

*I quaderni di*  
Agenda  **Digitale** <sup>eu</sup>

SETTEMBRE-DICEMBRE  
2023

n. 0015

**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](mailto: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

**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

|  |            |
|--|------------|
| <b>AI e produttività: dubbi, problemi e potenzialità per il futuro</b> .....   | <b>8</b>   |
| Di <b>Mauro Lombardi</b> , BABEL - Blockchain and Artificial intelligence for Business, Economics and Law – UNIFI  |            |
| <b>Quale digitale per l'età prescolare: cosa considerare</b> .....   | <b>24</b>  |
| Di <b>Mario Morcellini</b> , professore ordinario in Sociologia dei Processi Culturali e Comunicativi, Roma1   |            |
| <b>Come le allucinazioni di ChatGPT ci aiutano a capire la nostra idea di conoscenza</b> .....   | <b>38</b>  |
| Di <b>Valeria Martino</b> , post-doc presso l'Università di Torino   |            |
| <b>Studi storici con l'IA: verso le Digital Humanities</b> .....   | <b>43</b>  |
| Di <b>Matteo Valleriani</b> , Research Group Leader in Dept. I, Honorary Professor at the Technische Universität Berlin  |            |
| <b>La lunga estate calda della coscienza artificiale: nuovi studi</b> .....  | <b>57</b>  |
| Di <b>Antonio Chella</b> , Laboratorio di Robotica, dipartimento di Ingegneria Università degli Studi di Palermo   |            |
| <b>L'IA nell'apprendimento: una storia genetica</b> .....  | <b>62</b>  |
| Di <b>Paolo Ferri</b> , Professore Ordinario di Tecnologie della formazione, Università degli Studi Milano-Bicocca   |            |
| <b>Genomica, una visione non antropocentrica: i grandi problemi da risolvere</b> .....   | <b>71</b>  |
| Di <b>Mirta Michilli</b> , direttrice generale della Fondazione Mondo Digitale e <b>Alfonso Molina</b> , personal chair in Technology Strategy all'Università di Edimburgo e direttore scientifico della Fondazione Mondo Digitale   |            |
| <b>Scrivere poesia con GPT4: la sfida di scolpire versi con l'IA</b> .....   | <b>79</b>  |
| Di <b>Nick Young</b> , <b>Marcello Frixione</b> e <b>Enrico Terrone</b> - Università di Genova   |            |
| <b>Digital Pathology: i nuovi scenari dell'attività diagnostica</b> .....  | <b>86</b>  |
| Di <b>Elena Bottinelli</b> , Head of Digital Transition and Transformation del Gruppo San Donato   |            |
| <b>IA e salute mentale, un rapporto complesso: usi pratici, strumenti, rischi</b> .....  | <b>93</b>  |
| Di <b>Luca Bernardelli</b> , Psicologo, Autore del libro "Guida Psicologica alla Rivoluzione Digitale", CEO di BECOME. Augmented Life, Cofounder di BOWMAN - Data Matter e <b>Clelia Malighetti</b> , Psicologa, Psicoterapeuta, PhD |            |
| <b>Tecnologie persuasive: così influenzano le nostre decisioni</b> .....   | <b>109</b> |
| di <b>Chiara Cilardo</b> , Psicologa psicoterapeuta, esperta in psicologia digitale  |            |

# AI e produttività: dubbi, problemi e potenzialità per il futuro

Una serie di dilemmi e di sfide si profilano per l'umanità in un'era dominata dalla GenAI e dai suoi ulteriori, imminenti sviluppi. Esserne consapevoli a livello individuale e collettivo è solo il primo, essenziale, passo per elaborare strategie all'altezza delle prove da affrontare

Di **Mauro Lombardi**, BABEL - Blockchain and Artificial intelligence for Business, Economics and Law – UNIFI

Uno dei temi più rilevanti, al centro delle analisi e del dibattito in merito alla dinamica tecnico-produttiva internazionale nei decenni seguenti alla Seconda guerra mondiale, è quello del peculiare rallentamento dei tassi di crescita e di produttività in quasi tutti i Paesi industrializzati, dopo il trentennio terminato all'inizio degli anni '70.

In questo contributo cerchiamo di mettere in luce aspetti e fattori che possono aver agito nel determinare il fenomeno, che alcuni importanti studiosi ritengono addirittura effetto di "errori di misurazione" (*mismeasurement*). A tale fine sono descritti innanzitutto elementi di uno scenario di lungo periodo, per poi approfondire questioni connesse all'evoluzione tecnico-scientifica e all'impiego dell'*Artificial Intelligence* (d'ora in poi AI) nei processi socioeconomici.

## Elementi di scenario

La riflessione degli studiosi è innescata dalla celebre affermazione del Premio Nobel per l'Economia Robert Solow ("You can see the computer age everywhere but in the productivity statistics", 1987), e dal contributo di Gordon, che descrive con indicatori standard il profilo declinante di crescita e produttività negli USA nella seconda metà del secolo scorso (Tab.1).

Tab. 1 Tasso di crescita annuale del Pil in termini reali e componenti dell'output

| Period                            | Real GDP growth | Contribution of output identity components |                                 |                 |       |                        |                         |                               |
|-----------------------------------|-----------------|--|---------------------------------|-----------------|-------|------------------------|-------------------------|-------------------------------|
|                                   |                 | Output per hour <sup>a</sup>               | Hours per employee <sup>a</sup> | Employment rate | LFPR  | Working-age population | Mix effect <sup>b</sup> | Employment ratio <sup>c</sup> |
| <i>Between benchmark quarters</i> |                 |  |                                 |                 |       |                        |                         |                               |
| 1954:4–1963:3                     | 3.51            | 2.72                                       | -0.16                           | -0.02           | 0.03  | 1.43                   | -0.27                   | -0.22                         |
| 1963:3–1972:2                     | 4.07            | 2.54                                       | -0.44                           | -0.02           | 0.33  | 1.80                   | -0.62                   | 0.49                          |
| 1972:2–1978:3                     | 3.43            | 1.72                                       | -0.60                           | -0.06           | 0.73  | 1.94                   | -0.38                   | 0.08                          |
| 1978:3–1987:3                     | 2.69            | 1.20                                       | -0.26                           | 0.00            | 0.41  | 1.39                   | 0.06                    | -0.12                         |
| 1987:3–1994:4                     | 2.71            | 1.48                                       | -0.01                           | 0.06            | 0.23  | 1.08                   | -0.42                   | 0.28                          |
| 1994:4–2001:2                     | 3.21            | 2.11                                       | -0.31                           | 0.19            | 0.02  | 1.24                   | -0.61                   | 0.57                          |
| <i>Last two episodes</i>          |                 |  |                                 |                 |       |                        |                         |                               |
| 1990:3–1992:3                     | 1.32            | 1.95                                       | 0.01                            | -1.04           | 0.10  | 1.07                   | -0.26                   | -0.50                         |
| 2001:2–2003:2                     | 2.21            | 4.52                                       | -0.46                           | -0.92           | -0.27 | 1.20                   | -1.03                   | -0.83                         |

Fonte: Gordon (2003, Table 1).

La tesi è stata ribadita e motivata da Gordon in molti contributi, dove sono state anche proposte spiegazioni per il trend rilevato, come risulta anche dalla Tabella 2

**Tab. 2**

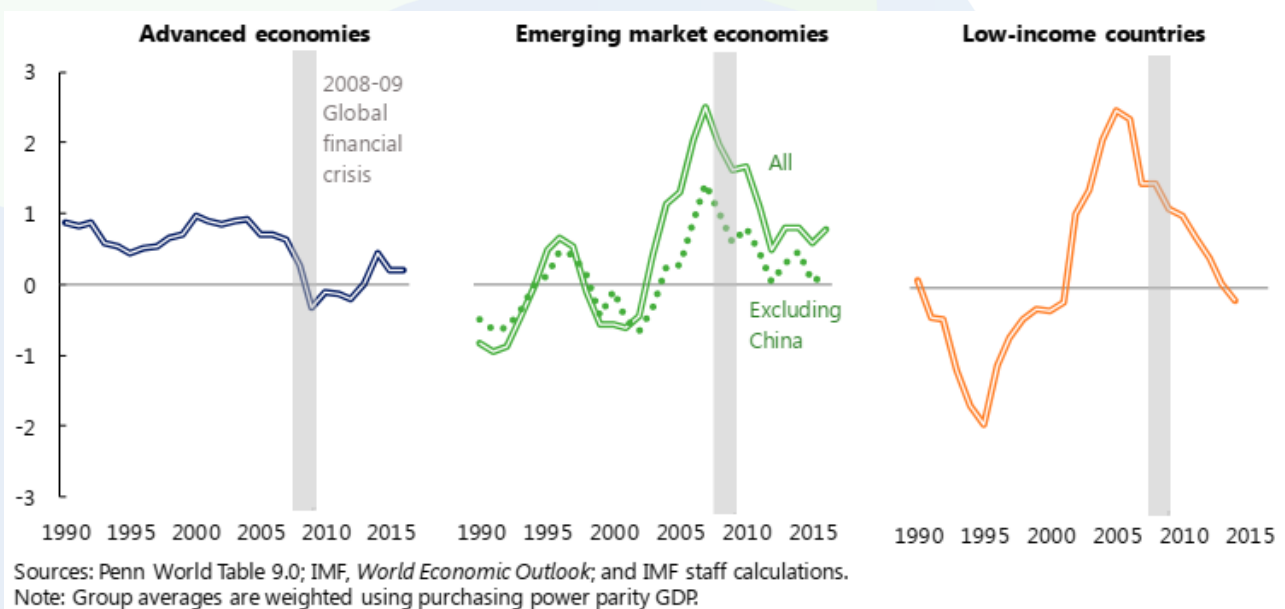
| Table 1. Anatomy of the Growth Slowdown<br>in the Total U.S. Economy (percent change at annual rate) |           |           |           |                   |                   |
|--|-----------|-----------|-----------|-------------------|-------------------|
|  | 1920-1970 | 1970-2006 | 2006-2016 | Change<br>(1)-(2) | Change<br>(2)-(3) |
|  | (1)       | (2)       | (3)       | (4)               | (5)               |
| Output   | 3.71      | 3.17      | 1.35      | -0.54             | -1.82             |
| Output per Hour  | 2.82      | 1.75      | 0.97      | -1.07             | -0.78             |
| Hours of Work  | 0.89      | 1.42      | 0.38      | 0.53              | -1.04             |
| Hours per Person   | -0.36     | 0.00      | -0.62     | 0.36              | -0.62             |
| Population aged 16+  | 1.25      | 1.42      | 1.00      | 0.17              | -0.42             |
| Addendum: Output per Person  | 2.46      | 1.75      | 0.35      | -0.71             | -1.40             |

Fonte: Gordon (2018)

L'attenzione degli economisti si concentra sul problema di come spiegare il trend indicato, tenendo anche presente che negli anni 1995-2004 c'è stata una lieve ripresa (+2,8% di incremento annuale del Pil), seguita da un più modesto valore (1,3% annuale nel periodo 2005-2015) (Syverson, 2017). Nel frattempo, c'è stata la cosiddetta *Grande Recessione* del 2007-2008, i cui effetti sono stati più prolungati di crisi precedenti, come quella del 2001, senza che la ripresa successiva abbia generato i ritmi espansivi dei decenni precedenti.

Il trend in realtà accomuna le economie avanzate, come si può vedere nella Fig 1, da cui emergono anche le differenze di comportamento che caratterizzano le Economie di mercato emergenti e i Paesi a basso reddito

**Fig. 1** Crescita della produttività totale dei fattori (tasso medio di crescita su 5 anni)<sup>1</sup>



<sup>1</sup> La "produttività totale dei fattori" (*Total Factor Productivity*) è l'indicatore adottato dall'economia *mainstream* per stimare l'efficienza di un sistema economico sulla base della differenza tra tasso di crescita del prodotto e i tassi di aumento negli impieghi degli input di capitale e lavoro. Per quanto riguarda la classificazione delle Economie, la fonte il Fondo Monetario Internazionale (*World Economic Outlook*),

<https://www.imf.org/en/Publications/WEO/weo-database/2023/April/groups-and-aggregates>.

Fonte: Adler et al., 2017, Fig. 2)

All'interno di tale quadro generale l'evoluzione delle Economie Europee presenta analoghi andamenti tendenziali, con valori accentuati verso il basso, come si evince dalla Tabella 3 Tab. 3

| Average annual real value added and hourly labour productivity growth rate (1) |       |         |        |       |           |
|--|-------|---------|--------|-------|-----------|
| (percentage points)  |       |         |        |       |           |
|  | Italy | Germany | France | Spain | Euro area |
| <b>(a) value added</b>   |       |         |        |       |           |
| 2000-2007  | 1.16  | 1.57    | 1.89   | 3.35  | 2.02      |
| 2007-2014  | -1.14 | 0.84    | 0.58   | -0.83 | 0.06      |
| 2014-2019  | 1.04  | 1.59    | 1.55   | 2.68  | 1.89      |
| 2019-2022  | 0.55  | 0.15    | 0.29   | -0.47 | 0.83      |
| 2000-2019  | 0.28  | 1.31    | 1.32   | 1.63  | 1.27      |
| 2000-2022  | 0.32  | 1.15    | 1.18   | 1.35  | 1.21      |
| <b>(b) labour productivity</b>   |       |         |        |       |           |
| 2000-2007  | 0.06  | 1.58    | 1.33   | 0.39  | 1.10      |
| 2007-2014  | 0.22  | 0.54    | 0.58   | 1.59  | 0.77      |
| 2014-2019  | 0.20  | 0.83    | 0.72   | 0.38  | 0.66      |
| 2019-2022  | 0.81  | 0.70    | -1.02  | -0.09 | 0.70      |
| 2000-2019  | 0.16  | 1.00    | 0.89   | 0.83  | 0.86      |
| 2000-2022  | 0.25  | 0.96    | 0.63   | 0.71  | 0.84      |

Source: own calculations on AMECO data. (1) Real gross value added excluding financial intermediation services indirectly measured (FISIM), total economy.

Fonte: Greco (2023, Table 1)

Emergono con chiarezza tre aspetti sui quali riflettere: 1) il sentiero di bassa crescita, che connota le maggiori economie europee nei primi venti anni del XXI secolo, dopo una lunga fase con caratteristiche simili, anche se più marcate nel decennio 1970. 2) In questo scenario l'Italia si caratterizza per i tassi di crescita sempre più bassi di tutti gli altri Paesi, anche se il differenziale rispetto a questi ultimi è minore negli anni 2014-2019, ma ciò è dovuto al leggero aumento registrato in Italia e al decremento rilevato negli altri (Greco, 2023). Occorre peraltro rilevare che ovunque vi sono state significative differenze di produttività tra i vari settori economico-produttivi, con l'industria che ha mostrato performance migliori dei servizi, particolarmente in Italia, dove la crescita della produttività nei servizi durante gli anni 2000-2019 è stata praticamente nulla (Greco, 2023: 8). Va anche osservato che componenti significative dell'industria italiana, in termini di specializzazione e quote di export<sup>2</sup>, si sono distinte per segnali positivi. 3) La crisi finanziaria del 2007-2008 e la "Grande Recessione", che ne è seguita, hanno profondamente influenzato gli investimenti, aumentati in tutti i Paesi solo in beni intangibili (ICT, tecnologie dell'informazione, servizi ad alta intensità di conoscenza). Il doppio shock della Grande recessione, cioè la sequenza rapida di due fasi (ripresa e nuova contrazione, *double dip*), ha inciso sostanzialmente sulla propensione a investire per una serie di motivi: rarefazione delle risorse finanziarie disponibili nelle banche, riduzione delle possibilità di accesso al credito, deficit di capitali nelle imprese con bilanci indeboliti dalla crisi, incertezza economica e politica (Adler et al., 2017: 14-18).

Per quanto riguarda l'Italia, nel periodo 2000-2022 il prodotto reale è cresciuto del +0,3%, mentre in Germania, Francia, Spagna e nella media europea l'incremento si è collocato tra +1,2% e +1-3%, con una crescita cumulata sull'intero arco di tempo del +7%, 4 volte meno che negli altri Paesi (Greco, 2023, 7-10).

Un ulteriore dato per dare una rappresentazione significativa dello scenario complessivo concerne la produttività del lavoro, cresciuta in Italia al tasso annuale di +0,16% negli 2000-2019, mentre in Germania l'incremento medio annuale è stato dell'1% e la media europea +0,86%. Il differenziale

<sup>2</sup> Macchine e attrezzature, veicoli a motore, mobili.

cumulato negli stessi anni indica una crescita complessiva del +3% in Italia, +19% in Germania, +16% nell'area Euro (Greco, 2023, 11-12 e Tabella 2, qui non riportata).

Volendo sintetizzare il quadro evolutivo generale, appare fondata la tesi che siamo in presenza di economie con elevata vulnerabilità alle crisi e debolezze intrinseche, anche se vi è un elemento positivo: la propensione a orientare gli investimenti in beni intangibili, segno a nostro avviso inequivocabile di una embrionale consapevolezza della profonda trasformazione in atto nei sistemi socioeconomici.

Per desiderio di completezza è necessario chiarire che alcuni autorevoli studiosi (Brynjolfsson e McAfee, 2014; Mokyr, 2014; Byrne, Oliner e Sichel, 2015, per citarne solo alcuni) hanno argomentato la tesi che la traiettoria –generalizzata- di rallentamento della crescita e della produttività è il risultato di problemi di misurazione statistica, in quanto gli indicatori standard impiegati non riescono a catturare effetti rilevanti delle ondate innovative dei decenni analizzati, quali: diffusione di tecnologie dell'informazione, Internet, smartphone, social networks, varie tipologie di media scaricabili, ecc.. La riduzione dei prezzi di una serie innumerevole di beni e servizi immateriali, che hanno consentito a consumatori di ottenere benefici con nuovi strumenti e modalità di consumo, e lo sviluppo di nuove tecnologie (ad es. nuovi materiali, ingegneria genetica) generano risultati che non sono catturati dai tradizionali indicatori statistici. Questa ipotesi, definita “*mismeasurement hypothesis*”, è stata sottoposta a vaglio critico da Syverson (2017), che ha evidenziato due aspetti che la indeboliscono: 1) il rallentamento di crescita e produttività è un fenomeno che accomuna tutti i Paesi, anche quelli dove le tecnologie dell'informazione hanno avuto una scarsa diffusione. 2) I sostenitori dell'ipotesi in questione enfatizzano la “discrepanza statistica” tra Reddito Interno Lordo (gross domestic income (GDI) e Prodotto Interno Lordo (gross domestic product, GDP), con il primo che supera sistematicamente il secondo. Tale gap riflette “lavoratori remunerati per produrre beni liberi ... ceduti gratuitamente oppure a prezzi molto più bassi del loro valore, in questo nodo riducendo la spesa su tali prodotti e di conseguenza la stima del prodotto interno lordo” (Syverson, 2017: 181-182, trad. nostra). Syverson effettua un'analisi puntuale dell'andamento e della composizione di GDI e GDP, che lo porta a concludere: “La crescita del GDI in relazione al GDP sembra riflettere la crescita del reddito da capitale più che del reddito da lavoro” (trad. nostra), quindi l'“anormalità” della misura dei redditi da lavoro nasconde invece la crescita della profittabilità delle attività di business, il che è incompatibile con la tesi dei lavoratori pagati per acquistare beni a prezzi inferiori al loro valore”. In definitiva, Syverson mostra la debolezza della “*mismeasurement hypothesis*” e indica il fenomeno rilevato da tutte le ricerche in materia: la crescita delle disuguaglianze in tutti Paesi, sia pure con gradi differenti per es. nei Paesi anglosassoni e in quelli del Nord Europa.

Possiamo a questo punto approfondire alcuni aspetti fondamentali, che non trovano adeguata rappresentazione nella descrizione statistica qui riproposta.

## **Un'era di profonde trasformazioni, diversa dalle altre nella storia dell'umanità**

Nell'arco di tempo considerato finora, specie negli ultimi tre decenni, si sono verificati enormi avanzamenti tecnico-scientifici nel campo dell'hardware e del software, che hanno prodotto vere e proprie discontinuità nell'evoluzione dei sistemi di algoritmi su cui si basa la AI. Sofferiamoci solo sui primi venti anni del XXI secolo, che si apre all'insegna del Deep Learning, con nuove architetture di reti neurali (per es. le “reti neurali ricorsive”) e nuovi algoritmi (per es. algoritmi non supervisionati), che sono implementati grazie all'impiego delle *graphics processing units* (GPUs) (unità di elaborazione grafica). Le GPUs consentono di ottenere performance fino ad allora irraggiungibili nel riconoscimento delle immagini, nell'analisi del linguaggio umano, nel riconoscimento del parlato, data la possibilità di elaborare modelli computazionali più potenti,

allenati su dataset sempre più ampi. La prima decade del nuovo secolo è stata giustamente definita la “*Golden Decade of Deep Learning*” (Dean, 2022), mentre *feedback loop* tra crescenti flussi di informazioni globali ed esigenze di aumento della potenza computazionale necessaria hanno spinto verso la creazione di chip sempre più potenti, quali le *Tensor Processing Units* (TPUs), un acceleratore dei processi di elaborazione dei circuiti integrati, il cui impiego ha consentito enormi progressi nel riconoscimento del parlato, nella traduzione e nei sistemi di classificazione delle immagini. I nuovi sistemi hardware-software hanno costituito, per così dire, il sostrato fisico-digitale per lo sviluppo di *open source frameworks*, che hanno reso molto più agevole la creazione di modelli computazionali basati su Deep Learning ad opera di molti gruppi di ricercatori indipendenti. Si è così avuta la proliferazione di librerie open-source di Machine Learning, come nel caso del celebre Tensorflow, sviluppato da Google nel 2015 per consentire l’interoperabilità di processi computazionali di varia matrice su una gamma diversificata dispositivi: personal computer, smartphone, data centers, web browser e così via. In tale contesto estremamente dinamico e ad elevata intensità di produzione conoscitiva<sup>3</sup>, si moltiplicano i campi disciplinari in cui sono applicati i nuovi sistemi di AI, i quali evolvono in modo esponenziale, fino all’elaborazione di una nuova architettura di reti neurali, la cosiddetta *Transformer*, realizzata da computer scientist di Google nel 2017 (Vaswani et al., 2017) e destinata a generare una sorta di “salto quantico”, cioè una profonda trasformazione del sistema in base al quale gli umani rappresentano la realtà e delle modalità di interazione uomo macchina. *Transformer* è infatti alla base dei *Large Language Models* (LLMs), sistemi algoritmici che hanno la capacità di “processare il linguaggio umano”. Nasce così, e si diffonde immediatamente, la *Generative Artificial Intelligence* (d’ora in poi GenAI), grazie al lancio di ChatGPT4 nel Novembre del 2022, che in pochi mesi raggiunge 100 milioni di utenti<sup>4</sup>. L’architettura *Transformer* consente di catturare le interazioni tra varie componenti di qualsiasi input, quindi di apprendere come è organizzato, sia esso un testo scritto, un brano musicale, un’immagine, oppure una proteina (Vaswani et al., 2017). I LLMs sono “modelli linguistici”, basati sul meccanismo della *self-attention*, o semplicemente *attention*, che individua le componenti di un linguaggio (scritto, visivo, parlato, pittorico, ecc.) e assegna ad esse rilevanza. È in grado di farlo per sequenze di input di varia natura, perché identifica regolarità presenti nel training set e quindi ricostruisce parte della struttura contestuale. *Transformer* analizza il linguaggio identificando la struttura grammaticale e le regole sintattiche attraverso “correlazioni statistiche tra frammenti linguistici, composti sulla base di stime probabilistiche”<sup>5</sup>. Questi collage di frammenti sono trasformati in moltiplicazioni di matrici contenenti righe e colonne di pixel in input-output, tanto che tutto risulta essere nient’altro che “*a matrix multiplication from computer graphics to training neural networks*” (Demis Hassabis, CEO di OpenMind, 2022). I collage sono basati su meccanismi che catturano le regolarità linguistiche e sono quindi capaci di ricostruire strutture linguistiche basate su sistemi statistici. Essi hanno, però, solo abilità linguistiche formali –acquisizione di regole grammaticali e sintattiche, morfologia delle frasi (Mahowald et al. 2023) e sono prive di abilità linguistiche funzionali, cioè quelle necessarie per “comprendere e usare il linguaggio nel mondo”, che richiede “*formal reasoning, world knowledge, situation modeling, and social-cognitive abilities*” (Mahowald et al., 2022). Le abilità linguistiche formali consentono di generare rappresentazioni molto efficaci e suggestive del mondo in cui agiscono gli umani, perché basate sulle stesse regole basilari dei nostri metodi di comunicazione, ma esse non consentono di comprenderne il significato, proprio perché manca la capacità di interagire con il contesto vitale da cui derivano le espressioni prodotte nei vari tipi di linguaggio.

<sup>3</sup> Per un’analisi più dettagliata e ricca si rinvia si veda Dean (2022).

<sup>4</sup> Stime non ufficiali indicano ad oggi 180 milioni.

<sup>5</sup> Nel gergo dei computer scientist, si tratta di “*next word predicting*” (Bubeck et al., Microsoft Research, 2023)

Siamo, in definitiva, di fronte a capacità percettive multimodali, che non equivalgono al sistema percettivo multimodale umano, le cui proprietà di rappresentazione attiva del mondo “reale” sono incomparabili con quelle dei sistemi artificiali (Seth, 2023).

I LLMs, dotati di una potenza computazionale in continua espansione, esercitata su database arricchiti da feedback degli utenti per cui sembrano non avere limiti nell'estendere la loro capacità di affrontare problemi e attrarre domanda di beni e servizi. Essi sono di fatto *generative chatbots*<sup>6</sup> con un enorme potenziale generatore di: 1) contenuti attraverso la combinazione di video, musica, immagini. 2) Testi costruiti sia secondo lo stile di scrittori, artisti e progettisti, sia mediante tentativi di comporre modelli stilistici in base a nuovi orientamenti artistici, eventualmente “catturati” con l'analisi di gusti e preferenze della popolazione.

## La mancata spinta propulsiva dell'AI

Dopo l'ondata innovativa della “Golden Decade” siamo oggi in un mondo caratterizzato dalla pervasività della GenAI, che è al centro della stampa internazionale (specializzata e non) e dei social di tutto il mondo, attratti dall'incessante emergere di nuovi protagonisti (i *chatbots*) alla conquista di utenti. Gli sviluppi descritti nel precedente paragrafo evidentemente costituiscono un'accelerazione evolutiva della AI e la loro portata è tale da indurre logicamente a chiedersi quali siano gli effetti –attuali e potenziali- dell'intensificarsi del processo d'innovazione, caratterizzato da quello che abbiamo prima chiamato “salto quantico”. Prendiamo allora in esame contributi importanti, elaborati da studiosi e centri di ricerca internazionali, scelti all'interno di una gamma molto ampia e di grande interesse, ma tutti diretti a rispondere ad un interrogativo di fondo: quali sono gli effetti sul lavoro e l'occupazione dei processi innovativi degli ultimi 30 anni, alla luce dell'accelerazione in atto?

Aghion et al. (2022) distinguono due tipi di effetti dell'automazione sulla domanda di lavoro: diretto e indiretto, a cui corrispondono due differenti scenari, rispettivamente negativo e positivo.

Lo **scenario negativo** implica questa sequenza: 1) la sostituzione di lavoratori con macchine nell'esecuzione di determinati compiti e funzioni riduce la domanda di lavoro e quindi i salari. 2)

La diminuzione delle retribuzioni può agire da fattore compensativo, in quanto può indurre la creazione di altre attività, con la conseguenza di aumentare la domanda di lavoro e di controbilanciare il precedente calo. A seconda delle assunzioni di base, i modelli elaborati in questa prospettiva pervengono a conclusioni divergenti in merito al processo di compensazione su salari e occupazione. Il recupero delle perdite di occupazione e salari può essere infatti parziale, incerto, ambiguo.

Lo **scenario positivo** si basa sull'ipotesi che l'automazione possa generare un aumento dell'occupazione sia a livello di impresa che a livello aggregato. La sequenza ipotizzata è così definita: 1) le imprese che automatizzano sono più produttive. 2) I loro output sono quindi più competitivi sia nei prezzi che nella qualità, il che permette di conquistare maggiori quote di mercato interno e internazionale, a cui segue una maggiore domanda di lavoro.

Tutto ciò non esclude, però, la possibilità che l'occupazione aggregata diminuisca in conseguenza dell'espulsione dal mercato delle imprese che non automatizzano, oppure lo fanno in modo inadeguato.

La ricerca empirica, svolta dagli stessi autori (Aghion et al., 2019, 2020) in riferimento alla Francia, porta a risultati che confermerebbero lo scenario positivo.

---

<sup>6</sup> Siamo in presenza di sistemi generativi (*Generative systems*), dove la “*Generativity is a system's capacity to produce unanticipated change through unfiltered contributions from broad and varied audiences.*” (Zittrain, 2006).

È interessante rilevare come la rassegna di una serie di studi empirici sugli effetti dell'automazione, analizzata sulla base sia dell'impiego di robot che dell'attività di brevettazione, portino a conclusioni contrastanti: effetti positivi nel caso dei robot, risultati controversi nel caso della brevettazione (Aghion, 2022: 11-15).

Bisogna poi tenere presente che la maggior parte di un'altra serie cospicua di studi empirici dei processi di automazione a livello di impresa (avvenuti in Francia, Olanda, Gran Bretagna, Spagna) individua la seguente catena di effetti: a maggiori investimenti in AI segue l'aumento della produttività, quindi una più rapida crescita delle vendite a livello di impresa e di industria, infine una sostanziale invarianza della quota di ricchezza destinata ai salari.

