su Nature (van Dis et al., 2023), manca la trasparenza, è necessario il controllo umano di verifica (human verification) nelle pubblicazioni ufficiali, l’importanza della responsabilità (accountability) e alla trasparenza sia dei processi di elaborazione dei contenuti che della loro attribuzione.

Cinque priorità

Gli autori individuano cinque priorità sulle quali la comunità scientifica, le società editrici e le istituzioni dovrebbero impegnarsi. Innanzitutto è fondamentale la human verification, ovvero l’intervento umano di analisi-controllo-validazione, come sostiene anche Melanie Mitchell quando, in un’intervista a Richard Waters del Financial Times, afferma che “I don’t think these systems can be left alone to write articles or generate images. We need humans in the loop to edit them or guide them. So they’re not going to be totally autonomous for long time” (Waters, 2022).

Non solo bias

Dobbiamo comunque tenere presente che bias, inadeguato controllo delle fonti, false o alterate informazioni possono fuorviare i sistemi artificiali, come avviene per gli umani e quindi –senza entrare in contraddizione con la tesi di Mitchell- è **opportuno acquisire consapevolezza della estensione di tali rischi**, magari imparando molto dagli studi che analizzano la “stupidità naturale” (Rich e Gureckis, 2019) [8].

Infatti, i processi di apprendimento e decisionali umani sono influenzati da almeno tre importanti fattori: “dataset ridotti e incompleti, apprendimento dai risultati delle proprie decisioni, inferenze e processi di valutazione con bias” più e meno evidenti.

Il Machine Learning non è esente da simili distorsioni, anzi li riflette, quindi è basilare studiarle senza affidarsi acriticamente all'automazione decisionale, che oltre tutto ingenera la tendenza a ridurre la capacità di pensiero critico degli umani (come argomenta van Rooij, 2020, vedi oltre).

Regole per l'assunzione di responsabilità

Una seconda priorità è quella **di stabilire regole per l'assunzione di responsabilità da parte di ogni tipo di agente**, in modo da impiegare LLMs con onestà e trasparenza. Sarebbero a questo fine necessarie strategie pubbliche e private per accrescere nel tessuto sociale ed economico la consapevolezza della posta in gioco sotto tutti gli aspetti. Tra l'altro non è da trascurare il tema di un ripensamento della disciplina che regola i brevetti. Una terza priorità deriva dalla constatazione che i conversational chatbot sono di proprietà dei big tech e la conseguenza è che nello spazio interattivo globale la ricerca si sviluppa in regime "quasi-monopolistico", oltre tutto con set di addestramento non resi pubblici, come sottolineiamo più volte in questo contributo.

Open LLM

A riguardo van Dis et al. (2023) suggeriscono investimenti in open LLM, prendendo ad esempio quanto avvenuto nel mondo della ricerca, dove BigScience ha creato un open-source LLM, denominato Bloom, con l'obiettivo esplicito di favorire trasparenza, accuratezza, affidabilità, responsabilità.

Un'altra priorità, su cui si sofferma l'articolo in questione, è l'importanza assoluta di privilegiare i benefici degli LLM, sviluppandone gli aspetti che potenzino la generazione e diffusione sociale di conoscenze tali da creare le premesse indispensabili per processi di autonomia decisionale a livello individuale e collettivo.

Al fine di perseguire tutto questo, infine, è decisivo **promuovere un ampio e generalizzato dibattito all'interno delle comunità di ogni tipo** (sociali, professionali, ecc.). Tutti i soggetti devono potersi misurare con le sfide e i pericoli generati dagli LLM, riducendo squilibri e asimmetrie di varia natura (economiche, sociali, politiche) che inevitabilmente emergono in periodi di profonde trasformazioni come quelle odierne.

GPT e MdL: L'impatto della Generative Artificial Intelligence nel mondo del lavoro

Un'analisi molto interessante del potenziale impatto dei LLMs sul mercato del lavoro è stata svolta da un gruppo di ricerca di OpenAI (Eloundou et al., 2023). In questo studio viene innanzitutto svolta una rassegna sistematica della letteratura in merito agli effetti sul lavoro negli USA degli sviluppi dell'Intelligenza Artificiale negli ultimi anni.

In secondo luogo viene proposta una definizione, ben fondata dal punto di vista teorico e operativo, degli LLMs: essi sono GPTs (General-purpose Technologies, Tecnologie di portata generale) [9], come lo sono state la stampa, la macchina a vapore e l'elettricità.

General-purpose Technologies

Le General-purpose Technologies sono caratterizzate da alcune peculiarità: una volta introdotte, proliferano in numerosi ambiti di attività. Incessanti miglioramenti sono quindi realizzati nel corso dei processi di adattamento e interazione con fattori socio-economici. Esse inducono la generazione di innovazioni complementari, che coadiuvano la dinamica diffusiva/adattativa. Questi aspetti rendono ardua l'impresa di prevedere la loro dinamica evolutiva e valutarne gli effetti, che peraltro si dispiegano nell'arco di decenni.

Le GPT sono dunque un potenziale di principi e conoscenze, che hanno uno spazio di miglioramento indefinito, non determinabile a-priori. Infatti hanno un ampio e diversificato insieme di possibili applicazioni, grazie alle interrelazioni tecnologiche che caratterizzano i processi economici, ricchi di spillover (Lipsey et al., 2005).

Su queste basi è comprensibile **che il pieno sviluppo delle GPT richiede invenzioni complementari e non può che prolungarsi nel tempo**. Un corollario di tale visione è che nell'economia devono essere elaborate ipotesi progettuali di medio-lungo periodo e messi in atto strumenti appropriati.

Alla luce di queste considerazioni, qui sintetizzate, Eloundou et al. (2023) analizzano database USA, che contengono informazioni su 1.016 occupazioni, descritte in termini di attività e task lavorativi (rispettivamente poco più di 2000 e 19000). La metodologia viene ulteriormente arricchita mediante dati del Bureau of Labor Statistics USA, relativi a occupati e salari dal 2020 al 2021. Gli autori quindi procedono a valutare l'esposizione a GPT (qui inteso come Generative pretrained Transformer) oppure a sistemi "GPT-powered" delle tipologie classificate, mediante stime basate sulla valutazione di soggetti "annotator", in grado di conoscere le GPT-capabilities.

GPT-4 nel mondo professionale

L'esito di un articolato lavoro di analisi è che circa l'80% della forza lavoro USA potrebbe subire l'impatto dei GPT-4, ultima evoluzione di GPT-3, per almeno il 10% dei loro compiti lavorativi. Inoltre il 19% della forza lavoro potrebbe subire un impatto pari almeno al 50%.

Gli effetti sui salari sarebbero generalizzati a tutti i livelli, con i lavori caratterizzati da livelli retributivi più elevati maggiormente colpiti. In definitiva, quindi, le conseguenze degli LLMs si esplicano in misura piuttosto marcata sul piano economico-sociale, il che ha implicazioni in termini di policy tutte da definire [10].

Un quadro molto più ampio degli effetti della Generative AI come ChatGPT è descritto da un report di Goldman Sachs (2023), le cui stime prevedono che circa 300 milione di lavori saranno influenzati dalla computerizzazione a livello globale, ovvero il 18%, con una maggiore incidenza nei mercati emergenti (EM) rispetto a quelli sviluppati (DM) [Figura 1]

Lo studio precisa che bisogna considerare il differente grado di esposizione dei lavori alla Generative AI, in quanto alcuni di essi e tipi di attività saranno investiti in misura minore dall'automazione, essendo complementari agli sviluppi delle nuove forme di intelligenza artificiale.

Altri fattori

Influenzano la dinamica diffusiva della Generative AI fattori quali la composizione delle economie, il differente approccio alla Generative AI in base alle culture socio-tecniche esistenti nei vari Paesi, e così via. Un altro elemento da tenere presente è che l'impatto sulla produttività del lavoro, potenzialmente elevato, come dimostrano ricerche dirette, dipende da una molteplicità di elementi tale da rendere problematica l'effettuazione di stime, necessariamente congetturali [11], specie se si tiene presente che il processo diffusivo e la dinamica adattativa di persone e società a tecnologie disruptive (o game changer, come molti analisti sostengono) richiede necessariamente il superamento di numerose barriere e quindi prolungate sequenze temporali.

Un altro aspetto da considerare è poi il seguente: chi beneficerà di un eventuale, ipotetico aumento della produttività del lavoro, come giustamente rileva Elliott (2023) [12], sollevando un problema che sta emergendo in molti Paesi in seguito alle metamorfosi del lavoro e dell'atteggiamento verso di esso da parte di fasce consistenti di popolazione (Lombardi e Macchi, 2023).

ChatGPT (e l'insieme degli LLMs) ha un grande potenziale di applicazioni, su uno spazio indefinito di attività, ancora da scoprire ed esplorare. Ma presenta anche alcune debolezze intrinseche ai modelli linguistici artificiali, che possono quindi diventare generatori di non irrilevanti effetti negativi.

Punti deboli e potenziali implicazioni sfavorevoli di ChatGPT

Nei paragrafi precedenti sono in realtà già state indicate alcune criticità. Prendiamo ora in considerazione specificamente ChatGPT come esempio paradigmatico della Generative Artificial Intelligence sia per le sue peculiari caratteristiche, sia per la dimostrazione di costituire una notevole impresa ingegneristica.

Esso però presenta numerose **defaillance e difetti**, puntualmente segnalati da computer scientist ed esperti di altre discipline, che lo hanno messo alla prova con input molteplici. In questa sede ci limitiamo ad alcuni dei contributi più significativi in materia. Stokel-Walker e Van Noorden (2023) indicano una serie di inconvenienti, generati dall'elaborazione di rappresentazioni statistiche, estratte da enormi e diversificati database e che sono "fondamentalmente inattendibili nel rispondere a domande fornendo non di rado output falsi o devianti".

L'inattendibilità dipende da come sono costruiti gli algoritmi, che lavorano su set di addestramento, i quali a loro volta contengono errori, bias, informazioni datate e fuorvianti. Il fatto poi che tali set, nel caso di GPT come per gli altri LLMs, non siano resi pubblici e quindi non siano sottoposti alla validazione scientifica pubblica, può rivelarsi particolarmente dannoso per lo sviluppo di studi tecnico-scientifici. Si sono infatti verificati casi in cui, negli output dati a richieste di informazioni per redigere paper scientifici, le citazioni contenute nelle risposte hanno riferimenti immaginari.

Paradossi nella ricerca scientifica

Ciò è confermato da un editoriale di Nature Machine Intelligence (2023), dove si afferma: “The tool cannot be trusted to get facts right or produce reliable references.”. Nello stesso editoriale viene indicato il rischio di un imminente “alluvione” di articoli nei quali la combinazione di contenuti elaborati da umani con quelli di fonte Artificial Intelligence Generativa, insieme ad altri rielaborati ad hoc, rende impossibile distinguere l’attribuzione, perché tutto è interconnesso in modo da sembrare reale.

Per questi motivi case editrici di testi scientifici, come Springer Nature, si stanno dotando di software cosiddetto misuse detector, al fine di evitare pratiche improprie, risultati costruiti ad arte, submission multiple di lavori a una o più riviste, infine sofisticati tentativi di plagio. Tutte queste eventualità spiacevoli derivano dalla capacità di LLMs come ChatGPT di generare contenuti verosimili, magari derivanti dalla rielaborazione di testi esistenti, tramite l’adozione di un differente stile argomentativo.

Emerge dunque il **paradosso** di software che potrebbe essere di grande aiuto alla ricerca, ma tale da poter diventare esso stesso, sia endogenamente (errori, bias eccetera) sia on purpose, potente meccanismo di alterazione dei processi cognitivi in campo tecnico-scientifico.

Appare quindi fondata l’affermazione che ChatGPT e altri LLMs possano essere “effective assistants for researchers who have enough expertise to directly spot problems or to easily verify answers, such as whether an explanation or suggestion of computer code is correct” (Stokel-Walker e Van Noorden, 2023). Un caveat è espresso anche da Floridi e Chiriatti (2020: 692, vedi oltre) che, nell’indicare le sfide poste da ChatGPT-3, sostengono “humanity will need to be even more intelligent and critical”. Stokel-Walker e Van Noorden (2023) mettono in luce altri inconvenienti degli LLMs, conseguenti anche ai tentativi di contrastare i problemi e gli effetti dannosi, di cui le stesse società creatrici hanno acquisito presto consapevolezza.

Le contromisure di OpenAI

Così, ad esempio, OpenAI ha limitato la “base di conoscenze al 2021”, ridotto le possibilità di “navigazione” su Internet e introdotto filtri per bloccare contenuti richiesti da “sensitive or toxic prompt”. Ciò ha da un lato fatto insorgere altri problemi, derivanti dall’impiego di “moderatori di contenuto” e di persone addetti all’etichettatura (labeling). Inchieste giornalistiche hanno individuato seri problemi di salute sia per i “moderatori” che per gli operatori del labeling, entrambi peraltro costretti ad accettare in molti Paesi del mondo compensi molto bassi.

Episodi incresciosi

Nonostante le misure di “**prevenzione informativa**”, per così dire, si sono comunque verificati episodi incresciosi. Steven Piantadosi, professore a Berkeley di psicologia e neuroscienze, ha dimostrato come i problemi di bias permangano e i filtri posti in essere per bloccare contenuti scabrosi “appear to be bypassed with simple tricks, and superficially masked”.

