

I quaderni di

Agenda  **Digitale** ^{eu}



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, Direttore responsabile Agendadigitale.eu

Direttori scientifici

Paolo Ferri, Professore Ordinario di Tecnologie della formazione, Università degli Studi Milano-Bicocca

Mario Morcellini, Professore ordinario emerito in Sociologia della Comunicazione e dei Media digitali, Sapienza Università di Roma

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 editoriale

Giovanni Boccia Artieri, Marco Centorrino, Antonio Opromolla, Mario Pireddu, Luca Toschi

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

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

Le nuove frontiere dell'Intelligenza Artificiale: dalle reti neurali all'IA etica	10
Di Antonio Chella, Laboratorio di Robotica – Dipartimento di Ingegneria Università degli Studi di Palermo	10
Apprendimento neurale: il contributo dei premi Nobel Hinton e Hopfield all'IA moderna	20
di Alessandro Acciai, Unicusano e Alessio Plebe, Università degli Studi di Messina	20
Co-intelligence: il futuro è dell'IA democratica	29
Di Antonio Cisternino Università di Pisa.....	29
La sfida dell'IA generativa multimodale.....	41
Di Bernardo Magnini, Fondazione Bruno Kessler, Trento.....	41
IA “trasparente”, solo così l'uomo torna al centro: come fare	46
Di Arianna Pipitone, Università degli studi di Palermo, National Research Council ICAR – CNR Palermo site	46
AI autonoma: il ruolo degli agenti intelligenti e dei sistemi multi-agente	53
Di Stefania Costantini, Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, L'Aquila, Italy	53
Agenti intelligenti: il cammino verso l'AI che comprende e supporta	64
Di Valeria Seidita, Dipartimento di Ingegneria, Università degli Studi di Palermo.....	64
Ragionamento artificiale: come cambia l'intelligenza delle macchine	76
Di Antonio Chella, Laboratorio di Robotica, dipartimento di Ingegneria Università degli Studi di Palermo	76
Pensare come una macchina: il percorso evolutivo verso un'IA che ragiona.....	81
Di Vito Roberto, Docente Senior, Dipartimento di Informatica, Università di Udine	81
L'inganno umano e la vulnerabilità degli algoritmi	91
Di Raffaella I. Rumiati, SISSA Trieste.....	91
Robot più consapevoli e trasparenti grazie all'IA.....	97
Di Silvia Rossi, Università di Napoli Federico II.....	97
Il futuro dei sentimenti: intelligenza artificiale, etica e nuovi robot	104
Di Pietro Perconti, Dipartimento di scienze cognitive, Università di Messina	104
L'IA oltre la mente: sfide filosofiche nell'era delle reti neurali	115
Di Marcello Frixione, Università di Genova.....	115
Coscienza artificiale: perché (e se) investire su un'IA senziente	121

Di Michele Farisco, Unità di Bioetica, Biogem scarl, Ariano Irpino (AV), Centre for Research Ethics and Bioethics, Uppsala University, Uppsala, Sweden	121
L'economia dell'IA: come la sostituzione cognitiva cambierà il lavoro.....	129
Di Sebastiano Bavetta Università di Palermo e Co-founder@SandB	129
Processi decisionali in un mondo caotico: strategie per imprese e società	141
Di Mauro Lombardi, Università di Firenze, BABEL – Blockchain and Artificial intelligence for Business, Economics and Law	141
Educare all'empatia: la soluzione contro l'analfabetismo emotivo sui social	153
Di Marino D'Amore, Docente di Sociologia generale presso Università degli Studi Niccolò Cusano.....	153
Assistenti virtuali in classe: così favoriscono benessere emotivo e apprendimento	160
Di Angelamaria De Feo, Ricercatrice presso Università Telematica Pegaso	160
Adolescenti sui social, vittime o carnefici? Le molte facce della violenza digitale	171
Di Giorgia Di Iorio, Dipartimento Neuroscienze Umane, Sapienza Università di Roma, UOC Neuropsichiatria Infantile, Policlinico Umberto I.....	171
Dario Esposito, Dipartimento di Neuroscienze Umane, Sapienza Università di Roma	171
Sara Romano, Dipartimento di Neuroscienze Umane, Sapienza Università di Roma, Unità Operativa di Neuropsichiatria Infantile, Policlinico Umberto I	171
Bazar e baroque economy: l'evoluzione dello shopping online su TikTok	179
Di Brigida Orria, Post-doc research fellow, Università degli Studi di Napoli Federico II.....	179
Sabrina Bellafronte, PhD student, dipartimento di scienze sociali all'Università degli studi di Napoli Federico II	179
Nuove frontiere del sapere: integrazione tra uomo e tecnologia	192
Di Alessandro Ugo Imbriglia, Sociologo del conflitto e dell'industria culturale	192