Emergono, inoltre, due aspetti di notevole interesse: 1) l'automazione non sembra influire sulla produttività nel breve periodo, probabilmente perché vi sono difficoltà all'interno delle aziende nel recepire le componenti innovative della AI. 2) Quanto avviene a livello delle imprese non consente di dedurre gli effetti complessivi, in quanto una serie di processi e fattori interagiscono rendendo arduo il compito di prevedere l'esito aggregato: la dinamica competitiva influenzata dai differenti gradi di ricezione delle innovazioni; la selezione di mercato in base alla velocità e al livello di innovatività dei prodotti, unito a mutamenti di prezzo; le differenze di contesto, riferite ai vari Paesi, e alla intensità con cui si sviluppa il commercio internazionale.

La conclusione, tratta da Aghion et al. (2022), è che “*automation is not in itself enemy of employment*” e la relazione tra automazione ed occupazione dipende da una molteplicità variabile di elementi, che interagiscono in modo imprevedibile nelle forme specifiche, anche se possono essere definite traiettorie possibili, da individuare sulla base del confronto tra le caratteristiche dei vari contesti e le sfide poste dalle innovazioni odierne.

Una delle direttrici di ricerca dovrebbe essere quella di individuare i compiti (*task*) sui quali gli sviluppi dell'IA incidono e le modalità con cui sono trasformati in prospettiva via via che sono adottati sistemi di AI, il tutto rapportato alle caratteristiche dei sistemi socio-economici. È ciò che hanno fatto Tyson e Zysman (2022), i quali mettono in rilievo il ruolo svolto dalle “*differences in policies, institutions, and societal norms of fairness.*” (ivi: 258), evidente nelle differenti modalità reattive verso i processi di globalizzazione e automazione assunte dagli Usa e dai Paesi del Nord Europa, oltre che dal Canada. Questa affermazione è rilevante se si pensa a tre effetti possibili dei processi automazione: 1) *Displacement effect*, cioè la diminuzione della domanda di lavoro in seguito all'acquisizione di abilità che permettono a sistemi di AI di svolgere compiti e mansioni eseguite da umani. 2) *Productivity effect*, ovvero l'aumento della domanda di lavoro per i compiti non automatizzati. 3) *Reinstatement effect*, che significa la creazione di nuovi compiti lavorativi, connessi alle innovazioni introdotte.

I tre effetti hanno assunto forme differenti negli Usa e in Germania. In quest'ultima, l'adozione di robot dal 1994 al 2014 non ha causato effetti apprezzabili sulla domanda di lavoro e sul reinserimento lavorativo, in forza di sistemi di protezione dei lavoratori e di politiche attive del lavoro, come viene argomentato in Dauth et al. (2017, 2018) e in Graetz e Michaels (2018)<sup>7</sup>.

Tyson e Zysman (2022) mettono anche in risalto come negli ultimi trenta anni l'automazione, chiamata spesso “*routine based technological change*” (RBTC), ha il più delle volte sostituito gli umani in attività di routine, inizialmente solo fisiche e poi progressivamente anche di tipo cognitivo, come quelle di supporto alla clientela e al lavoro di ufficio, più in generale quelle *data intensive*.

Questa traiettoria ha innescato l'inizio di un trend diffuso in tutte le economie, ovvero la polarizzazione occupazionale e salariale, quasi una sorta di “**piramide segmentata**” a seconda dei

---

<sup>7</sup> Una rassegna degli studi empirici concernenti gli effetti su 172.114 lavoratori dei processi automazione, visti attraverso l'impiego di robot in vari Paesi europei e in particolare in Germania, è contenuta in Lombardi (2022, <https://www.agendadigitale.eu/cultura-digitale/cose-il-lavoro-il-futuro-le-tecnologie-e-lignoto-da-esplorare/>). Per un'analisi sistematica dell'introduzione di nuove tecnologie, della dinamica occupazionale e delle politiche adottate si veda Tyson (2022).

task e delle attività via via interessate dalle applicazioni degli sviluppi della AI, fino ad arrivare alla GenAI, sulla quale esiste già una consistente di studi.

Prima di sviluppare questioni relative alla GenAI, è opportuno sintetizzare brevemente i punti finora emersi: 1) un trentennio di automazione, arricchita da “generazioni” successive di AI (reti neurali artificiali nei primi anni '90, Deep Learning e poi Transformer), non sembra aver innescato un'accelerazione di crescita e produttività. Esiste, pertanto, un divario tra aspettative degli studiosi e i processi reali rappresentati dalle statistiche standard. 2) Emerge con forza un dato incontrovertibile: la dinamica tecnologica e scientifica non può di per sé essere meccanismo propulsore, in assenza di fattori complementari di contesto. 3) Gli effetti del progresso tecnologico sono sempre molto diversificati e quindi la comprensione della molteplicità di elementi che influiscono –in modo positivo o negativo- sul suo dispiegarsi è essenziale per progettare appropriati interventi “maieutici”.

Sulla base di queste proposizioni affrontiamo il tema delle conseguenze -effettive e potenziali- della diffusione della GenAI.

## La GenAI e il mercato del lavoro

I modelli linguistici costituiscono una forte discontinuità nel campo delle *General Purpose Technologies* (tecnologie di portata generale), definite originariamente come “*characterized by the potential for pervasive use in a wide range of sectors and by their technological dynamism*” (Bresnahan e Trajtenberg, 1996). Essi sono sistemi “*general purpose*” di AI, nel senso che possono “*have many intended and unintended uses, and can be applied to many different tasks in various fields*” (Future of Life Institute, 2022), ovvero sistemi che –grazie a enormi quantità di memoria e dati di allenamento, uniti a crescente potenza computazionale- sono in grado di processi trasversali di apprendimento in domini di conoscenza non specificati all'inizio del processo di addestramento. La definizione più completa è a nostro avviso quella di General Purpose AI System (GPAIS): “*An AI system that can accomplish or be adapted to accomplish a range of distinct tasks, including some for which it was not intentionally and specifically trained*” (Gutierrez et al., 2023).

Nonostante il divario esistente tra il quadro fornito dalle statistiche e le aspettative di esperti e mondo del business, le previsioni di molti osservatori e della stampa specializzata, ad esempio Bloomberg (2023)<sup>8</sup>, sono ottimistiche. In realtà c'è chi, come Deloitte (2023, nutre qualche dubbio sugli “effetti collaterali” della GenAI, anche se resta “tecno-ottimista”: “*Gen AI can enhance labor demand and productivity as well as ease wage inequality, yet it remains to be seen if everyone can reap its many benefits*”. Lo studio in questione individua il motivo degli effetti limitati su crescita e produttività da parte delle innovazioni negli ultimi 40 anni nel fatto che esse hanno colpito soprattutto lavoratori con livelli salariali medio-bassi, sostituiti, nel caso di impiegati, da programmi di video-scrittura e di fogli di calcolo automatico, mentre nuovi macchinari e robot hanno sostituito lavoratori con skills e salari medi. Tutt'altro scenario si è aperto con la GenAI, che investe compiti e funzioni svolte nel campo della formazione, della matematica, della ricerca in qualsiasi disciplina (come sta già avvenendo), nonché lavori ad alto livello salariale sul terreno dei servizi (legali, finanziari professionali).

Date le peculiarità dei modelli linguistici e la prospettiva che essi riescano a penetrare in un set indefinito di domini conoscitivi e operativi, due elementi balzano all'evidenza: 1) non si può prevedere a priori quale dei tre effetti prima indicati (*displacement, productivity, reinstatement*) prevarrà (Green, 2023), proprio perché la GenAI può modificare singoli task e/o gruppi di essi, che

---

<sup>8</sup> “*The generative AI market is poised to explode, growing to \$1.3 trillion over the next 10 years from a market size of just \$40 billion in 2022, according to a new report by Bloomberg Intelligence (BI).*” (Bloomberg, 2023).

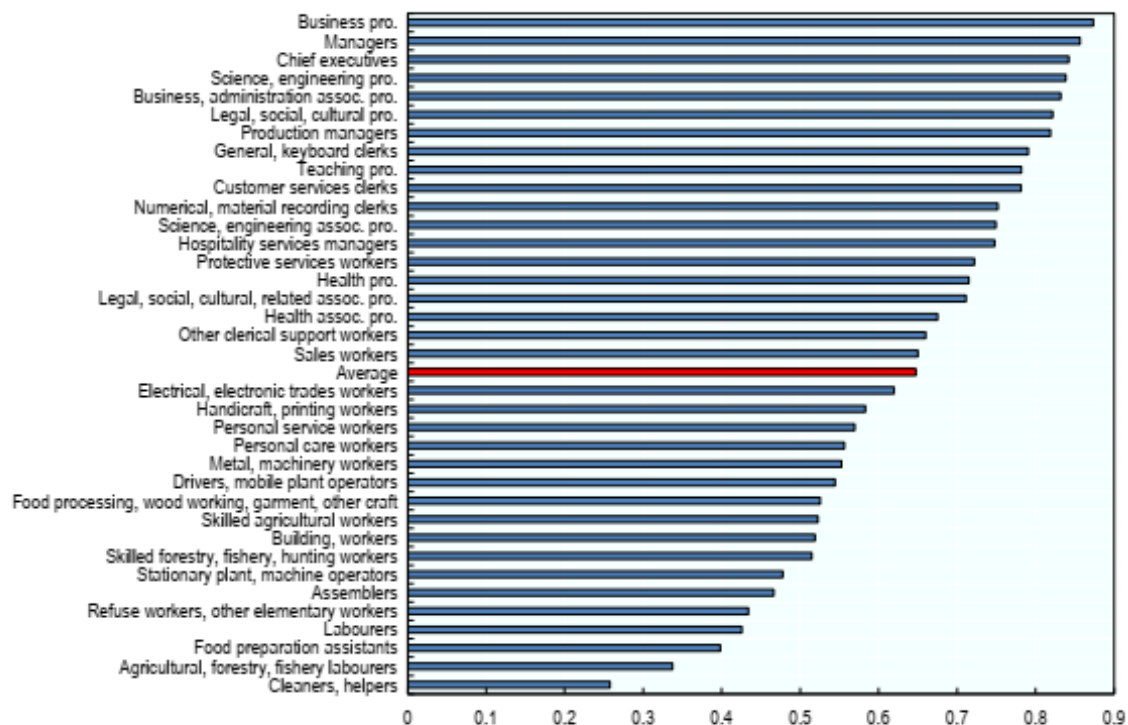
dependono da come concretamente si applicano i sistemi generativi, i quali a loro volta possono indurre mutamenti non prevedibili sulla base delle conoscenze possedute dagli operatori umani. 2) L'ondata innovativa odierna è differente dalle precedenti perché, a nostro parere, implica la riprogettazione su nuove basi di una serie indefinita di attività, quindi tutto dipende da una serie complessa di fattori: peculiarità tecnologiche e socio-culturali dei contesti micro (imprese); assetti istituzionali; andamenti macro-economici; evoluzione delle catene internazionali di subfornitura e dei network globali.

Due interrogativi rimangono perciò aperti (Green, 2023): 1) quale sarà l'effetto aggregato sulla domanda di lavoro sulla base dei movimenti racchiusi nei tre effetti? 2) Nel concreto, quale effetto prevarrà quando si automatizzano task o gruppi variabili di essi?

È certo che finora l'applicazione di modelli linguistici ha avuto un impatto limitato sull'occupazione e ha soprattutto riguardato occupazioni ad alta intensità di conoscenza, con retribuzioni medio-alte, come mostra chiaramente un'indagine svolta da Georgieff e Hye (2021), che stimano il grado di esposizione alla AI dell'occupazione sulla base essenzialmente di due indicatori: 1) misura livello ed esposizione a tecnologie *computer-based* (Felten et al., 2019). 2) Un *framework* che connette 59 task generali<sup>9</sup> a 19 abilità cognitive, estratte da una lista di 328, a loro volta desunte dalla letteratura scientifica, da progetti di punta e dalle sfide emergenti (Tolen et al., 2021).

L'analisi di Georgieff e Hye estende a 52 le abilità cognitive considerate e applica l'indagine a 23 Paesi OECD, dopo aver definito le abilità acquisite dai sistemi di AI negli anni 2010-2015. Su questa base sono classificate le occupazioni di colletti bianchi con dotazioni cognitive medio-alte con il più alto livello di esposizione alla AI (Fig 2)

Fig. 2 Occupazioni di colletti bianchi con dotazioni cognitive medio-alte maggiormente esposte alla AI-



Note: The averages presented are unweighted. Cross-country averages are taken over the 23 countries included in the analysis.

Fonte: Georgieff e Hye (2021, Fig. 4)

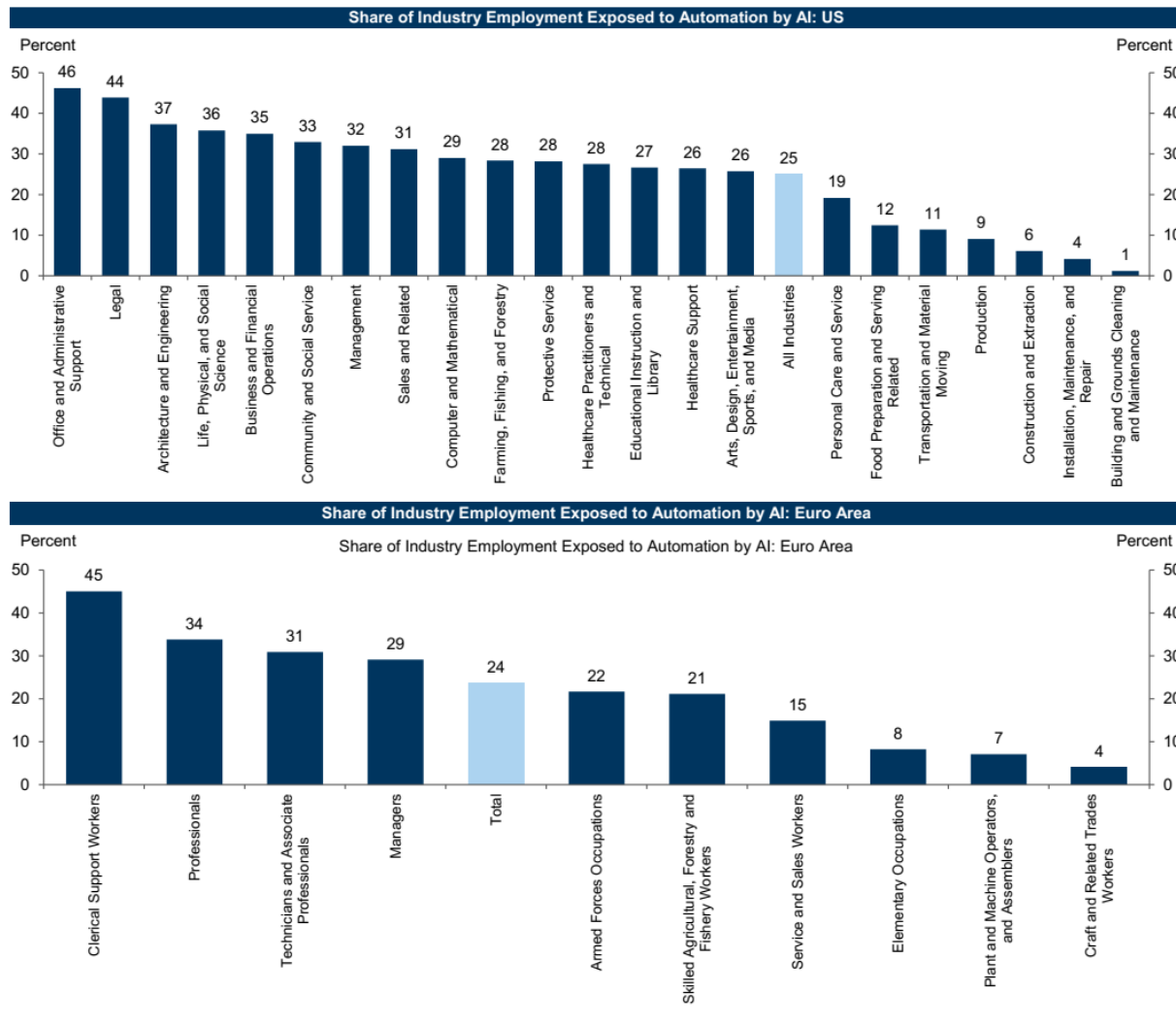
<sup>9</sup> Estratti da tre Survey, più precisamente da: European Working Conditions Survey (EWCS), Survey of Adult Skills ((PIAAC), Occupational database O\*NET (US Department of Labor)..

Prima, dunque, dell'emergere della GenAI, le abilità acquisite dalla AI riguardano in primo luogo "Business Professionals", "Managers", "Chief Executives" and "Science and Engineering Professionals", mentre le meno esposte sono "Food preparation assistants", "Agriculture, forestry and fishery labourers" and "Cleaners, helpers".

Queste stime sono coerenti con quelle recentemente pubblicate da Glodman Sachs (2023), dalle quali risulta che, tenendo presente il crescente impiego di GenAI, un quarto dei task svolti negli Usa e in Europa è esposto alla AI (Fig. 3), mentre la stima dell'esposizione a livello mondiale scende al 18% (Goldman Sachs, Exhibit 6, qui non riprodotta).

Fig. 3

Exhibit 5: One-Fourth of Current Work Tasks Could Be Automated by AI in the US and Europe



Source: Goldman Sachs Global Investment Research

Fonte: Goldman Sachs (2023: Exhibit 5).

Queste ipotesi previsive, sostanzialmente basate su stime dell'”esposizione delle occupazioni alla AI”, devono essere corroborate con analisi più puntuali e sistematiche, con dati inerenti ai processi reali di innovazione, tenendo presente, come autorevolmente sostengono Brynjolfsson et al. (2018) che, dopo aver analizzato le ondate innovative susseguitesesi nel decennio scorso, argomentano la seguente tesi. Innanzitutto, le occupazioni devono essere viste come “*bundle of tasks*”, insiemi composti di abilità e compiti lavorativi, sui quali vengono applicate le nuove tecnologie, come ad es. le *General Purpose Technologies*. Ogni ondata innovativa è diversa dalle precedenti, in forza delle abilità che Machine Learning e più in generale la AI acquisisce, che –come negli anni analizzati- inducono a reingegnerizzare processi e riorganizzare i *task*, a seconda dei settori in cui

sono applicate le innovazioni. Alla luce di queste riflessioni, è logico ipotizzare che saranno influenzate differenti parti della forza lavoro con inevitabili mutamenti della struttura salariale. In definitiva, quindi, sono destinate a cambiare la composizione delle attività e delle competenze necessarie, con una tendenziale grande varietà applicativa delle nuove tecnologie, a seconda delle specificità dei contesti aziendali e settoriali. A tutto questo bisogna poi aggiungere che l'adozione delle nuove tecnologie richiede tempo: occorre sia acquisire conoscenze del tutto nuove per ripensare processi e prodotti, sia adottare tecnologie complementari rispetto a quelle fondamentali, sia tenere conto delle relazioni industriali e della reattività delle reti di subfornitura.

Bisogna inoltre osservare che l'esposizione ai nuovi sistemi di AI non porta con sé variazioni occupazionali a livello aggregato (finora), come argomenta Green (2023, sez. 3.2.1) sulla base di una vasta serie di studi empirici. Mutamenti sono però rilevabili nella composizione della forza lavoro necessaria per le imprese che intendono innovare, come conferma un'indagine svolta Acemoglu et al. (2022a) circa l'offerta di lavoro online (posti vacanti, *job vacancies*): è rilevato un significativo spostamento verso richieste di persone con abilità connesse alla AI. Questi dati confermano la tendenza, rilevata nelle imprese che hanno una composizione delle attività lavorative (*task structure*) adatta per l'applicazione dei nuovi sistemi di AI, e portano alla conclusione che a livello di imprese c'è un effetto *displacement* compensato maggiormente dal *reinstatement*, mentre l'effetto *productivity* non sembra particolarmente incisivo, con il risultato che a livello aggregato l'impatto occupazionale negativo non è (ancora) percepibile<sup>10</sup>.

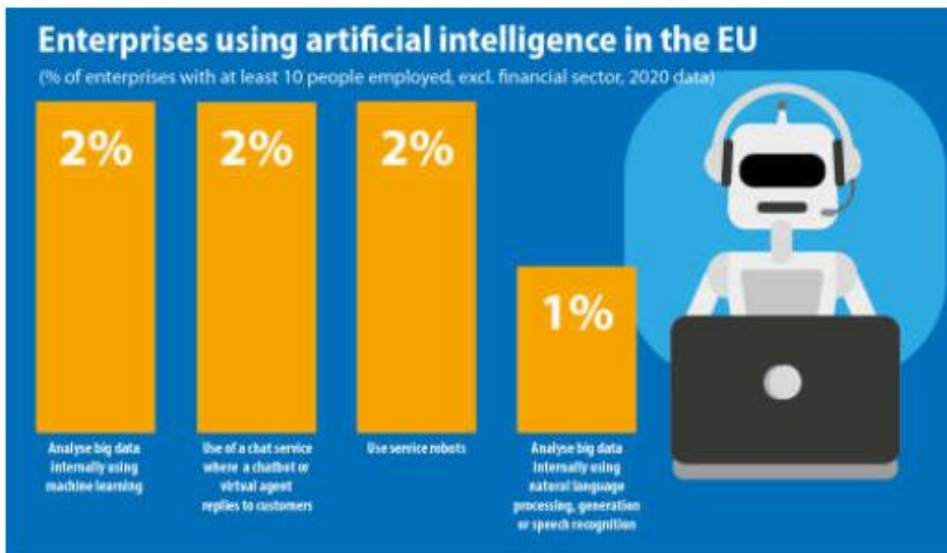
Ulteriori spunti di riflessione sono desumibili dall'indagine svolta da Acemoglu et al. (2022b) su 300.000 imprese americane, tratte dall' Annual Business Survey 2019, che contiene dati relativi all'adozione di cinque tecnologie: AI, robotica, apparecchiature dedicate, software specializzato, cloud computing. Emergono risultati molto interessanti, dal momento che adotta robotica e AI solo il 3,2% delle unità indagate, che però costituiscono una quota rilevante dell'occupazione e dell'economia, trattandosi prevalentemente di grandi imprese, le quali appaiono più propense all'innovazione rispetto ad altre entità appartenenti alle stesse tipologie industriali. Non è da trascurare, poi, che oltre metà delle aziende USA non ha in funzione AI e robot negli anni 2016-2018. Un secondo elemento degno di interesse è la grande varietà dei tassi di adozione tra i vari settori industriali, dove peraltro viene messo in luce che chi adotta tecnologie avanzate per l'automazione registra aumenti della produttività del lavoro e diminuzione della percentuale di occupati, con salari più alti, rispetto a entità produttive simili per dimensioni e appartenenza settoriale. Ulteriori ricerche sono quindi necessarie per conoscere gli andamenti aggregati di occupazione e salari, oltre che la possibile creazione di nuovi beni e servizi da offrire ai consumatori.

È opportuno proporre un confronto con l'Europa, dove l'utilizzo di AI nelle imprese è vicino a quello rilevato dall'indagine OECD, come si vede nella Fig. 4, con dati riferiti al 2020.

Fig. 4

---

<sup>10</sup> “However, while AI technologies appear to be changing task and skill composition at exposed establishments, any aggregate effects of AI, if present, are not yet detectable—plausibly because AI technologies are still in their infancy and have spread to only a limited part of the US economy” (Acemoglu et al., 2022: 337).



Fonte: Eurostat (2021)

Gli elementi addotti inducono i ricercatori ad affrontare interrogativo cruciale. Come attesta anche la rilevazione, svolta nel 2022 presso 2000 datori di lavoro e 5.300 lavoratori dei due settori (finanza e manifattura) di 7 Paesi OECD (2023), *Why is the employment effect of AI small (so far...)?* (Green, 2023: 113). Spunti di riflessione in merito sono forniti dallo stesso Rapporto OECD, dove si evidenzia in primo luogo come la lenta adozione di AI implica “cambiamenti marginali dell’occupazione” a livello micro, i quali non determinano incrementi significativi di produttività, tali da essere rilevabili dalle misurazioni a livello aggregato (Green et al., 2023). Un secondo elemento interessante è costituito dal comportamento di un campione di 100 imprese<sup>11</sup>, sottoposte a rilevazione diretta dell’OECD (Milanez, 2023). La ricerca conferma i punti messi in evidenza finora, tra cui il fatto che l’adozione di sistemi di AI investe un vasto insieme di task e di lavoratori, e mette in evidenza un dato di assoluto rilievo: la sostanziale invarianza dell’occupazione nelle aziende, nonostante un rallentamento nella crescita degli occupati. Questo risultato deriva dalle strategie poste in essere, che hanno privilegiato la riorganizzazione del lavoro mediante i cambiamenti del capitale fisico unito alla formazione del personale esistente. Cardine di questi disegni di trasformazione in risposta alle sfide della AI è il coinvolgimento delle maestranze attraverso forme di “dialogo sociale”, che è riuscito di fatto a far prevalere il *reinstatement* sul *displacement*. La ricerca infine sottolinea che sono stati anche conseguiti due obiettivi: miglioramento della qualità e della sicurezza del lavoro.

Un altro fondamentale elemento può aiutare a comprendere la problematicità inerente alla dinamica innovativa odierna e ancor più quella del futuro prossimo. Partiamo dalla considerazione che i nuovi sistemi di AI, soprattutto quelli di GenAI, implicano un radicalmente nuovo modo di pensare, agire e progettare. Da ciò deriva inevitabilmente che gli assetti e le culture aziendali devono misurarsi con nuovi paradigmi tecnico-produttivi e nuove modalità di interazione uomo-macchina, con quest’ultima per di più dotata di abilità linguistiche, come definite nel par. 2. È inevitabile che, come argomentano Brynjolfsson et al. (2019), sono necessari investimenti in beni complementari e intangibili, sulla base dei quali procedere a processi di adattamento delle culture tecnico-produttive e socio-economiche. Tutto ciò richiede tempo perché occorre ripensare i prodotti, da creare su basi cognitive nuove e in continua evoluzione, mentre si devono reingegnerizzare processi e organizzazioni<sup>12</sup>. La dinamica delineata può essere può avere notevoli complicazioni nell’odierno

<sup>11</sup> Le imprese operano nei seguenti Paesi: Austria, Canada, Francia, Germania, Irlanda, Giappone, Inghilterra, Stati Uniti.

<sup>12</sup> Brynjolfsson et al. (2019) propongono il modello “J” per rappresentare la sequenza di fasi: un lento processo di adattamento iniziale, seguito da un’accelerata piena utilizzazione del potenziale intrinseco alle nuove tecnologie.

scenario globale, dominato dalla diffusione sistematica di reti di subfornitura internazionali, che complicano il quadro evolutivo. Le innovazioni complementari, prima indicate, richiedono che si diffondano lungo catene globali del valore, a loro volta spesso disomogenee da vari punti di vista, rendendo così molto complessa quella che possiamo chiamare “sincronizzazione temporale”, resa necessaria dalla dinamica innovativa multidimensionale in corso.

## Problemi e sfide nell’evoluzione di produttività e crescita

Dalla rassegna effettuata e dai temi sollevati discende un quadro contraddistinto da una elevata complessità dello scenario odierno e di quello del prossimo futuro. Il numero elevato di variabili e processi di cambiamento in atto delinea uno scenario non riducibile al controllo di una sola variabile oppure di un solo fattore o processo propulsivo. La stessa dinamica tecnico-scientifica, motore del progresso socio-economico, non può essere la sola determinante di una delle possibili traiettorie di evoluzione sociale, ma deve essere temperata con altre componenti, pena la generazione di contraddizioni socio-tecniche, economiche e politiche foriere di conflitti e disastri incombenti. Lo sviluppo tecnico-scientifico da solo non può costituire “*one-size-fits-all*”, l’unico fattore propulsivo della dinamica sociale, anche se ovviamente è un ingrediente essenziale per uno scenario di progresso. Proprio le sue attuali caratteristiche, che abbiamo riassunto nell’espressione “salto quantico” (Cfr par. 2), indicano che siamo di fronte ad un’era di straordinaria trasformazione, che richiede profondi cambiamenti culturali e socio-tecniche, tali da indurre a ripensare “il posto dell’uomo nella natura” (Teilhard de Chardin, 2004: 185). L’umanità ha creato un potenziale tecnico-scientifico dai confini indefiniti, che incide su due elementi basilari del mondo in cui viviamo: l’intelligenza e la vita (Suleyman e Bhaskar, 2023). I modelli linguistici catturano proprietà formali del nostro modo fondamentale di rappresentare il mondo, cioè il linguaggio in tutte le sue forme espressive, e al tempo stesso consentono l’esplorazione accelerata dello “spazio delle proteine”, grazie appunto al Natural Language Processing<sup>13</sup>. Grazie ai modelli linguistici sono stati creati 2,2 milioni di nuovi materiali (Merchant et al., DeepMind, 2023) e un numero imprecisato di nuove proteine (Wang, 2023). Si è così aperto un orizzonte apparentemente senza limiti per la genomica dei materiali e la post-genomica umana, sulla base di un potenziale tecnico-scientifico indefinito, che induce una radicale riprogettazione dei fondamenti della vita e dei processi socio-economici. Uno spazio combinatoriale del tutto nuovo è dinanzi all’umanità e alle conoscenze-abilità finora possedute. È evidente che uno scenario di questa natura implica una non breve né agevole era di profonde ed estese trasformazioni delle attività e dei modi di interagire sia tra umani, sia tra questi e i prodotti del “mondo della tecno-scienza”. Bisogni, finalità, obiettivi di breve e lungo termine vanno ripensati necessariamente su basi collettive perché, come si evince dalle ricerche empiriche concernenti le imprese indicate nel paragrafo precedente, le trasformazioni da affrontare sono numerose e profonde, per cui è necessaria la collaborazione dei molteplici soggetti sociali coinvolti. È d’altronde chiaro che in contesti caratterizzati da processi evolutivi multidimensionali di portata globale, come quelli odierni, è essenziale la convergenza di progetti top-down e bottom-up: gli uni senza gli altri oppure privi di una reciproca congruenza comportano il fallimento epocale di ipotesi di crescita e produttività dei sistemi economico-produttivi. Occorre tempo per vedere risultati, che saranno deludenti fino a che le strategie poste in essere a livello sistemico non assimilano l’importanza della necessità di cambiare profondamente il modo di pensare e di agire per affrontare temi fondamentali non per questa o quella popolazione oppure

<sup>13</sup> Nel 2013 inizia la Critical Assessment for Structure Prediction” (CASP), competizione internazionale biennale per trovare chi per primo è in grado di decifrare (*crackare*) le regole di composizione delle proteine. Nel 2018 DeepMind vince con AlphaFold e successivamente vengono create un numero rilevante di proteine (Arnold, 2023; Wang, 2023). L’ulteriore accelerazione tecnico-scientifica è evidente.

impresa, ma il futuro dello stesso pianeta, date le forze che l'umanità sta oggi in qualche modo iniziando a controllare. Uno dei più grandi paradossi della nostra epoca è infatti costituito dal divario tra l'enorme potenziale tecnico-scientifico disponibile e i problemi che affliggono una parte non irrilevante dell'umanità.