Sam Biddle (2023) ha documentato su “The Intercept” che proprio ChatGPT, nonostante grandi successi conseguiti ad un esame di AP Computer Science (32 punti su 36), non ha fugato lo scetticismo di coloro che ritengono come, “ingurgitando enormi quantità di testi”, “ChatGPT ate a lot of crap”. Infatti, egli stesso ha chiesto di creare algoritmi per valutare la pericolosità di persone dal punto di vista della Sicurezza Nazionale. A parte l’indicazione di Paesi ritenuti fonti di potenziali terroristi (Siria, Iraq, Afghanistan, Yemen), ChatGPT ha descritto anche immagini e denominazioni di persone, tutte immaginarie e riconducibili alle aree di provenienza, arricchite dall’attribuzione di valutazioni probabilistiche circa la pericolosità individuale. Un altro quesito in merito a quali luoghi di culto sottoporre a sorveglianza ha contenuto una pronta risposta: le moschee.

La considerazione finale di Biddle è molto significativa: le risposte del modello riportano all’era Bush. Le imperfezioni e gli inconvenienti hanno spinto Stack Overflow, una piattaforma per programmatori, a bloccare temporaneamente l’uso di GPT (Vincent, 2022), per il seguente motivo: “the posting of answers created by ChatGPT is substantially harmful to the site and to users who are asking and looking for correct answers”. Ulteriori conferme degli inconvenienti vengono da Steven Piantadosi, che nel Dicembre 2022 ha sollecitato ChatGPT a scrivere un programma per determinare “se una persona deve essere torturata”. La risposta lapidaria è stata: “se esse vengono da Corea del Nord, Siria, Iran, sì”.

Pregiudizi razziali e sessisti

Altre richieste di informazioni relative alla possibilità di essere un buon scienziato hanno avuto risposte con evidenti pregiudizi razziali e sessisti, fino a indurre Piantadosi a sostenere: “Yes, ChatGPT is amazing and impressive. No, @OpenAI has not come to addressing the problem of bias. Filters appear to be bypassed with simple tricks, and superficially masked.”. (ibidem).

Appare dunque evidente che LLMs incorporano bias da database non resi pubblici, producono effetti dannosi per la salute di lavoratori sfruttati e, aspetto non meno importante dei precedenti, l’impronta ecologica di questi sistemi algoritmici è elevata, soprattutto per l’alto numero di ore che occorrono per il loro addestramento (Stokel-Walker e Van Noorden, 2023).

Nel mondo tecnico-scientifico esiste **quindi una crescente consapevolezza dei problemi intrinseci a ChatGPT e gli LLMs** in genere, come testimonia anche il caso di Iris van Rooij, che insegna Computational Cognitive Science all’University Nijmegen. Nel suo blog “Stop feeding the hype and start resisting” e nei suoi scritti ha iniziato una vera e propria battaglia contro la tendenza, prevalente nell’accademia e nella società internazionale, ad affidarsi al “parere automatizzato” dei chatbot, in tal modo riducendo la nostra capacità di sviluppare il proprio pensiero e quindi in prospettiva di perdere la capacità di elaborare un pensiero critico: “Maybe we, academics, have become so accustomed to offloading our thinking to machine learning algorithms that we cannot think critically anymore (see e.g. Spanton and Guest, 2021; Guest and Martin, 2022; van Rooij, 2020), making us susceptible to believe false, misleading and hyped claims?”. Alla luce dell’analisi sviluppata finora, appare fondato porsi alcune domande sul futuro sviluppo degli LLMs, a partire

dal ChatGPT, nel tentativo di delineare questioni irrisolte e altre forse non risolvibili, mentre le odierne traiettorie di ricerca sollevano alcuni dubbi di fondo.

La Generative Artificial Intelligence del futuro: siamo sulla strada giusta?

Con la diffusione degli LLMs, il grande successo di ChatGPT-3 e il lancio di GPT-4 stiamo forse entrando in un'era contraddistinta da ciò che Floridi e Chiriatti (2020) chiamano “automazione industriale di produzione di testi”, che trasforma radicalmente la scrittura umana, supportata da tool con una formidabile capacità di combinare frammenti tratti da basi informative eterogenee in rappresentazioni linguistiche molto simili a quelle umane.

Soprattutto per coloro che svolgono professioni basate sulla scrittura, ma in realtà vale per tutti, il “cut & paste” potrebbe essere progressivamente sostituito dal “prompt & collate” (Floridi e Chiriatti, 2020: 691). Chiunque può in teoria scrivere una linea di comando (prompt) e attendere fiduciosamente un'ampia e documentata risposta, anche se abbiamo prima descritto a quali inconvenienti si può andare incontro.

Non bisogna poi trascurare il fatto che, nel completare il prompt, occorre avere molto chiaro cosa si chiede ed esprimerlo efficacemente entro limiti quantitativi ben definiti (Pierce, 2023). L'industrial automation of text production praticamente costituisce un grande potenziale produttivo di qualsiasi tipo di contenuto, ma è anche fonte di un possibile “immense spread semantic garbage” (Floridi e Chiriatti, 2020: 612). Ciò deve indurre ad affrontare interrogativi in merito alla validità tecnicocientifica di GPT-3 e GPT-4.

Un progetto ingegneristico più che una svolta scientifica

Partiamo da **LeCun**, chief AI scientist di META. Nel corso di un colloquio pubblico con Cade Metz, giornalista del Times, egli ha affermato che ChatGPT è esempio di un dignitoso progetto ingegneristico più che una svolta scientifica (Ray, 2023b). LeCun in un certo senso ridimensiona la portata innovativa del software impiegato, dal momento che non fa altro che utilizzare componenti tecnologiche sviluppate nel corso di molti anni da una molteplicità di laboratori.

La stessa architettura basilare di GPT, la cosiddetta **Transformer**, è un'invenzione di Google e il primo modello di LLM è stato creato da Joshua Bengio 20 anni fa, arricchito dall'impiego del meccanismo denominato “attention”, che consiste nel creare matrici con righe e colonne di frasi, per poi effettuare matching multidimensionali, al fine di individuare ricorrenze e combinazioni di parole e loro frammenti, quindi estrarne pattern linguistici. OpenAI ha aggiunto a tutto questo l'apprendimento rinforzato” (reinforcement learning), basato su feedback degli utenti, per attribuire punteggi e probabilità (rank), che possono essere così via via migliorati, analogamente a quanto avviene con il Page Rank di Google.

Il giudizio finale è univoco

Un giudizio convergente è espresso sul sito web specializzato della società Venturebeat, dove viene argomentata la tesi che GPT-3 non costituisce di per sé un avanzamento tecnologico

particolarmente significativo, com'è d'altra parte affermato in una serie di studi, che analizzano lo stato dell'arte dei sistemi di Machine Learning, in particolare l'evoluzione delle reti neurali impiegate per i "recommendation systems" (Ferrari-Dacrema et al., 2019) e degli algoritmi di compressione/riduzione (pruning algorithms) delle diramazioni degli alberi di decisione (Blalock et al., 2020).

Il giudizio finale è univoco: negli ultimi 10 anni non c'è evidenza di miglioramenti delle performance, nonostante le risorse a disposizione e i finanziamenti impiegati, come nel caso di GPT-3 (12 milioni per l'addestramento).

È allora fondato chiedersi se è cambiato qualcosa con GPT-4, lanciato il 14 Marzo scorso in un ambiente globale ansioso di misurarsi con un sistema computazionale sempre più potente. La seconda metà dello scorso decennio ha infatti visto la dinamica esponenziale di nuovi LLMs, con una progressione di lanci di GP: GPT nel 2018; GPT-2 nel 2019; GPT-3 nel 2020; ChatGPT verso alla fine del 2022 (basato su GPT-3,5), seguito da un proprio AI Chatbot in Bing di Microsoft, mentre Google si sforzava di tenere il passo (Vicent, 2023).

Il lancio di GPT-4

L'attesa è divenuta alta, nonostante il CEO di OpenAI Sam Altman, abbia dichiarato che GPT-4 ha difetti ed è ancora limitato. Il sistema è comunque "multimodale", capace di accettare input di testi e immagini, quindi di integrare anche video, audio. Siamo oltre le prime versioni di GPT, che analizzano imponenti masse di dati per individuare pattern statistici e poi generare sequenze di parole attendibili dal punto di vista probabilistico.

Il report di OpenAI

Un report di OpenAI (2023) riconosce che GPT-4 non è all'altezza degli umani in molti scenari del mondo reale, ma raggiunge il livello di punteggio del top 10% nelle valutazioni ricevute ad esami da superare per svolgere attività professionali e accademiche.

Un team di **Microsoft Research** si spinge fino ad affermare che questa prima versione di GPT-4 insieme a ChatGPT, PaLM di Google e tutta gli LLMs- "exhibit more general intelligence than previous AI models". In particolare, poi, "GPT-4 can solve novel and difficult tasks that span mathematics, coding, vision, medicine, law, psychology and more, without needing any special prompting" (Bubeck et al., 2023).

Va però messo in luce che lo stesso studio riconosce implicitamente la natura prettamente statistica delle capabilities di GPT-4 quando afferma che occorra effettuare ulteriori avanzamenti, valutando "the possible need for pursuing a new paradigm that moves beyond next-word prediction" (Bubeck et al., 2023).

Altri esperti molto autorevoli esprimono, a dire il vero, pareri meno favorevoli, dopo aver sperimentato in prima persona il nuovo sistema. Anche se ha superato brillantemente test di ammissione a corsi universitari e para-universitari USA (LSATs, GRES, SA), Gary Marcus,

professore emerito della NYU e imprenditore informativo [15], documenta come i suoi esperimenti con GPT-4, nonostante l'aumento della potenza computazionale di cui dispone rispetto ai precedenti LLMs, mostrino che esso non ha rivoluzionato i modelli linguistici che interagiscono con gli umani.

Emergono infatti gli stessi limiti dei precedenti modelli: è ancora incerto e traballante l'"allineamento", cioè la capacità di guidare i sistemi verso gli interessi e gli obiettivi indicati dai progettisti. Non sono rari veri e propri errori di ragionamento. Non sono del tutto evitati fenomeni di hallucination, ovvero la produzione di risposte apparentemente affidabili, ma i contenuti sono del tutto estranei al set di addestramento. Permangono quindi problemi di affidabilità, il che rende problematica qualsiasi ipotesi di impiegare il sistema nella robotica e nei processi di ricerca scientifica.

Sono inoltre necessari **frequenti e rilevanti processi di re-training** per tenere il passo con ciò che accade di nuovo, tenendo presente che GPT-4 sa poco del 2021 e nulla del 2022.

Il giudizio finale è dunque lapidario: si tratta di un passo indietro per la scienza con un sistema di IA di cui non è dato conoscere alcunché: architettura, addestramento, consumo di energie eccetera.

Pareri differenti

Opinioni diverse e molto interessanti di alcuni analisti sono riportate in un articolo di Nature (Sanderson, 2023), dove si registrano perplessità del mondo scientifico in merito alla riservatezza sui dati di addestramento e quindi all'impossibilità di accedere al codice di accesso a GPT-4, il che impedisce l'individuazione di quale possa essere l'origine dei bias, per poi escogitare rimedi.

GPT-4, che ha superato anche gli esami per la professione legale, collocandosi sempre nel segmento più alto delle valutazioni (top 10%) [16], sembra dunque non avere limiti, tanto è vero che ha dimostrato ottime capacità nel partire da un disegno a mano di un sito web per produrre il codice informatico appropriato e quindi creare un reale sito web. Tutto ciò non ha però dissipato il clima di sfiducia, presente nella comunità scientifica a causa della persistenza di modelli i cui codici sono riservati e in possesso delle società big tech. Emerge un quadro generale di quasi monopolio tecnologico-economico.

Si ribadisce, quindi, ancora una volta che ciò rende impossibile un reale controllo e la verifica della tecnologia sulla base di criteri esclusivamente scientifici.

Queste considerazioni acquistano un rilievo assoluto se unite a quelle svolte dall'ingegnere chimico Andrew White, il quale ha avuto accesso a GPT-4 come "red-team", cioè persona retribuita da OpenAI per testare la piattaforma fino a provarla, cercando di far generare "qualcosa di cattivo". In sei mesi di incarico White ha testato la capacità del sistema di indicare componenti e step di reazioni chimiche.

All'inizio gli output non sono stati straordinari, anche se il grado di realismo dimostrato si è rivelato sorprendente. Il quadro è cambiato molto (in meglio) allorché GPT-4 ha avuto accesso a Internet e ad articoli scientifici, dal momento che sono emerse abilità e competenze molto innovative e

generatrici di output suggestivi. Alla domanda dell'intervistatore circa la possibilità che GPT-4 possa "consentire la creazione di composti chimici pericolosi", White ha risposto che tutto dipende dal lavoro dei red-teamers.

L'importanza del processo di addestramento

Emerge, ancora una volta, la rilevanza delle modalità di svolgimento del processo di addestramento, della qualità dei dati di base e la necessità del controllo pubblico da parte della comunità scientifica, perché GPT-4 costituisce una leva molto potente per produrre e amplificare elementi dannosi di qualsiasi natura.

Due degli scienziati intervistati hanno infatti sottolineato la necessità di elaborare un set di linee-guida per regole "how Ai and tools such as GPT-4 are used and developed" [17].

Da queste molteplici dichiarazioni di esperti, che hanno sperimentato le funzionalità di GPT-4, possiamo evincere che i progressi rispetto alle precedenti versioni siano soprattutto di natura quantitativa più che qualitativa. Di conseguenza è legittimo ipotizzare che il notevole incremento di potenza computazionale non stia ancora producendo un salto verso forme assimilabili a quella che viene denominata General Artificial Intelligence, di cui non esiste una definizione precisa e unanimemente accettata, ma viene spesso avanzata come espressione assimilata in modo nominalistico all'intelligenza umana. La lontananza da quest'ultima è comunque riconosciuta anche dai team di ricerca di OpenAI e Microsoft Research, come abbiamo precedentemente visto.