Le nuove frontiere dell'Intelligenza Artificiale: dalle reti neurali all'IA etica

L'IA sta evolvendo verso sistemi sempre più autonomi e sofisticati, integrando auto-apprendimento, multimodalità e nuove capacità computazionali. Questa convergenza promette di espandere le capacità cognitive umane, sollevando al contempo questioni etiche e filosofiche significative

Di **Antonio Chella**, Laboratorio di Robotica – Dipartimento di Ingegneria Università degli Studi di Palermo

L'intelligenza artificiale sta attraversando una fase di trasformazione radicale che sta ridefinendo non solo le capacità tecniche dei sistemi computazionali, ma anche la comprensione dell'intelligenza e della coscienza. Questa evoluzione si manifesta attraverso molteplici direzioni di sviluppo che stanno convergendo verso la creazione di sistemi più sofisticati, autonomi e anche potenzialmente coscienti.

In questo panorama in rapida evoluzione, emergono alcune tendenze fondamentali che stanno plasmando il futuro dell'IA. L'auto-apprendimento permette ai sistemi di cercare ed elaborare autonomamente le informazioni necessarie per il loro sviluppo. La multimodalità sta avvicinando l'IA alla ricchezza dell'esperienza sensoriale umana, integrando diverse forme di percezione e comprensione. I progressi nell'efficienza computazionale stanno liberalizzando l'accesso a queste tecnologie, rendendo possibile l'addestramento di sistemi avanzati anche con risorse limitate.

La convergenza tra neuroscienze e IA sta inoltre aprendo nuove frontiere nella comprensione della coscienza e della cognizione, portando allo sviluppo di sistemi che non solo emulano comportamenti intelligenti, ma potrebbero anche avvicinarsi a forme semplificate di esperienza cosciente. Parallelamente, lo sviluppo di sistemi più trasparenti e capaci di spiegare i propri ragionamenti affronta una delle critiche più significative mosse all'IA: la sua natura di “scatola nera.”

L'emergere di agenti intelligenti sempre più sofisticati apre nuove possibilità nell'automazione e nell'interazione uomo-macchina, mentre i nuovi paradigmi di ragionamento stanno superando i limiti tradizionali dell'apprendimento profondo, introducendo capacità di ragionamento causale e analogico più avanzate. L'integrazione di principi etici nello sviluppo dell'IA è diventata una priorità fondamentale. E' ormai

riconosciuto che la tecnologia deve essere non solo potente ed efficiente, ma anche allineata con i valori umani.

In questo articolo esploreremo alcune direzioni fondamentali dello sviluppo dell'IA, analizzando come si intersecano e si rafforzano reciprocamente, e come stanno contribuendo a plasmare un futuro in cui l'intelligenza artificiale potrebbe eguagliare ed espandere le capacità cognitive umane.

L'auto-apprendimento: quando l'IA diventa autonoma nella ricerca della conoscenza

L'auto-apprendimento rappresenta una delle frontiere più affascinanti dell'intelligenza artificiale. Tradizionalmente, i sistemi di IA vengono addestrati su basi di dati accuratamente selezionati e preparati da esseri umani, un processo che richiede tempo, risorse e una significativa supervisione umana. Tuttavia, gli ultimi sviluppi stanno portando alla creazione di sistemi capaci di cercare autonomamente i dati necessari per il proprio apprendimento, aprendo nuove prospettive nel campo dell'intelligenza artificiale [1, 2]. Questi sistemi utilizzano algoritmi di esplorazione che permettono loro di navigare in vasti spazi informativi, selezionando autonomamente i dati più rilevanti per il loro addestramento. Un esempio è rappresentato dai sistemi di *web crawling* intelligente, che non si limitano a raccogliere informazioni in modo indiscriminato, ma valutano la qualità e la pertinenza dei dati rispetto ai loro obiettivi di apprendimento [3]. Questa capacità di auto-selezione dei dati rappresenta un passo importante verso sistemi di IA autonomi e adattivi. L'auto-apprendimento non si limita alla raccolta di dati, ma include anche la capacità di identificare schemi e relazioni significative senza la supervisione umana. Questi sistemi possono scoprire autonomamente nuove categorie e classificazioni, evolvendosi continuamente in base alle informazioni che incontrano. Tale approccio potrebbe rivoluzionare la ricerca scientifica, dove i sistemi di IA possono esplorare enormi database di pubblicazioni scientifiche, identificare collegamenti nascosti e suggerire nuove direzioni di ricerca.