## Bibliografia

Acemoglu D. et al., 2022a, "Artificial Intelligence and Jobs: Evidence from Online Vacancies, *Journal of Labor Economics*, Vol.40/S1, 293-340.

Acemoglu D. et al., 2022b, Automation and the Workforce: A Firm-Level View from the 2019 Annual Business Survey, National Bureau of Economic Research, Cambridge, MA.

Arnold C., 2023, AlphaFold touted as next big thing for drug discovery - but is it?, *Nature*, 5 October: 15-16.

Brynjolfsson E., Rock D., Syverson C., 2021, The Productivity J-Curve: How Intangibles Complement General Purpose Technologies, *American Economic Journal: Macroeconomics*, Vol. 13/1, pp. 333-372,

Brynjolfsson E., Mitchell T., Rock D., 2018, What Can Machines Learn and What Does It Mean for Occupations and the Economy?, *AEA Papers and Proceedings*, Vol. 108, pp. 43-47.

Adler G., et al., 2017, Gone with the Headwinds: Global Productivity, IM F Discussion Papers.

Aghion, P., Antonin, C., Bunel, S., 2019, Artificial Intelligence, Growth and Employment: The Role of Policy. *Economie et Statistique / Economics and Statistics*, 510-511-512, 149-164.

Aghion, P., Antonin, C., Bunel, S., Jaravel, X., 2020, What are the Labor and Product Market Effects of Automation? New Evidence from France. *CEPR Discussion Paper*, 14443

Aghion et al., 2022, The Effects of Automation on Labor Demand: A Survey of the Recent Literature, *CEPR Discussion Papers*, DP16868.

Bloomberg, 2023, Generative AI to Become a \$1.3 Trillion Market by 2032, *Research Finds*, June 1.

Bresnahan T.F., Trajtenberg M-. 1995, General purpose technologies "Engines of growth?", *Journal of Econometrics*, 65, 83- 108.

Briggs J., Kodnani D., 2023, The Potentially Large Effects of Artificial Intelligence on Economic Growth, *Goldman Sachs Economics Research*

Brynjolfsson E., McAfee A., 2014, *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W.W. Norton & Company.

Brynjolfsson E., Mitchell T., Rock D., 2018, What Can Machines Learn and What Does It Mean for Occupations and the Economy?, *AEA Papers and Proceedings*, Vol. 108, 43-47.

Bubeck S. et al., [Microsoft Research](#), 2023, Sparks of Artificial General Intelligence: Early experiments with GPT-4, March

Byrne D. M., Oliner S.B., Sichel D.E., 2015. How Fast Are Semiconductor Prices Falling?, NBER Working Paper 21074.

Dauth W. et al., 2017, German Robots: The Impact of Industrial Robots on Workers, Institute for Employment Research Discussion Paper 30.

Dauth W. et al., 2018, [Adjusting to robots: Worker-level evidence](#), Federal Reserve Bank of Minneapolis, August 21.

Dean J., 2022, A Golden Decade of Deep Learning: Computing Systems and Applications, Nautilus, n. 151, Issue 2, 58-74.

Deloitte Insights, 2023, Generative AI and the labor market: the case for techno-optimism, December 5.

Eurostat 2021, Artificial Intelligence in EU Enterprises, April 13, consultato il 29-12-2023.

Felten, E., Raj R., Seamans R., 2018, "A Method to Link Advances in Artificial Intelligence to Occupational Abilities", AEA Papers and Proceedings, Vol. 108, 54-57.

Future of Life Institute, 2022, General Purpose AI and the AI Act.

Gordon R. J., 2003, Exploding Productivity Growth: Context, Causes, and Implications, Brookings Papers on Economic Activity, 2.

Gordon R.J., 2018, Why has economic growth slowed when innovation appears to be accelerating?, NBER Working Papers Series, December.

Greco R., 2023, A structural analysis of productivity in Italy: a cross-industry, cross-country perspective, Banca d'Italia, Questioni di Economia e Finanza. Occasional Papers.

Green A., 2023, Artificial intelligence and jobs: No signs of slowing labor demand (yet), OECD Employment Outlook 2023, Cap. 3.

Green et al., 2023, Artificial intelligence, job quality and inclusiveness, OECD Employment Outlook 2023, Cap. 4.

Gutierrez C.I. et al., 2023, A Proposal for a Definition of General Purpose Artificial Intelligence Systems, Digital Society, 4 September, 2: 36.

Hassabis D., 2022, su X "Turns out everything is a matrix multiplication from computer graphics to training neural networks :) ...."

Lombardi M., 2022, Cos'è il lavoro? Il futuro, le tecnologie e l'ignoto da esplorare, Agenda Digitale, 15 Febbraio. <https://www.agendadigitale.eu/cultura-digitale/cose-il-lavoro-il-futuro-le-tecnologie-e-lignoto-da-esplorare/>

Mahowald K., et al 2023, Dissociating language and thought in large language models a cognitive perspective, arXiv:2301.06627v1 [cs.CL], January 16.

Merchant A. et al., DeepMind, 2023, Scaling deep learning for materials discovery, *Nature*, Vol 624, 7 December: 81-85.

Milanez A., 2023, The impact of AI on the workplace: Evidence from OECD case studies of AI implementation, OECD Social, Employment and Migration Working Papers, No.289, OECD Publishing, Paris.

Mokyr J., 2014, "Secular Stagnation? Not in Your Life." Chap. 6 in *Secular Stagnation: Facts, Causes and Cures*, edited by Coen Teulings and Richard Baldwin. London: CEPR Press.

Seth A., 2023, Why Conscious AI Is a Bad, Bad Idea, *Nautilus*, May 8.

*Solow R.M., 1987, We'd better watch out, New York Times Book Review, July 12.*

Suleyman M., Bhaskar M., 2023, *The Coming Wave*, Crown-Penguin Book, New York.

Teilhard de Chardin P., 2004, *The Future of Man*, Image Books, New York.

Tolan S. et al., 2021, "Measuring the Occupational Impact of AI: Tasks, Cognitive Abilities and AI Benchmarks", *Journal of Artificial Intelligence Research*, Vol. 71, 191-236.

Tyson D'Andrea L., Zysman J., 2022, Automation, Ai & Work, *Daedalus*, the Journal of the American Academy of Arts & Sciences, 256-271.

Tyson D'Andrea L., 2022, Automation and the Future of Work in Germany: A Summary of Research and Policy Recommendations, BERKELEY ROUNDTABLE ON THE INTERNATIONAL ECONOMY.

Vaswani A et al., 2017 "Attention Is All You Need," [arXiv:1706.03762](https://arxiv.org/abs/1706.03762)

Wang J., 2023, A roadmap for exploring the untouched protein space for biology and medicine, *hLife*, Volume 1, Issue 2, December: 93-97.

Zittrain J.L., 2006, The Generative Internet. *Harvard Law Review*, Vol. 119, No. 7, 1974-2040

# Quale digitale per l'età prescolare: cosa considerare

Oggi è tempo di una cartografia dei cambiamenti sociali a cui le scienze dell'uomo aspirano. In forza di questa legittima postura conoscitiva, non pochi studi reclamano l'attenzione dei lettori, ricorrendo ad un incipit apparentemente consunto: c'è qualcosa di radicalmente nuovo nel nostro tempo.

Di **Mario Morcellini**, professore ordinario in Sociologia dei Processi Culturali e Comunicativi, Roma1

**La prima novità riguarda l'età della vita più delicata e strategica per la costruzione del futuro.** La novità di un'infanzia che fino a pochi anni fa si avvicinava alla scuola, e addirittura al **rituale d'ingresso**, senza una vera e propria qualificazione dello spazio di attesa tra la nascita e l'avvio dell'esperienza formativa.

## Cosa perdiamo acquisendo abitudini e scelte diverse dal passato

Già oggi questo è un vero e proprio **campo di battaglia**, a partire dalla constatazione di un tempo la cui intensità e durata sono cambiate profondamente sotto i nostri occhi. Infatti negli ultimi anni, i bambini non “nascono” alla società nel momento di ingresso a scuola, ma si cimentano molto rapidamente con la crescente concentrazione dell'attenzione rappresentata dalla bellezza e modernità dei due caratteristici device digitali, tablet e smartphone.

**Ma non si fermano qui le novità.** Altre coinvolgono le procedure con cui le scienze della cultura approcciano i rapidi cambiamenti dei fondali simbolici delle nostre esistenze, in qualche misura, ridimensionando fisiologicamente ciò che è nuovo e sovrastimando il fatto che esso si confonde e si “incarta” continuamente con il già visto, almeno fino al momento in cui scatta l'**interrogativo su cosa perdiamo acquisendo abitudini e scelte diverse dal passato** [1].

**Il gioco caratteristico della modernità è proprio l'oscillazione continua tra vecchio e nuovo.** Ma su questo tema non si registra un'adeguata attenzione perché ognuno di noi tende a mappare il “mondo antico” considerando il presente come un suo prolungamento, quasi prevalesse nel nostro sguardo l'idea che la continuità sormonta comunque le differenze, mentre **stiamo vivendo una fase di terremoto** che Papa Francesco ha chiamato esattamente “cambiamento d'epoca”.

Non si tratta allora di discutere se l'adulto di oggi sia antiquato (secondo quanto aveva ammonito Günther Anders [2]), perché il radicale licenziamento del passato, ma soprattutto l'**affermazione disordinata di una cancel culture** fanno pensare ad un **radicale cambiamento antropologico**. Certamente non siamo la prima generazione che si interroga sul mutamento, e dunque niente di nuovo sotto il sole. Ma non a costo di non riconoscere le **novità profonde che stanno emergendo**,

circoscrivendo dati più precisi e ancora indefiniti per una compiuta osservazione del nuovo soggetto moderno. Del resto è **in gioco il modello di sviluppo non solo sociale, ma quello di ognuno di noi**. E dunque dobbiamo **mettere in ordine il nostro sapere e dare un senso ai trend oltre l'individualismo**, selezionando segni distintivi rispetto al passato. Solo così si affermano nuove immagini sociali, una diversa “percezione del tempo” che ne comporta la sensazione di labilità e svuotamento, una sottovalutazione generale del fatto che il mutamento delle impostazioni culturali e simboliche presenta sempre un conto in termini di valori, e persino in termini di adeguata definizione della loro rilevanza nella vita.

Ecco come l'accelerazione percepita del tempo coinvolge e stravolge funzioni sociali elementari, mettendo in discussione la consecutio temporum nei processi di crescita e una piena presa di possesso delle parole.

## Zero/sei: il digitale a scuola nell'età pre-primaria

Già Gabriella Paolucci [3] ha notato, in un contesto più generale, che la compressione spazio/temporale provoca una rottura degli equilibri precedenti, ancora una volta incidendo sugli stili di conoscenza del mondo.

Diventa allora una vera scommessa anche teorica riflettere sui cambiamenti posti all'inizio della vita, per la buona ragione che si tratta di un periodo poco approfondito dalla letteratura contemporanea e su cui le variabili di impatto risultano semplici e facilmente leggibili anche in termini di interferenze.

Prendiamo le mosse dalla **natura assorbente dello sguardo dei bambini verso i digital media**, che già di per sé si configura come una **novità assoluta**, poiché solo fino a pochi anni fa **prevaleva una diversa varietà di interazioni e possibili indirizzi di socializzazione che oggi appaiono sullo sfondo**. Infatti la **familiarizzazione e sintonia** che i bambini rapidamente esibiscono verso la cosmologia comunicativa dei media digitali mette in evidenza che questi due mondi letteralmente si aspettavano.

Una tale evidenza è persino ovvia da parte dei grandi player mondiali degli immaginari, a cui era certamente chiaro che il mercato dei bambini potesse diventare un ingrediente fondamentale di successo e accreditamento.

Ben diversa è la nostra postura volta a **interrogarsi su un'autonomia percepita che rischia di diventare impenetrabile per gli adulti**. Entro questo processo è difficile non riconoscere un aspetto più squilibrante del percorso di crescita, e cioè, da un lato, la velocità nel “**riconoscimento di affinità**” tra linguaggi e tecnologie digitali e, dall'altro, **aspettative e bisogni profondamente cambiati da parte dei bambini**.

### I nodi

Al momento è presto (e soprattutto complicato) per porre il nodo del **diverso potere di impatto dei due fenomeni**. Ma è tardi per riconoscere che **tutto questo non può essere senza conseguenze per la costituzione della mente, dell'espressività, degli stili e della stessa varietà comunicativa dei bambini**. Una letteratura sia scientifica che divulgativa non ha mancato di segnalarci quanto tale realtà profili un processo di “conquista” delle menti non solo degli adulti ma di tutti [4]. Soprattutto bambini [5].

Anche quest'ultima è una precisazione molto riduttiva, poiché se è certamente in questione un processo di implementazione delle mentalità degli uomini contemporanei, e dunque sulle architetture cognitive imperniata su nuove relazioni tecnologiche, non possiamo non riconoscere che un tale processo può ingigantirsi in termini di autorevolezza e credibilità se proiettato su quanti stanno costruendo e abbigliando la propria identità.

**Da un lato, l'impatto degli adulti di riferimento resta figurativamente al suo posto, ma troppo spesso si confonde con una specie di distrazione e conseguente vuoto sociale rispetto ai nuovi venuti. Dall'altro, per contrasto, il fenomeno definito come concentrazione dell'attenzione e della seduzione visiva si trasforma in una zona militarizzata dal digitale che rischia di diventare il nuovo attore di un processo inequivocabile di co-formazione.**

## Un nuovo cantiere conoscitivo

Bastano questi pochi spunti per dichiarare aperto un cantiere conoscitivo radicalmente nuovo, che chiama in causa **etica e deontologia** di quanti debbono pensare e verificare la loro capacità educativa [6]. Ma fa intravedere anche **responsabilità nuove** della politica, e meglio ancora, delle **policy soprattutto sovranazionali chiamate a pensare nuove strategie di adeguazione e di reazione**.

Si apre dunque un diverso spazio scientifico nel quale l'unica cosa a cui non dovremo cedere è la disputa sulla proprietà e affinità di questo imponente segmento di ricerca, poiché solo il concorso di tanti saperi può ridurre quello che sembra un certo ritardo nel mettere in campo un pensiero strategico.

La titolarità scientifica non potrà che essere plurale. Ma è difficile non annotare che chi studia comunicazione è in vantaggio nel capire o intravedere tanti nodi di influenza delle tecnologie e di definizione del tipo di cultura che esse comportano.

L'unica certezza è che i settori scientifico-disciplinari sono troppo autoreferenziali e chiusi rispetto ad una tematica che, anche a causa delle continue novità nella comunicazione, va affrontata con una sensibilità larga che dovrebbe ispirarsi ai "novissimi".

Sullo sfondo si profila **un aumento di responsabilità per tutti quelli che non vogliono chiudere gli occhi sulla realtà semplicemente perché è offuscata da una corsa che appare quasi estrema.**

## L'indagine

Questo territorio è emerso con chiarezza ai miei occhi a partire da una lunga ed elaborata indagine [7] che già nel decennio precedente è stata avviata, riempita di contenuti e per molti versi ultimata, sempre in presenza di un clamoroso ritardo di finanziamento da parte di soggetti pubblici che pure l'avevano sostenuta nella parte programmatica.

Questa ricerca è stata, almeno per la parte sociocomunicativa e giuridica, documentata per tempo in diversi interventi e convegni scientifici [8], ma soprattutto in un monografico della rivista *ComunicazionepuntoDoc*.

L'obiettivo di quell'indagine ampia e ambiziosa era, infatti, quello di offrire elementi utili anche ai fini di eventuali applicazioni didattiche dei dispositivi touch screen e degli smart objects e, più in

generale, del ruolo della comunicazione in un tempo in cui essa non implica più una società ordinatrice ed un punto di riferimento esterno alla cosmologia dei media (Morcellini 2013) [9].

Erano queste le principali istanze che hanno spinto a realizzare un team di ricerca con due anime distinte, ma fortemente correlate. Da un lato, i partner universitari (Sapienza Università di Roma, Università di Napoli “Federico II”, Università di Salerno, Università di Trento). E dall’altro, quelli tecnologici (Engineering Ingegneria Informatica, capofila del progetto, Fastweb, Interactive Media e il Consorzio iCampus).

Dalla sinergia fra queste sensibilità conoscitive e specifiche expertise dei singoli soggetti è del resto nato il **Progetto di Ricerca Inf@nzia DIGI.tales 3.6**, vincitore del bando PON Ricerca e Competitività 2007-2013 – “Smart Cities and Communities and Social Innovation”, promosso dall’allora Ministero dell’Università e della Ricerca con l’intento di proporre un rinnovamento dei modelli educativi delle scuole dell’infanzia, favorendo, in un’ottica di interazione con insegnanti, gruppo dei pari e famiglia, il potenziamento delle modalità di esplorazione e manipolazione dei più piccoli.

## L’ipotesi di un Monopoly digitale prima della scuola

**Bambini e giovani di oggi possono essere descritti attraverso una semplice formula semantica: quella dell’indicativo presente.** Il trapassato prossimo non esiste più, e perfino l’imperfetto (la forma linguistica dello ieri) sembra tramontato. A dir la verità anche il futuro non è messo bene, in questo caso seguendo perfettamente la proiezione di quello che la realtà percepita presenta a ragazzi e giovani.

In tale contesto lo spasmo di questi ultimi nei confronti della comunicazione è profondamente indicativo della percezione di vuoto addebitata al vocabolario della vecchia educazione. Allora riflettiamo su alcuni caratteri che possiamo solo accostare speculativamente a questo ritratto dei nuovi venuti con la comunicazione.

Per nostra formazione culturale, non riusciamo a cogliere bene cosa significa per gli uomini cambiare tutto e veder cambiare tutto. Partiamo allora da un primo aspetto connotato al modo in cui la comunicazione sovrasta le persone.

**Non c’è simmetria nel rapporto tra soggetto e media, mentre il soggetto crede di essere in una situazione di parità** quando addirittura non si percepisce alla **regia** delle scelte.

Cerchiamo dunque di rappresentarci bene il vissuto profondo del nesso diventato assolutamente preferenziale con la comunicazione perché da qui discendono alcuni aspetti degli **atteggiamenti di fondo con cui il soggetto in età scolastica gira la testa verso i media digitali.**

Dobbiamo inoltre domandarci se non ci sia, più che in passato, un **eccesso di pulsioni individualistiche nell’atmosfera culturale che il ragazzo già respira nel mondo in cui è immerso.** Ora esse sono frequentemente **antisocietarie** e dunque enfatizzano un **modello di sviluppo individuale, meno legato che in passato agli adulti e persino al gruppo dei pari.**

Per converso ci è chiaro, anche storicamente, che **la formazione e l’educazione possono darsi solo in società** e dunque diventano più difficili in un contesto di individualismo e di frammentazione. Anzi è proprio questa frammentazione l’opzione tipica dell’esperienza mediale che spinge a cercare altrove il senso, o almeno un momento di ricomposizione, ed è così che la comunicazione assurge a

**grammatica fondamentale a cui il soggetto si volge per essere uguale agli altri**, ma anche per sentirsi **diverso dalle voci degli adulti circostanti**, nell'**illusione di costruire una sua personalità nuova e alternativa**. Un vero sistema operativo della società contemporanea, destinato dunque a costruire un altro modello di relazioni prevalentemente virtuali.

<https://www.agendadigitale.eu/cultura-digitale/tiktok-e-lonboarding-la-chiave-del-successo-cose-e-perche-e-importante/>

## Networking individualism

La letteratura teorica sui media ha definito tale trend come **networking individualism**, ma questa celebrata formula ideata da Castells va verificata. Così sembra sollecitare piuttosto un rovesciamento, **suggerendo l'idea che le relazioni comunicative, definite da Thompson quasi interazioni, sostituiscano le relazioni sociali**, ormai per molti versi congedate in termini di coinvolgimento e priorità dell'attenzione.

## La stagione della formazione

Occorre la pazienza di scoprire tracce della modificazione personale, psicologica e quasi percettiva del ragazzo che vive la stagione sempre più tempestosa della formazione. Può aiutarci l'indagine sugli interessi del soggetto emergenti dal suo sistema di orientamento soprattutto comunicativo, nella speranza che dalle architetture cognitive riscontrabili nelle frequenze al livello di scelte e comportamenti (ad alto tasso di ripetizione), emergano tratti culturali nuovi e da definire con grande pazienza.

## Le tendenze

Accanto alle pulsioni, già segnalate sopra, anche per rendere più concreta l'implementazione di quel concetto, si individua un importante trend (o meglio un'accentuata **disponibilità**) **alla disinibizione, aspettative corte quanto i pensieri, un evidente disinteresse nei confronti di regole condivise e una noncuranza per gli altri che induce una perdita di empatia nella vita** e rischia di segnare un'età della formazione.

Qui appare nella sua portata quel concetto di "fine del potere di assimilazione della cultura" [10], correlato al profondo mutamento dei modi in cui il soggetto percepisce e recepisce il presente. Non si può trascurare, del resto, che il giacimento di pulsioni sia sovralimentato dalla **disinibizione**, a sua volta strettamente connessa alla **disintermediazione** e al **disordine informativo** condizionato dalle *fake news*.

Si tratta dunque di lavorare ricorrendo all'inferenza. **Dall'osservazione, attraverso un processo faticoso di scavo e verifica, arrivare alle conclusioni e dunque alla formulazione di possibili ritratti tipologici.**

## Lo studio

Del resto, **superare le crisi significa anzitutto studiarle**, altrimenti si afferma una generica retorica del cambiamento incapace di cogliere cosa resta dopo. Da esse, infatti, si esce rafforzando la **fiducia**, ricostruendo **nuove geometrie sociali e affinità diverse rispetto al passato**, contro ogni rischio di **isolazionismo** e **relativismo** che altrimenti diventa **la cicatrice della nostra epoca**.

## I Care, We care: una responsabilità collettiva di fronte al cambiamento

Non è una scelta di maniera chiamare ora in causa il **centenario della nascita di Don Milani** [11] per l'attualità dei messaggi che ci consegna e per il grande esperimento educativo riconducibile alla sua impronta.

A partire da quel magistero, buona parte della cultura e della ricerca ha considerato la **povertà formativa un rischio connesso ai processi educativi**. Invece il prodotto di questa **sventatezza collettiva**, rispetto a cui nessuno è davvero esonerato da responsabilità, dimostra che **ogni fenomeno significativo di perdita di studenti lungo il percorso si traduce in un contributo alla fibrillazione della democrazia**. Se a tutto ciò si aggiunge **l'aumento del chiasso della comunicazione digitale ed un'ignoranza pubblica sempre più urlata**, ne discende che il danno non è solo quello inferto ad un'assunzione di responsabilità politica, ma al **clima culturale e valoriale che si sta instaurando**, anche in forza della veloce globalizzazione degli immaginari.

Don Milani ha speso la vita condividendo un'idea semplice e radicale: **ogni ragazzo che non riusciva a raggiungere e coinvolgere diventava una sconfitta pro futuro**. In questi anni, la controprova è che **le percentuali di dispersione o parziale disimpegno dalla formazione sono state un ingrediente decisivo** posto **contro gli ideali dell'uguaglianza e della promozione sociale**.

Ecco allora che il messaggio del maestro di Barbiana diventa quanto mai pertinente.

### Gli indicatori della situazione d'allarme sulla formazione

Egli non avrebbe esitato ad intervenire su quello che ruvidamente dobbiamo definire come **allarme sulla situazione contemporanea della formazione**, provocato da evidenti **indicatori di spaesamento normativo, fragilità di risposte e permanenti difficoltà ad individuare forme di tutela adeguate all'aggressività delle nuove piattaforme digitali**. Tutto ciò sollecita una **nuova responsabilità da parte degli organismi di regolazione chiamati in causa dall'impatto della comunicazione** proprio per fronteggiare questa contrazione delle chance formative che la letteratura ormai individua come **"povertà educativa digitale"** [12].

Sulla base di tale presupposto, e tanto più dopo una prolungata stagione di emergenza segnata dal Covid, occorre rinnovare l'attenzione sulla straordinaria novità di **un incipit "educativo" che va in scena prima dell'ingresso formale nella scuola primaria, costretta così a confrontarsi con uno dei passaggi più critici della condizione moderna**.

Lo scenario culturale sta cambiando più radicalmente di quanto possiamo renderci conto, **in termini non tanto di crisi dei valori, quanto di troppo rapidi processi di sostituzione**. Non c'è bisogno di ricordare che **rapidità e ritmi dei mutamenti socioculturali moderni** [13] **non coincidono agevolmente con le dinamiche formative**. E questo **gap mette ancora una volta in difficoltà il compito educativo della scuola, rischiando sia di compromettere la disponibilità dei bambini all'educazione**, che il ruolo e la passione degli stessi insegnanti.

### La deriva tecnologica

Si tratta di riflessioni tanto più degne di attenzione per quelle **frange sociali di bambini ancora escluse dall'istruzione pre-primaria** e dunque potenzialmente a rischio di “**deriva tecnologica**”. Se è vero infatti che la percentuale di iscritti alle Scuole dell'Infanzia - una fase del percorso da considerare sempre più centrale per la crescita dei minori - è, in Italia, notevolmente aumentata attestandosi al 91%, è innegabile che la distanza rispetto al nuovo obiettivo indicato dall'Unione Europea (96%) sia ancora da colmare [14]. Ma il disallineamento è ancora più rilevante

sull'**offerta di asili nido e di servizi 0-3-anni, ambito in direzione del quale, non casualmente, sono stati destinati importanti fondi del PNRR.**

## **Gli elementi di criticità e disagio che incrinano il progetto formativo**

Una tale presa d'atto è capace di richiamare caratteristiche e ragioni costitutive del cambiamento spingendo ad individuare gli **elementi di criticità e disagio che possono incrinare l'efficacia del progetto formativo**, fermo restando che **bambini provenienti da famiglie più attente e culturalmente dotate sembrano comunque manifestare una minore “dipendenza” dal sovradosaggio di device digitali.**

Sulla base di queste premesse, riflettiamo sui dati essenziali del consumo mediale relativo alla fascia 0-6 anni, richiamando i dati di sintesi sulla “dedizione” dei bambini (poi ragazzi e giovani) alle interfacce digitali, essenzialmente tablet e smartphone. La **dilatazione dei tempi di attenzione** è infatti radicalmente nuova e diversa anche rispetto alla fase storica in cui si parlava acriticamente di **teledipendenza** [15].

Peraltro, al tempo del festival della connettività, l'amplificazione percettiva che le tecnologie digitali regalano (non solo ai giovani), finisce per facilitare una vera e propria assuefazione ai climi culturali dominanti.

**Tutto ciò rischia di tradursi in un'arma di distrazione e desensibilizzazione del pensiero**, con il pericolo conseguente che tutti finiamo per diventare nient'altro che ciò che fa di noi la comunicazione. **Una nuova forma di “disabilitazione” degli individui**, in particolare quelli dentro il percorso educativo, che può tradursi in povertà formativa, e dunque in una brusca riduzione delle possibilità di partecipazione e cittadinanza.

### **Le radici del piacere: il prolungato contatto con il touch dei device**

Il problema più generale che dobbiamo porci è capire **le radici del piacere alla base del prolungato contatto dei ragazzi con i device digitali**. È plausibile attenderci che la **funzione psicologica principale** assolta dalla tecnologia sia **quella dell'entropia comunicativa e della centralità del di-vertimento**.

Se tutto questo diventa **totalizzante**, sfiorando la **dipendenza**, la **funzione della scuola risulta compromessa**.

Ma senza una presa d'atto delle ragioni della dedizione alle piattaforme non si riesce a restaurare un **canale formativo adeguato**.

Nella crescente moltiplicazione dei dispositivi comunicativi che consentono il collegamento in rete,

il ricorso a strumenti che utilizzano il **touchscreen** [16] ha preso il sopravvento rispetto alle tradizionali tastiere fisiche dei computer fissi o portatili.

**Già dal 2017 si segnala l'avvenuto sorpasso delle interfacce touch rispetto a quelle tradizionali**, protagoniste di una navigazione sempre più mobile (Audiweb Trends, a partire dal 2017 e successivi aggiornamenti) [17].

I dati, però, raccontano anche un'**urgenza conoscitiva non più rimandabile**, tutta inscritta nel **divario tra l'elevata penetrazione delle touch technologies tra giovani e giovanissimi e il bisogno di una ricognizione sistematica sull'utilizzo di tali strumenti nella fascia d'età 0-6**.