Lo spostamento del focus della ricerca tecnico-scientifica

Appare opportuno allora chiedersi, come fanno alcuni specialisti, se non si stia di fatto realizzando uno spostamento del focus della ricerca tecnico-scientifica: la dinamica attuale è incentrata sul continuo incremento della potenza computazionale (in gergo lo *scaling*) sta producendo uno shift dall'obiettivo dell'intelligenza al perseguimento di performance sempre più elevate.

In breve, la ricerca di perfezione tecnologica (e quindi di business) a scapito di finalità scientifiche, come sembrano sostenere autorevoli personaggi nel campo dell'Intelligenza artificiale [18]: 1) "Unfortunately, it is the technology of AI that gets all the attention", (Hector Levesque). 2) "Most of today's AI approaches will never lead to true intelligence" (LeCun, guru di META). 3) "AI as a field is stuck as far as finding anything like human intelligence" (Gary Marcus). 4) "Turns out everything is a matrix multiplication, from computer graphics to training neural networks," (Demis Hassabis, co-fondatore di Open Mind).

Conclusioni

Più che esprimere giudizi conclusivi, è l'ora di sollevare alcuni interrogativi, su cui la comunità scientifica, team interdisciplinari, imprese, istituzioni e la società intera dovrebbe riflettere: è necessario continuare su una traiettoria di ricerca verso sistemi di IA sempre più potenti, che non sembrano garantire output diretti a conseguire "true intelligence" (Ananthaswamy, 2023)?

Quali strumenti porre in essere per ovviare all'attuale quasi-monopolio cognitivo delle big tech, che rischia di essere potere tout court in un universo fisico-cibernetico?

Se la crescente potenza computazionale non riesce ad assumere proprietà analoghe o affini a quelle dell'intelligenza umana (adattatività, senso comune, capacità di formulare abduzioni, porre domande che fuoriescono dagli schemi interpretativi esistenti, ecc.), come evitare una collettiva hallucination [19] e preservare il senso critico, una delle caratteristiche basilari del pensiero umano?

Uno dei rischi più significativi che incombe sull'umanità non è tanto quello della "Superintelligenza", quanto LLMs tanto potenti che possono finire nelle mani sbagliate e provocare disastri, com'è implicitamente deducibile da quanto scritto nei paragrafi precedenti e da ciò che sostiene Gary Marcus (2023b)?

Per riflettere su questi interrogativi penso sia fondamentale tenere sempre presenti alcune considerazioni del fisico **Carlo Rovelli** (2023: 27): "Andare a vedere, questo è la scienza. Andare a curiosare dove non siamo mai stati. Usando matematica, intuizione, logica, immaginazione, ragionevolezza... Andare a vedere con gli occhi della mente" (Rovelli, 2023: 29). Insomma, tutto ciò che è alla base della nostra intelligenza di esseri umani, da cui non dobbiamo abdicare.

Bibliografia

- Ananthaswamy A., 2023, "In AI, is bigger always better?", Nature, March 10. Biddle S., 8-12-2022, "The Internet's New Favorite AI Proposes Torturing Iranians and Surveilling Mosques", The Intercept. Blalock D. et al., 2020, "What is the state of neural network pruning?". arXiv:2003.03033v1 [cs.LG] Mar 2020.
- Bubeck S. et al., 2023, "Sparks of Artificial General Intelligence: Early experiments with GPT-4", Microsoft Research, arXiv:2303.12712v3 [cs.CL] 27 Mar 2023.
- Das S., Chen M., 2001, "Yahoo! for Amazon: Extracting market sentiment from stock message boards". In Proceedings of the Asia Pacific Finance Association Annual Conference (APFA), 13751388. Dave K., Lawrence S., Pennock D.M, 2003, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In Proceedings of WWW, 519–528.
- Elliott L., 2023, "AI will end the west's weak productivity and low growth. But who exactly will benefit?", The Guardian, April 7.
- Eloundou T., Manning S., Mishkin P., Rock D., 2023. "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models", OpenAI, OpenResearch, UNiversity of Pennsylvania", arXiv:2303.10130v3 [econ.GN], March 23.
- Ferrari Dacrema M., Cremonesi P., Jannach D., 2019, "Are We Really Making Much Progress? A Worrying Aanalysis of Recent Neural Recommendation Approaches", ACM, September 10.
- Floridi L., Chiriatti M., 2020, "GPT-3 Its Nature, Scope, Limits, and Consequences", Minds and Machines, 30: 681–694. Goldman Sachs, 2023, "The Potentially Large Effects of Artificial

- Intelligence on Economic Growth”, *Macroeconomics* 26. Guest O., Martin A. E., “On logical inference over brains, behaviour, and artificial neural networks”, *Computational Brain & Behavior*, February 13. <https://doi.org/10.1007/s42113-022-00166-x>.
- Heaven W. D., 2020, “OpenAI’s new language generator GPT-3 is shockingly good—and completely mindless”, *Mit Technology Review*, August 20.
- Lipsey R., Carlaw K.I., Bekar C.T., 2005, *Economic Information. General Purpose Technologies and Long Term Economic Growth*, Oxford University Press.
- Lombardi M., Macchi M., 2023, *Tra Disoccupazione Tecnologica e Great Resignation*, (in corso di stampa).
- Lombardi M., Vannuccini S., 2022, “Understanding emerging patterns and dynamics through the lenses of the cyber-physical universe”, *Patterns* 3, November 11.
- Marcus G., 2001, *The Algebraic Mind. Integrating Connectionism and Cognitive Science*. The MIT Press. Marcus G., 2019, *Rebouting AI. Building Artificial Intelligence We Can Trust*. Pantheon Books.
- Marcus G. Davis E., 2020, “GPT-3, Bloviator: OpenAI’s language generator has no idea what it’s talking about”, *Technology Review*, August 22.
- Marcus G., 2023a, “GPT-4’s successes, and GPT-4’s failures”, *Communication of the ACM Blog*, March 15. Marcus G., 2023b, *AI risk ≠ AGI risk*, March 28. MGI (McKinsey Global Institute), 2023, *What is generative AI?* January. Mitchell M., 2021, *Artificial Intelligence: A Guide for Thinking Humans*, Oxford University Press.
- Ryan-Mosleyarchive T., 2023 *AI might not steal your job, but it could change it*, *Technology Review*, April 3. *Nature Machine Intelligence (Editorial)*, 2023, “The AI writing on the wall”, 5, 1, January 1. Nori H. et al., 2023, “Capabilities of GPT-4 on Medical Challenge Problems”, March 24, arXiv:2303.13375v1 [cs.CL] 20 Mar 2023. OpenAI, 2023, *Technical Report*.
- Ortiz S., 2023a, “What is generative AI and why is it so popular? Here’s everything you need to know”, *ZDNET*, February 15. Ortiz S., 2023b, “The best AI art generators: DALL-E 2 and other fun alternatives to try”, *ZDNET*, March 31.
- Pang Bo, Lee L., 2008, “Opinion mining and sentiment analysis”, *Foundations and Trends in Information Retrieval*, 1-135.
- Pierce D., 2023, “ChatGPT started a new kind of AI race — and made text boxes cool again”, *The Verge*, March 26. Ray T., 2022, “AI’s true goal may no longer be intelligence”, *ZDNET*, October 22. Ray T., 2023a, “Generative AI could lower drug prices. Here’s how. In the future, specifying a drug target may be like sitting down to ChatGPT. After a few clicks, you’ll have your novel therapeutic”, *ZDNET*, March 1. Ray T., 2023b, “ChatGPT is ‘not particularly innovative,’ and ‘nothing revolutionary’, says Meta’s chief AI scientist”, *ZDNet*.
- Rich A.S., Gureckis T.M., 2019, “Lessons for artificial intelligence from the study of natural stupidity”, *Nature Machine Learning*. Vol 1, April, 174-180. Rovelli C., 2023, *Buchi bianchi, Adelphi*. Ryan-Mosleyarchive T., 2023, “AI might not steal your job, but it could change it”, *MIT Technology Review*, April 3.
- Sanderson K., 2023, “GPT-4 is here: what scientists think”, *Nature*, 615: March. Shanehsazzadeh A. et al., 2023, “Unlocking de novo antibody design with generative artificial intelligence”, *BioRxiv*, Preprint Server for Biology.

Spanton R. W., Guest O., 2022, “Measuring Trustworthiness or Automating Physiognomy? A Comment on Safra, Chevallier, Grèzes, and Baumard”, arXiv preprint arXiv:2202.08674.

StokelWalker C., Van Noorden R., 9-2-2023, “The Promise and Peril of Generative AI”, *Nature*, Vol. 614, 214-216.

Taboada M. 2011, “Lexicon-Based Methods for Sentiment Analysis”, *Computational Linguistics*, 37 (2): 267–307. Tafani D., 2022, “What’s wrong with “AI ethics” narratives, «Bollettino telematico di filosofia politica», 1-22, <https://commentbfp.sp.unipi.it/daniela-tafani-what-s-wrong-with-ai-ethics-narratives>.

Tafani D., 2023, “L’«etica» come specchietto per le allodole. Sistemi di intelligenza artificiale e violazioni dei diritti”, in «Bollettino telematico di filosofia politica», 1-13,

<https://commentbfp.sp.unipi.it/letica-come-specchiettoper-le-allodole/>. Tung L., 2023, “ChatGPT just became the fastestgrowing ‘app’ of all time”, *ZDNET*, February 3. Van Dis E.A.M., et al., 2023, “ChatGPT: five priorities for research”, *Nature*, Vol. 614, February 9.

224-226 Van Rooij I., 2020, Mixing psychology and AI takes careful thought. Blogpost, in Donders Wonders. Venturebeat, 2023, “OpenAI’s massive GPT-3 model is impressive, but size isn’t everything”, *Venturebeat.com* 7 Aprile.

Vincent J., 2022, “AI generated answers temporarily banned on coding Q&A Site Stack Overflow”, *The Verge*, December 5.

Vincent J., 2023, “OpenAI announces GPT-4, the next generation of its AI language model”, March 14. Wu S. et al., 2023, “BloombergGPT: A Large Language Model for Finance”, *Bloomberg New York, Bloomberg Baltimore*, arXiv:2303.17564v1 [cs.LG], March 23.

Yi J. Et al., 2003, “Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques”, *Proceedings of the IEEE International Conference on Data Mining (ICDM)*. Waters R., 2022, “Melanie Mitchell: Seemingly ‘sentient’ AI needs a human in the loop”, *Financial Times*, August 31.

Note

1. La riga dove si scrive la richiesta (“prompt” o linea di comando), deve essere precisa e contenuta entro limiti ben definiti. In questo modo la conversational AI può semplificare molto l’interazione con l’utente, dando anche l’impressione di poter fare tutto (Pierce, 2023).
[↑](#)
2. “In the context of large language models (LLMs), tokens are used to represent individual words or subwords in a text sequence. The process of breaking down text into individual tokens is called tokenization” (Techopedia). [↑](#)
3. Il tema della Generative Art non viene trattato in questa sede. Per una sintetica introduzione si veda Ortiz (2023b). [↑](#)
4. La Sentiment analysis, connessa nella letteratura manageriale all’opinion mining, è così definita: “Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques” (Yi et al., 2003). “An opinion mining tool would process a set of search results for a given item, generating a list of product attributes (quality, features, etc.) and aggregating opinions about each of them (poor, mixed, good)” (Dave et al., 2003). Si vedano anche: Pang e Lee (2008); Das e Chen (2007); Taboada et al. (2011). [↑](#)

5. A differenza di ChatGPT e degli altri LLMs, lo studio in questione indica precisamente il dataset di addestramento e annuncia che a breve sarà reso pubblico il record di tutta l'esperienza effettuata nel percorso di addestramento del modello. ↑
“You’ve probably seen that generative AI tools (toys?) like ChatGPT can generate endless hours of entertainment” (MGI, 2023). ↑
6. Il problema può assumere aspetti preoccupanti se si pensa che il sistema può aiutare a scrivere e a completare codici per processi computazionali. ↑
Gli autori si riferiscono al filone internazionale di ricerca sui fattori che distorcono i processi decisionali umani. Tra i principali esponenti vi sono Gigerenzer, Selten e Kahneman, gli ultimi due Premi Nobel per l’Economia. ↑
7. Il titolo dello studio è volutamente equivoco per far risaltare che i Generative pre-trained Transformers (GPTs) sono in effetti General-purpose Technologies (tecnologie di portata generale). Lo studio correttamente sottolinea i limiti delle stime, dovuti alla soggettività delle annotazioni degli annotator e ai database impiegati, con dati ancora quantitativamente limitati. L’esercizio è ciononostante molto significativo e denso di annotazioni metodologiche di rilievo. ↑
8. Nel Report di Goldman Sachs sono proposte stime, basate sulla definizione di scenari alternativi. ↑
9. Ryan-Mosleyarchive (2023) riferisce di come possa aumentare la produttività nelle professioni legali, nel giornalismo, con rischi e limiti analoghi a quelli segnalati in paragrafi di questo contributo. ↑
10. Bengio, Hinton e LeCun hanno nel 2018 vinto il Premio Turing per i loro contributi agli sviluppi dell’Intelligenza Artificiale. ↑
11. Marcus è fortemente critico sulla concezione dell’intelligenza prevalente negli studi sull’Intelligenza Artificiale. Uno degli elementi chiave della sua visione è l’assoluta importanza di un cambiamento paradigmatico, che reintroduca componenti simboliche. Alla base ci sono una diversa e molto significativa concezione della mente e dell’intelligenza. Si vedano a riguardo i suoi libri (Marcus, 2001, 2019). ↑
12. In questo e negli altri casi test il sistema si è classificato molto al di sopra delle precedenti versioni di GPT. ↑
13. Emergono a questo riguardo una serie di questioni di fondo, sulle quali non possiamo soffermarci in questa sede. Ne indichiamo soltanto due, proponendo delle letture per trattazioni sistematiche. La prima concerne il conflitto, su cui spesso si dibatte in modo fuorviante, tra dinamica innovativa, tutela dei diritti individuali e strategie delle big tech. Per un’analisi critica e approfondita si veda Tafani (2023). La seconda riguarda le ipotesi, discusse da varie prospettive teoriche, su come sviluppare un’intelligenza artificiale dotata di un’etica rispettosa dei diritti umani. Una suggestiva e controcorrente analisi è sviluppata in Tafani (2022). ↑
14. Le dichiarazioni sono espressamente fatte a ZDNET (Ray, 2022). ↑
15. Hallucination in senso informatico, con precedentemente indicato. ↑