Tuttavia, l'auto-apprendimento solleva importanti sfide. La qualità dei dati è un aspetto cruciale: un sistema che apprende autonomamente potrebbe basarsi su informazioni errate o distorte, perpetuando o amplificando pregiudizi esistenti. Inoltre, la capacità di questi sistemi di evolversi autonomamente solleva questioni sulla prevedibilità e controllabilità del loro comportamento. Sarà quindi fondamentale sviluppare meccanismi di controllo e validazione che garantiscano che l'apprendimento autonomo avvenga entro confini sicuri ed eticamente accettabili.

La multimodalità: verso un'intelligenza artificiale che integra diversi sensi

La multimodalità rappresenta un salto fondamentale nello sviluppo dell'intelligenza artificiale, avvicinando i sistemi artificiali alla complessità dell'esperienza umana. I sistemi multimodali sono sistemi in grado di integrare ed elaborare simultaneamente diverse forme di input - testo, immagini, suoni, dati tattili e altri tipi di sensori - creando una comprensione ricca e contestualizzata dell'ambiente [4,5,6].

Questa evoluzione porterà alla creazione di sistemi di IA più sofisticati e versatili. Ad esempio, i moderni assistenti robotici potranno combinare la visione artificiale con il feedback tattile per manipolare oggetti con maggiore precisione, mentre i sistemi di diagnosi medica potranno integrare immagini diagnostiche, dati clinici testuali e parametri biometrici per fornire valutazioni più accurate. La multimodalità non si limita alla semplice raccolta di dati da diverse fonti, ma implica una fusione intelligente di informazioni che permette di cogliere sfumature e correlazioni che potrebbero sfuggire analizzando ogni singola modalità separatamente.

Un aspetto particolarmente promettente della multimodalità è la sua applicazione nel campo dell'interazione tra persone e robot [7]. I sistemi multimodali possono interpretare non solo le parole pronunciate, ma anche il tono della voce, le espressioni facciali e il linguaggio corporeo, permettendo interazioni più naturali e intuitive. Questo approccio potrebbe rivoluzionare settori come l'assistenza sanitaria, l'educazione e il supporto alle persone con disabilità, dove la capacità di comprendere e rispondere a diversi tipi di segnali è fondamentale.

La sfida principale nella progettazione di sistemi multimodali risiede nella necessità di gestire e integrare efficacemente flussi di dati molto diversi tra loro. Ogni modalità ha le sue caratteristiche specifiche in termini di formato, velocità di acquisizione e rumore di fondo. Inoltre, le diverse modalità potrebbero fornire informazioni apparentemente contraddittorie, rendendo necessario lo sviluppo di sofisticati algoritmi di fusione e risoluzione dei conflitti.

IA democratica: sistemi efficienti con risorse limitate

Uno degli sviluppi più significativi nel campo dell'intelligenza artificiale riguarda la creazione di sistemi che possano essere addestrati ed eseguiti su infrastrutture computazionali modeste. Questa tendenza verso la “democratizzazione” dell'intelligenza artificiale sta aprendo nuove possibilità ai ricercatori, alle piccole imprese e agli sviluppatori individuali che non hanno accesso a grandi centri di calcolo o risorse costose. Un esempio in questo campo è il sistema OpenCog Hyperon [8] sviluppato dal gruppo di ricerca di Ben Goertzel, uno dei padri della AGI - Artificial General Intelligence.

Le innovazioni in questo campo si concentrano su diverse strategie di ottimizzazione. Una direzione promettente è lo sviluppo di architetture neurali efficienti, che richiedono pochi parametri pur mantenendo prestazioni competitive. Queste “reti leggere” utilizzano tecniche come la potatura dei collegamenti neuronali meno significativi e la quantizzazione dei pesi, riducendo significativamente le risorse necessarie per l'addestramento e l'inferenza [9].

Un altro approccio interessante è basato sull'ottimizzazione dell'apprendimento, dove modelli pre-addestrati su grandi quantità di dati sono adattati efficacemente per compiti specifici con limitate risorse computazionali [10]. Questa tecnica permette di sfruttare la conoscenza accumulata da modelli più grandi senza dover ripetere l'intero processo di addestramento da zero.