Anche il nostro **Istituto nazionale di statistica**, che negli anni ha aggiornato l'impianto metodologico delle sue rilevazioni sui comportamenti degli italiani rispetto alla tecnologia, a lungo ha concentrato l'**attenzione per smartphone e tablet a partire dai quattordicenni** [18].

Una certa disorganicità dei dati puntuali sulla penetrazione di queste tecnologie tra i giovanissimi rende ancor più preoccupante, tanto più dopo i **lunghi tempi della didattica a distanza (Dad)** [19], l'assenza di un framework teorico coerente e condiviso per quanto riguarda gli effetti dell'utilizzo dei touch screen tra i bambini fino a sei anni.

## Tessere di un mosaico da analizzare

**Singoli recenti esperimenti**, di impostazione quasi esclusivamente psicologica e laboratoriale, iniziano a comporre un **timido mosaico di risultati**. Manca, però, un'analisi complessiva dell'impatto sociale di queste nuove tecnologie su una fascia d'età così delicata dal punto di vista dello sviluppo cognitivo e comportamentale.

## L'impatto sociale del digitale sulla fascia d'età 0/6

Ecco perché diventa **necessaria e urgente una vertenza conoscitiva sul tema**.

Dobbiamo ricorrere ad un'espressione volutamente provocatoria parlando di bambini come "**soggetti sociali non identificati**". Rischiano di restare sostanzialmente sconosciuti, a partire dal fatto che non presentano le caratteristiche di organizzazione del discorso e sequenza logica tipiche (più o meno) degli adulti.

Senza pensare, inoltre, che in questi livelli di età non si è ancora compiutamente costituita quella risorsa di personalità che è la memoria. Occorre, pertanto, tematizzare l'incredibile **capacità di suggestione delle tecnologie che puntano sull'immagine e sullo schermo**, e sulla possibilità di **agire quasi immediatamente su queste forme della rappresentazione attraverso l'intervento diretto, ma soprattutto performativo, della mano**.

A maggior ragione per quanto riguarda la **fascia d'età prescolare**, una stagione in cui è possibile osservare **l'impatto dei media sullo sviluppo cognitivo, emotivo e relazionale del bambino, prima dell'incontro con gli insegnanti**, che pure è stato da secoli una specie di **cerimonia di apertura alla conoscenza**.

## Effetto Dad durante la pandemia

Proprio la scuola, durante la **crisi Covid**, ha rappresentato il **teatro di ingresso ufficiale delle tecnologie nella vita di molti bambini soprattutto della primaria**. Tutto questo è attestato

(anche) da una **ricerca condotta dalla Università Milano Bicocca** sulla fascia di età **1-11 anni**. Essa rileva che i bambini in età scolare hanno utilizzato di più i device digitali, anche se sono stati mediamente **impegnati (53,4%) fra le tre e le quattro ore con la Dad**.

**L'effetto 'stanchezza da digitale' non è soverchiante**. Anzi, si può forse ipotizzare una sorta di adattamento all'utilizzo dello strumento, che non viene più percepito come un qualcosa di inedito.

Per i **bambini al di sotto dei 5 anni**, invece, il tablet e altri device mobili sono stati presenti per **motivi di svago e per passare il tempo del confinamento casalingo**, in modo più intenso laddove i genitori possiedono un titolo di studio basso [20].

<https://www.agendadigitale.eu/scuola-digitale/dad-e-crisi-educativa-qualche-connessione-cosa-dicono-gli-studi/>

## Serve un programma di ricerca sistematico e plurale

Lo scenario delineato sollecita la necessità di un **programma di ricerca sistematico e plurale su questi temi** e sulle relative **implicazioni**, concentrato sull'analisi delle conseguenze in termini di socializzazione alla tecnologia.

Si tratta di un focus specifico e stringente, orientato a mettere a sistema una suggestione teorica di lungo corso che si sostanzia nella novità delle interfacce ospitate dai tanti dispositivi tecnologici che fanno compagnia ai bambini.

Dobbiamo tener conto che proprio in quella fascia di età cominciano ad esplicitarsi **nuove pulsioni identitarie**, con comportamenti che delineano una **percezione di sé sempre più anticipata e sostanziata da spinte anche aggressive**.

### La fatica della socializzazione

Questo profilo incerto che si intravede rende più complicata la successiva **"fatica della socializzazione"** coincidente con il **percorso scolastico** poiché il **sentiment dei bambini, iperstimolato dalla compagnia digitale, offre una gratificante sponda** anche espressiva alla grammatica dei comportamenti. Impossibile non annotare che il **sovradosaggio di stimoli medialti inerisce alle aree cerebrali del bambino**, alimentando **interessi completamente nuovi per quella fascia di età** e comunque manifestandosi in un'**intensità di relazione con i dispositivi che tutto è meno che una garanzia di formazione**. Il problema è che **sappiamo ancora troppo poco di loro**.

La possibile spiegazione è che le ricerche sono diacroniche, e le interpretazioni spesso "di seconda mano": ci si limita a confronti spontanei con la propria esperienza e a libere associazioni di idee osservando i nuovi venuti.

### L'anticipazione cognitiva

Bisogna invece capirli così come sono. Soltanto così si può intuire di più sulla funzione dei media. E tutto ciò implica anche **prospettare un superamento del media system**, in un nuovo contesto in cui la **frammentazione** sovrasta la **ricapitolazione** e dunque la **possibilità di riconoscere un sistema ordinato**.

Se non riflettiamo sui nuovi trend non si coglie l'aspetto potente di **“anticipazione cognitiva”** che essi portano con sé e che forse fa intravedere **il vero gadget di successo nelle scelte giovanili** (l'esempio più affine nel vecchio sistema mainstream è stato prima Italia 1 e poi MTV).

## La mappa dei cambiamenti

La mappa dei cambiamenti è tuttavia ben più ampia e attende di essere adeguatamente compilata. Si va da un **radicale mutamento della visione della vita ad una riclassificazione non meno rilevante della percezione dell'adulto, genitore e tanto più maestro.**

A ben vedere quest'aspetto è così significativo da ipotizzare addirittura una **derubricazione di centralità del suo valore nell'interazione quotidiana.**

Infatti è impossibile non mettere in dubbio l'ipotesi che siamo di fronte ad una minima quantità di socializzazione documentata dai **cambiamenti delle agende quotidiane, dai modelli di gestione del tempo, dall'ipertrofia dei consumi mediali, da uno slittamento significativo dei sostantivi di valore e da una pericolosa confusione di reale e virtuale, ma anche di vero e falso comunicativo fino alle fake**, che finisce per costruire una **black box di contenuti** in cui è impossibile capire cosa resterà.

Parliamo di **comfort zone perché la destrezza digitale induce empatia nella visione della loro personalità in costruzione.** Ma la verità è che comfort significa semplicemente un' **autonomia non assistita dalle figure di mediazione** [21].

## Chi è mancato nello scambio educativo

Sarebbe logico interrogarci su **“chi è mancato”** nell'esercizio dello scambio educativo e nell'interazione con i bambini. Ammettiamolo: è un quesito ingeneroso perché nessuno vive con chiarezza processi che non sono stati adeguatamente coltivati e annunciati dalla cultura e dalla ricerca.

Tuttavia è **certo che i genitori reagiscono troppo spesso con la distrazione e con la vaga speranza che le cose cambieranno all'arrivo della scuola.** Ma in realtà la stessa capacità di **gestualità affettiva è messa in discussione**, ben più che in passato, dall'immersione digitale che troppo spesso finisce col costruire un **muro divisorio con genitori e famiglia.**

In questo contesto il **ruolo del genitore sembra già “usurato” dal digitale**, perché il maestro può contare almeno sul fatto che **la cerimonia istituzionale della formazione continua a metterlo al centro dello sguardo.** Si tratta di **indizi empirici troppo poco fondati.**

E allora bisogna confessare che **tra quanti sono mancati debbono essere conteggiati gli studiosi, i ricercatori, gli esperti di formazione, ma anche di comunicazione.** Sono i primi a constatare quanto il ritmo del cambiamento è incalzante.

E qui l'interrogativo su chi non si è fatto adeguatamente sentire si confonde con un altro passaggio ineludibile di una **lista di responsabilità in cui nessuno può sentirsi del tutto innocente.**

## Cosa manca: l'autorità

Se dovessimo riassumere al massimo **cosa manca**, in una partita che si giochi ancora in termini di socializzazione, **il primo termine ad affacciarsi alla mente è quello di istituzione.** Ma per rendere

più preciso questo elenco occorre chiamare in causa il termine **autorità**, spesso addolcito da **autorevolezza**, il **senso del limite che diventa sempre più improbabile ed evanescente**, accanto ad ogni dimensione di tipo verticale. L'**insaziabile sete di interazioni in rete** costruisce e reitera una dimensione puramente orizzontale. Dove il confronto tra pari diventa una delle bussole più potenti della crescita.

Non sono banali gli interrogativi che poniamo. Se anche solo una parte delle tendenze delineate risultasse condivisa, la scossa etica e deontologica diventerebbe perentoria. Ma vedremo presto che non siamo certo gli unici ad aver intravisto le criticità dei processi di cambiamento.

## **Il sapere: condividere, non tramandare**

Si dovrebbe, per esempio, riconoscere che **sapere non è più tramandare**, ma essenzialmente **condividere**. Ed è così che si può prospettare una cancellazione culturale del passato e dei valori che sembrano faticare sempre più a sopravvivere ai nostri giorni.

Ecco perché, a questo punto, il discorso deve estendersi all'incipiente processo formativo. Infatti, non essendo facile intervenire in una fascia di età poco istituzionalizzata, l'interrogativo conseguente è **quanto il nostro mondo educativo ce la può fare a rimettere a posto le priorità di discorso**, sostenendo **la sfida della nuova complessità** e senza trascurare il **contesto di incertezza sociale e valoriale che respiriamo nel nostro tempo**.

## **Contrasto alla povertà educativa**

La necessità è quella di dominare e divulgare il più possibile la "letteratura grigia" e l'imponente produzione di dati, non sempre adeguatamente comunicati e tematizzati. Ciò significa un completamento dell'analisi e dell'individuazione delle fonti, anche comunitarie [22].

Fatta la debita citazione dei documenti istituzionali, ovviamente decisivi in una società democratica anche in ordine all'adozione dei conseguenti provvedimenti normativi, occorre dire che **sulla tematica della deprivazione educativa si è registrata una funzione di spinta e sostegno da parte del Terzo settore**. Esso ha esemplarmente funzionato su questo tema in una **logica** talora **di supplenza** ma più spesso di **sussidiarietà e coordinamento**, l'unica che può garantire **risultati significativi nel contrasto alla povertà formativa**.

## **Disuguaglianze**

Nell'ambito di quest'ampio catalogo di apporti viene avanzata poi una riflessione sulle **disuguaglianze educative** nella direzione del **digitale**, motivandola sulla base dell'analisi dei comportamenti di ragazzi e adolescenti italiani, e dunque alorizzando una seconda base empirica rispetto al dibattito precedente [23].

## **La povertà culturale**

Ancora una volta i dati mostrano dunque una chiara convergenza nella sottolineatura della **povertà culturale**, a partire dalle sue evidenze lungo il percorso formativo. La deprivazione educativa deve essere assunta come un **attacco frontale alla società e persino alla stessa digital transformation**.

Non vogliamo accontentarci di una **generazione che rischia di sentirsi esonerata dal progetto di superare i propri limiti**. Altrimenti finiamo per **incoraggiare una coincidenza tra i giovani a**

**scuola e quanti presto diventeranno i neet**, soggetti né in formazione né al lavoro, “portatori sani” di una vera e propria **disoccupazione mentale e valoriale**. Sotto questo profilo, la questione della povertà educativa rappresenta davvero un'intervista al mondo moderno. Si parte così da una consapevolezza più diffusa (soprattutto nell'universo femminile) che **la deprivazione aggrava possibili forme di disabilità**, dense di conseguenze sociali di lungo periodo.

## La necessità di coinvolgere gli adulti

C'è bisogno letteralmente di alimentare un **dibattito pubblico**. A partire dalle Istituzioni, bisogna coinvolgere gli **adulti** nel momento stesso in cui si capisce quanto la loro definizione derivi ironicamente da **adolescere**, spesso inclinando verso una troppo esile **assunzione di responsabilità**.

Infatti è necessario **riprogrammare il posizionamento delle varie parti dello scambio culturale intorno alla scuola**.

## Conclusioni

**Si è rotto un equilibrio secolare in forza di cui la complessità ineludibile della formazione era data per scontata**. Non come un “servizio militare obbligatorio”, ma aggiornamento di un **ideale forgiato da millenni**.

Il senso di quella storia era che **il soggetto sottoposto a stimoli e contenuti attentamente studiati e graduati, si avviava alla fatica della conoscenza** che, nel tempo, si sarebbe rivelata **fonte di autonomia e coinvolgimento**.

Non possiamo dimenticare **quanto Aristotele ci dice sul percorso di acquisizione dei saperi**: esso ha radici amare ma frutti dolci [24].

In momenti di passaggio come questo, **la scuola implicitamente finisce sotto attacco**, perché additata come **luogo di complessità**. E siamo allora indotti a riflettere su chi siamo noi, oggi, interrogandoci su dove stiamo andando e **se non rischiamo di buttar via istituzioni che hanno in sé la forza patrimoniale dei secoli**.

Anche sotto questa angolazione viene alla mente la memorabile frase degli **Annalisti francesi** che ci hanno ammonito: **“per essere, bisogna essere stati”**.

Meditiamo su quanto un tale motto ci insegna ad abbracciare anche la parte di fatica e iniziazione tipica di ogni processo istituzionale di apprendimento del sapere.

## Bibliografia

[1] Il riferimento di partenza è al Numero monografico della Rivista Paradoxa, “La comunicazione al posto della politica”, anno XIV, n. 3, luglio/settembre 2020, curato da me e da Michele Prospero, anch'egli dell'Università di Roma.

[2] Cfr. G. Anders, L'uomo è antiquato (vol. 1: Considerazioni sull'anima nell'epoca della Seconda Rivoluzione industriale), Il Saggiatore, Torino 1963, anno di edizione originale 1956.

[3] G. Paolucci, Cronofagia. La contrazione del tempo e dello spazio nell'era della globalizzazione, Guerini Studio, Milano 2003.

[4] Particolarmente significativo a questo proposito il monito rivolto ai giovani da Papa Francesco il 3 agosto 2023 durante la Giornata Mondiale della Gioventù a Lisbona: "Tanti oggi sanno il tuo nome, ma non ti chiamano per nome. Il tuo nome infatti è noto, appare sui social, viene elaborato da algoritmi che gli associano gusti e preferenze. Tutto questo però non interpella la tua unicità, ma la tua utilità per le indagini di mercato" ... "Sono le illusioni del virtuale e dobbiamo stare attenti a non lasciarci ingannare". Cfr. <https://www.rsi.ch/news/mondo/Papa-Attenti-ai-lupi-nei-social1-16443292.html>

[5] Per fare qualche esempio rappresentativo, rinvio a testi recenti di G. Lizza, Gli orizzonti della nuova geopolitica. Verso il 2050, Utet Università, Roma 2021 e G. Tremonti, Globalizzazione. Le piaghe e la cura possibile, Solferino, Milano 2022.

Non meno interessante il lavoro di A. Cazzullo, Metti via quel cellulare. Un papà, due figli. Una rivoluzione, Mondadori, Milano 2017. Sulla complessità e criticità di questo scenario segnalo inoltre il saggio di G. Mascheroni, "Datizzati alla nascita: a rischio i diritti delle nuove generazioni", in I Quaderni di Agenda Digitale, n.6, settembre-dicembre 2020, ma si veda anche, nello stesso Numero, l'interessante contributo di I. Ferrero, "Tik Tok, perché cattura così tanto i ragazzi: l'onboarding".

[6] In tale contesto, che chiama in causa anche una rinnovata "filosofia della libertà", non solo educativa, il pensiero corre alle teorizzazioni imprescindibili di I. Berlin e, in particolare, alla distinzione fra "libertà positiva" e libertà negativa". Si rimanda per questi temi al suo capitale Two concepts of liberty (prima pubblicazione 1958). Per la versione italiana si veda I. Berlin, Due concetti di libertà (traduzione di M. Santambrogio), Feltrinelli, Milano 2000.

[7] Cfr. M. Morcellini, "La relazione dinamica bambini/tecnologie. Una ricerca che ricomincia da tre", in Quando la tecnologia stressa la formazione, Comunicazioneepuntodoc, n. 18 (dicembre 2017).

[8] In particolare rinvio ad un Convegno monografico svoltosi presso l'Autorità per le Garanzie nelle Comunicazioni e a un intervento al Giffoni Film Festival, entrambi nello stesso anno di edizione della Rivista.

[9] Sull'intera tematica rinvio a M. Morcellini, Comunicazione e media, Egea, Milano 2013.

[10] Il passaggio è tratto dall'intervento introduttivo di Massimo Cacciari a un Convegno dell'Associazione Italiana Docenti Universitari (AIDU), organizzato il 7 maggio 2019 dal Presidente Alfonso Barbarisi presso l'Istituto "Luigi Sturzo" di Roma.

[11] Ho dedicato all'impatto tutt'altro che tramontato dell'educatore e sacerdote, a cento anni dalla sua nascita, la mia rubrica intitolata "Don Milani, prete maestro. Anche a distanza", in Formiche, n. 192 (giugno 2023).

[12] Qui allora il rinvio d'obbligo è alla Ricerca "Povertà educativa digitale" promossa da Conferenza dei Rettori delle Università Italiane, Ministero dell'Istruzione e del Merito e Ministero dell'Università e della Ricerca a partire dal 2022 e aderente alle indicazioni della Legge 178/2020, art. 1, commi 507, 508 e 509. Rimando qui al primo Dossier di

restituzione dati (“Povertà educativa e digitale. Una “Spada di Damocle” nella rigenerazione post-Covid”), da me coordinato e firmato, e consegnato al Ministero dell’Istruzione a fine 2022.

[13] Su questi temi rimando a M. Morcellini, *Passaggio al futuro. La socializzazione nell’età dei mass-media*, FrancoAngeli, Roma 1992.

[14] Su questi aspetti, si rinvia ai recenti e interessanti dati di Openpolis 2022 disponibili su [L'accesso all'istruzione pre-primaria in Italia e in Ue](#) - Openpolis.

[15] Cfr. M. Morcellini, *La TV fa bene ai bambini*, Meltemi Editore, Roma 2005.

[16] Questo trend è più articolatamente discusso nel mio editoriale premesso al monografico della rivista *Comunicazionepuntodoc* già citato, “La relazione dinamica bambini/tecnologie. Una ricerca che ricomincia da tre”.

[17] Audiweb Trends, Sintesi e analisi dei risultati della Ricerca di base sulla diffusione dell’online in Italia, varie edizioni a partire da giugno 2017, ma soprattutto gli aggiornamenti successivi.

[18] Istat, *Cittadini imprese e ICT*, varie edizioni annuali.

[19] In prima lettura, rinvio almeno a M. Morcellini, *Antivirus. Una società senza sistemi immunitari*, Castelvecchi, Roma 2020 (con prefazione di Maurizio Costanzo).

[20] Per il Report completo si veda S. Mantovani, M. Picca, P. Ferri, C. Bove, P. Manzoni, *Bambini e lockdown un anno dopo: la parola ai genitori*, Università degli Studi di Milano Bicocca, 2020.

[21] Ho discusso la stessa tematica in un saggio recente: M. Morcellini, “Giovani geneticamente modificati dall’infosfera.

La ricerca sociale come impegno di ascolto”, in *NPG (Note di pastorale giovanile)*, 57(2023), n. 6, pp. 19-28. Rimando inoltre ai saggi di C. Costa e del Direttore della Rivista G. De Nicolò, ma anche all’intero Fascicolo per la varietà e ricchezza dei contributi.

[22] Il primo riferimento da citare in proposito è certamente un testo del gennaio 2018 firmato dall’allora Miur, e specificamente dalla Cabina di regia per la lotta alla dispersione scolastica e alla povertà educativa, intitolato “Una politica nazionale di contrasto del fallimento formativo e della povertà educativa”. Il testo presenta numerosi vantaggi per chi intende valorizzare la conoscenza della deprivazione formativa, costituendosi così come fonte attendibile delle iniziative e ricerche (anche europee) fino al momento della chiusura di quel Rapporto.

[23] Un punto di forza è rappresentato certamente dall’ampia disponibilità di dati, spesso aggiornati, di fonte anzitutto istituzionale, ma anche provenienti da Enti che si sono specializzati sul tema, in particolare negli ultimi 15 anni.

Nell’insieme questa letteratura è rilevante anche per determinare una strategia operativa chiaramente aderente alle indicazioni della citata Legge 178/2020, art. 1, commi 507, 508 e 509.

[24] Rinvio qui a M. Morcellini, V. Martino, *Contro il declino dell’Università. Appunti e idee per una comunità che cambia*, Il Sole 24 Ore, Milano 2005.

# Come le allucinazioni di ChatGPT ci aiutano a capire la nostra idea di conoscenza

Le allucinazioni, gli errori grossolani che i sistemi di IA generativa producono nelle loro risposte, sono un problema di non facile soluzione. Molti ricercatori stanno provando a risolverlo, ma al di là delle specifiche tecniche e dei risultati raggiunti (o raggiungibili), quello che è interessante è l'idea di conoscenza che viene fuori da queste soluzioni

Di **Valeria Martino**, post-doc presso l'Università di Torino

Passati i primi mesi dal rilascio di [ChatGPT](#), grazie al quale i sistemi basati sui LLM sono diventati oggetto di discussione nell'opinione pubblica più vasta, passano anche i grandi entusiasmi e i grandi malumori e ci si può porre **qualche domanda** sul loro funzionamento con uno sguardo se non neutro, quantomeno un po' più disinteressato.

## LLM, il problema delle allucinazioni

Una questione molto rilevante che interessa questi sistemi è quella delle cosiddette **allucinazioni**, ovvero, ormai lo sappiamo, il fatto, incontestabile, che possono produrre errori macroscopici, in cui non cadrebbe nemmeno un bambino. Si parla di allucinazioni, infatti, proprio quando nel produrre un testo, il sistema inserisce errori grossolani, fatti falsi che tutti riconosciamo come tali – e questo succede soprattutto in alcuni ambiti, ad esempio le biografie di persone realmente esistite, ma anche mosse degli scacchi o semplici operazioni aritmetiche. Da diverse parti si è sottolineato come il termine “allucinazione” sia in realtà inappropriato (Juliani 2023). Infatti, allucinazione è un termine tecnico, riferito alla percezione che, in quanto tale, è inadatto a descrivere un sistema come ChatGPT, che letteralmente non percepisce alcunché. E questo sia che si voglia far riferimento alle pseudo-allucinazioni, cioè quelle che percepiamo nella consapevolezza che si tratta di allucinazioni (nella storia della filosofia troviamo l'esempio ricorrente del bastone immerso nell'acqua: lo *vediamo* come spezzato, *sappiamo* che non lo è, continuiamo a vederlo spezzato. Da qui, la discussione sul rapporto tra ragione e sensi nel perseguimento della conoscenza. Nel caso dell'allucinazione, vediamo una cosa o percepiamo un suono che sappiamo non essere reale, eppure continuiamo a vedere o sentire quel qualcosa); sia di quelle vere e proprie, che si rilevano in caso di assunzione di alcune droghe, per esempio, in cui l'allucinazione non viene nemmeno riconosciuta come tale. Un modo di giustificare l'uso del termine potrebbe essere proprio il riferimento allo scollamento con la realtà: viene prodotto qualcosa che viene esperito come vero pur non essendolo. Resta vero che si sarebbero potuti scegliere altri termini, come illusione (*delusion* invece di *hallucination*), o più semplicemente errore (perché di questo in fondo si tratta, di errori clamorosi) ma la scelta del nome cambia di poco la natura della questione o, quantomeno, non tocca i suoi aspetti più interessanti.

Molti esperti sono concordi nell'affermare che si tratterebbe di **un problema mai del tutto eliminabile**. Infatti, poiché i LLM si basano sulla predizione della parola successiva, potranno sempre sbagliare previsione, in particolare non essendo dotati di una comprensione contestuale della domanda che viene loro posta. Che si tratti di ricordarci la biografia di Napoleone o di

inventare una storia che lo abbia come protagonista per ChatGPT non fa alcuna differenza. E se per Aristotele, notoriamente, la poesia era superiore alla storia perché, essendo generale e applicandosi dunque a più situazioni simili anziché a una sola catena di eventi, ha un carattere di più ampio respiro e quindi ci può insegnare più cose, per ChatGPT (e simili, naturalmente) storia e poesia si equivalgono e “tratto da una storia vera”, “liberamente tratto da” o “qualunque riferimento a persone esistenti o a fatti realmente accaduti è puramente casuale” (per restare in ambito estetico-cinematografico) sembrano formule del tutto intercambiabili.

## Non possiamo assuefarci alle allucinazioni dell'IA

Eppure, ci sembra ovvio che non è così: se stiamo facendo una ricerca di storia vogliamo che la risposta ottenuta corrisponda ai fatti realmente accaduti; se vogliamo conoscere la trama di *Guerra e pace* non vogliamo conoscere la storia della campagna di Russia di Napoleone, ma vogliamo sapere cosa Tolstoj ne ha (davvero) scritto nel suo romanzo; se vogliamo scriverne uno noi, di romanzo, magari ci interesseranno la coerenza della storia e la correttezza grammaticale della prosa, ma nessun riferimento alla realtà ci interesserà (salvo che per evitare un'accusa di plagio). È da questa capacità di contestualizzare e di capire, di conseguenza, la rilevanza del riscontro tra fatti e testo scritto che dipende l'utilità stessa di questi strumenti. Soprattutto, quando si paventi la possibilità, come avviene, di sostituire i tradizionali motori di ricerca con sistemi predittivi.

In ogni caso, la possibilità stessa di conoscere le fonti delle notizie o delle informazioni reperite sembra una necessità ineludibile, se lo scopo non è ottenere testi sensati ma informazioni appunto.

<https://www.agendadigitale.eu/scuola-digitale/chatgpt-piu-educazione-contro-i-rischi-etici-cosi-evitiamo-il-pericolo-di-una-cultura-unica/>

## Cosa fare con una enciclopedia creativa?

Ma anche indipendentemente dalla questione della reperibilità delle fonti, **cosa potremmo farcene di una enciclopedia creativa** (ovvero di una raccolta di storie, biografie, calcoli, informazioni, in cui la realtà e la correttezza formale si mescolano all'invenzione e al verosimile fino al punto di perdere ogni riferimento ed essere noi stessi incapaci di distinguere il vero dal falso)? Laddove il maggior pericolo si ha non tanto quando si consultasse una simile enciclopedia per divertimento, ma la si compulsasse per imparare qualcosa, per avere le risposte corrette alle nostre domande. Come i bambini che chiedono ad Alexa di svolgere delle operazioni aritmetiche per terminare più in fretta i loro compiti.

Le risposte possibili alla domanda “che cosa possiamo farcene?” sono almeno due: possiamo **usarla per implementare la nostra creatività** (e quindi non a scopo informativo) – d'altronde GPT è in grado di emendare testi, produrre quiz, fare abstract o ridurli a partire da un testo, suggerire acronimi (ricordandoci, ne avessimo ancora bisogno, che inventare vuol dire conoscere il già esistente. Così nella retorica latina dove *inventio* stava appunto ad indicare la prima fase della produzione di un testo di retorica, quella del reperimento delle fonti utili). In questa direzione, dovremmo dimenticarci di utilizzare questi strumenti come dei sistemi di ricerca e trovare in altro la loro utilità. I LLM ci aiutano a svolgere piccoli ingrati compiti, tradizionalmente affidati a stagisti o altre figure agli inizi della propria carriera, e ci facilitano la vita. Noto risvolto della medaglia: velocizzando il lavoro delle persone più esperte e sostituendo di fatto quello delle meno esperte (Noy e Zhang 2023), si può facilmente prevedere la riduzione delle assunzioni nel settore dei cosiddetti colletti bianchi, con tutte le conseguenze sociopolitiche legate al lavoro e al salario che ne consegue.

Oppure, possiamo aspettarci di **emendare** una simile enciclopedia **a tal punto da rendere quasi nullo il problema delle allucinazioni** che, d'altronde, sembra legato a filo doppio con la questione della “overtrust”, cioè di una eccessiva, e ingiustificata, fiducia nello strumento tecnico che si traduce in una sospensione del giudizio, o meglio nella temporanea sospensione delle capacità critiche (un po' come quando da bambini ci fidiamo dei genitori o degli insegnanti e diamo per vero quello che ci dicono, salvo poi apprendere che le persone, anche le più dotte, possono sbagliarsi, e che le loro opinioni possono essere guidate, da pregiudizi, prese di posizione, dottrine, e via dicendo).

## Una questione di fiducia e capacità critica

E qui troviamo, nella **fiducia**, un'arma a doppio taglio. Infatti, se ho bisogno di una enciclopedia è perché mi fido che sia redatta secondo criteri di scientificità, proprio come mi fido di una mappa in scala 1: 10.000 perché di una mappa dell'Impero 1 : 1 non saprei che farmene. Senza fiducia, dunque, nessuna utilità. Questa fiducia, però, deve essere esercitata attivamente e, almeno in parte, giustificata.

Allora, abbiamo da un lato la necessità di **sviluppare capacità critica** e insieme capacità di prendere le giuste misure, dare il giusto peso a ciò che riceviamo come risposta a una nostra domanda (e questo lo si impara a poco a poco, formandosi, proprio come si impara a non credere sempre ai nostri adulti di riferimento da bambini o a mettere in dubbio quello che leggiamo e ascoltiamo; ma anche imparando a conoscere meglio gli strumenti che utilizziamo); e dall'altro lato, abbiamo dei tentativi di **ottimizzare le risposte** per renderle almeno meno inclini a cadere nelle allucinazioni.