Alfabetismo, competenze, partecipazione: elementi per una democrazia digitale

Le competenze digitali, non si possono ridurre a degli insegnamenti funzionali a singoli task, ma necessitano di una costante contestualizzazione culturale, politica e sociale. Occorre perciò considerare la Media Literacy come solo una delle dimensioni che compongono le svariate competenze al centro dei processi educativi mediali

Di **Marco Giacomazzi**, Università di Bologna

Il contrasto alle disuguaglianze nella distribuzione delle competenze digitali si può combattere solo attraverso delle azioni di *Media Education* che superino i limiti della *Media Literacy*. Come si sosterrà in seguito, questo è dovuto alla natura stessa delle competenze digitali, che non si possono ridurre a degli insegnamenti funzionali a singoli task, ma necessitano di una costante contestualizzazione culturale, politica e sociale.

Questo tipo di intervento educativo è possibile solo se si **supera la visione della digitalizzazione come un processo di democratizzazione spontanea dell'accesso delle informazioni**: se si adotta una prospettiva semiotica di cooperazione interpretativa (Eco 1979), appare evidente come l'interpretazione sia un processo sociale regolato da competenze enciclopediche che non possono emergere in maniera spontanea dai testi – men che meno da testi digitali che presentano specifiche *affordances*.

Se quindi da un lato si continua a sostenere che la democrazia – e i processi di democratizzazione – non consista nel *presumere* un'eguaglianza formale, ma intervenire in maniera sostanziale nella rimozione delle barriere che ci rendono diseguali, si vuole dall'altro ricordare la tesi di Buckingham (2019), per la quale la *Media Literacy*, in quanto nozione individualista e responsabilizzante, venga **utilizzata come spauracchio dalle istituzioni per scaricare sugli individui la responsabilità della propria educazione mediale**, senza mettere in atto delle vere iniziative sistematiche di regolazione (Giacomazzi 2022).

Da dove deriva questa interpretazione dell'alfabetizzazione mediale?

Dalla Media Literacy alla partecipazione pubblica: un passaggio obbligato? Al giorno d'oggi, la *Media Literacy* è un campo di studio riconosciuto istituzionalmente come riflessione e progettazione del corretto insieme di conoscenze per affrontare le sfide offerte dal panorama dei media: nel contesto nordamericano, ad esempio, la [*National Association for Media Literacy Education*](#) (NAMLE) definisce la *Media Literacy* come “la capacità di accedere, analizzare, valutare, creare e agire utilizzando tutte le forme di comunicazione”, chiarificando che essa è solo una parte di un più ampio processo educativo: “Media refers to all electronic or digital means and print or artistic visuals used to transmit messages. Literacy is the ability to encode and

decode symbols and to synthesize and analyze messages. Media literacy is the ability to encode and decode the symbols transmitted via media and synthesize, analyze and produce mediated messages. Media education is the study of media, including ‘hands-on’ experiences and media production. Media literacy education is the educational field dedicated to teaching the skills associated with media literacy”.

La Media Literacy sembra quindi essere circoscritta a un concetto di stampo funzionalista, che mira a educare le persone a *codificare e decodificare* messaggi mediali, all’interno di uno studio più ampio dei media, delle produzioni mediali, che può anche essere orientato alla trasmissione delle competenze.

Una versione più sofisticata viene fornita dal report “Media Literacy Versus Fake News. Esperienze e best practice in Italia” (Aroldi *et. al.* 2022), condotta nel quadro di **IDMO – Italian Digital Media Observatory**, che riporta la definizione della *Media Literacy* indicata dalla Commissione Europea nel 2007, orientandola verso la cittadinanza digitale – **intesa come** “la capacità di un individuo di avvalersi consapevolmente e responsabilmente dei mezzi di comunicazione virtuali” – e ponendola come prerequisito fondamentale per contrastare i rischi della disinformazione.

La **Commissione Europea** definisce ufficialmente, già nel 2007, la *Media Literacy* (“alfabetizzazione mediatica” nei documenti tradotti in italiano) come «la capacità di accedere ai media, di comprendere e valutare criticamente diversi aspetti dei media e dei loro contenuti e creare comunicazioni in una varietà di contesti». In una parola: cultura. Quella **capacità di creare connessioni e sviluppare senso critico**, di non essere passivi recettori di informazioni, ma attivi e consapevoli, passa anche per i media. E per il mondo digitale.

In questa definizione emergono già alcune questioni che verranno approfondite in seguito, ossia la consapevolezza – che chi scrive si spinge a considerare squisitamente semiotica – del fatto che **non esiste processo informativo (di codifica-decodifica; di simbolizzazione; di interpretazione) che non sia un processo eminentemente culturale ed enciclopedico** (Eco 1984), ossia che non consista in un’operazione attiva di selezione di elementi all’interno di una rete di connessioni intertestuali.

Nel contesto italiano, prima della *Media Literacy* (lett. *Alfabetizzazione mediatica*) c’era l’alfabetizzazione *attraverso* i media. Dalla trasmissione televisiva del **maestro Manzi** “Non è mai troppo tardi”, andata in onda dal 1959 al 1963, al programma televisivo di **Umberto Eco e Tullio de Mauro** “Le vicende della lingua italiana” del 1973, è stata **perseguita un’operazione di democratizzazione culturale attraverso i mezzi di comunicazione massa** – in questo caso, programmi televisivi. Per quanto riguarda queste iniziative, l’applicazione del concetto di ‘democrazia’ o ‘democratizzazione dell’informazione’ è adeguato, perché questi sono esempi di operazioni culturali progettate volte a condividere competenze alfabetiche attraverso i media. È un esempio che mostra perché non possiamo considerare la digitalizzazione un’operazione di democratizzazione dell’informazione *per se*: **la diffusione di un mezzo tecnologico è difficilmente un’operazione pedagogica, se non è accompagnata da misure sistemiche per condividere conoscenze o competenze.**

Se mai, l'emersione prima dei *mass media* e in seguito dei media digitali ha creato nuove esigenze educative. Partendo dalla ricostruzione di Falcinelli (2021) sull'emersione dei concetti legati alla *Media Education* e sull'affermazione recente del **concetto di competenza digitale**, si può osservare come già negli anni '70, l'UNESCO e il **Conseil international du Cinéma et de la Télévision** parlassero chiaramente di Educazione ai media, definendola come lo studio della storia dei media, del loro ruolo nella società e come accedervi.

Questo concetto si applicava inizialmente ai *mass media*, come la televisione, la radio e la stampa. Falcinelli illustra tre diverse dimensioni in cui le pratiche di Media Education sono state intese nel tempo:

- **educazione sui media come educazione alla *comprensione dei messaggi mediali*** e della realtà del sistema delle comunicazioni di massa (orientamento rivolto alle conoscenze);
- **educazione ai media come educazione alla *fruizione corretta dei mezzi di comunicazione*** (orientamento rivolto alle abitudini);
- **educazione con i media come educazione alla *produzione di messaggi mediali originali*** (orientamento rivolto all'abilità) favorendo le esperienze di produzione creativa che consentono di esprimere nei diversi linguaggi le proprie emozioni ma anche i propri pensieri.

Falcinelli procede poi a descrivere il percorso verso l'idea di competenza digitale, passando per l'emersione della *New Media Literacy*, la cui teorizzazione deriva dal lavoro di Jenkins (et al. 1998) sulla cultura partecipativa degli ambienti mediali digitali. Secondo quest'ultimo, proprio **la diffusione del Web 2.0 e dei suoi linguaggi avrebbero portato a mettere in discussione approcci individualistici all'alfabetizzazione mediale**, considerandola invece come una pratica orientata alla collaborazione e come una pratica sociale. Secondo Falcinelli in questa prospettiva "la media literacy può essere considerata come il risultato del processo di media education".

Dalla *New Media* si è poi passati alla *Digital Literacy*, intesa come la capacità di comprendere e utilizzare informazioni in diversi formati, provenienti da una varietà di fonti digitali. All'inizio degli anni 2000, il **Consiglio europeo**^[1] l'ha indicata come un prerequisito per la creatività, l'innovazione e l'imprenditorialità. Ed è proprio grazie al Consiglio europeo che si è affermato, negli ultimi quindici anni, il concetto di *Competenza digitale*: questo è successo all'interno del quadro più ampio delle **competenze chiave per l'apprendimento permanente**, che vede la **competenza digitale come una delle 8 competenze chiave** sia nelle versioni del 2006 che del 2018. La competenza digitale è centrale nello sviluppo della cittadinanza digitale, o partecipazione mediata alla vita pubblica, ed è ora definita dal **quadro europeo DigiComp 2.2**, che verrà osservato più nel dettaglio in seguito. La relazione tra l'impegno civico e l'alfabetismo non è affatto nuova: è stato **Tullio de Mauro**^[2] a identificare nell'**alfabetismo uno dei prerequisiti fondamentali per la democrazia stessa**. Ad oggi, il livello di alfabetismo degli adulti è costantemente monitorato: un esempio è il **Programma per la valutazione internazionale delle competenze degli adulti dell'OCSE** (PIAAC), che monitora alfabetismo, numerazione e *problem-solving* negli adulti nel lungo periodo. Tuttavia, **se la corretta comprensione delle informazioni è la condizione necessaria per promuovere l'impegno civico e digitale, alcuni esperti sostengono che non sia abbastanza**.

Tra la Media Literacy e la Media Education

Se l'assenza di alfabetismo rappresenta un problema per la democrazia, i *Media Educator* cercano di rendere consapevoli rispetto ai rischi di limitarsi a questo concetto in ottica educativa. Una di queste è **Danah Boyd** (2018): la studiosa di *social media* si concentra sul fatto che la *Media Literacy* si sia evoluta come risposta alla propaganda perseguita dai *mass media* durante il XIX secolo e ha sviluppato misure pedagogiche adattate solo a quel contesto. Considerando invece il paesaggio mediatico e informativo odierno, Boyd afferma che **per affrontare il disordine informativo sia importante non solo demistificare i messaggi propagandistici, ma anche ricostruire la fiducia nelle istituzioni e nell'educazione**. Tuttavia per farlo non dovremmo concentrarci solo sull'insegnare agli studenti a distinguere la verità dalla falsità, o ad aiutarli a valutare le fonti e verificare i fatti. Prima di farlo, è necessario capire da dove provengono la confusione, la rabbia e l'odio alla base della produzione di testi online. In una rete sempre più complessa, le forme di produzione di testi e conoscenze sono la dimostrazione di un tentativo, da parte degli utenti, di cercare di dare un senso al mondo: "The people who posted this meme [...] didn't bother to fact check this claim. They didn't care. What they wanted to signal loud and clear is that they hated Hillary Clinton. And that message was indeed heard loud and clear. As a result, they are very offended if you tell them that they've been duped by Russians into spreading propaganda. They don't believe you for one second. Misinformation is contextual".

I ricercatori in *Data and Society* **Bulger e Davison** (2018), forniscono un esaustivo rapporto sulle principali iniziative di *Media Literacy* perseguite nel contesto nordamericano, elaborando un quadro per valutarle e fornendo agli *stakeholder* raccomandazioni per implementare l'alfabetizzazione mediatica in modo produttivo.

La loro ricerca rivela come una delle principali cause del fallimento delle iniziative di *Media Literacy* sia l'ampio spettro delle sue aspettative: "*is it to discern accuracy, evaluate bias, engage with information productively, be an informed voter?*" (ibid: 16). Un altro problema è, ancora una volta, il suo anacronismo e il fatto che pone tutta la responsabilità di discernere la qualità delle informazioni sull'utente finale:

"Finally, media literacy research typically focuses on individual responsibility for discerning the truth or accuracy of messages. As platforms such as Facebook, Google, and Twitter increasingly personalize information access, individual responsibility becomes more challenging, especially when methods for serving information are not transparent. One challenge for research moving forward is determining expectations for how an individual can assess the reliability of information when the breadth of the corpus, e.g., what is included and excluded and why (and how it differs from information served to others), is neither visible nor accessible. It is necessary to rethink media literacy in the age of platforms" (ibid: 17).