La ricerca sta anche esplorando approcci distribuiti che permettono di suddividere il carico computazionale su reti di dispositivi più piccoli. Questi sistemi collaborativi possono sfruttare la potenza aggregata di multiple unità di calcolo modeste, creando una sorta di “supercalcolatore virtuale” accessibile anche a realtà con risorse limitate [11].

Questa IA più democratica ha importanti implicazioni sociali ed economiche. Permette una più ampia partecipazione all'innovazione nell'intelligenza artificiale, promuovendo la diversità delle applicazioni e delle prospettive. Inoltre, rende possibile lo sviluppo di soluzioni più personalizzate e contestualizzate, adatte alle esigenze specifiche di comunità e mercati locali.

Trasparenza: verso un'IA comprensibile e verificabile

La trasparenza dei sistemi di intelligenza artificiale e la loro capacità di fornire spiegazioni comprensibili per le loro decisioni stanno diventando sempre più cruciali man mano che l'IA assume un ruolo più importante in decisioni che impattano la vita delle persone. Il campo dell'IA spiegabile (XAI - Explainable Artificial Intelligence) sta facendo grandi progressi nello sviluppo di sistemi che non si limitano a fornire risultati, ma sono in grado di articolare il processo di ragionamento che li ha portati a determinate conclusioni [12, 13].

I sistemi che implementano un dialogo interiore (inner speech) rappresentano un esempio interessante di questo approccio [14]. Questi sistemi simulano un dialogo interno simile al pensiero umano, esplicitando i passaggi logici e le considerazioni che portano a una determinata conclusione. Questo non solo rende più comprensibile il processo decisionale dell'IA, ma permette anche di identificare potenziali errori o pregiudizi nel ragionamento.

La trasparenza non si limita alla spiegazione post-hoc delle decisioni, ma include anche la possibilità di esaminare e comprendere la struttura interna dei modelli di IA. Nuove tecniche di visualizzazione e analisi permetteranno di aprire la scatola nera dei sistemi neurali, mostrando come diverse parti della rete si attivano in risposta a diversi input e come le informazioni vengono elaborate attraverso i vari strati del modello [15].

Questi sviluppi sono importanti in settori critici come la medicina, la finanza e il sistema giudiziario, dove le decisioni dell'IA devono essere non solo accurate ma anche giustificabili e verificabili. La capacità di un sistema di spiegare le proprie decisioni in modo chiaro e comprensibile è fondamentale per costruire la fiducia degli utenti e garantire l'identificazione delle responsabilità delle applicazioni di IA.

Agenti Intelligenti: verso sistemi autonomi e interattivi

L'evoluzione degli agenti intelligenti rappresenta una delle frontiere più promettenti dell'intelligenza artificiale. Gli agenti sono sistemi software in grado di percepire l'ambiente circostante attraverso sensori, prendere decisioni autonome e agire per raggiungere obiettivi specifici. La loro peculiarità risiede nella capacità di operare in modo continuativo e adattivo, apprendendo dalle interazioni con l'ambiente e con altri agenti, come descritto ad esempio in [16, 17].

Gli sviluppi recenti nel campo degli agenti intelligenti hanno portato alla creazione di sistemi sempre più sofisticati, capaci di gestire compiti complessi in ambienti dinamici e

incerti. Un esempio significativo è rappresentato dagli agenti conversazionali avanzati, che non si limitano a rispondere a domande specifiche, ma mantengono una comprensione del contesto della conversazione e possono adattare il loro comportamento in base alle esigenze dell'interlocutore [18].

Un aspetto particolarmente interessante è lo sviluppo di sistemi multi-agente, dove diversi agenti intelligenti collaborano e competono per raggiungere obiettivi comuni. Questi sistemi stanno trovando applicazione in campi come la gestione del traffico urbano, la robotica collaborativa e la simulazione di scenari complessi [19]. La capacità degli agenti di coordinarsi e negoziare tra loro apre nuove possibilità per la risoluzione di problemi distribuiti e la gestione di sistemi complessi.

La ricerca sugli agenti intelligenti sta anche esplorando l'integrazione di capacità cognitive avanzate, come la pianificazione a lungo termine, il ragionamento causale e l'apprendimento per trasferimento di conoscenze tra diversi domini [20]. Questi sviluppi stanno portando alla creazione di agenti sempre più autonomi e versatili, capaci di affrontare situazioni impreviste e di adattarsi a nuovi contesti.