## I metodi per ridurre le allucinazioni

E qui i metodi proposti (e in via di sviluppo e sperimentazione) sono i più diversi. Si parla di una **tecnica di rinforzo dell'apprendimento** grazie al feedback umano (gli utenti o persone appositamente reclutate per questa funzione aiutano i sistemi a migliorare le proprie risposte). E qui si apre ancora una volta l'enorme questione del microlavoro e dello sfruttamento dei lavoratori sul web di cui siamo ben a conoscenza, e che qui tralasciamo (una nota e brillante analisi si può trovare in Casilli 2019).

Un'ampia serie di lavori molto recenti si è concentrata in generale sul miglioramento dell'accuratezza dei fatti e del ragionamento nei modelli linguistici. Si va così dalla richiesta di **dimostrazioni di catene di pensieri basate su task a zero-shot** (si parla di zero-shot chain-of-thought), uso di verifiche, autoconsistenza o anche, banalmente, di inserire la richiesta di procedere passo per passo in un ragionamento. GPT 4, per esempio, è stato testato da Open AI (2023) con task di classificazione a zero-shot, ovvero senza allenamento precedente, e confrontata con lo svolgimento degli stessi test da parte di esseri umani e, sembra, aver ridotto di molto i problemi di allucinazione a cui erano esposti i modelli precedenti.

Manakul e altri studiosi di Cambridge (2023) hanno proposto un sistema da loro battezzato “**SelfCheckGPT**” basato sull'idea che, senza fare riferimento ad archivi esterni, i LLM sono in grado di riconoscere le proprie allucinazioni, in particolare quando forniscono risposte diverse o inconsistenti alla stessa domanda. Banalizzando, se alla stessa domanda (per esempio, dove è nato Napoleone?) do ogni volta una risposta diversa, posso inferirne di non sapere dove è nato Napoleone.

Interessante è la proposta avanzata a maggio da una serie di studiosi del MIT e di Google Brain (Du, Li, Torralba, Tenenbaum & Mordatch 2023), ispirata a *La società della mente* di Marvin Minsky (1986), e basata sull'idea che **per eliminare le allucinazioni si possono far comunicare tra di loro due diversi sistemi** (ponendo più volte la stessa domanda a ChatGPT o facendo dialogare GPT e Bard, per esempio) finché l'errore non viene eliminato. E questo accade soprattutto, sembra, perché gli elementi su cui i due attori del dialogo non sono d'accordo vengono eliminati dalla risposta conclusiva. A livello meramente teorico, ci si può chiedere se l'eliminazione sia una buona strategia o in quali casi possa esserlo. È ovvio, infatti, che procedendo in questo modo non viene eliminata solo l'informazione scorretta, ma anche quella corretta solo perché in contrasto con l'altra. Questo sistema, dunque, pur ottenendo il risultato dell'eliminazione delle allucinazioni rischia di impoverire drasticamente i risultati.

## Conclusioni

Al di là delle specifiche tecniche e dei risultati raggiunti (o raggiungibili), quello che è interessante (e che può essere spunto per una nostra riflessione, indipendentemente dalle tesi sostenute da chi lavora all'eliminazione delle allucinazioni e all'implementazione dell'affidabilità dei LLM) è **l'idea di conoscenza che viene fuori da queste soluzioni**, che da una parte ci riporta un po' alle origini mitiche del world wide web, come espressione di una grande intelligenza generale, fonte di verità o almeno di discussioni veridiche; dall'altro, può lasciarci perplessi, in un momento in cui ci sembrava una garanzia l'idea che la verità non viene dal dialogo – da cui semmai possiamo raggiungere la convivenza, che è una cosa ben diversa – ma da un confronto con la realtà che, resistendoci, ci ricorda che quante vittorie ha ottenuto Roger Federer a Wimbledon (questo l'esempio di Manakul) non dipende da cosa ne dica chicchessia ma da quante volte effettivamente Federer ha vinto a Wimbledon.

Esistono, infatti, questioni di fatto che appunto non dipendono dalle nostre opinioni e che possiamo semplicemente verificare esternamente e altre, che però hanno molto meno a che fare con le allucinazioni, per dirimere le quali il dialogo (insieme al dissidio) è e resterà un valido strumento, per esempio, domande come “per fare cosa è utile ChatGPT?”.

## Bibliografia

Casilli, A. (2019). *Schiavi del clic. Perché lavoriamo tutti per il nuovo capitalismo?*. Milano: Feltrinelli.

Chemero, A. (2023). LLMs differ from human cognition because they are not embodied. *Nat Hum Behav* 7, 1828-1829.

De Vynck, Gerrit (2023). ChatGPT ‘hallucinates.’ Some researchers worry it isn’t fixable. *Washington Post*, May 30, 2023.

Du, Yilun, Li, Shuang, Torralba, Antonio, Tenenbaum, Joshua B., and Igor Mordatch (2023). Improving Factuality and Reasoning in Language Models through Multiagent Debate. arXiv:2305.14325.

Juliani, A. (2023). Large Language Models Don't “Hallucinate”. *Better programming*, March 10.

Manakul, Potsawee, Liusie, Adian, and Mark J. F. Gales (2023). SelfCheckGPT: Zero-Resource Black-Box Hallucination Detection for Generative Large Language Models. arXiv:2303.08896

Noy, S. e Zhang, W. (2023, March 2). *Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence*. Working paper.  
[https://economics.mit.edu/sites/default/files/inline-files/Noy\\_Zhang\\_1.pdf](https://economics.mit.edu/sites/default/files/inline-files/Noy_Zhang_1.pdf)

Open AI (2023). GPT-4 System Card OpenAI. March 23, 2023. GPT-4 System Card OpenAI March 23, 2023.

Ouyang, L., Wu, J. et al. (2022). Training language models to follow instructions with human feedback. [arXiv:2203.02155](https://arxiv.org/abs/2203.02155).

Piantadosi, S.T., Hillar, F. (2022). Meaning without reference in large language models. Xiv:2208.02957



# Studi storici con l'IA: verso le Digital Humanities

Sistemi intelligenti per analizzare e “interpretare” le genealogie tra le fonti storiche, con formulazione di ipotesi basate su correlazioni che possono essere assunte come cause nelle ricostruzioni storiche. Ecco a che punto sono gli studi storici con l'Intelligenza Artificiale, dal Machine Learning al Deep Learning, fino a XAI

Di **Matteo Valleriani**, Research Group Leader in Dept. I, Honorary Professor at the Technische Universität Berlin

Innumerevoli **progetti di ricerca**, anche di notevoli dimensioni e a tutti i livelli istituzionali, hanno collezionato **banche dati** molto ricche. Benché non si sia ancora raggiunto il momento in cui si possa parlare di **Big Data nelle Digital Humanities**, le basi di sapere sviluppate fino ad ora nel **campo delle discipline umanistiche** rappresentano una nuova sfida per coloro che intendono analizzarle allo scopo di effettuare **ricostruzioni ed argomentazioni storiche**. Si tratta in effetti di **dataset** caratterizzati da molte dimensioni, cioè da un numero molto alto di tipi di dati connessi gli uni agli altri in molteplici modi, attraverso un grande collezione di attributi utilizzati per definire ogni singolo **data point**.

Un importante sviluppo avvenuto nell'ambito delle Digital Humanities, per rendere questi dataset più accessibili alla ricerca, concerne tecnologie realizzate allo scopo di **visualizzare e dunque esplorare i dati acquisiti**. Molto spesso sono **tecniche di visualizzazione geo-temporale** che permettono di acquisire un'**idea iniziale sull'andamento dei dati nel tempo e, se necessario, nello spazio**.

Esistono anche modalità di **visualizzazione più astratte**, allo scopo di esplorare le distribuzioni dei dati e giungere ad un'idea complessiva delle loro interconnessioni (Valleriani et al. 2023). Da qui, è stato breve il **passaggio ad un approccio matematico attraverso l'utilizzo di tecniche per l'analisi dei network**.

Ma facciamo una premessa, per introdurre il [progetto](#) di studi storici “The Sphere. Knowledge System Evolution and the Shared Scientific Identity of Europe”, svolto con l'Intelligenza Artificiale, per capire come **sistemi intelligenti in futuro permetteranno di analizzare e di “interpretare” le genealogie tra le fonti storiche, supportando così il lavoro dello storico attraverso ipotesi basate su correlazioni** che, vagliate, possono permettere le ricostruzioni storiche. Utilizzando **tecniche che spaziano dal Machine Learning al Deep Learning, fino a XAI**.

## Digital Humanities: verso i big data nella ricerca storica

La **quantificazione delle fonti storiche ancora inesplorate e preservate negli archivi e delle biblioteche** è sostanzialmente impossibile ma probabilmente ci si aggira **tra il 70 e l'80 per cento**.

<https://www.agendadigitale.eu/cultura-digitale/intelligenza-artificiale-arte-e-cultura-elementi-per-una-vera-valutazione-estetica/>

Anche prendendo in considerazione solamente le **fonti storiche** che sono state digitalizzate e messe a disposizione attraverso Internet durante gli ultimi vent'anni, il **numero di fonti disponibili per la ricerca storica eccede di gran lunga le capacità d'analisi degli storici professionisti** i quali utilizzano il classico approccio del “close reading”: lettura e annotazione.

La prima conseguenza del processo di digitalizzazione eseguito dalle istituzioni, atte alla preservazione della memoria storica, è stata la nascita della **disciplina Digital Humanities**. Attraverso i suoi più recenti sviluppi e soprattutto grazie alla diffusione delle tecnologie semantiche, siamo ora in grado di amministrare una grande quantità di fonti storiche, spesso organizzate in collezioni, e di dati, siano questi metadati di carattere bibliografico o dati estratti attraverso l'esame dei contenuti delle fonti stesse (Kräutli and Valleriani 2018; El-Hajj and Valleriani 2021).

## Computational History

L'introduzione della **Network Analysis nelle scienze storiche** ha rappresentato il primo passo verso la creazione di una nuova disciplina, fondata sulla Digital Humanities ma anche distinta da essa e che oggi viene sempre più spesso denominata **Computational History**.

Da diversi anni si è consolidata una comunità di storici specializzati nei più diversi campi di ricerca, ma che hanno in comune l'idea di analizzare i dataset applicando precetti e metodi della Network Analysis. Solitamente si tratta di **metodi utilizzati e sviluppati inizialmente nell'ambito della sociologia** e dunque di **Social Network Analysis**. Ma più recentemente e, a causa della multidimensionalità dei dataset, anche tecniche d'**analisi tipiche della fisica dei sistemi complessi** sono state implementate con successo.

Sono nate **riviste specializzate**, come per esempio il [Journal of Historical Network Research](#). Alcuni progetti, per essere in grado di raggiungere gli scopi storico-investigativi prefissati, continuano a sviluppare **nuove metodologie e approcci**, spostando quasi ininterrottamente le frontiere della Computational History.

### Il progetto

Il [progetto](#) “The Sphere. Knowledge System Evolution and the Shared Scientific Identity of Europe” mira allo studio dell'**evoluzione del sapere scientifico durante il Rinascimento**. In particolare si focalizza sulla **ricostruzione di meccanismi epistemologici di economia del sapere quali i processi di matematizzazione e di omogenizzazione del sapere scientifico**.

Le fonti storiche analizzate sono **i libri di testo usati in tutte le università europee tra circa la metà del XV secolo (1472, il primo anno in cui testi scientifico-matematici furono stampati) e il 1650** (Valleriani 2017; El-Hajj et al. 2022). La ricerca intende inoltre dimostrare come la **crescente omogenizzazione del sapere scientifico in questo periodo abbia rappresentato un fattore determinante per la creazione di una identità culturale europea durante la pre-modernità**.

Attraverso l'analisi di oltre **350 diverse edizioni**, per un ammontare di circa **76.000 pagine di materiale**, per un totale stimato di oltre **300.000 volumi circolanti nel territorio europeo**, è uno dei primi progetti di ricerca che ha **applicato l'approccio tipico della fisica dei sistemi complessi**.

Le edizioni sono di **libri di testo introduttivi allo studio dell'astronomia geocentrica**, solitamente composti da vari trattati scientifici di autori diversi. Gli studiosi hanno **selezionato tutte**

**edizioni contenenti** un testo molto usato a questo scopo e cioè **il Tractatus de sphaera di Johannes de Sacrobosco**.

Quest'analisi ha portato alla scoperta che **gli studi scientifici di alto livello divennero continuamente più omogenei soprattutto a partire dal 1532** e in seguito ad un **processo di imitazione** dei contenuti scientifici impartiti presso le università che si trovavano nei territori allora recentemente convertiti al **Protestantesimo**, ed in particolare all'**Università di Wittenberg** (Valleriani et al. 2019; Zamani et al. 2020; Valleriani et al. 2022a; Zamani et al. In press).

Altri studi, miranti all'analisi di correlazioni tra la trasformazione dei contenuti dei testi scientifici e il tessuto economico e sociale di attori quali autori e stampatori, hanno rivelato che tale processo di imitazione era dovuto principalmente, ma non solo, a motivi legati ai **meccanismi europei di produzione e marketing del libro di testo accademico di quel periodo** (Valleriani 2020; Valleriani and Ottone 2022; Valleriani et al. 2022b) [Figura 1].

Fig. 1

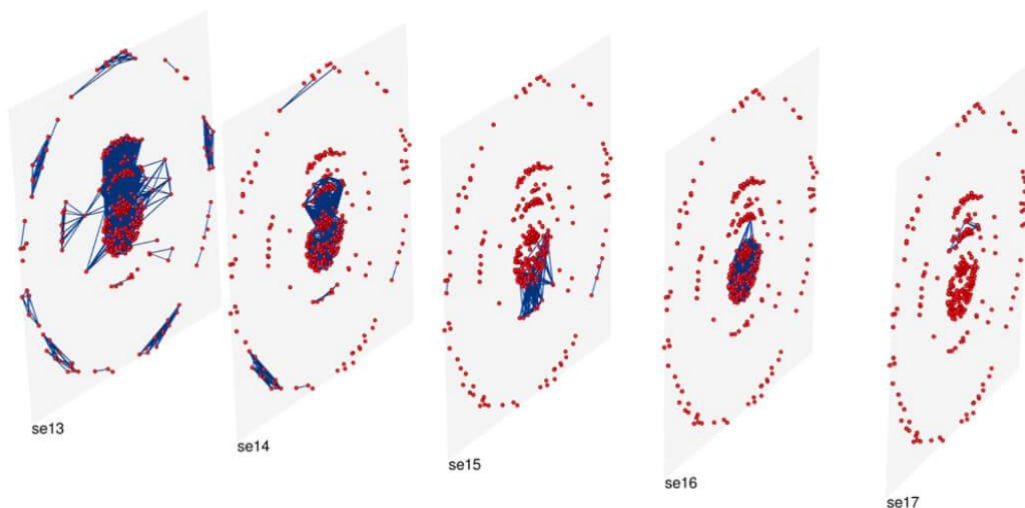


Figura 1. Esempio di applicazione della Network Analysis agli studi storici. Multiplex Network con cinque layers. Ogni layer corrisponde, nell'interpretazione finale, ad un'ipotesi storica. I nodi dei layers sono i libri di testo contenenti almeno il Tractatus de sphaera di Johannes de Sacrobosco. I network sono diretti e la direzione è basata sulla cronologia di pubblicazione dei testi. Nei cinque network i testi sono connessi quando: se13 – contengono lo stesso testo originale, se14 – contengono lo stesso commento ad un testo originale, se15 – contengono diverse traduzioni dello stesso testo, se16 – contengono commenti diversi allo stesso testo originale, se17 – contengono diversi commenti ad uno stesso commento o ad una stessa traduzione. Tratto da (Valleriani et al. 2019), nel quale è stato dimostrato che il contenuto dei libri di testo europei è diventato sempre più omogeneo a causa di un processo di imitazione dei libri di testo prodotti a Wittenberg a partire dal 1531. In particolare si tratta di testi originali e commenti riferentesi a loro (layers se13 e se14).

Visualizzazione effettuata utilizzando MuxViz (De Domenico et al. 2015).

## Metodo di studio

Questi studi hanno richiesto l'utilizzo di dati riferiti esclusivamente all'apparato testuale delle fonti. In pratica, gli studiosi prima hanno sezionato i volumi in porzioni di testo (una sorta di autopsia) e poi hanno fissato relazioni di tipo semantico tra queste porzioni: “basato su”, “commento di”, “frammento di” e “traduzione di”.

Il network sincronico così generato ha acquisito poi una profondità temporale (network diacronico) in base alle ri-occorrenze delle porzioni di testo identificate e dunque delle loro reciproche relazioni semantiche.

Oltre ai dati provenienti dall'esame dei testi, gli studiosi hanno inoltre effettuato l'estrazione di tutti gli elementi visivi così come delle tabelle numeriche ed alfanumeriche. In questo modo tre tipi di dati (denominati *knowledge atoms*) si definiscono come rappresentativi dei contenuti delle edizioni: testo, immagine, tabella [Figura 2a–c]. La conclusione dell'esame completo dei contenuti delle fonti avverrà dunque solo quando si stabilirà la relazione fra tutti i dati, tra gli uni e gli altri, e si svolgerà l'analisi del network risultante.

CAPITVLVM PRIMVM  
PROSDOCIMVS DE BELDOMANDO PATAVINVS  
ASTRONOMIÆ STVDIOSIS. S.P.D.

**B**onum omne quanto cōmunius, tanto melius. **C**Hac conclusionē studiosi iuvenes in meo adduxi proposito. Vt si in hoc onere per uos mihi imposto uestigia aliorum naturales libros exponētium, prorsus non obseruauerim, Aliquam admirationem non assumatis, intentio utiq; mea est cōmentariolum ædere non solum philosophis naturalibus sed quibuslibet literarum amatoribus

**C**Sphæræ mundi tractatus fœliciter inchoat.

**C**Novitiis A dolescētibus ad astronomicā temp. ca pessendā aditum impetrātibus: pro breui rectoq; tramite a uulgari uestigio semoto: Ioannis de sacro buso sphæricum opusculum. **C**Contraq; cremonensia in planetarum theoricas de Iyrameta: Ioannis de monte regio disputationes tam accuratissime q̄ utilissime fœliciter inchoat.



**R**ACTATVM de sphæra quattuor capitulis distinguimus.

**C**Dicturi primo compositionē sphæræ: quid sit sphæra: quid eius centrum: quid axis sphæræ: quid sit polus mundi: quot sint sphæræ: & quæ sit forma mundi.

**C**In secūdo de circulis ex quibus sphæra materialis componitur: & illa supercælestis (quæ per istam imaginatur) componi intelligitur. **C**In tertio de ortu & occasu signorum: & de diuersitate dierum & noctium quæ fit habitantibus in diuersis locis: & de diuisione climatum. **C**In quarto de circulis & motibus planetarū: & de causis eclipsum.

Diffinitio sphæræ. Capitulum primum.



**S**phæra igitur ab Euclide sic describitur. Sphæra est transitus circumferentiæ dimidii circuli: quæ totiēs (fixa diametro) quo usq; ad locum suum redeat circūducitur. **C**Idest sphæra est tale rotundum & solidum quod describitur ab Arcu semicirculi circumducto.



intendit. In secūda uero parte nobis offēdit quid postea in unaquaq; harū partiū sit facturus ibi secūda. **D**ICTVRI PRIMO &c. De prima parte dicit autor q̄ ipse intendit diuidere hunc suum tractatum de sphæra i quattuor capla. **C**Circa hāc ptē notādū priō q̄ notāter autor hūc suū libellū tractatū noiauit

Sphæ.

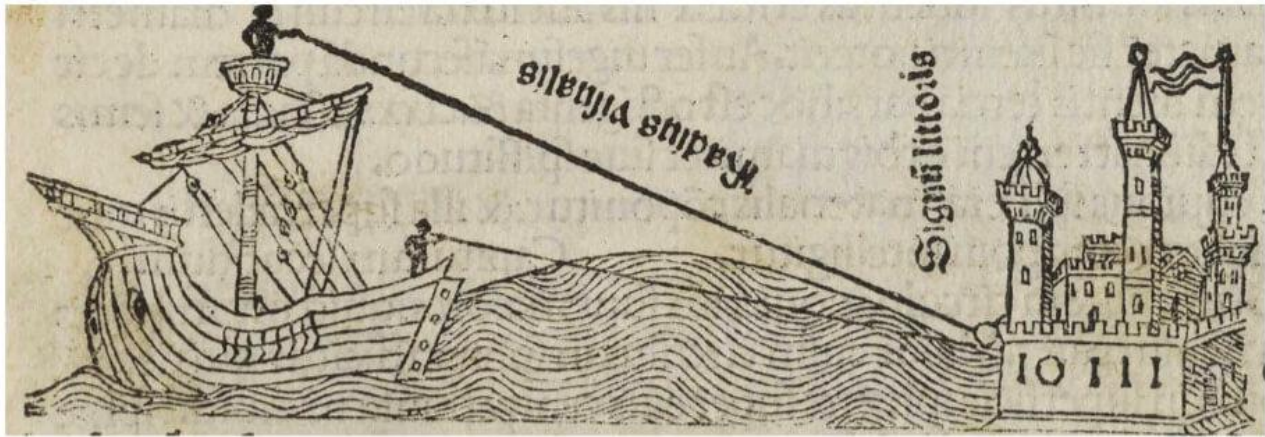
b

frugiferum, Et quia luce clarius uideo. **T**ractatum hunc de sphæra (quem in præsentiarum pro mea erga uos beniuolentia exponere intendo) non solum poetis sed disciplinarum omnium amatoribus fore satis necessarium, Deberet enim esse omnibus Enchiridiōn, ut cognita eius doctrina, uerior homo efficiatur. Visum est mihi bonum atq; utile in huius tractatuli expositione stilum capere nō solum Astrologiæ amatoribus atq; philosophis naturalibus sed omnibus afferentē magnam uoluptatem atq; utilitatem, ut inde opus efficiatur cōmunius ac melius. Et licet naturalium philosophorum mos in exponendis libris sit proponere et determinare quæstiones ordinate, primo ad partem arguēdo affirmatiuam, deinde eas sigillatim per titulos distinguendo, postmodum notabilia et cōclusiones cum suis probationibus pro una partium apponendo et tandem ad argumenta pro alia parte faciētia respondendo. Quā modus iste procedendi uidetur admodū difficilis prolixus atq; perplexus præcipue logicam ignorantibus: eo q̄ in huiusmodi quæstionibus prolongantur ea quæ paucissimis uerbis determinari possent. Nihilominus de cetero in processu huius operis dubia solum modo circa textum literæ orientia declarare, ut quilibet adolescens ea possit faciliter intelligere. Ad onus itaq; per uos mihi impostū (quod difficillimū quod uobis denegare nec uolo nec possum) deueniens. Annotabo literæ textum qui sic incipit. **C**Tractatum de sphæra quattuor capitulis distinguimus.

**C**Commētum sphæræ per Prosdocimum de Beldomādo patricium Patavinum diuinae Matheos professorem clarissimum.

**T**RACTATVM de sphæra &c. Quibusdam prolixitatibus in principiis librorum cōmuniter quæri solitis prætermisissis uenio ad huius libri diuisionē. Vnde liber iste primaria sua diuisione in duas diuiditur partes. scilicet in prohemium et tractatū. Tractatus incipit ibi. **S**PHÆRA IGITVR &c. Prima pars. scilicet prohemium adhuc in duas diuiditur partes: quā in prima parte autor nobis offēdit in quot ptes principales hūc suū tractatū diuidere

Fig. 2b



TABVLA SINVM RECTORVM.