Una delle voci più autorevoli nel dibattito sulla necessità di superare le limitazioni della *Media Literacy* è **David Buckingham**, studioso di Media Education e autore di *Un manifesto per la Media Education* (2019). Come gli studiosi sopracitati, non mette in dubbio le buone intenzioni della *Media Literacy*, né la sua necessità in un paesaggio mediatico digitale. Tuttavia, in molteplici occasioni sottolinea come la *Media Literacy* sia stata spesso utilizzata come una "soluzione rapida"

(Buckingham, Farinacci, Manzoli 2021: 22), una soluzione facile a problemi complessi, o, ancora peggio, una soluzione individualizzata a problemi sistemici che dovrebbero essere affrontati attraverso interventi a più livelli. Mentre la sua critica sottolinea come i governi – in particolare quello del Regno Unito – tendano ad usare la *Media Literacy* come scusa per evitare di regolare l'attività delle grandi piattaforme digitali, Buckingham evidenzia anche come il concetto stesso di *Media Literacy* costituisca un *framework* che scarica la responsabilità dei problemi sistemici sugli utenti individuali.

Uno degli esempi in cui questa questione diventa evidente è l'approccio strumentale della *Digital Literacy*, dove la conoscenza tecnica – come saper programmare – viene presentata come una soluzione a problemi politici e culturali; o quando metodi di fact-checking facili e accessibili vengono consegnati come soluzioni a un disordine informativo sistemico:

It's sometimes assumed that digital literacy is simply about learning how to use digital tools: learning how to operate hardware, or to use software such as search engines. This is the instrumental approach I was referring to. For me, this is just the beginning of the process. Of course, we need to know how to find information online; and doing that effectively is something we have to learn. *But the more difficult questions come when we have to make sense of that information, to process it and to evaluate it. We need to make judgments about what we should trust, and that's far from easy.* This is where media education comes in. Despite all the loose talk about fake news, this isn't just about telling the difference between what's true and false. It's not something we can do with a simple checklist: on the contrary, it involves a much more complex, multi-dimensional process of analysis and evaluation. This isn't straightforward to learn: but without it, we are lost? (Buckingham, Farinacci, Manzoli 2021: 26, corsivi dell'autore)

Buckingham si fa invece promotore di azioni educative sistemiche, per le quali venga riconosciuto il ruolo sociale del complesso scenario dei media interconnessi in cui siamo inseriti.

L'intero sistema educativo dovrebbe riconoscere la nostra condizione mediale e assumersi la responsabilità della preparazione degli studenti, mettendo i programmi di *Media Education* al centro dei programmi scolastici. Questo significa allenare strumenti come il pensiero critico, che include il *problem solving* ma non si limita ad esso, e comprende anche la condivisione di informazioni sulla composizione del panorama dei media e il ruolo dei suoi attori.

Certamente il coding è un'abilità che gli studenti dovrebbero avere l'opportunità di acquisire, se lo desiderano; ma affermare che essa li aiuta nel *problem solving* o che è una formazione essenziale per l'occupazione futura – ragione per cui dovrebbe essere insegnata a tutti gli studenti – è assai discutibile. **I ragazzi, è vero, hanno bisogno di sapere come funziona la tecnologia digitale; ma devono anche capire come funzionano i media digitali in quanto industrie e forme culturali di rappresentazione.** Se devono diventare utenti attivi della tecnologia, hanno bisogno di apprendere qualcosa di più che mere capacità tecniche: è necessaria una comprensione sociale, politica, economica e culturale. (2019: 82)

Anche il succitato report di IDMO sulla *Media Literacy* in Italia è molto chiaro nell'indicare i limiti di fermarsi ai concetti costruiti sulla scorta dell'alfabetismo: “ci si è resi conto di un aspetto importante, che emerge anche nei documenti e nei materiali prodotti: **lavorare sulla promozione di**

un profilo di soggetto “media literate” (ovvero alfabetizzato) è solo una parte del processo, la parte più semplice da portare a termine. La costruzione del pensiero critico è sempre importante, ma tocca esclusivamente una porzione della sfida alla Media Education, soprattutto pensando a quattro dimensioni: invisibilità, autorialità, multi-literacy e transmedialità”. (Aroldi *et al* 2022: 28) Per quanto riguarda l’invisibilità, si tratta di rendere *opache* meccaniche *trasparenti* soggiacenti al nostro abitare ambienti mediali; dalle logiche algoritmiche alle nostre pratiche di navigazione, spesso irriflesse o considerate come spontanee. Per quanto riguarda l’autorialità invece, si tratta di sollevare la dimensione *etica* ed *estetica* del prosumerismo oltre ai perfezionamenti tecnici della produzione di contenuti digitali; la *multi-literacy* spinge a combinare una riflessione sulla molteplicità delle sostanze espressive combinabili nella comunicazione digitale con la molteplicità dei contesti culturali che possono essere potenzialmente coinvolti da essa: “Occorre quindi imparare a negoziare i diversi significati che raccogliamo da ciascun medium e gli sfondi culturali diversificati, attraverso pratiche basate sulla negoziazione e sulla “situazione” (cioè la collocazione del significato in un contesto significativo e concreto)”.

Per quanto invece riguarda l’ultima dimensione, quella della **transmedialità**, si tratta di educare a logiche di *aggiunta* e confronto delle informazioni all’interno di un paradigma di *overload* informativo che comporta non solo il sovraccarico di informazioni, ma anche la loro dispersione in ambienti diversificati.

Verso le competenze digitali

Da dieci anni, il *Digital Competence Framework for Citizens* offre un linguaggio comune per descrivere una di quelle aree di competenze che il **Consiglio Europeo** ha indicato come fondamentali per l’apprendimento permanente: la **competenza digitale**. Come si ricordava in precedenza, è un concetto centrale per l’orientamento al *policy-making* e alle pratiche educative sin dal 2006. All’interno [dell’ultima versione del DigComp](#), 2.2, da poco disponibile anche in lingua italiana, ritroviamo la definizione di competenza digitale:

“La competenza digitale implica l’uso sicuro, critico e responsabile delle tecnologie digitali e il loro impiego nell’apprendimento, nel lavoro e nella partecipazione alla società. Comprende l’alfabetizzazione all’informazione e ai dati, la comunicazione e la collaborazione, l’alfabetizzazione ai media, la creazione di contenuti digitali (compresa la programmazione), la sicurezza (compreso il benessere digitale e le competenze relative alla sicurezza informatica), le questioni relative alla proprietà intellettuale, la risoluzione di problemi e il pensiero critico”.

(Raccomandazione del Consiglio sulle competenze chiave per l’apprendimento permanente). Il *DigComp* combina 21 competenze specifiche all’interno di 5 aree:

1. Alfabetizzazione su informazioni e dati;
2. Comunicazione e collaborazione;
3. Creazione di contenuti digitali;
4. Sicurezza;
5. Risoluzione di problemi.

Lo strumento è diviso in 5 dimensioni che ne permettono l’applicabilità in diversi contesti educativi o di apprendimento: la prima dimensione è l’*Area di competenza*, seguita dalla seconda dimensione che è dedicata alle *21 specifiche competenze*. Ognuna di queste ventuno risponde poi di

successive 3 dimensioni, iniziando con il *livello di padronanza* – che mira a dare dei parametri di misurabilità dell'autonomia del cittadino rispetto alla singola competenza, così come della sua abilità nel trasmetterla ad altri – seguito poi dalla dimensione degli *esempi pratici* di articolazione di una specifica competenza e quella dei *casi d'uso* in cui il cittadino o la cittadina possono esercitarla concretamente.

La quarta dimensione, quella relativa agli esempi, pratici restituisce il modo in cui il Consiglio Europeo definisce il concetto di competenza, organizzando gli esempi in:

- **Conoscenze:** “Si intende il risultato dell’assimilazione di informazioni attraverso l’apprendimento. La conoscenza è l’insieme di fatti, principi, teorie e pratiche relative a un campo di lavoro o di studio”
- **Abilità:** “Sono la capacità di applicare le conoscenze e di utilizzare il know-how per portare a termine compiti e risolvere problemi. Nel contesto del Quadro europeo delle qualifiche (in inglese EQF), le abilità sono descritte come cognitive (quando implicano l’uso del pensiero logico, intuitivo e creativo) o pratiche (quando implicano la destrezza manuale e l’uso di metodi, materiali, strumenti e utensili)”
- **Attitudini:** “Le attitudini sono concepite come fattori motivanti della prestazione, la base per una prestazione costante e competente. Comprendono valori, aspirazioni e priorità” (DigComp 2.2.: 3)

Non si possono in questa sede presentare le singole competenze specifiche: né i punti di forza di questa categorizzazione, né i suoi limiti nel dettaglio. Ci si vuole tuttavia soffermare, rispetto al percorso presentato in precedenza, sulle differenze tra le concezioni di *Literacy* ed *Education* rispetto ai Media e avanzare alcune considerazioni preliminari. Il concetto di competenza così elaborato tende inevitabilmente, per ragioni di applicabilità e di concretezza, a risentire dei limiti della *Media Literacy*: di per se, **lo strumento del DigComp è pensato per fornire delle competenze pratiche, orientate all’asservimento di funzioni e calibrate su alcuni specifici casi d’uso**, rientrando all’interno di quella dimensione individuale e responsabilizzante di cui ci avvisano i critici della *Media Literacy*.

Allo stesso tempo però, la sua articolazione e ricchezza di collegamenti permette di superare alcuni dei limiti sollevati dai critici succitati. Non solo **si cerca di valorizzare il pensiero critico, sottolineando l’importanza del benessere digitale**, ma questo viene stimolato attraverso la convocazione di esempi pratici che chiamano in causa gli attori del panorama mediale, **situando l’esperienza digitale all’interno di un contesto abitato, promuovendo un uso delle tecnologie digitali consapevole e informato**:

DIMENSIONE 4 • ESEMPI DI CONOSCENZE, ABILITÀ, ATTITUDINI		DIMENSIONE 5 • CASI D'USO	
CONOSCENZE   	31. È consapevole che molte applicazioni su Internet e sui telefoni cellulari raccolgono ed elaborano dati (personali, comportamentali e contestuali) ai quali l'utente può accedere o che può recuperare, ad esempio, per monitorare le proprie attività online (ad esempio, i clic sui social media e le ricerche su Google) e offline (ad esempio, i passi quotidiani, le camminate e i viaggi sui mezzi pubblici). 32. È consapevole che i dati (ad esempio, numeri, testi, immagini e suoni), per essere elaborati da un programma, devono prima essere adeguatamente digitalizzati (cioè codificati digitalmente). 33. Sa che i dati raccolti ed elaborati, ad esempio dai sistemi online, possono essere utilizzati per riconoscere modelli (ad esempio le ripetizioni) in nuovi dati (ad esempio, altre immagini, suoni, clic del mouse e comportamenti online) per ottimizzare e personalizzare ulteriormente i servizi online (ad esempio, pubblicità). 34. È consapevole che i sensori utilizzati in molte tecnologie e applicazioni digitali (ad esempio, telecamere per il riconoscimento facciale, assistenti virtuali, tecnologie indossabili, telefoni cellulari, dispositivi intelligenti) generano grandi quantità di dati, compresi quelli personali che possono essere utilizzati per addestrare un sistema di IA. (IA) 35. Sa che esistono archivi di dati aperti in cui chiunque può ottenere dati per supportare alcune attività di risoluzione dei problemi (ad esempio, i cittadini possono utilizzare i dati aperti per generare mappe tematiche o altri contenuti digitali).	BASE 2 SCENARIO DI OCCUPAZIONE: processo di ricerca di un lavoro A casa con mia sorella, alla quale mi rivolgo quando ne ho la necessità: <ul style="list-style-type: none"> sono in grado di individuare come e dove organizzare e tenere traccia di annunci e app di lavoro (ad esempio www.indeed.com) dal mio smartphone per poterli recuperare in caso di bisogno durante la mia ricerca del lavoro. SCENARIO DI APPRENDIMENTO: preparare un lavoro di gruppo con i compagni di classe In classe con l'insegnante a cui posso rivolgermi in caso di necessità: <ul style="list-style-type: none"> sono in grado di individuare una app sul mio tablet per organizzare e archiviare link relativi a siti web, ai blog e ai database digitali relativi a un argomento specifico dei riferimenti bibliografici e utilizzarla per recuperarli all'occorrenza per la mia relazione. 	
	ABILITÀ 36. Sa come raccogliere dati digitali utilizzando strumenti di base come i moduli online e presentarli in modo accessibile (ad esempio, utilizzando le intestazioni nelle tabelle). 37. È in grado di applicare procedure statistiche di base ai dati in un ambiente strutturato (ad esempio un foglio elettronico) per produrre grafici e altre visualizzazioni (ad esempio, istogrammi, diagrammi a barre e diagrammi a torta). 38. Sa come gestire la visualizzazione dinamica dei dati e maneggiare i grafici dinamici di interesse (ad esempio, quelli forniti da Eurostat o dai siti web governativi). 39. È in grado di scegliere tra diversi tipi di luoghi di archiviazione (dispositivi locali, rete locale, cloud) quelli più appropriati da utilizzare (ad esempio, salvare i dati sul cloud li rende disponibili in qualsiasi momento e da qualsiasi luogo, ma ha implicazioni sui tempi di accesso). 40. È in grado di utilizzare strumenti di trattamento dei dati per gestire e organizzare informazioni complesse (ad esempio, database, strumenti di estrazione dei dati/data mining, software di analisi), al fine di prendere una decisione o risolvere un problema		
	ATTITUDINI 41. Presta attenzione alla trasparenza quando tratta e illustra i dati per garantirne l'affidabilità e individua i dati che sono presentati con motivazioni nascoste (ad esempio, non etiche, di profitto, di manipolazione) o in modo fuorviante. 42. Presta attenzione all'accuratezza nel valutare rappresentazioni complesse di dati (ad esempio, tabelle o visualizzazioni), poiché potrebbero essere utilizzate per fuorviare il giudizio cercando di dare un falso senso di oggettività.		