Nuovi Paradigmi di Ragionamento: Oltre l'Apprendimento Profondo

Il campo dell'intelligenza artificiale sta vivendo una significativa evoluzione nei modi in cui i sistemi elaborano informazioni e giungono a conclusioni. Mentre l'apprendimento profondo continua a dominare molte applicazioni, i paradigmi di ragionamento simbolico stanno trovando nuova linfa per affrontare le limitazioni degli approcci basati su reti neurali [21].

Un'innovazione significativa è rappresentata dai sistemi di ragionamento neuro-simbolico, che combinano l'apprendimento profondo con il ragionamento logico [22]. Questi sistemi integrano la capacità di apprendimento delle reti neurali con la precisione e la verificabilità del ragionamento simbolico, permettendo di affrontare problemi che richiedono l'elaborazione di dati grezzi e l'applicazione di regole logiche complesse.

Particolarmente promettente è lo sviluppo di architetture che implementano forme di ragionamento causale [23]. A differenza dei sistemi tradizionali che si basano principalmente su correlazioni statistiche, questi nuovi approcci cercano di comprendere e modellare le relazioni causa-effetto nel mondo reale. Questo permette non solo di fare previsioni più accurate, ma anche di ragionare sugli effetti di potenziali interventi e di rispondere a domande controfattuali.

I sistemi di ragionamento analogico rappresentano un'altra direzione che sta trovando nuovo vigore [24]. Questi sistemi cercano di emulare la capacità umana di trasferire conoscenze e soluzioni da un dominio all'altro attraverso il riconoscimento di schemi e strutture comuni [25]. Tale approccio si sta rivelando particolarmente utile in campi come la risoluzione creativa dei problemi e la scoperta scientifica.

IA e Neuroscienze: verso la coscienza artificiale

L'intersezione tra intelligenza artificiale e neuroscienze rappresenta uno dei campi più affascinanti e promettenti della ricerca. Mentre l'IA continua a fare progressi significativi

nell'emulare comportamenti intelligenti, le neuroscienze offrono preziose intuizioni sui meccanismi fondamentali della cognizione biologica, aprendo nuove prospettive per lo sviluppo di sistemi artificiali più avanzati e, potenzialmente, coscienti [26, 27].

La ricerca sulla coscienza artificiale sta emergendo come un campo distintivo che combina approcci computazionali con teorie neuroscientifiche della coscienza. Teorie come la Global Workspace Theory stanno fornendo framework concettuali per comprendere e potenzialmente replicare aspetti della coscienza in sistemi artificiali [28]. Questi approcci non si limitano a cercare di riprodurre comportamenti esterni, ma mirano a comprendere e replicare i meccanismi interni che potrebbero dare origine all'esperienza cosciente.

Un aspetto particolarmente interessante è lo studio dei correlati neurali della coscienza (NCC) e il tentativo di tradurre queste conoscenze in architetture computazionali [29]. I ricercatori stanno analizzando reti neurali artificiali che incorporano caratteristiche chiave dei circuiti cerebrali coinvolti nella generazione della coscienza, come i meccanismi di feedback ricorrente e l'integrazione di informazioni attraverso diverse aree specializzate [30].

Un tema cruciale in questo campo è la questione della fenomenologia artificiale: se e come i sistemi di IA potrebbero sviluppare esperienze soggettive simili alla coscienza umana. Questo solleva non solo sfide tecniche ma anche profonde questioni filosofiche sulla natura della coscienza e sulla possibilità di replicarla artificialmente [31].

L'integrazione tra IA e neuroscienze sta anche portando a progressi significativi nella comprensione e nel trattamento di disturbi neurologici: ad esempio, i modelli computazionali ispirati al cervello stanno aiutando a comprendere meglio gli stati di coma e alcune patologie come l'Alzheimer e la schizofrenia [32].

L'IA etica: valori e responsabilità nei sistemi artificiali

L'integrazione di principi etici nei sistemi di intelligenza artificiale rappresenta una delle sfide più complesse e importanti del settore. L'IA etica non si limita all'implementazione di semplici regole di comportamento, ma richiede lo sviluppo di sistemi capaci di comprendere e applicare principi morali in contesti complessi e ambigui [33].

Un aspetto fondamentale dell'IA etica è la capacità di bilanciare diversi valori e obiettivi, talvolta in conflitto tra loro. Ad esempio, un sistema di IA nel settore sanitario deve bilanciare l'efficienza operativa con la privacy dei pazienti, o un sistema di guida autonoma deve gestire situazioni dove diversi principi etici potrebbero suggerire azioni contrastanti [34].