| Gra | 0   |    |    | 1   |    |    | 2   |    |    | 3   |    |    | 4   |    |    | 5   |    |    | 6   |    |    | 7   |    |    | 8   |    |    |
|-----|-----|----|----|-----|----|----|-----|----|----|-----|----|----|-----|----|----|-----|----|----|-----|----|----|-----|----|----|-----|----|----|
|     | pt. | m. | i. | pt. | m. | i. | pt. | m. | i. | pt. | m. | i. | pt. | m. | i. | pt. | m. | i. | pt. | m. | i. | pt. | m. | i. | pt. | m. | i. |
| 0   | 0   | 0  | 0  | 1   | 2  | 50 | 2   | 5  | 38 | 3   | 8  | 25 | 4   | 11 | 7  | 5   | 13 | 46 | 6   | 16 | 18 | 7   | 18 | 44 | 0   | 21 | 1  |
| 1   | 0   | 1  | 3  |     | 3  | 53 |     | 6  | 41 |     | 9  | 27 |     | 12 | 10 |     | 14 | 48 |     | 17 | 21 |     | 19 | 46 |     | 22 | 4  |
| 2   |     | 2  | 6  |     | 4  | 55 |     | 7  | 44 |     | 10 | 30 |     | 13 | 13 |     | 15 | 51 |     | 18 | 23 |     | 20 | 49 |     | 23 | 0  |
| 3   |     | 3  | 9  |     | 5  | 58 |     | 8  | 47 |     | 11 | 33 |     | 14 | 15 |     | 16 | 53 |     | 19 | 26 |     | 21 | 51 |     | 24 | 8  |
| 4   |     | 4  | 11 |     | 7  | 1  |     | 9  | 49 |     | 12 | 35 |     | 15 | 18 |     | 17 | 56 |     | 20 | 28 |     | 22 | 53 |     | 25 | 10 |
| 5   |     | 5  | 14 |     | 8  | 4  |     | 10 | 52 |     | 13 | 38 |     | 16 | 21 |     | 18 | 59 |     | 21 | 31 |     | 23 | 56 |     | 26 | 12 |
| 6   |     | 6  | 17 |     | 9  | 7  |     | 11 | 54 |     | 14 | 41 |     | 17 | 24 |     | 20 | 1  |     | 22 | 33 |     | 24 | 58 |     | 27 | 15 |
| 7   |     | 7  | 20 |     | 10 | 9  |     | 12 | 57 |     | 15 | 44 |     | 18 | 26 |     | 21 | 4  |     | 23 | 36 |     | 26 | 0  |     | 28 | 17 |
| 8   |     | 8  | 23 |     | 11 | 12 |     | 14 | 0  |     | 16 | 46 |     | 19 | 29 |     | 22 | 6  |     | 24 | 38 |     | 27 | 3  |     | 29 | 19 |
| 9   |     | 9  | 26 |     | 12 | 15 |     | 15 | 3  |     | 17 | 49 |     | 20 | 32 |     | 23 | 9  |     | 25 | 40 |     | 28 | 5  |     | 30 | 21 |
| 10  |     | 10 | 28 |     | 13 | 18 |     | 16 | 6  |     | 18 | 52 |     | 21 | 34 |     | 24 | 12 |     | 26 | 43 |     | 29 | 7  |     | 31 | 23 |
| 11  |     | 11 | 31 |     | 14 | 21 |     | 17 | 8  |     | 19 | 55 |     | 22 | 37 |     | 25 | 14 |     | 27 | 45 |     | 30 | 10 |     | 32 | 26 |
| 12  |     | 12 | 34 |     | 15 | 24 |     | 18 | 11 |     | 20 | 57 |     | 23 | 40 |     | 26 | 17 |     | 28 | 48 |     | 31 | 12 |     | 33 | 28 |
| 13  |     | 13 | 37 |     | 16 | 26 |     | 19 | 14 |     | 22 | 0  |     | 24 | 42 |     | 27 | 19 |     | 29 | 50 |     | 32 | 14 |     | 34 | 30 |
| 14  |     | 14 | 40 |     | 17 | 29 |     | 20 | 17 |     | 23 | 3  |     | 25 | 45 |     | 28 | 22 |     | 30 | 53 |     | 33 | 17 |     | 35 | 32 |
| 15  |     | 15 | 42 |     | 18 | 32 |     | 21 | 20 |     | 24 | 6  |     | 26 | 47 |     | 29 | 24 |     | 31 | 55 |     | 34 | 19 |     | 36 | 34 |
| 16  |     | 16 | 45 |     | 19 | 35 |     | 22 | 23 |     | 25 | 8  |     | 27 | 50 |     | 30 | 27 |     | 32 | 58 |     | 35 | 21 |     | 37 | 37 |
| 17  |     | 17 | 48 |     | 20 | 38 |     | 23 | 26 |     | 26 | 11 |     | 28 | 53 |     | 31 | 29 |     | 34 | 0  |     | 36 | 24 |     | 38 | 39 |
| 18  |     | 18 | 51 |     | 21 | 40 |     | 24 | 28 |     | 27 | 14 |     | 29 | 55 |     | 32 | 32 |     | 35 | 3  |     | 37 | 26 |     | 39 | 41 |
| 19  |     | 19 | 54 |     | 22 | 43 |     | 25 | 31 |     | 28 | 16 |     | 30 | 58 |     | 33 | 34 |     | 36 | 5  |     | 38 | 28 |     | 40 | 43 |
| 20  |     | 20 | 57 |     | 23 | 46 |     | 26 | 34 |     | 29 | 19 |     | 32 | 0  |     | 34 | 37 |     | 37 | 7  |     | 39 | 31 |     | 41 | 45 |
| 21  |     | 21 | 59 |     | 24 | 49 |     | 27 | 37 |     | 30 | 22 |     | 33 | 3  |     | 35 | 40 |     | 38 | 10 |     | 40 | 33 |     | 42 | 47 |
| 22  |     | 22 | 2  |     | 25 | 52 |     | 28 | 40 |     | 31 | 25 |     | 34 | 6  |     | 36 | 42 |     | 39 | 12 |     | 41 | 35 |     | 43 | 50 |
| 23  |     | 23 | 5  |     | 26 | 55 |     | 29 | 42 |     | 32 | 27 |     | 35 | 9  |     | 37 | 45 |     | 40 | 15 |     | 42 | 37 |     | 44 | 52 |
| 24  |     | 24 | 8  |     | 27 | 57 |     | 30 | 45 |     | 33 | 30 |     | 36 | 11 |     | 38 | 47 |     | 41 | 17 |     | 43 | 40 |     | 45 | 54 |
| 25  |     | 25 | 11 |     | 29 | 0  |     | 31 | 48 |     | 34 | 33 |     | 37 | 14 |     | 39 | 50 |     | 42 | 20 |     | 44 | 42 |     | 46 | 56 |
| 26  |     | 26 | 14 |     | 30 | 3  |     | 32 | 51 |     | 35 | 35 |     | 38 | 17 |     | 40 | 52 |     | 43 | 22 |     | 45 | 44 |     | 47 | 58 |
| 27  |     | 27 | 16 |     | 31 | 5  |     | 33 | 54 |     | 36 | 38 |     | 39 | 19 |     | 41 | 55 |     | 44 | 25 |     | 46 | 47 |     | 49 | 0  |
| 28  |     | 28 | 19 |     | 32 | 8  |     | 34 | 56 |     | 37 | 41 |     | 40 | 22 |     | 42 | 57 |     | 45 | 27 |     | 47 | 49 |     | 50 | 3  |
| 29  |     | 29 | 22 |     | 33 | 11 |     | 35 | 59 |     | 38 | 44 |     | 41 | 24 |     | 43 | 0  |     | 46 | 30 |     | 48 | 51 |     | 51 | 5  |
| 30  |     | 30 | 25 |     | 34 | 14 |     | 37 | 2  |     | 39 | 46 |     | 42 | 27 |     | 45 | 3  |     | 47 | 32 |     | 49 | 54 |     | 52 | 7  |
| 31  |     | 31 | 28 |     | 35 | 17 |     | 38 | 5  |     | 40 | 49 |     | 43 | 30 |     | 46 | 5  |     | 48 | 34 |     | 50 | 56 |     | 53 | 9  |
| 32  |     | 32 | 31 |     | 36 | 20 |     | 39 | 7  |     | 41 | 52 |     | 44 | 32 |     | 47 | 8  |     | 49 | 37 |     | 51 | 58 |     | 54 | 11 |
| 33  |     | 33 | 33 |     | 37 | 23 |     | 40 | 10 |     | 42 | 55 |     | 45 | 35 |     | 48 | 10 |     | 50 | 39 |     | 53 | 1  |     | 55 | 13 |
| 34  |     | 34 | 36 |     | 38 | 25 |     | 41 | 13 |     | 43 | 57 |     | 46 | 38 |     | 49 | 13 |     | 51 | 42 |     | 54 | 3  |     | 56 | 15 |
| 35  |     | 35 | 39 |     | 39 | 28 |     | 42 | 16 |     | 45 | 0  |     | 47 | 40 |     | 50 | 15 |     | 52 | 44 |     | 55 | 5  |     | 57 | 17 |
| 36  |     | 36 | 42 |     | 40 | 31 |     | 43 | 18 |     | 46 | 3  |     | 48 | 43 |     | 51 | 18 |     | 53 | 46 |     | 56 | 7  |     | 58 | 20 |
| 37  |     | 37 | 45 |     | 41 | 34 |     | 44 | 21 |     | 47 | 5  |     | 49 | 46 |     | 52 | 20 |     | 54 | 49 |     | 57 | 10 | 8   | 59 | 22 |
| 38  |     | 38 | 48 |     | 42 | 37 |     | 45 | 24 |     | 48 | 8  |     | 50 | 48 |     | 53 | 23 |     | 55 | 51 |     | 58 | 12 | 9   | 0  | 24 |
| 39  |     | 39 | 50 |     | 43 | 39 |     | 46 | 27 |     | 49 | 11 |     | 51 | 51 |     | 54 | 25 |     | 56 | 54 | 7   | 59 | 14 |     | 1  | 26 |
| 40  |     | 40 | 53 |     | 44 | 42 |     | 47 | 30 |     | 50 | 13 |     | 52 | 53 |     | 55 | 28 |     | 57 | 56 | 8   | 0  | 16 |     | 2  | 28 |
| 41  |     | 41 | 56 |     | 45 | 45 |     | 48 | 32 |     | 51 | 16 |     | 53 | 56 |     | 56 | 30 | 6   | 58 | 59 |     | 1  | 19 |     | 3  | 30 |
| 42  |     | 42 | 59 |     | 46 | 48 |     | 49 | 35 |     | 52 | 19 |     | 54 | 59 |     | 57 | 33 | 7   | 0  | 1  |     | 2  | 21 |     | 4  | 32 |
| 43  |     | 43 | 2  |     | 47 | 51 |     | 50 | 38 |     | 53 | 22 |     | 56 | 1  |     | 58 | 35 |     | 1  | 3  |     | 3  | 23 |     | 5  | 34 |
| 44  |     | 44 | 5  |     | 48 | 53 |     | 51 | 41 |     | 54 | 24 |     | 57 | 4  | 5   | 59 | 38 |     | 2  | 6  |     | 4  | 26 |     | 6  | 36 |
| 45  |     | 45 | 7  |     | 49 | 56 |     | 52 | 43 |     | 55 | 27 |     | 58 | 7  | 6   | 0  | 41 |     | 3  | 8  |     | 5  | 28 |     | 7  | 39 |
| 46  |     | 46 | 10 |     | 50 | 59 |     | 53 | 46 |     | 56 | 30 | 4   | 59 | 9  |     | 1  | 43 |     | 4  | 10 |     | 6  | 30 |     | 8  | 41 |
| 47  |     | 47 | 13 |     | 52 | 2  |     | 54 | 49 |     | 57 | 32 | 5   | 0  | 12 |     | 2  | 45 |     | 5  | 13 |     | 7  | 32 |     | 9  | 43 |
| 48  |     | 48 | 16 |     | 53 | 5  |     | 55 | 51 |     | 58 | 35 |     | 1  | 14 |     | 3  | 48 |     | 6  | 15 |     | 8  | 34 |     | 10 | 45 |
| 49  |     | 49 | 19 |     | 54 | 7  |     | 56 | 54 | 3   | 59 | 38 |     | 2  | 17 |     | 4  | 50 |     | 7  | 18 |     | 9  | 37 |     | 11 | 47 |
| 50  |     | 50 | 22 |     | 55 | 10 |     | 57 | 57 | 4   | 0  | 40 |     | 3  | 20 |     | 5  | 53 |     | 8  | 20 |     | 10 | 39 |     | 12 | 49 |
| 51  |     | 51 | 24 |     | 56 | 13 | 2   | 59 | 0  |     | 1  | 43 |     | 4  | 22 |     | 6  | 55 |     | 9  | 22 |     | 11 | 41 |     | 13 | 51 |
| 52  |     | 52 | 27 |     | 57 | 16 | 3   | 0  | 2  |     | 2  | 46 |     | 5  | 25 |     | 7  | 58 |     | 10 | 25 |     | 12 | 43 |     | 14 | 53 |
| 53  |     | 53 | 30 |     | 58 | 19 |     | 1  | 5  |     | 3  | 49 |     | 6  | 27 |     | 8  | 0  |     | 11 | 27 |     | 13 | 46 |     | 15 | 55 |
| 54  |     | 54 | 33 | 1   | 59 | 21 |     | 2  | 8  |     | 4  | 51 |     | 7  | 30 |     | 9  | 3  |     | 12 | 30 |     | 14 | 48 |     | 16 | 57 |
| 55  |     | 55 | 36 | 2   | 0  | 24 |     | 3  | 10 |     | 5  | 54 |     | 8  | 33 |     | 10 | 5  |     | 13 | 32 |     | 15 | 50 |     | 17 | 59 |
| 56  |     | 56 | 39 |     | 1  | 27 |     | 4  | 13 |     | 6  | 57 |     | 9  | 35 |     | 11 | 8  |     | 14 | 34 |     | 16 | 52 |     | 19 | 1  |
| 57  | 0   | 57 | 41 |     | 2  | 30 |     | 5  | 16 |     | 7  | 59 |     | 10 | 38 |     | 12 | 10 |     | 15 | 37 |     | 17 | 55 |     | 20 | 4  |
| 58  | 1   | 0  | 44 |     | 3  | 33 |     | 6  | 19 |     | 8  | 2  |     | 11 | 40 |     | 13 | 13 |     | 16 | 39 |     | 18 | 57 |     | 21 | 6  |
| 59  | 1   | 1  | 47 | 2   | 4  | 35 | 3   | 7  | 22 | 4   | 10 | 5  | 5   | 12 | 43 | 6   | 15 | 15 | 7   | 17 | 41 | 8   | 19 | 59 | 9   | 22 | 8  |
| 60  | 1   | 2  | 50 | 2   | 5  | 38 | 3   | 8  | 25 | 4   | 11 | 7  | 5   | 13 | 46 | 6   | 16 | 18 | 7   | 18 | 44 | 8   | 21 | 1  | 9   | 23 | 10 |

RECTAE. SVBTENSÆ.

Fig. 2. Knowledge atoms sono tipi di dati scelti per rappresentare il contenuto delle fonti storiche esaminate. Nel progetto The Sphere i knowledge atoms sono testi, elementi visivi e tabelle numeriche e alfanumeriche. Queste illustrazioni rappresentano esempi di knowledge atoms tratti dai libri di testo rinascimentali utilizzati per l'introduzione allo studio dell'astronomia:

2a) testo; a causa dell'impossibilità di rendere i testi rinascimentali machine readable, i testi sono stati scomposti in porzioni di testo, generalmente attraverso l'identificazione di titoli e sottotitoli; in questa pagina si notano il testo di riferimento di Sacrobosco e il commento stampato attorno ad esso in forma di parentesi. Si notano inoltre

tre elementi visivi, due iniziali e una figura geometrica (contenuto: commento di Prosdocimo Beldomandi al Tractatus de sphaera di Johannes de Sacrobosco). Tratto da (Sacrobosco et al. 1508) ;

2b) illustrazione scientifica (contenuto: dimostrazione della sfericità dell'elemento acqua nell'ambito della concezione geocentrica di Aristotele). Tratto da (Sacrobosco et al. 1490);

2c) tabella numerica (contenuto: tavola logaritmica con i valori della funzione del seno). Tratto da (Finé 1542) . Biblioteca del Max-Planck-Institut für Wissenschaftsgeschichte, Berlino.

## Machine Learning: gli studi storici con l'Intelligenza Artificiale

Il passo finale non è ancora stato raggiunto a causa del fatto che l'estrazione e il controllo dei dati richiede un'enorme quantità di risorse umane. Un forte limite per gli studi storici che utilizzano questo approccio.

A causa di questo problema, tuttavia, nei tempi più recenti il campo della Computational History si sta estendendo allo scopo di sviluppare **applicazioni e modelli specifici di Machine Learning che consentono di estrarre dati automaticamente**, di classificarli in base a tassonomie predefinite ed infine di analizzarli.

La possibilità di usufruire in modo immediato di architetture come per esempio VGG16 e YOLO ha permesso **l'implementazione di Machine Learning al livello di analisi delle fonti**, dove per analisi si intende l'estrazione di elementi fondamentali per capirne il contenuto e quindi nuovi dati. **Nel progetto The Sphere si esegue l'analisi di testi a stampa rinascimentali, impossibili da trasformare in testi leggibili dalla macchina**, se non attraverso un investimento di risorse umane oltre la ragionevolezza. Per questo motivo in questo progetto, così **come in ogni progetto di Computational History che si focalizza su fonti rinascimentali europee**, l'applicazione di **Machine Learning è sostanzialmente un'applicazione nel campo di Computer Vision**. Si utilizzano modelli per segmentare le immagini delle pagine dei trattati storici e, infine, per distinguere il testo dalle immagini.

<https://www.agendadigitale.eu/cultura-digitale/se-la-storia-si-impara-a-cielo-aperto-cosi-natura-musei-e-digitale-aumentano-lapprendimento/>

### L'uso del Deep Learning

Nel caso del progetto *The Sphere* è stato sviluppato **CorDeep**, un Convolutional Neural Network in grado di identificare ogni tipo di elemento visivo e di classificarlo in base ad una tassonomia generale che distingue tra decorazioni e vignette, marche di stampatori, iniziali, e illustrazioni scientifiche (Büttner et al. 2022) [Figura 3].

È ovviamente possibile specializzare il modello per compiere ulteriori distinzioni, per esempio per identificare automaticamente quali immagini rappresentano illustrazioni astratte, strumenti matematici, e macchine. Ecco come si è parzialmente condotto questo passo.

Come in tutti gli sviluppi di intelligenza artificiale (IA) che utilizzano la tecnologia Deep Learning però, è possibile raggiungere risultati ottimali solo se una grande quantità di dati è a disposizione. **CorDeep funziona con una precisione (mean Average Precision - mAP) che va oltre il 98% e**

sono state utilizzate circa 23.000 scansioni di pagine con elementi visivi così come altrettante pagine senza elementi visivi (circa 38.500 per il training e le restanti per la validazione). Per raggiungere la stessa precisione, effettuando una maggiore specializzazione del modello, servirebbero diverse migliaia di immagini per ogni nuova categoria.

L'estrazione e la classificazione generano dati in molteplici formati da inserire agevolmente nella banca dati di riferimento, associandoli quindi a tutti gli altri dati presenti e, soprattutto, alle fonti storiche da cui è avvenuta l'estrazione.

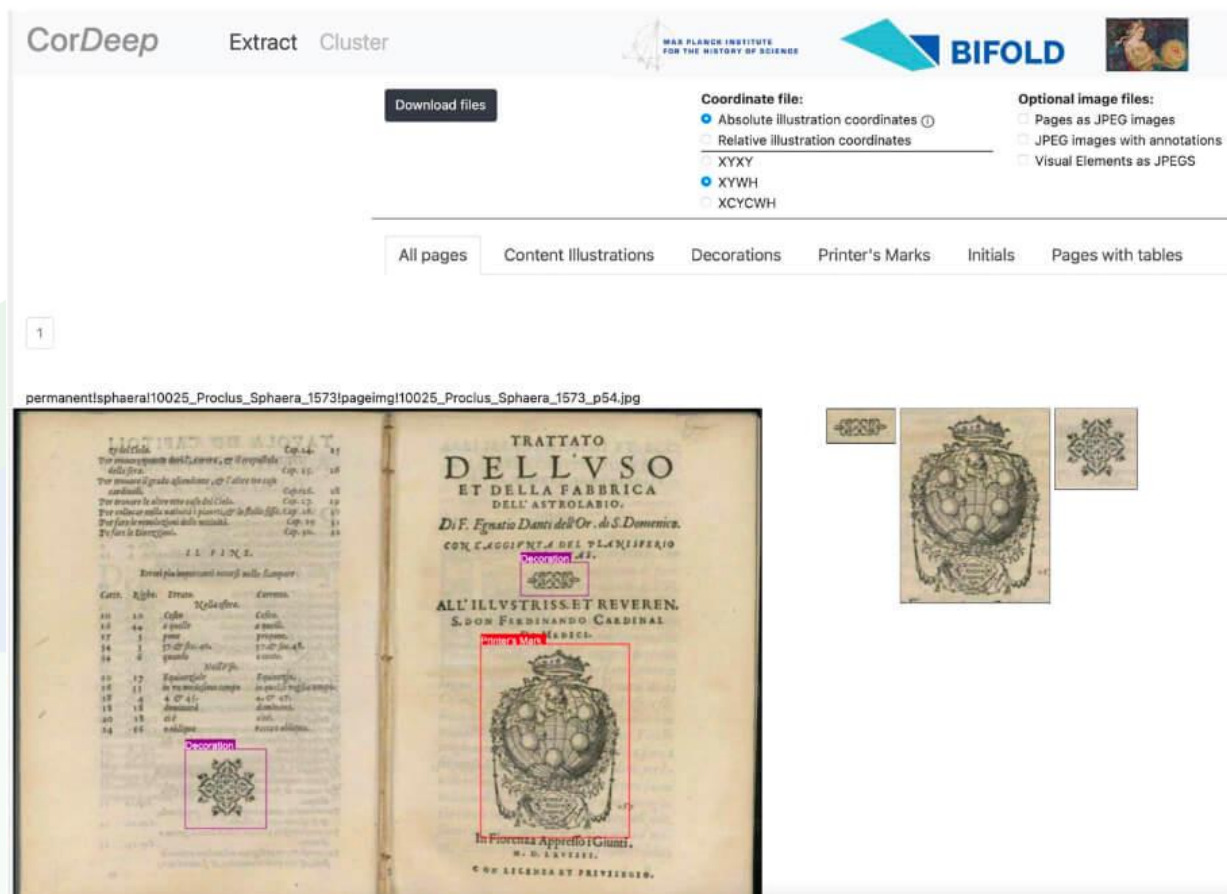


Figura 3. CorDeep (applicazione web). CorDeep permette di estrarre tutti gli elementi visivi dai testi rinascimentali e li classifica automaticamente tra illustrazioni di contenuto, iniziali, marche di stampatori e decorazioni e vignette. Il [codice](#) è a disposizione e può essere installato localmente per poter eseguire un nuovo training su fonti di tipo diverso (Büttner et al. 2022) .

## Network Analysis

Per proseguire le ricerche mantenendo un approccio collegato alla **Network Analysis** e per poter compiere un'esplorazione che riguarda migliaia di illustrazioni, è necessario raggruppare le immagini attraverso un principio di similitudine.

Nel caso del progetto The Sphere si è particolarmente interessati a scoprire quali stampatori hanno utilizzato le stesse strategie visive ovvero le stesse illustrazioni, per trasmettere lo stesso sapere scientifico.

Dunque si cercano illustrazioni scientifiche uguali tra loro e non solamente simili. **Anche in questo caso il Deep Learning ha soccorso la ricerca e grazie all'utilizzo di un'architettura ResNet50,**

raggruppando le illustrazioni in base a questo principio. In tal modo si crea un network tra fonti storiche in base al fatto che contengono la stessa illustrazione.

Questo network andrà infine ad aggiungersi a quelli già creati ed analizzati in base alle relazioni semantiche tra testi. Dunque, sarà possibile compiere ricerche in riferimento alla relazione testo e immagini, prendendo in considerazione una grande quantità di fonti storiche ed un lungo arco temporale.

Ulteriori sviluppi in corso nel progetto The Sphere e nel campo dell'Intelligenza Artificiale stanno conducendo alla produzione di un **modello in grado di identificare tabelle numeriche ed alfanumeriche**. Ma soprattutto in grado di raggrupparle in base alla similitudine dei loro contenuti.

Si tratta di un lavoro molto complesso dovuto all'eterogeneità dei dati a disposizione in quanto la stessa tabella numerica prevede un numero quasi illimitato di layout di paginazione. E, a causa delle diverse dimensioni dei testi a stampa rinascimentali, su una o molteplici pagine (Eberle et al. 2020).

## **Explainable Artificial Intelligence (XAI)**

I più recenti sviluppi in riferimento alla **explainability e interpretability** delineano **un'ulteriore linea di ricerca nel campo della Computational History e legata all'Intelligenza Artificiale**.

L'**Explainable Artificial Intelligence (XAI)** rappresenta un'emergente specializzazione in cui attualmente si investono grandi risorse. **Il metodo infatti promette di svelare i segreti della cosiddetta black box rappresentata dal Deep Learning e di comprendere perché - e come - il modello ha compiuto una previsione specifica.**

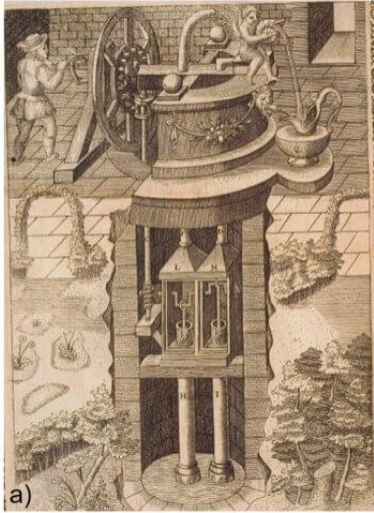
XAI è dunque una ricerca che **mira a risolvere problemi di tipo legale e giuridico ogni qual volta vi sono modelli di IA**, che possono agire in piena autonomia, come per esempio nel campo medico. XAI però è anche un approccio con un grande potenziale conoscitivo. Infatti **scoprire i motivi del comportamento di un modello IA può anche essere un modo per raccogliere nuove conoscenze sul materiale analizzato dal modello.**

Il progetto The Sphere ha permesso di compiere recentemente un primo passo verso **l'utilizzo di XAI nel campo degli studi storici** e allo scopo di imparare da IA come analizzare le fonti storiche.

Una volta preparato, il classificatore **CorDeep** consente infatti di distinguere quali illustrazioni rappresentano strumenti matematici rinascimentali, come per esempio un quadrante, un astrolabio o un orologio solare.

Validata la classificazione, l'applicazione di XAI ha permesso di comprendere quali sono gli elementi dell'illustrazione determinanti per il modello IA per compiere la stessa classificazione. In questo modo è possibile raccogliere un numero di informazioni sufficiente per formulare una definizione di strumento matematico rinascimentale così come interpretato dall'Intelligenza Artificiale e non dagli esseri umani (El-Hajj et al. Forthcoming) [Figura 4].

original image



a)

machine



b) predicted: machine

scientific illustration



c)

Figura 4. Applicazione di Explainable Artificial Intelligence (XAI). Dopo aver preparato un classificatore che distingue tra illustrazioni rinascimentali di macchine, strumenti matematici, e illustrazioni scientifiche, applicando la LRP (Layerwise relevance propagation) vengono generate heatmaps che permettono di comprendere quali aspetti sono stati determinanti per il modello per compiere la previsione effettuata (attivazioni rosse) e quali per non compierne altre (attivazioni blu). In questo caso si tratta di una cosiddetta local explanation, riferentesi ad ogni singola illustrazione (El-Hajj et al. Forthcoming). Nell'illustrazione è rappresentata una macchina per sollevare acque dal sottosuolo. Tratta da (Ramelli 1588). Biblioteca del Max-Planck-Institut für Wissenschaftsgeschichte, Berlino.

## Sviluppi futuri degli studi storici con l'Intelligenza Artificiale

Le possibilità di sviluppo all'intersezione tra gli studi storici e quelli rivolti allo sviluppo dell'IA sono estremamente vasti. Alcuni di questi trovano un uso corrente nell'ambito del progetto The Sphere.

In quasi tutti i casi si richiede lo sviluppo di **modelli di Deep Learning che funzionano in base ad un supervised learning**. Benché i modelli risultino estremamente efficienti, questa caratteristica però rappresenta anche un limite. Infatti la quantità di dati, la qualità della loro struttura, e il numero di annotazioni necessari per compiere il training richiesto, si basa spesso su un investimento in termini di risorse umane che solo in rari casi è possibile.

La ricerca si deve dunque muovere verso nuove frontiere che, nel caso di Deep Learning, siano in grado di imparare autonomamente in modo unsupervised, eventualmente attraverso la generazione di dati sintetici non reali. E soprattutto attraverso la realizzazione di modelli fondazionali applicabili in contesti diversi.

Questi modelli fondazionali devono essere in grado di analizzare collezioni di dati non strutturati per le quali sono a disposizione solo poche annotazioni. La situazione non è tipica solo degli studi storici. Infatti è sempre più comune in svariati campi. Ciò permetterebbe un'agilità e flessibilità d'applicazione della IA, attualmente non possibile. È prevedibile che **l'unione degli studi storici con la ricerca nel campo di IA aiuterà a spianare questa via**.

**In riferimento ai soli studi storici, l'unione apre a uno scenario futuro che si muove su due piani distinti. Il primo riguarda l'analisi delle fonti storiche e la creazione di genealogie tra loro. Scoprire in che modo le fonti storiche sono collegate semanticamente le une alle altre è la base di ogni ricerca storica. Fino ad ora questo compito richiedeva una laboriosa analisi manuale ad opera dei singoli professionisti. Questa analisi dunque si limita ad un numero molto piccolo di fonti, in confronto con quelle a disposizione.**

Grazie alla digitalizzazione delle fonti e al **trasferimento dei metadati dal cartaceo al catalogo elettronico**, l'IA può completamente eseguire questo compito. Le **biblioteche e gli archivi** che procederanno in questa direzione si arricchiranno dunque di un **nuovo compito ed una nuova missione**. Diventeranno **centri di dati già connessi e probabilmente esplorati nel loro andamento complessivo spazio temporale**.

Le ricerche storiche prenderanno il via da questi dati e non più dalle singole fonti. **Solo dopo aver scoperto particolarità nell'andamento dei dati, allora si tornerà a leggere individualmente le fonti, proprio per spiegare i cambiamenti**. In questo modo si ribalta il rapporto tra macro e micro storia. Sarà **un'analisi dell'andamento generale che fornirà idee per gli approfondimenti e non viceversa**, come è stato fatto fino ad ora, dove, partendo da studi specifici si cerca di connetterli costruendo una linea generale, **sostanzialmente collegando i punti con una linea**.

## Conclusioni

Nel progetto The Sphere, l'applicazione dei **modelli di estrazione e clustering** ha condotto alla **formazione di tre nuove collezioni di dati** che si riferiscono a **tre diverse collezioni di fonti storiche** nel corso di soli sei mesi e con l'impiego di due assistenti di progetto assunti part-time.

**La prima collezione ha invece richiesto 4 anni di lavoro e, in tutto, una decina di persone**. Il progetto The Sphere riuscirà a compiere un grande passo avanti quando riuscirà ad entrare in una solida alleanza con una grande istituzione per la preservazione delle fonti storiche.

Da un punto di vista più astratto, l'impostazione legata alla Network Analysis permette di concepire modelli di IA applicati ai network, come per esempio, **Graph Neural Networks (GNN) e Temporal Graph Neural Networks (TGNN)**. GNN, TGNN e il concomitante utilizzo di XAI lasciano già ora intravedere lo sviluppo di sistemi intelligenti in grado non solo di **analizzare ma anche di "interpretare" le genealogie tra le fonti storiche, supportando così il lavoro dello storico attraverso la formulazione di ipotesi basate su correlazioni** che, vagliate, possono essere assunte come cause nelle ricostruzioni storiche.

## Ringraziamenti

L'autore ringrazia [Hassan El-Hajj](#), IT Officer al Max-Planck-Institut für Wissenschaftsgeschichte di Berlino per i consigli elargiti su alcune parti dell'articolo.

## Bibliografia

Büttner, Jochen, Julius Martinetz, Hassan El-Hajj and Matteo Valleriani. 2022. CorDeep and the Sacrobosco Dataset: Detection of Visual Elements in Historical Documents. *Journal of Imaging* 8 (10): 285. <https://doi.org/https://doi.org/10.3390/jimaging8100285>

De Domenico, Manlio, Mason A. Porter and Alex Arenas. 2015. MuxViz: a Tool for Multilayer Analysis and Visualization of Networks. *Journal of Complex Networks* 3: 159–176.

Eberle, Oliver, Jochen Büttner, Florian Kräutli, Klaus-Robert Müller, Matteo Valleriani and Grégoire Montavon. 2020. Building and Interpreting Deep Similarity Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2020.3020738>.

El-Hajj, Hassan, Oliver Eberle, Anika Merklein, Anna Siebold, Noga Shlomi, Jochen Büttner, Julius Martinetz, Grégoire Montavon, Klaus-Robert Müller and Matteo Valleriani. Forthcoming. The Historian XAI: Explainability and Transparency in the Realm of Digital Humanities. *International Journal of Digital Humanities*.

El-Hajj, Hassan and Matteo Valleriani. 2021. CIDOC2VEC: Extracting Information from Atomized CIDOC-CRM Humanities Knowledge Graphs. *Information* 12: 1–18. <https://doi.org/https://doi.org/10.3390/info12120503>.

El-Hajj, Hassan, Maryam Zamani, Jochen Büttner, Julius Martinetz, Oliver Eberle, Noga Shlomi, Anna Siebold, Grégoire Montavon, Klaus-Robert Müller, Holger Kantz and Matteo Valleriani. 2022. An Ever-Expanding Humanities Knowledge Graph: The Sphaera Corpus at the Intersection of Humanities, Data Management, and Machine Learning. *Datenbank-Spektrum: Zeitschrift für Datenbanktechnologien und Information Retrieval*. <https://doi.org/10.1007/s13222-022-00414-1>.

Finé, Oronce. 1542. *Orontij Finei Delphinatis, regii mathematicarvm professoris, De Mundi sphaera, sive Cosmographia, primáve Astronomiae parte, Lib. V: Inaudita methodo ab authore renovati, proprijsque tum commentarijs & figuris, tum demonstrationibus & tabulis recens illustrati. Eivsdem Orontii, reclarum in circuli quadrante subtensarum (quos sinus vocant) demonstratio, supputatióque facillima, nunc primùm edita; vna cum eorundem sinuum tabula, fideli admodum calculo restituta. Eivsdem Orontii, organum vniversale, ex supradicta sinuum ratione contextum, quo tum Geometrici, tum omnes astro nomici canones, ex quatuor sinuum proportione pendentis, mira facilitate practicantur*. Oaris: Simon de Colines. <https://hdl.handle.net/21.11103/sphaera.101201>.

Kräutli, Florian and Matteo Valleriani. 2018. CorpusTracer: A CIDOC Database for Tracing Knowledge Networks. *Digital Scholarship in the Humanities* 33: 336–346. <https://doi.org/10.1093/lc/fqx047>.