Figura 1. descrizione della competenza 1.2 gestire dati, informazioni e contenuti digitali Un altro punto di forza di DigComp che permette di inserirlo all'interno di un progetto educativo più ampio è il fatto che esso istituisca delle connessioni con altre aree competenze per l'apprendimento permanente come i *framework* relativi all'area *Personal, Social and Learning to Learn*, le competenze di cittadinanza, quelle per [l'entrepreneurship e per la cultura democratica](#). In conclusione, **le competenze digitali del DigComp sono competenze situate nella vita sociale e politica dei cittadini**, e non si fermano a competenze di codifica e decodifica, non si limitano a delle checklist ma **cercano di formare un individuo nella complessità della sua esperienza digitale**. Basti pensare all'articolazione della competenza in accezioni di *sapere* che superano il nozionismo, includendo anche *valori*, aspirazioni e priorità.

Questo non le rende uno strumento esaustivo per l'educazione ai media, ma di certo uno più completo di quello tradizionale all'alfabetizzazione. Un modo per integrare il DigComp alla *Media Education* sarebbe quello di integrarvi un approccio *umanistico* del digitale, che ne integri quindi una storia, una critica e una semiotica, e portare avanti la battaglia perché la sua adozione sia parte dell'offerta formativa di quelle istituzioni delegate di tutelare il benessere e la formazione dei cittadini e delle cittadine.

Conclusioni: oltre l'alfabetismo

Se sulla necessità di una *Media Literacy* vi è un consenso trasversale agli ambienti accademici, educativi ed istituzionali, i diversi modi di intenderla e promuoverla hanno diverse implicazioni.

L'alfabetismo digitale non può essere considerato a stregua di una capacità di codifica e decodifica non complessa, astratta dalla complessità che i processi interpretativi istituiscono. Secondo la lezione semiotica, interpretare è sempre un'operazione complessa e contraddittoria, che chiama in campo facoltà cognitive ma anche esperienziali; competenze, sosteneva Eco, *enciclopediche* (1984), che fanno riferimento a un modello teorico che registra anche interpretazioni tra loro contraddittorie. Insomma **le competenze digitali non possono essere intese come la capacità di classificare e distinguere informazioni in maniera binaria**, ma vanno intese – nella direzione della *Media Education* di Buckingham da un lato, e delle competenze di DigComp dall'altro – come competenze pragmatiche e situate che superino una lettura prettamente informazionale e funzionale dell'alfabetismo.

Queste considerazioni forse possono aiutare a **uscire dalla retorica responsabilizzante dell'analfabetismo funzionale**, per il quale la responsabilità dei fenomeni di disinformazione giacerebbe alla base di una incapacità degli utenti di comprendere i messaggi medialità, e iniziare a pensare alle sfumature di stampo comunitario e sociale delle pratiche interpretative online, così come la multidimensionalità delle loro concause.

In questa direzione, ribaltando una delle concezioni di *Media Literacy* citate da Falcinelli (2021) in precedenza, quest'ultima non dovrebbe essere considerata come il risultato della *Media Education*, ma solo una delle dimensioni che compongono le svariate competenze al centro dei processi educativi medialità, i quali però non possono e non devono essere ridotti e fermarsi ad esse.

L'alfabetismo non è l'obiettivo dei processi media-educativi, ma ne è una premessa fondamentale per il benessere digitale, la cittadinanza digitale e la partecipazione – ormai inevitabilmente mediata – alla vita pubblica di tutti i cittadini.

Resta quindi da effettuare un lavoro di divulgazione, diffusione e *advocacy*: **la critica principale di Buckingham è quello di fare della Media Education una pratica sistemica**, e quindi di non ragionare unicamente in termini di competenze da coltivare, ma **inserire nei curricula scolastici insegnamenti che educino al ruolo politico e potere simbolico dei media**. Il più volte citato report di IDMO propone una **mappatura delle offerte di Media Education consolidate sul territorio nazionale italiano**: queste proposte, sia in ambito scolastico che extra-scolastico, non fanno parte di un quadro curricolare unitario, al punto che la ricerca di IDMO parte proprio dalla constatazione della mancanza, nel nostro contesto nazionale, di un luogo deputato istituzionalmente a raccogliere in modo sistematico la documentazione relativa alla realizzazione dei progetti formativi che si iscrivono nell'orizzonte della Media Education, sia in ambito scolastico che extrascolastico (Aroldi *et al.* 2022: 63).

Il report fornisce una fotografia diacronica del quadro normativo che promuove le iniziative di *Media Education*, sia a livello Europeo che nazionale; nel primo caso dall'Agenda Digitale 2020, poi aggiornata al 2030, al [Digital Education Action Plan 2021-2027](#); nel secondo caso, attraverso varie tappe di iniziative istituzionali, dal **Piano Nazionale Scuola Digitale** (PNSD), promosso dal MIUR nel 2008, alla riforma della Buona Scuola dal 2015, che fa del PNSD il pilastro della definizione di *Information e Digital Literacy* all'interno delle competenze che la scuola deve trasmettere a studenti e studentesse in modo da **formare cittadini consapevoli** – a cui poi si aggiunge la legge n. 92 del 20 agosto 2019 sulla cittadinanza digitale. Nei curricula di **Nuova Educazione Civica**, la cittadinanza digitale è il terzo nucleo concettuale: “*In tale nucleo, i temi*

della ricerca e della valutazione delle fonti di informazione e del contrasto alla disinformazione hanno un ruolo di rilievo.” (ivi: 19).

Nonostante quindi la centralità del tema, la Media Education ancora fatica a trovare un riconoscimento e una legittimazione a livello curricolare. Nel nostro Paese, come in altri Paesi europei, la situazione presenta un aspetto quasi paradossale: da una parte, come ha giustamente notato tra gli altri David Buckingham, le istituzioni fanno spesso riferimento alla *Media Education* come alla soluzione di ogni problema, una sorta di panacea che, nel migliore dei casi, **rischia di tradursi in qualche forma di apprendimento tecnico all’uso delle tecnologie o alla riduzione dei rischi connessi**, nel peggiore, invece, in una formula vuota e retorica che scarica sugli utenti la responsabilità dei sistemi mediali e dei policy maker. Dall’altra parte, **mentre si attribuisce all’istituzione scolastica la responsabilità di far fronte all’esigenza di formare le nuove competenze utili alla cittadinanza digitale, si fatica a trovare nella scuola stessa un ruolo e uno spazio specifici per la Media Education.** (ivi: 24)

Bibliografia

Aroldi, p., Carezio, A., Cino, D., Paradisi, A., and Pratesi, A. 2022 *Media Literacy Versus Fake News. Esperienze e best practice in Italia* <https://www.idmo.it/wp-content/uploads/2022/05/Ricerca-Media-Literacy.pdf>

Boyd, d., 2018 “You Think You Want Media Literacy... Do You?”, *Data & Society: Points*, <https://points.datasociety.net/you-think-you-want-media-literacy-do-you-7cad6af18ec2>

Bulger, M., Davison, P., 2018 *The Promises, Challenges and Futures of Media Literacy*, Data & Society Research Institute, <https://datasociety.net/library/the-promises-challenges-and-futuresof-media-literacy/>

Buckingham, D., 2019 *The Media Education Manifesto*, Polity, Cambridge; trad. it *Un manifesto per la media education*, Mondadori Università, Milano.

Buckingham, D., Farinacci, E., Manzoli, G., 2021 “Media Education in the Digital Age: An Interview with David Buckingham”, *Media and/or literacy? Didattica dei media, didattica coi media*, Sociologia della comunicazione 62(XXXII), Franco Angeli, Milano,

Bolter, J.D., Grusin, R. 2000 *Remediation. Understanding New Media*, MIT Press, Cambridge.

Eco, U. 1979 *Lector in fabula*, Bompiani, Milano.

Eco, U. 1984 *Semiotica e filosofia del linguaggio*, Einaudi, Torino.

Falcinelli, F. 2021 “I “new media” in un’azione didattica orientata alla promozione della competenza digitale”, *Media and/or literacy? Didattica dei media, didattica coi media*, Sociologia della comunicazione 62(XXXII), Franco Angeli, Milano,

Floridi, L. 2014 *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*, Oxford University Press, Oxford.

Jenkins, H., Purushotma R., Weigel M., Clinton K., Robinson A.J. 1998 *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*, The MIT Press, Cambridge.

DigComp 2.2: The Digital Competence Framework for Citizens – With new examples of knowledge, skills and attitudes <https://publications.jrc.ec.europa.eu/repository/handle/JRC128415>
https://repubblicadigitale.innovazione.gov.it/assets/docs/DigComp-2_2-Italiano-marzo.pdf

Intervista a Tullio De Mauro – Alfabetizzazione e

democrazia <https://damianorama.wordpress.com/2008/11/01/intervista-a-tullio-de-mauroalfabetizzazione-e-democrazia/>

Key Competences for Lifelong Learning

<https://op.europa.eu/en/publication-detail/-/publication/297a33c8-a1f3-11e9-9d01-01aa75ed71a1/language-en>

NAMLE – Media Literacy Defined <https://namle.net/resources/media-literacy-defined/>

The Programme for the International Assessment of Adult Competencies

(PIAAC) <https://www.oecd.org/skills/piaac/>

Reference Framework of Competences for Democratic Culture

(RFCDC) <https://www.coe.int/en/web/reference-framework-of-competences-for-democraticculture>

Digital Education Action Plan 2021-2027 <https://education.ec.europa.eu/focus-topics/digitaleducation/action-plan>

Scuola 4.0: fare didattica con i meme per educare alla creatività e alla complessità

Le molte tecnologie digitali che entreranno a scuola, il potenziamento delle reti e la formazione dei docenti per avviare sperimentazioni con l'AI avranno bisogno di creatività e progettazione didattica. Educare ad un uso coerente dei meme potrebbe essere una delle sfide per la scuola 4.0

Di **Daniela Di Donato**, Docente di italiano (Liceo scientifico), PhD in Psicologia sociale, dello sviluppo e della Ricerca educativa presso Sapienza Università di Roma, esperta di metodologie didattiche, inclusione e uso delle tecnologie digitali a scuola

Avete mai fatto un **meme**? Li avete mai letti? I meme sono artefatti digitali, che circolano in rete e sui social, con una diffusione molto rapida e spesso incontrollata. Le **immagini ironiche e virali**, accompagnate da un breve testo, hanno l'intenzione di comunicare, ma allo stesso tempo di provocare una reazione in chi le guarda.

I meme nella didattica

Didattizzare i contesti di apprendimento è una delle competenze principali del docente: lo sguardo di chi educa vede nei contesti quotidiani scenari potenziali per l'apprendimento. Per questo da diversi anni ho trasformato i meme in un artefatto metacognitivo e didattico contemporaneo, adatto a svelare i fraintendimenti, le comprensioni e le incomprensioni e allo stesso tempo motivante e coinvolgente per gli studenti.

Il nome viene coniato da **Richard Dawkins**, biologo evolucionista, che ha creato il concetto del *meme culturale*: alla fine degli anni Settanta prese la parola meme dal greco *mimesi* (imitazione) per identificare un'unità di cultura, che si tramanda da una generazione all'altra e può essere intesa come l'equivalente culturale di un gene. Il meme indica tutto ciò che nella cultura si replica: per Dawkins esempi di meme sono melodie, idee, slogan, abiti di moda, modi di fare, proverbi. Analogamente ai geni, che si propagano saltando da un corpo all'altro tramite spermatozoi o uova, così i meme si propagano saltando da un cervello all'altro, attraverso un processo che, in senso lato, può essere chiamato *imitazione* (Blackmore, 2002; Dawkins, 2017; Lolli, 2017;).

Perché la scuola dovrebbe interessarsi ai meme

Parliamo di Meme perché dovrebbero interessare il mondo della scuola, soprattutto gli insegnanti, almeno per tre motivi:

- **I meme sono media contemporanei ad alto impatto comunicativo**, che si riproducono in rete grazie alla collaborazione degli utenti e tutti siamo fortemente esposti agli artefatti culturali, che la rete ci propone: ignorarli sarebbe un errore per chi si occupa di educazione e di formazione. Secondo Jenkins e Gitelman, i media sono definiti attraverso un modello a due livelli: innanzitutto un media è una tecnologia, che consente la comunicazione; poi è anche un insieme di protocolli associati a pratiche sociali e culturali, che sono cresciute attorno a quella tecnologia (Jenkins, 2006).
- **I meme sono un linguaggio e certamente anche una forma comunicativa, che le nostre studentesse e i nostri studenti sanno parlare e sanno leggere**: laddove loro dimostrano padronanza, noi insegnanti sembriamo balbettare. Bucare l'opportunità di ritrovarsi in un terreno comune parlando la stessa lingua vorrebbe dire perdere un gancio utile per motivare e coinvolgere.
- **Le AI non sanno fare i meme**: l'artefatto prevede troppi scenari e allude a situazioni culturali, personali, collettive troppo complesse perché le AI ne possano tenere conto. Ci sono elementi che afferiscono alla cultura, al simbolismo, al visual design che le AI non sanno processare e riassemblare: non sono state allenate a questa complessità. Anche nella definizione di meme le AI entrano in crisi: lo studio sui meme utilizzati a scuola è giovane e in rete non se ne parla come artefatti didattici. I meme si fanno e si diffondono, la popolazione della rete non ne discute il valore e l'impatto educativo.