La ricerca in questo campo sta esplorando diverse direzioni. Una direzione è lo sviluppo di sistemi etici formali che possono essere incorporati nei sistemi di IA, permettendo di valutare le implicazioni morali delle loro azioni. Un'altra direzione è lo sviluppo di sistemi capaci di apprendere principi etici dall'osservazione del comportamento umano, pur mantenendo la capacità di identificare e correggere eventuali pregiudizi presenti nei dati di addestramento [35, 36]. Una terza direzione infine mira alla fusione dei due approcci suddetti per creare un sistema ibrido di IA etica.

L'IA etica deve anche considerare la diversità culturale e la variabilità dei valori morali tra diverse società e contesti. Questo implica lo sviluppo di sistemi sufficientemente flessibili da adattarsi a diverse cornici etiche, pur mantenendo alcuni principi fondamentali universali come la non maleficenza e il rispetto dell'autonomia umana [37].

Un aspetto cruciale dell'IA etica è la trasparenza nelle decisioni etiche. I sistemi devono essere in grado di spiegare non solo come hanno raggiunto una determinata decisione, ma anche quali principi etici hanno guidato tale scelta [38]. Questo è fondamentale per costruire la fiducia degli utenti e permettere un dibattito informato sulle implicazioni etiche dell'uso dell'IA in diversi contesti.

Conclusioni

L'evoluzione dell'intelligenza artificiale sta procedendo su molteplici direzioni interconnesse, ciascuna delle quali contribuisce a una comprensione più profonda non solo delle capacità tecniche dei sistemi artificiali, ma anche della natura stessa dell'intelligenza e della coscienza. L'integrazione tra neuroscienze e IA sta aprendo nuove prospettive sulla possibilità di sviluppare sistemi intelligenti e potenzialmente coscienti, sollevando questioni fondamentali sulla natura dell'esperienza soggettiva e sulla possibilità di replicarla artificialmente.

L'auto-apprendimento, la multimodalità, l'efficienza computazionale, gli agenti intelligenti e i nuovi paradigmi di ragionamento stanno convergendo verso sistemi sempre più sofisticati e autonomi. La ricerca sulla coscienza artificiale sta aggiungendo una dimensione completamente nuova a questi sviluppi, suggerendo la possibilità che i futuri sistemi di IA potrebbero non solo elaborare informazioni ma anche sperimentare forme di consapevolezza.

Queste diverse direzioni di sviluppo si influenzano e si rafforzano reciprocamente in modi complessi. Le intuizioni dalle neuroscienze informano lo sviluppo di nuove architetture di IA, mentre i progressi nell'IA forniscono nuovi strumenti per comprendere il funzionamento del cervello. La trasparenza nel ragionamento diventa ancora più cruciale quando consideriamo la possibilità di sistemi con forme di coscienza artificiale, mentre l'etica assume una dimensione ancora più profonda nella prospettiva di macchine potenzialmente senzienti.

La sfida per il futuro sarà integrare questi diversi aspetti in modo responsabile ed etico. Dovremo bilanciare il desiderio di creare sistemi sempre più avanzati con la necessità di comprendere e gestire le implicazioni della possibile emergenza di forme di coscienza artificiale. Sarà fondamentale mantenere un dialogo continuo tra scienziati, filosofi, studiosi di etica e società civile per guidare questi sviluppi in direzioni benefiche per l'umanità. Il campo dell'intelligenza artificiale sta entrando in una nuova era dove le questioni tecniche si intrecciano indissolubilmente con quelle filosofiche ed etiche. La possibilità di creare sistemi non solo intelligenti ma potenzialmente coscienti ci obbliga a ripensare fundamentalmente il rapporto tra umani e macchine, e a considerare con ancora maggiore attenzione le implicazioni etiche e sociali di questi sviluppi. Solo attraverso un approccio

olistico e responsabile potremo garantire che questa evoluzione contribuisca positivamente al progresso della società nel suo complesso.