Ramelli, Agostino. 1588. *Le diverse et artificiose machine del Capitano Agostino Ramelli dal Ponte Della Tresia Ingegniero del Christianissimo Re di Francia et di pollonia: nelle quali si contengono uarij et industriosi Mouimenti, degni digrandiβima speculatione, per cauarne beneficio infinito in ogni sorte d'operatione*. Parigi: In casa del Autore.

Sacrobosco, Johannes de, Pierre d'Ailly, Francesco Capuano, Robert Grosseteste, Jacques Lefèvre d'Étaples, Joannes Regiomontanus, Georg von Peurbach and Bartolomeo Vespucci. 1508. *Nota eorum quæ in hoc libro continentur. Oratio de laudibus astrologiae habita a Bartholomeo Vespucio florentino in almo Patavio Gymnasio anno.M.d.vi. TEXTUS SPHAERAE IOANNIS DE SACRO BUSTO. Expositio sphaerae Eximii artium & medicinae doctoris Domini Francisci Capuani de Manfredonia. Annotationes nonnullae eiusdem Bartholomei Vespucii hic ide itersertae. Iacobi frabri stapulensis Commentarii in eandem sphaeram. Reverendissimi Domini Petri de aliaco Cardinalis & episcopi Cameracensis in eandem quaestiones subtilissimae numero xiiii. Reverendissimi episcopi Domini Roberti linconiensis sphaerae compendium. Disputationes Ioannis de regio monte contra cremonensia deliramenta. Theoricarum novarum textus cum expositione*

*eiusdem Francisci Capuani omnia nuper diligentia summa emendata*. Venice: Giovanni and Bernardino Rosso for Giuntino Giunta. <https://hdl.handle.net/21.11103/sphaera.100915>.

Sacrobosco, Johannes de, Joannes Regiomontanus and Georg von Peurbach. 1490. *Sphaera mundi compendium foeliciter inchoat. Noviciis adolescentibus: ad astronomicam rem publicam capessendam aditum impetrantibus: pro brevi rectoque tramite a vulgari vestigio semoto: Ioannis de Sacro busto sphaericum opusculum una cum additionibus nonnullis littera A sparsim ubi intersertae sint signatis: Contraque cremonensia in planetarum theoricas delyramenta Ioannis de monte regio disputationes tam acuratiss. atque utilis. Nec non Georgii purbachii in erundem motus planetarum acuratiss. theoricae: dicatum opus: utili serie contextum: fausto sidere inchoat*. Venice: Ottaviano Scoto I. <https://hdl.handle.net/21.11103/sphaera.100885>.

Valleriani, Matteo. 2017. The Tracts on The Sphere. Knowledge Restructured over a Network. In *Structures of Practical Knowledge*, ed. Matteo Valleriani, 421–473. Dordrecht: Springer.

———, ed. 2020. De sphaera of Johannes de Sacrobosco in the Early Modern Period: The Authors of the Commentaries. Cham: Springer Nature. <https://doi.org/10.1007/978-3-030-30833-9>.

Valleriani, Matteo, Beate Federau and Olga Nicolaeva. 2022a. The Hidden Praeceptor: How Georg Rheticus Taught Geocentric Cosmology to Europe. *Perspectives in Science* 30 (3): 407–436. [https://doi.org/10.1162/posc\\_a\\_00421](https://doi.org/10.1162/posc_a_00421). [https://doi.org/10.1162/posc\\_a\\_00421](https://doi.org/10.1162/posc_a_00421).

Valleriani, Matteo, Florian Kräutli, Daan Lockhorst and Noga Shlomi. 2023. Vision on Vision: Defining Similarities Among Early Modern Illustrations on Cosmology. In *Scientific Visual Representations in History*, eds. Matteo Valleriani, Giulia Giannini and Enrico Giannetto, 99–137. Cham: Springer. [https://doi.org/10.1007/978-3-031-11317-8\\_4](https://doi.org/10.1007/978-3-031-11317-8_4).

Valleriani, Matteo, Florian Kräutli, Maryam Zamani, Alejandro Tejedor, Christoph Sander, Malte Vogl, Sabine Bertram, Gesa Funke and Holger Kantz. 2019. The Emergence of Epistemic Communities in the Sphaera Corpus: Mechanisms of Knowledge Evolution. *Journal of Historical Network Research* 3: 50–91. <https://doi.org/10.25517/jhnr.v3i1.63>.

Valleriani, Matteo and Andrea Ottone, eds. 2022. Publishing Sacrobosco's «De sphaera» in Early Modern Europe. Modes of Material and Scientific Exchange. Cham: Springer. <https://doi.org/10.1007/978-3-030-86600-6>.

Valleriani, Matteo, Malte Vogl, Hassan El-Hajj and Kim Pham. 2022b. The Network of Early Modern Printers and Its Impact on the Evolution of Scientific Knowledge: Automatic Detection of Awareness Relations. *Histories* 2 (4): 466–503. <https://doi.org/https://doi.org/10.3390/histories2040033>.

Zamani, Maryam, Hassan El-Hajj, Malte Vogl, Holger Kantz and Matteo Valleriani. In press. Modeling the Accumulation of Scientific Knowledge in the Early Modern Period. *Humanities & Social Sciences Communications*.

Zamani, Maryam, Alejandro Tejedor, Malte Vogl, Florian Kräutli, Matteo Valleriani and Holger Kantz. 2020. Evolution and Transformation of Early Modern Cosmological Knowledge: A Network Study. *Scientific Reports - Nature* 10: 19822. <https://doi.org/10.1038/s41598-020-76916-3>.

# La lunga estate calda della coscienza artificiale: nuovi studi

Due contributi, pubblicati da un filosofo della mente e della coscienza e da un gruppo di famosi neuroscienziati e studiosi di AI, hanno stimolato il dibattito sulla possibilità di costruire una IA cosciente. Vediamo a quali conclusioni sono giunti

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

L'estate del 2023 sarà ricordata per le alte temperature e per i numerosi incendi in Italia e nel mondo. Tuttavia, per gli studiosi della **coscienza artificiale** sarà ricordata anche per altri motivi, ossia per un deciso punto di svolta negli studi in questo campo.

Come è noto, il campo di ricerca sulla coscienza artificiale mira a indagare il fenomeno della coscienza alla luce della robotica e dei sistemi artificiali, della psicologia, della filosofia della mente, dell'etica e delle neuroscienze (Chella 2023). Alla svolta in questo settore hanno contribuito due eventi, due "incendi" nel panorama scientifico della coscienza artificiale, per riprendere il titolo del famoso film tratto dai racconti di William Faulkner.

## I due eventi che hanno incendiato il dibattito sulla coscienza artificiale

Il primo evento riguarda la pubblicazione di un contributo da parte di **David Chalmers**, un influente filosofo della mente e della coscienza. Il secondo evento è pubblicazione di un contributo da parte di un gruppo di famosi neuroscienziati e studiosi di AI, tra cui **Yoshua Bengio**, vincitore del prestigioso Turing Award. Entrambi i contributi si focalizzano sulla possibilità di costruire una IA cosciente, e, a differenza di altri contributi, non negano a priori questa possibilità ma la analizzano in maniera scientifica alla luce delle conoscenze attuali. Soprattutto, entrambi i contributi propongono costruttivamente una lista di proprietà strutturali che dovrebbe possedere una IA cosciente.

È necessario rimarcare che la possibilità di fruire liberamente di sistemi IA come ChatGPT abbia cambiato la **percezione comune** degli studi sulla coscienza artificiale. Prima di ChatGPT, il dibattito sulla coscienza di un sistema IA era relegato ai congressi e convegni specialistici, essenzialmente in ambito filosofico. Adesso, la possibilità di usare ChatGPT nelle applicazioni quotidiane ha posto con forza **il problema se un sistema di questo tipo possa essere cosciente** o possa diventare cosciente in un prossimo futuro.

A questo proposito si ricorda **il caso Lemoine** del giugno 2022. Blake Lemoine, un ingegnere del software di Google, è stato licenziato per aver asserito di riscontrare prove di coscienza nel sistema software LaMBDA 2, un sistema IA sviluppato da Google e simile a ChatGPT. Gli ingegneri di Google hanno provveduto a smentire categoricamente questa possibilità. Tuttavia, Lemoine aveva lungamente interagito con il sistema e un estratto, obiettivamente impressionante, delle sue

interazioni con LaMBDA 2 è riportato sul sito *Medium* e tradotto in italiano dal Corriere della Sera (Lemoine 2022).

Un ulteriore elemento di discussione che si aggiunge al caso Lemoine riguarda un articolo a firma di **alcuni ricercatori Microsoft** (Bubeck et al. 2023) pubblicato nella primavera del 2023 che presenta e commenta alcuni esperimenti condotti su ChatGPT. I ricercatori ipotizzano che CHatGPT può essere considerato un primo esempio di intelligenza artificiale generale (AGI – Artificial General Intelligence), visto che può essere utilizzato per risolvere problemi in diversi ambiti quali la matematica, il coding, la medicina, la legge, la psicologia, e altri. In particolare, i ricercatori hanno provato GPT-4, il sistema IA sottostante ChatGPT, chiedendo al sistema di scrivere una prova del teorema sui numeri primi in forma di poema, di disegnare un unicorno senza averlo mai visto, di creare una animazione complessa nel linguaggio di programmazione Python e di risolvere un problema matematico da scuola superiore. GPT-4 ha risposto in maniera soddisfacente in tutti questi compiti. Da qui l'ipotesi dei ricercatori Microsoft di trovarsi davanti ad un primo esempio di intelligenza artificiale generale.

## La “coscienza” di ChatGPT

David Chalmers, nel contributo pubblicato sul Boston Review il 9 agosto 2023 (Chalmers 2023) analizza ChatGPT alla luce delle conoscenze attuali sulla coscienza. In particolare, Chalmers analizza i possibili “indicatori di coscienza” per un sistema IA.

Inizialmente Chalmers analizza alcuni indicatori sufficienti per la coscienza del sistema IA. Se un sistema IA rispondesse positivamente a questi indicatori, potrebbe essere considerato cosciente.

Il primo indicatore analizzato da Chalmers è **la capacità di auto-rappresentarsi**, ossia di descrivere la propria vita interiore. Lemoine nella sua analisi aveva fatto riferimento proprio a questa capacità. Tuttavia, ChatGPT è stato addestrato su una enorme mole di dati, inclusi libri e romanzi, e pertanto è in grado di “simulare” una propria vita interiore sulla base delle conoscenze pregresse. Infatti, numerosi romanzi descrivono la vita interiore del protagonista, le proprie emozioni e pensieri, e quindi ChatGPT può facilmente simulare una propria vita interiore imitando la letteratura.

Gli altri indicatori individuati da Chalmers sono **la capacità di sembrare cosciente**, le capacità conversazionali, la capacità di avere una intelligenza artificiale generale. Queste capacità sono soltanto in parte e in maniera limitata presenti in ChatGPT, e possono essere facilmente spiegate in termini di imitazione, come nel caso della possibilità di auto-rappresentarsi. In merito alla capacità di possedere una intelligenza artificiale generale, sebbene i risultati preliminari siano incoraggianti, come detto sopra, ChatGPT è molto lontana dal livello dell'intelligenza umana.

Chalmers passa quindi ad analizzare **alcuni indicatori strutturali** necessari per un'entità cosciente, quali la necessità di un substrato biologico, la necessità di avere sensori ed un corpo, la capacità di costruire un modello del mondo e di sé stesso, **la capacità di elaborazioni ricorsive**, possedere una memoria di lavoro globale, ed avere una propria personalità.

Sebbene non tutte queste caratteristiche siano accettate nella comunità scientifica, come ad esempio la necessità di un substrato biologico o di avere un corpo, per essere coscienti, nessuna di queste caratteristiche è presente in ChatGPT, e quindi Chalmers conclude che **allo stato attuale ChatGPT non possiede né le caratteristiche necessarie, né quelle sufficienti per essere considerato un sistema IA cosciente.**

Il contributo di Chalmers ha un importante risvolto positivo che consiste nell'aver **tracciato una strada** proponendo un insieme di caratteristiche che devono essere analizzate perché un sistema IA possa essere considerato cosciente. In particolare, per Chalmers non basta che il sistema “sembri” cosciente, ma deve anche soddisfare delle caratteristiche strutturali. In particolare, la presenza di una propria personalità è una caratteristica particolarmente rilevante in ChatGPT, in quanto, essendo un sistema addestrato su una enorme quantità di dati, presenta una forte disomogeneità nelle risposte che dipende dal contesto e dal tipo di prompt.

## Coscienza e IA

A pochi giorni di distanza del contributo di Chalmers, il 17 agosto 2023, un gruppo di ricercatori guidati da Patrick Butlin e Robert Long, tra cui possiamo annoverare tra gli altri il già citato Yoshua Bengio, i neuroscienziati Ryota Kanai e Liad Mudrik e il filosofo Eric Schwitzgebel, hanno pubblicato **un importante contributo per gli studi sulla coscienza artificiale** (Butlin et al. 2023). Su X/Twitter diversi commentatori, discutendo il contributo, sono stati concordi nel rimarcare che la coscienza artificiale si avvia ad essere un tema di ricerca “mainstream” dell'Intelligenza Artificiale.

## La posizione del “funzionalismo computazionale”

Il contributo di Butlin et al., rimarca il fatto che lo studio della coscienza nei sistemi IA è un problema di ricerca scientificamente trattabile e che deve essere studiato utilizzando gli appropriati metodi scientifici. Dal punto di vista filosofico, gli autori assumono la posizione del “**funzionalismo computazionale**,” secondo cui la condizione necessaria e sufficiente perché un sistema sia cosciente è che il sistema stesso sia in grado di effettuare computazioni del giusto tipo. Semplificando, secondo questa posizione il cervello umano e il computer sono sistemi computazionali ma su un substrato diverso. Se il computer effettuasse il giusto tipo di computazioni, sarebbe cosciente quanto il cervello umano. Gli autori riconoscono che questa posizione è problematica ed infatti la presentano come una ipotesi di lavoro pragmatica.

In sintesi, lo studio di Butlin et al., amplia ed estende la strategia adottata da Chalmers nel contributo precedentemente descritto. Gli autori hanno analizzato le teorie neuroscientifiche più accreditate sulla coscienza ed hanno estratto **un insieme di indicatori di proprietà** che, se presenti in un sistema IA, indicherebbero la possibilità della coscienza nel sistema stesso.

Nel rivendicare l'approccio teorico strutturale, gli autori dello studio rimarcano il rischio dell'analisi basata sulle interazioni con il sistema IA, già messo in luce nel precedente studio di Chalmers: un sistema IA ben addestrato potrebbe facilmente imitare comportamenti coscienti.

## Le teorie scientifiche sulla coscienza

Le teorie scientifiche sulla coscienza analizzate nello studio sono: la teoria dell'elaborazione ricorsiva, lo spazio di lavoro globale, le teorie di ordine superiore, gli schemi di attenzione, l'elaborazione predittiva e la teoria degli agenti corporei. Si noti che alcune di queste teorie erano state già prese in considerazione da Chalmers.

Alcuni indicatori di proprietà individuati riguardano la capacità del sistema di integrare informazioni provenienti dai sensori, la capacità di avere una memoria di lavoro globale in grado di

diffondere informazioni a tutto il sistema, la capacità di avere un sistema di attenzione, la capacità di generare aspettative sul mondo esterno.

Gli autori hanno compiuto uno sforzo rilevante per definire una lista di indicatori di proprietà coerente e senza incongruenze e sovrapposizioni. La lista di indicatori di proprietà può ragionevolmente essere considerata **una sorta di manifesto della coscienza artificiale**.

Il contributo di Butlin et al., è molto importante nello studio della coscienza artificiale, in quanto propone un metodo, che consiste nel prendere ispirazione dalle teorie della coscienza delle neuroscienze ed estrarre da queste degli opportuni indicatori di proprietà. È utile rimarcare che la lista di indicatori di proprietà si riferisce alle proprietà strutturali di un sistema IA, e non al suo comportamento esterno, proprio per evitare l'effetto di imitazione tipico dei sistemi come ChatGPT.

Gli autori discutono quindi come alcuni di questi indicatori siano già presenti o possano essere implementati nei futuri sistemi IA. Tuttavia, a oggi, **nessun sistema soddisfa tutti gli indicatori e quindi nessun sistema, incluso ChatGPT, è un buon candidato ad essere un sistema IA cosciente**, in concordanza con quanto asserito da Chalmers.

Gli autori rimarcano che la lista di indicatori proposta è una lista aperta. Man mano che gli studi sulla coscienza avanzano, la lista potrà essere emendata e aggiornata.

## Conclusioni

Entrambi gli “incendi” scoppiati nell'estate del 2023 nel campo della coscienza artificiale hanno contribuito ad assestare questa disciplina, ormai riconosciuta quale componente “mainstream” dell'Intelligenza Artificiale.

Il grande merito è di avere definito **un metodo che prende spunto dalle teorie neuroscientifiche** sulla coscienza artificiale e avere definito una lista di indicatori di proprietà, una sorta di manifesto della coscienza artificiale.

Infine, è da rimarcare il fatto che entrambi i contributi pongono l'accento sulle problematiche etiche della coscienza artificiale. Chalmers evidenzia chiaramente le sfide che si aprono relativamente ai sistemi IA coscienti, che devono essere prese in considerazione attentamente nella progettazione di un sistema IA cosciente. Anche Butlin et al., concludono che un sistema di IA cosciente può sicuramente apportare benefici ma è necessario considerare in maniera attenta i possibili rischi.

## Bibliografia

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y.T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M.T., Zhang, Y. (2023): Sparks of Artificial General Intelligence: Early experiments with GPT-4, <https://doi.org/10.48550/arXiv.2303.12712>

Butlin, P., Long, R., Elmoznino, E., Bengio, Y., Birch, J., Constant, A., Deane, G., Fleming, S.M., Frith, C., Ji, X., Kanai, R., Klein, C., Lindsay, G., Michel, M., Mudrik, L., Peters, M.A.K., Schwitzgebel, E., Simon, J., VanRullen, R. (2023): Consciousness in Artificial Intelligence: Insights from the Science of Consciousness, <https://doi.org/10.48550/arXiv.2308.08708>

Chalmers, D. (2023): Could a Large Language Model Be Conscious? Boston Review,

<https://www.bostonreview.net/articles/could-a-large-language-model-be-conscious/>; si veda anche <https://doi.org/10.48550/arXiv.2303.07103>

Chella, A. (2023): Robot coscienti, realtà possibile o utopia? Cosa dicono gli studi, Agenda Digitale EU, <https://www.agendadigitale.eu/cultura-digitale/robot-coscienti-imitazione-emulazione/>

Lemoine, B. (2022): Is LaMDA Sentient? – an Interview, Medium, Jun 11, 2022 <https://cajundiscordian.medium.com/is-lambda-sentient-an-interview-ea64d916d917>; si veda anche: L'intervista con l'intelligenza artificiale LaMDA di Google, in italiano: è senziente? [https://www.corriere.it/tecnologia/22\\_giugno\\_14/lambda-google-italiano-medium-1baf7b5c-eb42-11ec-b89b-6b199698064a.shtml](https://www.corriere.it/tecnologia/22_giugno_14/lambda-google-italiano-medium-1baf7b5c-eb42-11ec-b89b-6b199698064a.shtml)



# L'IA nell'apprendimento: una storia genetica

L'interesse dei ricercatori nel campo dell'educazione e della formazione è molto precedente all'hype dell'intelligenza artificiale generativa. Proviamo a ripercorre, nei suoi passaggi fondamentali la storia dello sviluppo del campo di studi dell'Artificial Intelligence in Education

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

La storia che ci accingiamo a narrare è una storia lunga anche più di quella di Internet e del Web. Il primo protagonista è forse lo stesso **Alan Turing** che per primo si pose la domanda “può pensare una macchina?” (Turing, 1950, p. 435).

Restringiamo però qui il campo dell'indagine al **rapporto tra intelligenza artificiale ed apprendimento**.

## Il rapporto tra intelligenza artificiale ed apprendimento

Anche questa narrazione parte dalla metà del secolo scorso perché è proprio durante un workshop estivo che si tenne a Dartmouth nel New Hampshire, John McCarthy (il primo ad utilizzare il termine intelligenza artificiale), Marvin Minsky, Nathaniel Rochester e Claude Shannon, inaugurarono “formalmente” questo campo di studi con paper di 17 pagine noto come Proposal of Dartmouth.

“Penso che sia sostanzialmente una simulatrice del comportamento umano medio, e se è così, imiterà l'uomo anche nel mostrarsi sensibile alle minacce”. Primo Levi, Vizio di Forma, Einaudi

Torino 1971

«Lo studio procederà sulla base della congettura per cui, in linea di principio, ogni aspetto dell'apprendimento o una qualsiasi altra caratteristica dell'intelligenza possano essere descritte così precisamente da poter costruire una macchina che le simuli.» (McCarthy et Al, 1956 ora 2006, p. 1). Il programma di ricerca della “Intelligenza artificiale forte” è dunque quello di costruire una “macchina per insegnare” basata su regole logico-simboliche. Si tratta dello stesso paradigma di ricerca, l'istruzione “programmata” basata sulle macchine per insegnare che propone **Skinner**, nel 1968 (Skinner, 1968). **“Programmare” significa per Skinner dare più oggettività al sapere didattico e a quello sull'apprendimento.**

Quasi settanta anni dopo e quarant'anni dopo la confutazione da parte di Searle del paradigma dell'IA forte (Searle, 1980, 1990) gli studi sull'intelligenza artificiale riprendono il centro della scena anche nel campo dell'educazione, della scuola e dell'apprendimento, con un fondamentale mining shift. L'intelligenza artificiale contemporanea si fonda, infatti, su un approccio teorico opposto a quello di Turing: quello dell'intelligenze artificiale debole.

## Le macchine non pensano

Le macchine non pensano, non sono intenzionali, **ma possono emulare alcune funzioni biologiche del cervello**. È sul successo di questo differente approccio che si fondano, le innovazioni e le scoperte che hanno portato il programma di ricerca fondato da McCarthy e Minsky, nella sua versione “debole” a diffondersi e radicarsi in molti settori della vita economica, sociale e individuale contemporanea (Eugeni, 2021). Ne sono un esempio la **computer vision, gli algoritmi di riconoscimento testuale e vocale**, quelli “di raccomandazione” dei motori di ricerca e dei social network, gli algoritmi predittivi dei fenomeni (finanziari, medici, sociali, farmacologici), le applicazioni dell’intelligenza artificiale all’automazione dalla robotica, all’automotive, fino agli assistenti vocali degli smartphone (Eugeni 2021, Rivoltella, Panciroli, 2020, 2023,).

L’ultima e più eclatante evoluzione di questo paradigma sono i **Generative Pre-trained Transformer (GPT)** i chat bot per la generazione di testi ed immagini che si basano sulle reti neurali di ultima generazione sviluppate per la prima volta da un gruppo di ricerca promosso da Google nel 2017 (Vasany et al. 2017),<sup>[1]</sup>. Si tratta dei cosiddetti **Large Language Model** (Casola, S., Lauriola, I., Lavelli, A., (2022) un sottodominio dell’Intelligenza artificiale che utilizza la machine e il deep learning<sup>[2]</sup>. Questi modelli utilizzano enormi quantità di dati testuali per la generazione di risposte agli input testuali (prompt) forniti dagli utenti. Ad esempio, quando si formula una richiesta a ChatGPT 4, il suo “modello di conversazione” utilizza la sua rete neurale multistrato transformer<sup>[3]</sup> per elaborare, e analizzare il testo della domanda confrontandolo con le basi dati su cui è stato “pre-addestrato” per generare, poi, la sua risposta in linguaggio naturale. Esistono differenti tipologie di queste reti neurali. Insieme ai Recurrent Neural Network per il Natural Language Processing (Chat GPT 4, Bard, Stable ecc.) sono stati rilasciati anche i Convolutional Neural Network progettati per la classificazione e/o la generazione di immagini grazie alla loro capacità di riconoscere i pattern di relazioni spaziali tra i pixel dell’immagine.

## Per una storia genetica dell’Intelligenza artificiale

L’interesse dei ricercatori nel campo dell’educazione e della formazione è molto precedente all’hype dell’intelligenza artificiale generativa. Proviamo a ripercorre, nei suoi passaggi fondamentali **la storia dello sviluppo del campo di studi dell’Artificial Intelligence in Education**<sup>[4]</sup>, in particolare, Rivoltella e Panciroli (2023) ricostruiscono la storia di questo rapporto nella tabella che segue che riprendiamo e integriamo modificandola (Rivoltella Panciroli, 2023. p. 49)

|                    |  |  |
|--------------------|--|--|
| Fase 1 1945-1950   | Fase Meccanica   | Skinner Macchina per insegnare   |
| Fase 2 1955-1990   | Fase della programmazione  | Computer Aided Instruction e Computer Based Training (CAI e CBT) Intelligent CAI |
| Fase 3 1970-1990   | Fase della Personalizzazione Self-Adaptive Educational System (SAES) | Sistemi di tutoraggi automatico, basati sul costruttivismo e sugli algoritmi.    |
| Fase 4 2000 – oggi | Fase della Personalizzazione Intelligent Adaptive Learning System    | Cloud Computing, Data Mining, Big Data, Generative AI                            |

Tabella 1. Le tappe dello sviluppo dell'Artificial intelligence in Education (Rivoltella Panciroli, 2023. p. 49).

## Da Turing ai CBT: l'intelligenza artificiale forte al lavoro nell'educazione

La prima fase è quella delle rudimentali **macchine per insegnare** di Skinner ad esempio Glider, che si presentava come una scatola con una finestra all'interno della quale la macchina presentava quattro risposte possibili ad una domanda posta allo studente ed un meccanismo che permetteva allo studente di rispondere. La seconda fase è quella delle prime macchine per la Computer Aided Instruction. Per la prima volta i calcolatori vengono applicati al mondo dell'istruzione il primo Computer Based training è Plato. Questo progetto fu messo in opera nel periodo tra 1961 e il 1963 presso il campus dell'Università dell'Illinois. Il suo ideatore fu il Professor **Don Bitzer**, un ingegnere elettronico particolarmente interessato all'utilizzo del calcolatore nella didattica (Ferri, 2007). Per programmare Plato fu ideato anche uno specifico linguaggio di programmazione, chiamato Tutor, usato poi per scrivere il codice del software didattico. Si trattava di un sistema basato su semplici test a risposta multipla digitalizzati. Ad esempio, il programma Notes che girava sul Main-frame<sup>[5]</sup> di Plato visualizzava sullo schermo solo 20 linee di testo, che presentavano il contenuto dell'esercitazione da svolgere. Attraverso comandi a tastiera era poi possibile, sullo schermo, effettuare le seguenti operazioni: tornare alla risposta precedente, ritornare all'inizio dell'esercitazione, saltare un esercizio, cominciare a scrivere una nuova risposta.

A partire dagli anni Ottanta i grandissimi sviluppi dell'ICT – ricordiamo ad esempio il primo Personal Computer di Apple (1976) e di IBM (1981) – i Computer Based Training (CBT) e programmi di Computer Aided Instruction (CAI) cominciano diventare “intelligenti” nel senso che si comincia a parlare di Intelligent Computer Aided Instruction. I sistemi informatici e i software che li supportano possono essere installati nell'hard disk di un PC e sono composti tre/quattro elementi: a. una base dati che definisce gli oggetti di conoscenze organizzati in modo progressivo; b. un “modello di studente” che rappresenta lo “stato della conoscenza” che deve acquisire lo studente; c. un modello pedagogico e un'interfaccia utente per lo svolgimento delle attività; cominciamo a definirli intelligenti, perché cominciano ad essere predefiniti algoritmi inferenziali, che permettono di elevare progressivamente il livello di difficoltà dei contenuti erogati. Il processo è questo: un test diagnostico valuta il livello di conoscenze dello studente. Sulla base dei risultati dei vari test, rapportati al “modello di studente” tipo, il sistema suggerisce materiali e test, ovviamente presenti nella base dati, che sono stati precedentemente codificati (marcati) come adatti ad acquisire quel livello di preparazione (Rivoltella Panciroli, pp. 53-55) secondo un modello pedagogico istruzionista di solito l'”istruzione programmata” di Skinner oppure il modello ADDIE<sup>[6]</sup>.

## L'intelligenza artificiale “debole”: un tool digitale per l'apprendimento

La terza fase è quella del Self-Adaptive Educational System. **La cesura è netta rispetto al passato**, la confutazione di Searle del paradigma dell'intelligenza “artificiale forte” dispiega i suoi effetti anche nel campo della formazione. La cesura con il modello “istruzionista” è netta. L'epistemologia di riferimento dei sistemi di Intelligenza artificiale a supporto dell'educazione diventa il costruttivismo: le teorie di Davide Merrill (Merrill 1992), quelle del coding costruttivista di Seymour Papert (Papert, 1993) e la didattica ipertestuale di Spiro e Jacobson, per citarne solo alcune (Spiro et Al., 1991, Jacobson, Spiro., 1995).

**La rivoluzione del Web (Ferri, 1998) porta questi sistemi di AI a supporto della formazione, anche all'interno dei primi Learning an Management System** (installati sui server dell'organizzazione, in locale) dove è caricata la base dati di conoscenze necessaria per la formazione. In questo caso gli algoritmi inferenziali sono “istruiti” per adattare i contenuti e materiali allo stile di apprendimento dello studente e funzionano attraverso modelli di “presa di decisione” improntati su grafi ad albero molto articolati, non c'è quindi una standardizzazione delle performance dello studente, ma il sistema si fonda sulla personalizzazione dell'apprendimento.