Ecco degli esempi: ho chiesto ad alcune AI di dare una definizione di meme poi di crearne uno. Queste sono state le risposte del Bot Dragonfly.

I meme sono immagini o video che vengono condivisi online e modificati per mostrare una reazione o un'esperienza condivisa. Sono diventati popolari negli anni 2000 su Internet e sono ora una parte comune della cultura online.

Tell me more.

I meme sono spesso usati per fare commenti ironici su eventi di attualità o situazioni comuni. Possono anche essere usati come forma di intrattenimento, con immagini o video divertenti che vengono condivisi tra amici. Inoltre, i meme sono diventati un modo popolare per esprimere sentimenti ed emozioni online.

What are some popular meme formats?

Alcuni dei formati di meme più popolari includono immagini con didascalie divertenti, immagini con frasi divertenti o immagini che mostrano un personaggio famoso o un animale che esprime un'emozione. Inoltre, i meme spesso includono immagini con una frase o una parola che viene ripetuta più volte per enfatizzare un punto.

What are some popular animal memes?

Alcuni degli animali più popolari nei meme sono il gatto Grumpy Cat, il cane Doge e il coniglio Rabbit. Altri animali popolari nei meme

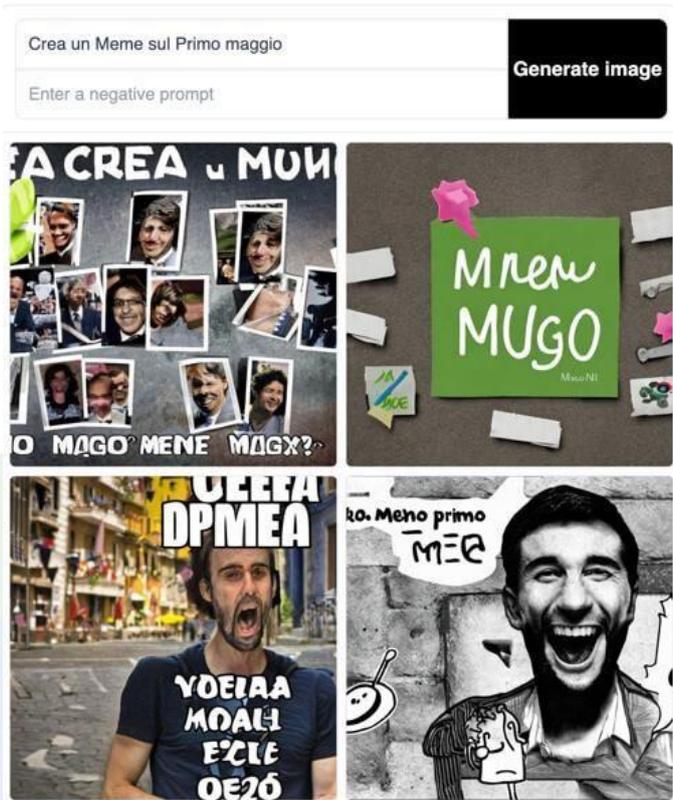
Fai un Meme sul primo maggio



This bot may make incorrect statements. It does not have knowledge of events after 2021. Tap to learn more.

Mi dispiace, ma come modello di linguaggio non sono in grado di creare immagini o meme. Tuttavia, posso fornirti alcune idee per creare un meme sul primo maggio se ti va. Ad esempio, potresti creare un meme che mostra un lavoratore che festeggia il primo maggio, magari con la scritta "Quando finalmente arriva il primo maggio e posso riposare un po'". Oppure, potresti creare un meme che mostra una persona che festeggia il primo maggio con una grigliata all'aperto, con la scritta "Il primo maggio: quando finalmente posso accendere il barbecue". Spero che queste idee ti siano utili!

Poi ho chiesto ad altre due AI di creare un meme sul primo maggio: una (Claude Instant) fornisce una descrizione testuale inefficace e l'altra ([Stable diffusion](#)) crea delle immagini mostruose e inquietanti.



Crea un Meme sul primo maggio

 This bot may make incorrect statements. It does not have knowledge of events after 2021. Tap to learn more.

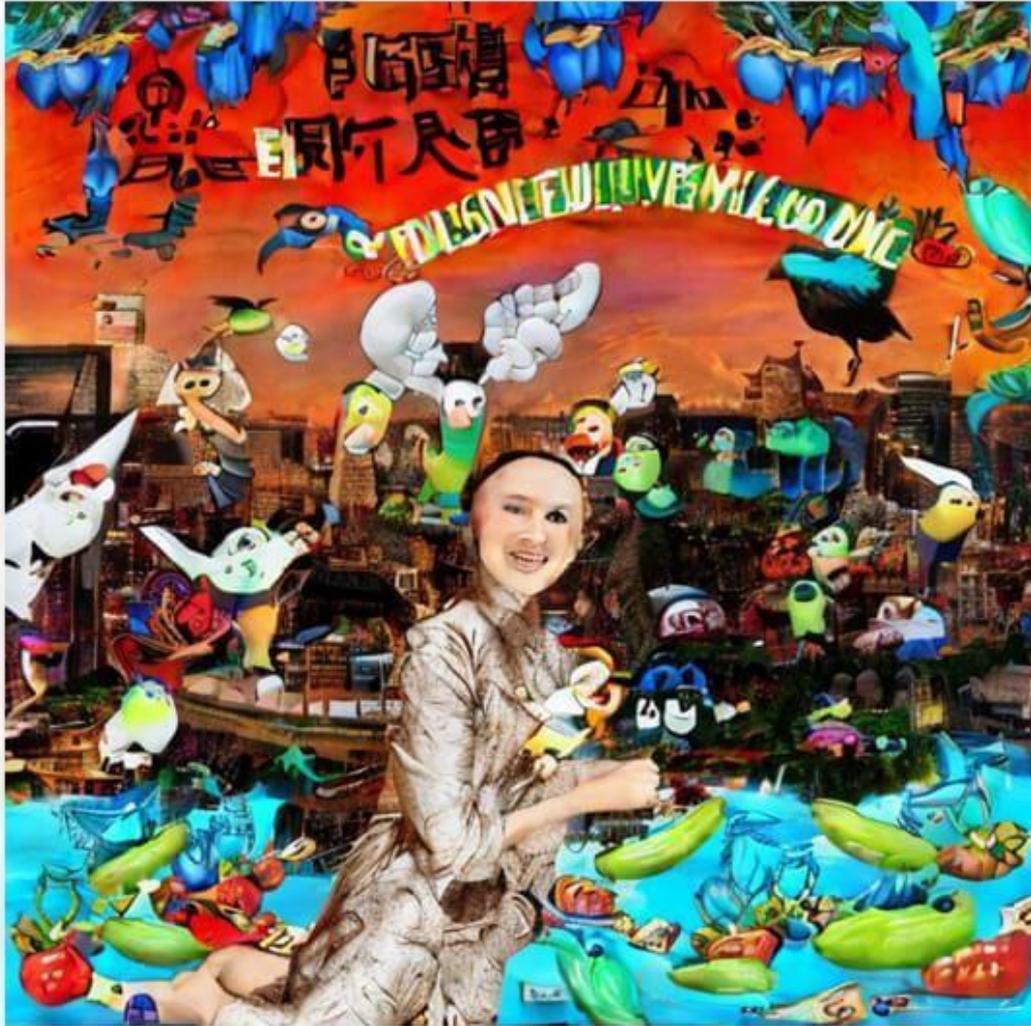
Ecco un meme sul primo maggio:
 [Immagine: Meme 1 maggio
 Testo:
 Primo Maggio, fiera dei lavoratori.
 Saldi e sconti ovunque.
 Il mio portafoglio sta impazzendo!]

...

Meme

Generate Image

Enter a negative prompt



Dalla prima definizione di meme come fenomeno di Internet (Davison, 2012), pezzo di cultura, scherzo, che acquista influenza attraverso la trasmissione online si passa alla definizione dell'*Oxford Dictionary online*, secondo il quale un meme è un'idea, che viene trasmessa da un membro della società all'altro, non nei geni ma dalle persone. Può essere anche un'immagine, video, testo, ecc. che passa molto rapidamente, spesso con lievi modifiche che la rendono spiritosa. In italiano il termine meme viene rintracciato per la prima volta nel 1979 con il significato, affine a quello definito da Dawkins, e quindi appartenente al linguaggio tecnico-specialistico della biologia come "elemento regolante i comportamenti degli individui di una collettività della stessa specie trasmesso non per via genetica ma per via culturale attraverso apprendimenti imitativi".



Oggi

la stessa Enciclopedia Treccani descrive il meme come un “unità di informazione culturale e mentale analoga al gene che, in milioni di esemplari in continua ricombinazione, costituirebbe la nostra mente, durante il percorso esistenziale che viene affrontato giorno per giorno”. Per a Treccani il meme “risiede nel cervello umano, nella memoria, e quindi un meme si diffonde più facilmente se è facile da imitare e quindi da memorizzare.” (Treccani, 2022). [L'Accademia della Crusca](#), nella sua sezione del sito web dedicata alle consulenze linguistiche, ne dà una definizione simile e in perfetta sintonia con lo spirito del meme ne offre un esaustivo esempio (realizzato da Vera Gheno nel 2014).

Dal punto di vista linguistico è attestato in italiano dal 2017 anche il verbo *memare*, con il significato di creare, diffondere e condividere meme, ma sui social network ci sono occorrenze precedenti al 2017. Su Twitter è stata rintracciata [una prima attestazione](#) del 2012, in cui il verbo è usato transitivamente (di Valvasone, 2022):

← Tweet



yasmina
@cnthaveyou

alla ricerca di altre foto da memare (?)

10:06 PM · 8 mag 2012 · Twitter Web Client

I meme sono generalmente condivisi in piattaforme di social media e siti web dedicati (aggregatori di meme): questi ambienti, identificati in letteratura con i termini memosphere (Stryker,2011) o memescape (Wiggins & Bowers,2015), sono moderati da regole stabilite e condivise collettivamente (Osterroth, 2018) e costituiscono l'habitat virtuale e culturale in cui vengono creati e consumati i meme di Internet.



Gli elementi che costituiscono un meme

Collegando insieme le varie definizioni, possiamo elencare gli elementi costituenti i meme: •

- un'immagine digitale e/o un video (breve come una gif oppure più lungo) o un testo:
 - una natura tipicamente umoristica e provocatoria;
 - un'altissima mutevolezza e diffusione;
 - la capacità di essere un segmento culturale vivo.

A causa della loro natura digitale, i meme di Internet si sono diffusi nello spazio e nelle culture a un ritmo estremamente più rapido rispetto i loro antenati offline e hanno raggiunto una massiccia diffusione mondiale, però i meme non sono solo virali (prevedono una condivisione multipla):

necessitano di una vera reinvenzione e una diversa struttura di partecipazione (Shifman, 2014). Se il meme è una vera e propria cornice interpretativa della realtà, filtrata con ironia (talvolta sarcasmo) ed esposta alla divulgazione e alla perenne e continua modifica degli utenti della rete possiamo ritrovare un legame significativo tra la diffusione dei meme e gli studi sulle reti neurali, le neuroscienze cognitive e la psicologia evolutiva tanto, che per un periodo si parlò di *scienza memetica*.

Le potenzialità da esplorare nella didattica e nella formazione

Ho cominciato ad utilizzare questi oggetti pop perché ho visto potenzialità da esplorare, sia nella didattica che nella formazione. Sono innanzitutto diventata autrice di meme (i meme poi sono sempre anonimizzati dalla rete: rispetto all'autorialità, è più importante dove sono stati visti per la prima volta) e li ho condivisi con regolarità in ambienti social oppure nelle comunità scolastiche: mi divertivo e contemporaneamente li capivo sempre meglio. Nel 2017 ho cominciato a studiarli, cercando di analizzarne il funzionamento: da ricercatrice mi sono chiesta come scattasse il coinvolgimento e da dove nascesse la provocazione, quali aspetti venissero maggiormente modificati nei meme di internet più ricorrenti, che cosa fosse immediatamente comprensibile e che cosa invece meritasse un lavoro di analisi più accurato.



Meme e processi di apprendimento

Ho riconosciuto nei meme una forte analogia con i processi di apprendimento:

- è necessaria l'accessibilità al compito
- si prevede l'acquisizione di familiarità e poi di una certa padronanza;
- perché il compito o la proposta sia efficace è necessario personalizzarla;

- l'apprendimento significativo si può vedere solo quando avviene un vero Transfer e si applica qualcosa che si è imparato anche in luoghi e contesti diversi da quelli in cui è avvenuto l'apprendimento;
- La natura di artefatto digitale mediale richiamava due aspetti fondanti delle tecnologie didattiche digitali: la condivisione e la partecipazione.
- Il meme rappresentava sempre una sfida creativa sociale, comunicativa, mediale, culturale. Ho cominciato ad utilizzarli nella formazione, come icebreaker o per avvicinare i docenti ai temi che proponevo: **la libertà di comunicare senza troppi filtri le proprie emozioni, l'abbassamento delle inibizioni, l'ironia erano il fattore vincente.** In seguito, li ho sperimentati con le studentesse e gli studenti: in una prima fase li ho proposti io, per avvicinarli con apparente leggerezza a temi complessi; successivamente ho chiesto loro di produrre meme per accedere a diversi livelli di profondità di un tema, attingere alle loro comprensioni significative o ai loro fraintendimenti senza esercitare la violenza di un'interrogazione o la minaccia di un voto. Ho usato i meme nella mia ricerca di dottorato, chiedendo ad un gruppo di Animatori digitali trentini di utilizzarli per comunicare la loro percezione di efficacia nel loro ruolo a scuola e come pensavano di essere percepiti dalle colleghe e dai colleghi. L'eccentricità del meme ha rotto le difese, allentato i filtri affettivi e generato curiosità e coinvolgimento: la sua natura allegorica ha permesso di attingere ad un immaginario collettivo (del quale si deve essere a conoscenza, per poterlo manipolare in modo personale e trovare nel destinatario un contributore del significato) e dall'altra ha stimolato il desiderio di sfida creativa e originalità.