Bibliografia

- [1] Silver, D., Schrittwieser, J., Simonyan, K. *et al.* (2017) Mastering the game of Go without human knowledge. *Nature* **550**, 354–359, Doi: 10.1038/nature24270
- [2] Ji, K., Chen, J., Gao, A., Xie, W., Wan, X., & Wang, B. (2024). LLMs Could Autonomously Learn Without External Supervision. *arXiv preprint arXiv:2406.00606*. <https://arxiv.org/abs/2406.00606>
- [3] Kontogiannis, A., Kelesis, D., Pollatos, V., Paliouras, G., & Giannakopoulos, G. (2021). Tree-based Focused Web Crawling with Reinforcement Learning. *arXiv preprint arXiv:2112.07620*. <https://arxiv.org/abs/2112.07620>
- [4] Baltrusaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423–443. Doi: 10.1109/TPAMI.2018.2798607
- [5] Durante, Z., Huang, Q., Wake, N. *et al.* (2024). Agent AI: Surveying the Horizons of Multimodal Interaction. *arXiv preprint arXiv:2401.03568*. <https://arxiv.org/abs/2401.03568>
- [6] Zhang, D., Yu, Y., Dong, J. *et al.* (2024) MM-LLMs: Recent Advances in MultiModal Large Language Models *arXiv preprint arXiv:2401.13601* <https://arxiv.org/abs/2401.13601>
- [7] Ferrari, D., Pupa, A., Signoretti, A., & Secchi, C. (2023). “Safe Multimodal Communication in Human-Robot Collaboration.” *arXiv preprint arXiv:2308.03690*. <https://arxiv.org/abs/2308.03690>
- [8] Goertzel, B., Bogdanov, V., Duncan, M. *et al.* (2023). OpenCog Hyperon: A Framework for AGI at the Human Level and Beyond. *arXiv preprint arXiv:2310.18318*. <https://arxiv.org/abs/2310.18318>
- [9] Bibi, U. *et al.*, (2024). Advances in Pruning and Quantization for Natural Language Processing. *IEEE Access*, vol. 12, pp. 139113-139128, 2024, Doi: 10.1109/ACCESS.2024.3465631
- [10] Lv, K., Yang, Y., Liu, T. *et al.* (2023). Full Parameter Fine-tuning for Large Language Models with Limited Resources *arXiv preprint arXiv:2306.09782*, <https://arxiv.org/abs/2306.09782>
- [11] X. Liu *et al.*, (2023) Distributed Intelligence in Wireless Networks. *IEEE Open Journal of the Communications Society*, vol. 4, pp. 1001-1039. Doi: 10.1109/OJCOMS.2023.3265425.
- [12] Minh, D., Wang, H.X., Li, Y.F. *et al.* Explainable artificial intelligence: a comprehensive review. *Artif Intell Rev* 55, 3503–3568 (2022). Doi: 10.1007/s10462-021-10088-y
- [13] Yang, W., Wei, Y., Wei, H. *et al.* Survey on Explainable AI: From Approaches, Limitations and Applications Aspects. *Hum-Cent Intell Syst* 3, 161–188 (2023). Doi: 10.1007/s44230-023-00038-y
- [14] Chella, A., Pipitone, A., Morin, A. & Racy, F. (2020). Developing Self-Awareness in Robots via Inner Speech. *Frontiers in Robotics and AI*, 7. Doi: 10.3389/frobt.2020.00016

- [15] Yosinski, J., Clune, J., Nguyen, A. et al. (2015). Understanding Neural Networks Through Deep Visualization. Deep Learning Workshop, 31st International Conference on Machine Learning, Lille, France.
- [16] Qian, X., Yuemaier, A., Liang, et al. (2022). Adaptive Multi-Agent Continuous Learning System, arXiv preprint arXiv:2212.07646, 2022.
<https://arxiv.org/abs/2212.07646>
- [17] K. Peng, T. Ma, X. Yu, H. Rong, Y. Qian and N. Al-Nabhan, GCMA: An Adaptive Multiagent Reinforcement Learning Framework with Group Communication for Complex and Similar Tasks Coordination, IEEE Transactions on Games, vol. 16, no. 3, pp. 670-682, Sept. 2024, doi: 10.1109/TG.2023.3346394.
- [18] Mariani, M.M., Hashemi, N., Wirtz, J. (2023). Artificial intelligence empowered conversational agents: A systematic literature review and research agenda, Journal of Business Research, 161, Doi: 10.1016/j.jbusres.2023.113838.
- [19] Azevedo, T., de Araújo, P. J.M., Rossetti R.J.F. & Rocha, A.P.C. (2014). JADE, TraSMAPI and SUMO: A tool-chain for simulating traffic light control, Proceedings of the 8th International Workshop on Agents in Traffic and Transportation, ATT'14, held at the Thirteenth International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS'14, 8-15.
- [20] Edmonds, M., Ma, X., Qi, S. et al. (2020). Theory-based Causal Transfer: Integrating Instance-level Induction and Abstract-level Structure Learning, Proc. AAAI-20, pp. 1283 – 1291.
- [21] Marcus, G., & Davis, E. (2019). "Rebooting AI: Building Artificial Intelligence We Can Trust." Pantheon.
- [22] Garcez, A.d., Lamb, L.C. Neurosymbolic AI: the 3rd wave. Artif Intell Rev 56, 12387–12406 (2023). <https://doi.org/10.1007/s10462-023-10448-w>
- [23] Pearl, J., & Mackenzie, D. (2018). "The Book of Why: The New Science of Cause and Effect in AI Systems." Basic Books.
- [24] Hofstadter, D., & Sander, E. (2013). "Surfaces and Essences: Analogy as the Fuel and Fire of Thinking in AI." Basic Books.
- [25] Webb, T., Holyoak, K.J. & Lu, H. Emergent analogical reasoning in large language models. Nat Hum Behav 7, 1526–1541 (2023). [Doi: 10.1038/s41562-023-01659-w](https://doi.org/10.1038/s41562-023-01659-w)
- [26] Dehaene, S., Lau, H., Kouider, S. (2021). What Is Consciousness, and Could Machines Have It?. In: von Braun, J., S. Archer, M., Reichberg, G.M., Sánchez Sorondo, M. (eds) Robotics, AI, and Humanity. Springer, Cham. https://doi.org/10.1007/978-3-030-54173-6_4
- [27] Chella, A. (2022). Robots and Machine Consciousness, in A. Cangelosi, M. Asada (eds), Cognitive Robotics, Cap. 23, MIT Press, Doi: 10.7551/mitpress/13780.003.0029.
- [28] Butlin, P., Long, R. et al (2023). Consciousness in Artificial Intelligence: Insights from the Science of Consciousness, arXiv preprint arXiv:2308.08708,
<https://arxiv.org/abs/2308.08708>

- [29] Seth, A. K. (2015). The Cybernetic Bayesian Brain - From Interoceptive Inference to Sensorimotor Contingencies. In T. Metzinger & J. M. Windt (Eds). Open MIND: 35(T). Frankfurt am Main: MIND Group. Doi: 10.15502/9783958570108
- [30] Kriegeskorte, N., Douglas, P.K. Cognitive computational neuroscience. *Nat Neurosci* 21, 1148–1160 (2018). Doi: 10.1038/s41593-018-0210-5
- [31] Chalmers, D. (2023). Could a Large Language Model be Conscious? Boston Review, August 9, arXiv:2303.07103, <https://arxiv.org/abs/2303.07103>
- [32] Jones, D., Lowe, V. et al (2022). A computational model of neurodegeneration in Alzheimer's disease. *Nature communications*, 13(1), 1643. Doi: 10.1038/s41467-022-29047-4
- [33] Dubber, M.D., Pasquale, F. & Das, S. (Eds.) (2020). *The Oxford Handbook of Ethics of AI*. Oxford University Press, Oxford, UK.
- [34] Evans, K., de Moura, N., Chauvier, S., Chatila, R. (2023). Automated Driving Without Ethics: Meaning, Design and Real-World Implementation. In: Fossa, F., Cheli, F. (eds) *Connected and Automated Vehicles: Integrating Engineering and Ethics*. Studies in Applied Philosophy, Epistemology and Rational Ethics, vol 67. Springer, Cham. Doi: 10.1007/978-3-031-39991-6_7
- [35] Poszler, F., Portmann, E. & Lütge, C. Formalizing ethical principles within AI systems: experts' opinions on why (not) and how to do it. *AI Ethics* (2024). Doi: 10.1007/s43681-024-00425-6
- [36] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). Doi: 10.1177/2053951716679679
- [37] Prabhakaran, V., Qadri, R. & Hutchinson, B. (2022). Cultural Incongruencies in Artificial Intelligence. First Workshop on Cultures in AI/AI in Culture (non-archival), NeurIPS 2022. https://ai-cultures.github.io/papers/Cultural_Compencies_in_Artificial_Intelligence_NeurIPS_2022_Culture_AI_Workshop.pdf
- [38] Jobin, A., Ienca, M. & Vayena, F. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence* 1, 9, 389–399. Doi: 10.1038/s42256-019-0088-2

I quaderni di

Agenda  Digitale ^{eu}

NETWORK **DIGITAL** 360

Il 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
TEL. 02 92852785
MAIL: MARKETING@DIGITAL4.BIZ
© ICT & Strategy