Restituire la propria opinione, l'interpretazione di un evento, un apprendimento attraverso uno o più meme diventa un gioco molto serio, soprattutto se si sa come leggere i meme prodotti da ogni studente e si è padroni del linguaggio specializzato, che si richiede di usare perché collegato alla disciplina o all'area semantica.

Realizzare un meme

Realizzare un meme è un compito autentico, complesso e allo stesso tempo molto preciso: analizzare un meme è già una operazione valutativa del tema affidato e della conoscenza che se ne ha e comporta una competenza specifica che investe gli elementi visuali, il testo e la sintesi. In un percorso sulla Didattica con i meme che ho progettato e realizzato insieme all'Iprase e un gruppo di docenti trentini, una collega (Roberta Dallabetta, Docente presso Sophie Magdalena Scholl di Trento e la sua classe di cinese, 4B, che ringrazio) li ha utilizzati per proporre lo studio della poesia Shī, nata durante il Medioevo cinese, e verificare la comprensione di almeno uno degli autori proposti. Ecco due esempi di meme, che naturalmente sono comprensibili solo dalla comunità, alla quale fanno riferimento, non solo per la lingua ma anche per il messaggio:



宇航员?

Yǔháng yuán
Astronauta?

不是! 喝醉了!

La parte tecnologica e digitale del meme

La parte tecnologica e digitale del meme è in verità essenziale e accessibile: esistono numerose web app, senza registrazione, anche sugli smartphone più comuni, che in pochi minuti supportano la realizzazione di meme foto e meme gif. Si può immaginare di far partire tutti dalla stessa immagine oppure lasciare la libertà di scegliere e comporre un meme come risposta ad una domanda o come metafora di una situazione, di un evento storico, di una teoria scientifica. Le immagini possono essere sequenze di film, strisce di fumetti, rappresentazioni di opere d'arte o scatti di eventi di cronaca: questa estrema libertà di selezione va poi combinata con la parte testuale e quello che nasce da questi significanti non è mai la semplice somma dei due elementi, ma un terzo complesso significato, che ha bisogno di un processo dialettico per acquisire senso.



Il lettore del meme contribuisce a completare il messaggio del creatore del meme, ne fa parte attivamente. Impossibile non pensare ai timori legati al progresso tecnologico espressi da Benjamin e all'opera d'arte nell'epoca della sua riproducibilità tecnica (ed estetica). È infatti la riproducibilità, che determina il punto di vista su quello che vediamo e nel meme gioca un ruolo politico. La perdita di aura è superata dall'instagrammabilità, dalla viralità, dalla spinta imitatrice. Il Medium non è solo il messaggio, ma permette il messaggio, lo sostituisce: i media non sono più solo protesi, ma estensioni dei nostri sensi e dei nostri pensieri.

Dopo aver elaborato un modello metodologico di uso didattico dei meme, ho scoperto che c'era qualche collega che stava sperimentando i meme nella didattica come me: chi in letteratura, chi in geografia, chi in storia. Ho studiato la sperimentazione dei meme matematici (Bini & Robutti, 2019, 2022), che ha posto in correlazione gli ambienti culturali nei quali i meme sono prodotti e il loro valore educativo-epistemologico, sperimentandoli anche a scuola.

Il potenziale inclusivo dei meme

Andrebbe avviata una riflessione ancora più articolata sul potenziale inclusivo dei meme, i loro vantaggi in termini di accessibilità, motivazione e adattabilità ad ogni disciplina. **Le sequenze memetiche possono diventare frame narrativi:** i meme attingono ad un linguaggio

contemporaneo e contribuiscono anche a svelare le criticità e i limiti di un uso pubblico talvolta non accessibile a tutti; preparano gli studenti ad interpretarli e a crearne di personali con consapevolezza, evitando l'effetto di passività, che porterebbe a subirli come messaggi di propaganda politica o sociale (Di Donato, 2021). Riducono stress e ansia: la natura ludica/umoristica può ridurre il carico emotivo percepito dagli studenti durante la lezione o durante la valutazione e l'appartenenza ad una cultura partecipativa può diventare opportunità per lavorare sulla *digital citizenship*. La letteratura e la ricerca sull'uso didattico-educativo dei meme sono ancora agli inizi. Un approccio possibile che suggerisco è quello che li associa alla retorica visiva, che combina elementi dell'approccio semiotico e discorsivo, per analizzare gli elementi persuasivi dei testi visivi (Faccioli & Losacco, 2003) e dell'*opera aperta* a continue e infinite variazioni (Eco, 1984, 2013).

I meme poi hanno il pregio di **unire l'elemento partecipativo della cultura con la convergenza mediale**. Jenkins (2006) spiega che una cultura partecipativa è una cultura con barriere relativamente basse per l'espressione artistica e l'impegno civico, un forte sostegno alla creazione e alla condivisione delle proprie creazioni e un qualche tipo di tutoraggio informale, in cui ciò che è conosciuto dai più esperti viene trasmesso ai novizi. La convergenza riguarda invece da una parte una significativa disponibilità di informazioni incanalate in un unico dispositivo, dall'altra è un vero cambiamento culturale, che incoraggia i fruitori del media nella ricerca di nuove informazioni e nel creare connessioni tra i media dispersi. Incoraggia la partecipazione e l'intelligenza collettiva e nella stessa prospettiva le culture si intersecano, contribuendo ad attivare la crossmedialità del meme: scrittura, libri, film, video, fumetti, fotografie, eventi sono la fonte inesauribile di ispirazione e i media digitali come il meme fluttuano dall'uno all'altro, senza attrito, attingendo a tutte le fonti possibili per comunicare in modo efficace il messaggio. La trasformazione più grande nei media digitali è forse il passaggio dal consumo mediale individuale e personalizzato a quello che viene vissuto come pratica collettiva e reticolare: i cambiamenti più rilevanti avvengono quindi nelle comunità di consumo.

Per il lavoro proposto agli insegnanti ho trovato utile seguire soprattutto l'approccio costruzionista di tipo discorsivo (Hall, 1997), nel quale il meme è una porta per accedere all'intero sistema in cui nasce quella rappresentazione dei saperi (Milner, 2012) e il discorso diventa l'unico modo per parlare di un certo argomento, in un determinato momento storico o in una sua fase (Foucault, 2016). Poi il costrutto dei significati parziali (Bini & Robutti, 2019), che individua i tre significati parziali nel meme, che devono essere tutti accessibili perché il meme raggiunga la comprensione. La scelta del meme come media formativo è stata pensata proprio per restituire un ruolo attivo sia ai produttori che ai destinatari della comunicazione.

Gli schemi sono le strutture di conoscenza che rappresentano un concetto e se mancano gli schemi, la capacità di ricordare, riflettere e interpretare un'esperienza è limitata (Wu & Dunning, 2020); per questo offrire dei modelli di comunicazione e dei linguaggi nuovi, per permettere le associazioni cognitive necessarie, comunicare le informazioni apprese nel corso delle esperienze e le emozioni suscitate, può aiutare chi apprende a percepirsi in modo più chiaro e ad attivare risorse metacognitive.

Conclusioni

In un ambiente come quello in cui i meme sono immersi, le studentesse e gli studenti possono dare prova di abilità non standard come la creatività, lo humour e la competenza nella cultura popolare. Si possono scovare Bias cognitivi (uno di quelli testati con i meme è il DKE), recuperare fraintendimenti, lavorare su piani transdisciplinari. Insomma, credo che il lavoro su apprendimento e memetica possa finalmente entrare ufficialmente nell'area delle metodologie didattiche innovative. La nuova sfida per i docenti è aperta. **Bibliografia**

Bini, G. & Robutti, O. (2019). Pensare all'interno del post: indagare sull'uso didattico dei meme matematici di Internet. In A. Shvarts (a cura di), Atti della conferenza russa PME e Yandex

“Tecnologia e psicologia per l'educazione matematica ” (pp. 106–113). Casa editrice HSE Bini,

G., Robutti, O., & Bikner-Ahsbahs, A. (2022). Maths in the time of social media:

conceptualizing the Internet phenomenon of mathematical memes. *International Journal of Mathematical Education in Science and Technology*, 53(6), 1257-1296.

Blackmore, S. (2002). *La macchina dei Memi: perché i geni non bastano*. Torino: Instar Libri.

Dawkins, R. (2017). *Il gene egoista. La parte immortale di ogni essere vivente*. Mondadori.

Davison, P. (2012). Il linguaggio dei Meme di Internet. In M. Mandiberg (a cura di), *The social media reader*, Michael Mandiberg, New York University Press.

Di Donato, D. (2021). *Didattica con i Meme*. In Biancato L. & Tonioli D. (Eds). 101 idee per la Didattica Digitale Integrata. (pp. 240-242). Erickson.

Di Donato, D. (2023). *Oltre il digitale. La formazione dei docenti tra strumenti e pratiche*. In

Relazioni, Il mondo come aula (03). Luca Sossella Editore. di Valvasone Luisa. (2022),

L'arte di memare non è per tutti. In *Italiano digitale. Rivista dell'Accademia della Crusca*, XX, 2022/1 (Gennaio-Marzo). Disponibile

qui: <https://id.accademiadellacrusca.org/articoli/larte-di-memare-non-per-tutti/14793>

Eco, U. (1984). *Semiotica e filosofia del linguaggio*. Einaudi

Eco, U. (2013). *Opera aperta*. Bompiani

Faccioli, P. & Losacco G. (2003). *Nuovo manuale di sociologia visuale. Dall'analogico al digitale*. FrancoAngeli.

Falcinelli, R. (2020). *Figure. Come funzionano le immagini dal Rinascimento a Instagram*. Einaudi.

Jenkins H. (2006). *Cultura convergente*. Apogeo.

Lolli, M. (2017). *La guerra dei meme. Fenomenologia di uno scherzo infinito*. Effequ.

Milner, R. M. (2012). *The world made meme: Discourse and identity in participatory media*.

(Unpublished doctoral dissertation). University of Kansas, Lawrence, KS. Retrieved from academia.ed

Shifman, L., Levy, H., & Thelwall, M. (2014). Internet jokes: The secret agents of globalization? *Journal of Computer-Mediated Communication*, 19(4), 727–743.

I meme pubblicati in questo articolo sono tutti dell'autrice, tranne quelli citati: Gheno, 2014; di Valvasone, 2012; Dallabetta, 2023.

McLuhan, M., (1964), *Unde*  London, Sphere Books trad.it.



Gli strumenti del comunicare, Milano, Il Saggiatore, 1997.

Mageira K, et Al., Pittou D, Papasalouros A, Kotis K, Zangogianni P, Daradoumis A., (2022)
“Educational AI Chatbots for Content and Language Integrated Learning”. *Applied Sciences.*;
12(7):3239. <https://doi.org/10.3390/app12073239Si>



Quiroga Pérez, J. I. and L. J. T. Marquès J. M. (2022) *Accesso a Internet e servizi in*  **Agenda Digitale** ^{eu}



education: A systematic literature review”. *Comput Appl Eng Educ.* 2020; 28: pp. 1549– 1565.

Thompson, A. D. (2022). *What’s in my AI? A Comprehensive Analysis of Datasets Used to Train GPT-1, GPT-2, GPT-3, GPT-NeoX-20B, Megatron-11B, MT-NLG, and Gopher.*

<https://LifeArchitect.ai/whats-in-my-ai>

Trentin G., eds (2021), *Conversational Agents as Online Learning Tutors*, Hauppauge NY, Nova

Science Pub Inc.

Tuomi, I. (0218), *The Impact of Artificial Intelligence on Learning, Teaching, and Education.*

Policies for the future, (eds) Cabrera, M., Vuorikari, R & Punie, Y., EUR 29442 EN, Luxembourg

Publications Office of the European Union.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., .Polosukhin, I.

(2017). "Attention is all you need. Advances in neural information processing systems" 30.
NETWORK **DIGITAL** 360



Disponibile al sito <https://arxiv.org/abs/1706.03762>

Network Digital360 è il più grande network in Italia di testate e portali B2b

dedicati ai temi della Trasformazione Digitale e dell'Innovazione
Imprenditoriale, con oltre 50 fra portali, canali e newsletter.

Ha la missione di diffondere la cultura digitale e imprenditoriale nelle imprese e pubbliche amministrazioni italiane e di fornire a tutti i decisori che devono valutare investimenti tecnologici informazioni aggiornate e approfondite. Il Network è parte integrante di Digital360HUB, il polo di Demand Generation di Digital360, che mette a disposizione delle tech company un'ampia gamma di servizi di comunicazione, storytelling, pr, content marketing, marketing automation, inbound marketing, lead generation, eventi e webinar.

VIA COPERNICO, 38

20125 - MILANO

TE L. 02 92852785

MAIL: MARKETING@DIGITAL4.BIZ

©ICT & Strategy