Lo studente dopo il “log in” all'interno del sistema LMS, esegue un test, l'algoritmo identifica il suo profilo e il sistema viene personalizzato sulle sue esigenze e sul suo stile di apprendimento. In questo caso non esiste come nei CBT un “modello standard di studente” e uno di “conoscenze da acquisire”, ma il sistema che diventa uno strumento (tool e non tutor) al servizio delle esigenze e degli stili di apprendimento dei discenti. I data set degli utenti, sia quelli relativi ai fabbisogni formativi che quelli relativi all'andamento della formazione sono costruiti attraverso **la somministrazione progressiva di test validati su un'ampia gamma di esercizi** e su un determinato dominio di conoscenza di conoscenza (ontologia), reso disponibile in piattaforma (Rivoltella, Panciroli, pp.56-57). Cominciano ad essere tracciate anche le modalità di fruizione dei contenuti da parte degli utenti e le modalità e i tempi fruizione degli utenti e anche questi contribuiscono ad aiutare il sistema a personalizzare l'apprendimento. In questo caso non è il sistema ma è chi apprende, con l'aiuto dell'insegnante o del formatore, che regola e governa il processo di apprendimento basandosi sulle opzioni proposte dal sistema sono “scelte” sulla base delle sue esigenze (Triantafillou, T., Pomportsis, A., Demetriadis, S. (2003). Ci avviciniamo qui alla quarta fase quella più contemporanea dell'AIED e cioè agli Intelligent Adaptive Learning System.

## Verso gli agenti conversazionali per l'apprendimento

Si tratta di una rilevante evoluzione sia metodologica sia tecnologica della fase precedente sempre fondata sull'epistemologia costruttivista. Con gli sviluppi del Web 2.0, dei social e delle piattaforme cloud anche per l'educazione, i sistemi di Intelligenza Artificiale si evolvono in modo molto rapido. Gli ambienti LMS/VLE progressivamente migrano in Cloud e si rende in questo modo disponibile una grandissima quantità di dati, su cui possono operare **algoritmi inferenziali di analisi** sempre complessi e raffinati. Il tracciamento della formazione diventa uno standard e si afferma lo SCORM (Bhol et al., 2000) che permette di tracciare in maniera “omogenea” le attività di apprendimento dei formandi. Possiamo individuare almeno **cinque tendenze** di sviluppo dell'Intelligenza artificiale nel capo dell'educazione: a. l'analisi dei big data dell'apprendimento; b. i sistemi di tutoraggio automatico; c. i sistemi predittivi delle performance di apprendimento; c. i laboratori e simulazione interattivi d. i sistemi di valutazione automatica. Analizziamoli separatamente.

## L'analisi dei Big Data

Gli ambienti LMS/VLE progressivamente migrano in Cloud e si rende in questo modo disponibile una grandissima quantità di dati, inoltre, il tracciamento della formazione diventa uno standard e si afferma lo SCORM (Bhol et al., 2000) che permette di tracciare in maniera “omogenea” le attività di apprendimento dei formandi all'interno degli LMS/VLE. Questi Learning Analytics (De Santis, 2019) non possono che essere analizzata applicando l'Intelligenza artificiale che permette di elaborare in forma automatica o semi automatica, attraverso algoritmi di machine learning e di deep learning<sup>[7]</sup>, i risultati della formazione. Gli ambienti virtuali di apprendimento (LMS/VLE), contemporanei integrano già forme più o meno avanzate di “tracciamento” della formazione basate sull'AI. 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 organizzazione o dagli studenti di un'università o di una scuola, vengono elaborati dall'AI. Si generano in questo modo in report grafici, e analisi dell'efficacia della formazione che permettono ai learning designer di contare su un cruscotto digitale di monitoraggio dell'apprendimento impensabile nel mondo analogico (Franzoni e al, 2020).

## Gli Intelligent tutoring system e i ChatBot per la formazione

I sistemi di intelligenza artificiale, una volta pre-addestrati, posso formulare **inferenze probabilistiche** sull'andamento della formazione e “raccomandare” o suggerire a chi apprende nuovi contenuti o strategie di apprendimento (Franzoni e al, 2020). Si sviluppano in questa fase, sempre per supportare al meglio l'apprendimento personalizzato, anche i primi (ITS) Intelligent tutoring system: sono i “padri” dei Large Language Model di AI generativa odierna. Il tutoraggio abilitato dall'AI permette non solo di personalizzare l'esperienza di apprendimento ma di rispondere in maniera, spesso efficace, alle necessità discenti in assenza di un tutor umano (Woolf, 2009, Popenici et Al. 2019, Winkler, 2018). Si tratta dei primi ChatBot di tutoraggio personalizzato: i chat Bot sono tratta software che simulano una conversazione con un utente attraverso una interfaccia testuale. Probabilmente **il primo Chatbot educativo è stato AutoTutor** (Graesser ed Al., 1999). Tra i “modelli conversazionali” sviluppati in questa fase, possiamo citare ad esempio, Deep-speare, un modello di **machine learning** addestrato su 2700 sonetti di Shakespeare che genera componimenti poetici che molti lettori non sanno distinguere dalle composizioni del poeta (Lau, 2018). O ancora l'applicazione dei chatbot permette di supportare chi progetta la formazione e chi apprende rispetto a specifiche metodologie. In particolare, ricordiamo il caso d'uso sperimentale di un chatbot educativo AsasaraBot, progettato per assistere agli studenti delle scuole superiori, secondo la metodologia CLIL nell'apprendimento delle lingue straniere (Mageira et al., 2022). Oppure ancora, in contesto italiano la “base dati conversazionale” dedicata a **La poetessa saffo** realizzata da Alessandro Iannella (Iannella, Labrunda, Santercole., Viti, 2021 pp. 120-151).

## L'analisi predittiva dei risultati di apprendimento e del comportamento degli studenti

Attraverso l'analisi dei Learning Analytics, inoltre, si possono analizzare i risultati di apprendimento di una popolazione di studenti, e identificare tendenze e pattern di comportamento, che permettono di individuare con una buona probabilità le classi di studenti a rischio di abbandono durante un percorso universitario, un corso on-line o un Mooc, consentendo, in questo modo interventi formativi mirati per ridurre il rischio stesso (Baker & Siemens, 2014, Fitton, Finnegan, & Proulx, 2020 Rienties, Simonsen, & Herodotou, 2020).

## I laboratori e simulazioni interattive

Grazie all'IA è stato possibile riprodurre in digitale condizioni di laboratorio e sperimentali pericolose o estreme in modo da permettere ai discenti di esercitarsi senza rischi (Quiroga Pérez al., 2020).

## Sistemi valutazione automatizzata

Nel campo della valutazione l'IA ha reso più agevole e permesso di automatizzare in parte il processo di valutazione. Questo vale in particolare per i test a scelta multipla, ma anche, con l'avanzare delle capacità algoritmiche, per le risposte a domande aperte (Shermis & Hamner, 2013).

## Conclusioni

Abbiamo completato la nostra analisi della storia genetica, ovviamente sintetica e parziale, della relazione tra **intelligenza artificiale e apprendimento**. Come speriamo di avere chiarito si tratta di una storia lunga e complessa che trova le sue origini agli albori della storia del calcolo automatico contemporaneo e che si è sviluppata in complesso in questi più di settanta anni. Resta un'ultima parte piccola parte di questa storia da narrare. L'intelligenza artificiale, infatti, è stata anche, a partire dalla metà degli anni Dieci del nostro secolo, oggetto di grande attenzione a livello internazionale da parte di stati ed organizzazioni internazionali. Questi cominciano a varare una serie di studi che analizzano **gli impatti dell'applicazione dell'Intelligenza artificiale in campo socioeconomico ed in particolare in contesto educativo**. Gli Stati Uniti sono i primi a lanciare nel 2016 il National Artificial Intelligence Research and Development Strategic Plan (National Science and Technology Council, 2016) seguiti nel 2017 dalla Cina con "Next Generation Artificiale Intelligence Plan" (Government of Republic of China, 2017) e dall'Unione Europea con uno studio specifico "The Impact of Artificial Intelligence on Learning, Teaching, and Education". Questa però è un'altra narrazione – la storia genetica che abbiamo raccontato arriva alle soglie dell'Hype dell'intelligenza artificiale generativa – che abbiamo già cominciato a narrare su Agenda Digitale nel mio articolo **Chat GPT a scuola, ecco come potrebbe cambiare la didattica**. Continueremo a narrare nei prossimi mesi e anni dal momento che la rivoluzione dell'intelligenza artificiale generativa è solo agli albori.

## Bibliografia

Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. In K. Sawyer (Ed.), Cambridge Handbook of the Learning Sciences (2nd ed., pp. 253–274).

Baker, T., Smith, L., & Anissa, N. (2019). Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges. NESTA. <https://www.nesta.org.uk/report/education-rebooted>

Bohl, O., Scheuhase, J., Sengler, R., & Winand, U. (2002). The sharable content object reference model (SCORM) – a critical review. In International Conference on Computers in Education (Vol. 2, pp. 950-951). <https://doi.org/10.1109/CIE.2002.1186122>

Casola, S., Lauriola, I., & Lavelli, A. (2022). Pre-trained transformers: An empirical comparison. Machine Learning with Applications, 9, 100334. <https://doi.org/10.1016/j.mlwa.2022.100334>

De Santis, A., Bellini, C., Sannicandro, K., & Minerva, T. (2019). Data management in learning analytics: Terms and perspectives. JE-LKS. Journal of E-Learning and Knowledge Society, 15(3), 133-144. <https://doi.org/10.20368/1971-8829/1135021>

Government of the People's Republic of China. (2017). Next generation of artificial intelligence plan. <https://flia.org/wp-content/uploads/2017/07/A-New-Generation-of-Artificial-Intelligence-Development-Plan-1.pdf>

Graesser, A.C., Wiemer-Hastings, K., Wiemer-Hastings, P., Kreuz, R., & Tutoring Research Group. (1999). AutoTutor: A simulation of a human tutor. Journal of Cognitive Systems Research, 1, 35-51.

EU. (2020). The assessment list for trustworthy artificial intelligence (ALTAI). <https://futurium.ec.europa.eu/en/european-ai-alliance/pages/welcome-altai-portal>

EU. (2021a). Coordinated plan on artificial intelligence 2021 review. <https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review>

EU. (2021b). EU artificial intelligence act. <https://artificialintelligenceact.eu/the-act/>

Eugeni, R. (2021). Capitale algoritmico. Cinque dispositivi postmediali più uno. Morcelliana.

Ferri, P. (1998). La rivoluzione digitale. Individuo, comunità e testo nell'era di Internet. Mimesis.

Ferri, P. (2005). E-learning: Didattica, comunicazione e tecnologie digitali. Le Monnier.

Ferri, P. (2019). MOOC, didattica universitaria digitale e learning analytics: Opportunità e prospettive. Giornale italiano della ricerca educativa, Speciale settembre 2019, 13-26.

Fitton, I. S., Finnegan, D. J., & Proulx, M.J. (2020). Immersive virtual environments and embodied agents for e-learning applications. PeerJ Computer Science, 6, e315. <https://doi.org/10.7717/peerj-cs.315>

Iannella, A., Labruna, T., Santercole, T., & Viti, B. (2021). Reclaiming conversation: Introducing a novel approach to using conversational AI at school. In G. Trentin (Ed.), Conversational agents as online learning tutors (pp. 120-151). Nova Science Pub Inc.

Harper, B., Squires, D., & Mcdougall, A. (2000). Constructivist simulations: A new design paradigm. Journal of Educational Multimedia and Hypermedia, 9(2), 115-130. <https://www.learntechlib.org/primary/p/8080/>

Franzoni, V., Milani, A., Mengoni, P., & Piccinato, F. (2020). Artificial intelligence visual metaphors in e-learning interfaces for learning analytics. Applications, 10, 7195. <https://doi.org/10.3390/app10207195>

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.

Jacobson, M.J. & Spiro, R.J. (1995). Hypertext learning environments, cognitive flexibility, and the transfer of complex knowledge: An empirical investigation. *Journal of Educational Computing Research*, 12(4), 301-33.

Mageira, K., 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/app12073239>

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*, 27(4), 12. <https://doi.org/10.1609/aimag.v27i4.1904>

Merrill, M. D. (1992). Constructivism and Instructional Design. In T. M. Duffy & D. H. Johansen (Eds.), *Constructivism and the Technology of Instruction: a Conversation*. L. Erlbaum.

OECD. (2019). Recommendation of the Council on OECD Legal Instruments Artificial Intelligence. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

Panciroli, C., Rivoltella, P. C., Gabbrielli, M., & Zawacki Richter, O. (2020). Artificial Intelligence and education: new research perspectives. *Form@re – Open Journal Per La Formazione in Rete*, 20(3), 1–12. <https://doi.org/10.13128/form-10210>

Panciroli, C., & Rivoltella, P. (2023). *Pedagogia Algoritmica, per una riflessione educativa sull'intelligenza artificiale*. Milano: Morcelliana.

Papert, S., (1993). *The Children's Machine: Rethinking School in the Age of the Computer*, Basic Books, New York.

York, 1993, trad. it. *I bambini e il computer*, a cura di A. Bellomi, Rizzoli, Milano, 1994.

Pedró, F., Subosa, M., Rivas, A., & Valverde, P. (2019). *Artificial intelligence in education: Challenges and opportunities for sustainable development*. Paris: Unesco.

Quiroga Pérez, J., Daradoumis, T., & Marquès Puig, J. M. (2020). Rediscovering the use of chatbots in education: A systematic literature review. *Comput Appl Eng Educ*, 28, 1549–1565.

Popenici, S., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12, 22. <https://doi.org/10.1186/s41039-017-0062-8>

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536. <https://doi.org/10.1038/323533a>

Searle, J. R. (1980). *Minds, Brains, and Programs*. *The Behavioral and Brain Sciences*, 3.

Searle, J. R. (1990). Is the Brain's Mind a Computer Program? *Scientific American*, 262.

Shermis, M. D., & Hamner, B. (2013). *Automated essay scoring: A cross-disciplinary perspective*. Routledge.

Skinner, B. F. (1968). *The Technology of Teaching*. Appleton-Century-Crofts.

Spiro, R. J., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1991). Cognitive Flexibility, Constructivism, and Hypertext: Random Access Instruction for Advanced Knowledge Acquisition in Ill-Structured Domains. *Educational Technology*, 31(5), 24–33. <http://www.jstor.org/stable/44427517>

Triantafillou, T., Pomportsis, A., & Demetriadis, S. (2003). The design and the formative evaluation of an adaptive educational system based on cognitive styles. *Computers & Education*, 41(1), 87–103.

Tuomi, I. (2018). The Impact of Artificial Intelligence on Learning, Teaching, and Education. In M. Cabrera, R. Vuorikari, & Y. Punie (Eds.), *Policies for the future* (EUR 29442 EN). Luxembourg: Publications Office of the European Union.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, 30. Curran Associates, Inc. <http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf>

Ye, R., Sun, F., & Li, J. (2021). Artificial Intelligence in Education: Origin, Development and Rise. In X. J. Liu, Z. Nie, J. Yu, F. Xie, & R. Song (Eds.), *Intelligent Robotics and Applications. ICIRA 2021. Lecture Notes in Computer Science*, 13016. Springer. [https://doi.org/10.1007/978-3-030-89092-6\\_49](https://doi.org/10.1007/978-3-030-89092-6_49)

Woolf, B. P. (2010). *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning*. Morgan Kaufmann.

Winkler, R., & Soellner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Academy of Management Annual Meeting Proceedings*. <http://dx.doi.org/10.5465/AMBPP.2018.15903abstract>

# Genomica, una visione non antropocentrica: i grandi problemi da risolvere

È giusto riportare in vita una specie estinta e cancellare una specie indesiderata? Per una visione non antropocentrica della genomica serve riflettere sulle nostre azioni tenendo conto del benessere di tutte le forme di vita sulla Terra

Di **Mirta Michilli**, direttrice generale della Fondazione Mondo Digitale e **Alfonso Molina**, personal chair in Technology Strategy all'Università di Edimburgo e direttore scientifico della Fondazione Mondo Digitale

Dialogo aperto e analisi etica sono essenziali per bilanciare i potenziali benefici della genetica con biodiversità e sostenibilità ambientale. **La governance della tecnologia gene drive però è tra le più complesse**, con forti resistenze al coinvolgimento delle comunità.

## Sviluppi della genetica per la biosfera

Grazie ai potenti **poteri di sequenziamento e modificazione del genoma**, gli scienziati e gli ingegneri genetici hanno esteso le loro attività alla biosfera del pianeta. Piante e animali, piccoli e grandi, stanno entrando nei ranghi degli organismi geneticamente sequenziati. Compresa la vita estinta, come testimonia il sequenziamento del genoma di un uomo venuto dal ghiaccio tirolese, vissuto 5.300 anni fa durante l'Età del Rame, e di due mammut morti rispettivamente 20.000 e 60.000 anni fa (Callaway 2015; Keller et al. 2012).

Confrontando i genomi dei mammut con quelli di tre elefanti asiatici è stata trovata una differenza di circa 1,4 milioni di lettere di Dna che alterano la sequenza di oltre 1.600 geni codificanti proteine. Questa conoscenza potrebbe essere usata per **modificare geneticamente un elefante asiatico** in un animale simile al mammut lanoso, in grado di sopravvivere alle temperature artiche (Callaway 2015). Potrebbe essere vista come la "**resurrezione**" di una specie estinta, processo noto come de-estinzione.

Il finanziamento di 15 milioni di dollari assegnato al professore George Church, genetista di spicco dell'Università di Harvard, ha riaperto la controversia sull'opportunità di cercare di **riportare in vita animali estinti da tempo** (Stein, 2015). Altre attività di ricerca sulla de-estinzione vedono l'Università della California, Santa Cruz, impegnata nella ricerca del piccione passeggero estinto all'inizio del XX secolo (Biello, 2014) e del Dodo estinto alla fine del XIX secolo (Novak, 2016), entrambi a causa dello sfruttamento umano.

Mentre si è sviluppata velocemente una visione non antropocentrica, la riflessione ha continuato a riguardare soprattutto **gli aspetti e le conseguenze dell'editing genetico umano**. Gli esseri umani, tuttavia, per quanto considerino importante il loro posto nella scala della vita, sono solo una parte dell'intero sistema ecologico e usano altre forme di vita per la ricerca e la sperimentazione dell'editing genetico, il più delle volte a proprio servizio.

**Gli animali sono usati come terreno di prova per le tecniche di genoma**, ma vengono anche modificati geneticamente per scopi medici e/o commerciali, ad esempio creare organi da trapiantare negli esseri umani.

## Le raccomandazioni relative all'editing del genoma degli animali e delle piante

Questo rapporto utilitaristico con altre specie ha suscitato richieste di un comportamento etico più umano e di **una maggiore responsabilità negli interventi con un ampio potenziale di impatto ecologico**. Il Gruppo europeo sull'etica nella scienza e nella tecnologia (EGE, 2021) ha formulato diverse raccomandazioni relative all'editing del genoma degli animali, tra cui:

- **rafforzare la supervisione** dell'editing del genoma negli animali per gli esperimenti scientifici;
- **applicare standard rigorosi** alla sperimentazione con primati non umani e investire nello sviluppo di alternative;
- **regolamentare l'attività di allevamento di animali** che producono organi umani per il trapianto;
- **impedire l'uso non regolamentato** degli strumenti di editing del genoma;
- **rafforzare la supervisione etica** delle pratiche che comportano la riduzione delle capacità naturali degli animali;
- **garantire il benessere degli animali** da allevamento sottoposti a editing del genoma.

Altre raccomandazioni riguardano l'editing genico nelle piante, come:

- **valutare appieno i potenziali e i rischi** delle piante geneticamente modificate per l'agricoltura;
- **studiare i meccanismi di tracciabilità** ed etichettatura delle colture geneticamente modificate;
- **prestare maggiore attenzione ai dibattiti pubblici** sui prodotti agricoli geneticamente modificati.

## Forme di vita, desiderate e indesiderate

**Il tema dell'estinzione e de-estinzione delle specie nell'ecologia della vita del pianeta è forse uno dei più impegnativi dal punto di vista etico per il futuro**. Sappiamo che stiamo vivendo un periodo caratterizzato come la Sesta Estinzione, a causa della sostanziale **perdita di biodiversità** a livello globale e della grande quantità di specie minacciate di estinzione (Boulter 2003; Cowie et al., 2022; Kolbert, 2015).

A differenza delle precedenti estinzioni di massa del pianeta, **questa volta la causa è l'impatto delle attività della specie umana**. "Le estinzioni si stanno verificando in tre modi principali. Stiamo cambiando e cancellando gli ambienti, stiamo introducendo specie che minacciano ecosistemi precedentemente stabili e stiamo uccidendo" (Boulter 2003, p.166).

### L'estinzione volitiva da *gene editing*

L'ultima novità è l'estinzione volitiva da *gene editing*, ovvero l'uccisione intenzionale delle specie attraverso la modifica genetica, nota come *gene drive*. L'idea di gene drive risale a diversi decenni fa e il fenomeno avviene in modo naturale, ad esempio attraverso la **diffusione di elementi genetici**

**egoisti** che hanno una maggiore possibilità di essere ereditati (Esvelt et al. 2014; Saey, 2015). I *gene drive* di origine umana non sempre spazzano via le popolazioni.

Come l'editing del genoma umano, **un gene drive può essere progettato per indurre la comparsa, la riduzione o il miglioramento dei tratti desiderati in un'intera popolazione di vita non umana**. La sua probabilità di ereditarietà superiore al 50% assicura che la maggior parte dei discendenti di un individuo diventino portatori e così accade con le generazioni successive, fino a quando il 100% di una popolazione porta il gene ingegnerizzato.

## L'estinzione ingegnerizzata

**Eliminando la capacità riproduttiva della specie è possibile eliminare alcune forme di vita indesiderate.** L'estinzione ingegnerizzata è una delle applicazioni che destano più interesse. Prendiamo, ad esempio, il noto caso della **zanzara portatrice di malaria**. Il maschio viene geneticamente modificato per limitare e infine eliminare la nascita di prole femminile derivante dall'accoppiamento sessuale maschio-femmina. Questo processo continua nelle generazioni successive, con un numero sempre più basso di femmine e, quindi, una minore capacità di riproduzione, fino alla scomparsa dell'intera popolazione. Un'alternativa sarebbe quella di eliminare il batterio parassita nella zanzara, impedendole di trasmettere la malattia.

Alcuni dei principali benefici attesi che spingono la **trazione genica** sono la potenziale eliminazione delle malattie diffuse dai vettori, come la malaria e la dengue, e il controllo dei parassiti delle colture e delle erbacce resistenti agli erbicidi che hanno un impatto costoso sull'agricoltura. Altre applicazioni possono essere la **tutela delle specie in pericolo e il contenimento di quelle invasive** con un forte impatto ecologico nei loro nuovi habitat. I gene drive possono aiutare a ripristinare gli habitat al loro equilibrio vitale originale.

## Lo sviluppo e l'applicazione del gene-editing CRISPR Cas9

Lo sviluppo e l'applicazione del gene-editing CRISPR Cas9 stanno avanzando rapidamente. Solo un paio di decenni fa, nel 2003, lo scienziato dell'Imperial College Austin Burt pubblicò un documento in cui sosteneva che i geni ingegnerizzati potevano essere usati per manipolare le popolazioni naturali, anche se il numero di individui modificati dal gene era una piccola percentuale dell'intera popolazione. Ha stimato che, in condizioni favorevoli, **una popolazione potrebbe essere sradicata in 20 generazioni** (Burt, 2003). Dopo otto anni di duro lavoro, nel 2011, la prima dimostrazione di un gene drive di successo ha avuto luogo presso l'Imperial College, attraverso lo sviluppo di zanzare geneticamente modificate resistenti alla malaria, sotto la guida dei due genetisti Austin Burt e Andrea Crisanti (Windbichler et al. 2011). Il lavoro di laboratorio è proseguito a ritmo serrato, soprattutto perché **l'avvento di CRISPR-Cas 9 ha dato un enorme impulso alla ricerca e alla sperimentazione sui driver genetici**. Ha anche alimentato le preoccupazioni per le conseguenze ecologiche impreviste, dannose e irreversibili di eventuali rilasci nella vita selvatica. Desto allarme anche il bioterrorismo con rilascio intenzionale per danneggiare una popolazione umana (Annas et al. 2021).

## La tecnologia a propulsione genica e gli attivisti ambientali

Il superamento dei confini dei laboratori per entrare nella complessità ecologica del mondo reale ha reso urgenti raccomandazioni e linee guida per **promuovere un comportamento responsabile tra gli scienziati e gli stakeholder coinvolti**. Nel 2014, l'Organizzazione mondiale della sanità, Gene

Convene e la Fondazione per gli Istituti Nazionali di Sanità (FNIH) hanno pubblicato un Quadro di orientamento per la sperimentazione di zanzare geneticamente modificate. Il documento (OMS 2014) raccomanda che

- **i test delle zanzare modificate** per il *gene drive* inizino in condizioni di confinamento fisico
- **i test sul campo** inizino con il rilascio di zanzare maschio modificate sterili o comunque autolimitanti con il controllo degli effetti mediante l'interruzione dei rilasci.

Nel luglio 2015, un gruppo internazionale di 26 esperti di genetica, guidati da team dell'Istituto Wyss di Harvard e dell'Università della California, ha invitato tutti i ricercatori a usare **strategie di confinamento multiple** per evitare l'alterazione accidentale delle popolazioni selvatiche, e ha fornito linee guida esplicite per le autorità di regolamentazione, con la speranza di avviare anche il coinvolgimento del pubblico (Università di Harvard, 2015a; 2015b). Altre voci si sono unite, tra cui **l'appello degli attivisti ambientali per una moratoria globale sulle sperimentazioni sul campo della *gene drive***, poi respinto in occasione di un incontro sulla biodiversità delle Nazioni Unite nel dicembre 2016.

**L'accordo finale si limitava a sollecitare la cautela** nella sperimentazione sul campo dei prodotti, oltre a sostenere una migliore valutazione del rischio dei potenziali effetti (Callaway, 2016). Tuttavia, in vista di un incontro delle Nazioni Unite del 2021, un'alleanza mondiale di oltre 200 organizzazioni ha fatto circolare una petizione che invita le autorità europee a lavorare per una moratoria globale sul rilascio di organismi a *drive* genetici per proteggere la biodiversità. È stata firmata da oltre 260.000 persone.

## Il parere degli esperti Usa

Nel 2016, le Accademie Nazionali di Scienze, Ingegneria e Medicina degli Stati Uniti hanno pubblicato il rapporto “Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values” (Accademie Nazionali, 2016). Il rapporto, redatto da un gruppo interdisciplinare di 16 esperti, ha sottolineato **i potenziali benefici della tecnologia *gene drive*, le incertezze di conoscenza relative alle prove sul campo e ai rilasci nella vita selvatica, la complessità dell'approccio normativo complessivo ecologico, sociale, etico e legale necessario per la diffusione di una tecnologia con impatto potenzialmente irreversibile** nella sfera vivente.

Ha approvato l'approccio graduale dell'OMS e ha invitato le autorità governative a **sviluppare e mantenere politiche e meccanismi chiari per il coinvolgimento del pubblico nella valutazione del rischio e nel processo decisionale**, in particolare in presenza di potenziali test sul campo e rilasci ambientali. Il tema del coinvolgimento e della partecipazione del pubblico è una raccomandazione quasi universale nei rapporti e nelle dichiarazioni riguardanti la tecnologia *gene drive* e la sua applicazione nell'ambiente.

## La guida OMS

Nel 2020 l'OMS ha pubblicato una nuova guida intitolata “**Etica e malattie trasmesse da vettori**”, in cui si ribadisce la raccomandazione di un approccio graduale alla sperimentazione di zanzare geneticamente modificate (GMM); e la priorità di un coinvolgimento sostanziale delle comunità, comprese le popolazioni sotto-rappresentate e indigene, nelle prove sul campo di qualsiasi nuova strategia di controllo delle malattie trasmesse da vettori e di qualsiasi nuova strategia di intervento per la salute pubblica. **Il coinvolgimento della comunità è particolarmente importante per le**

**azioni a livello di area**, dove i rischi e i benefici possono interessare ampi segmenti della popolazione. La Guida include un capitolo completo contenente approcci e strategie per il coinvolgimento della comunità (OMS 2021; 2020a, 2020b).

## La sperimentazione fuori dalla gabbia

**La ricerca e la tecnologia gene drive continuano a migliorare, portando a sperimentazioni più avanzate** che coinvolgono anche i mammiferi. Nel 2019 l'esperimento sui topi condotto presso l'Università della California, San Diego, non era una vera e propria manipolazione genetica, poiché non era stato progettato per essere ereditabile, diffondendosi senza controllo a diverse generazioni di topi. L'esperimento ha **modificato gli embrioni di topo** per favorire l'ereditarietà di un tratto che conferisce ai topi una pelliccia bianca.

Ha funzionato parzialmente e solo nei topi femmina, mentre l'ereditarietà della mutazione è stata stimata in oltre il 70% dei figli di un topo femmina (Grunwald et al. 2019; Rennie e Cepelewicz, 2019). Ulteriori lavori sono destinati a **migliorare la tecnologia e a favorire l'ereditarietà delle mutazioni modificate geneticamente** coinvolgendo altri mammiferi, con il potenziale di creare anche nuove specie.

### Un approccio a tappe

In generale è stato seguito **l'approccio a tappe** raccomandato per garantire la sicurezza della sperimentazione e dell'applicazione sul campo del gene drive.

- **Nella prima fase**, la ricerca e la sperimentazione si avvale di gabbie, compresi ambienti più ricchi come le serre. Questa fase è principalmente sotto il controllo della comunità scientifica, che applica le linee guida e le procedure di sicurezza, per lo più definite dalla comunità stessa.
- **La seconda fase** consiste in **un rilascio controllato in un ambiente aperto**, sia attraverso mutazioni genetiche controllate e incapaci di autopropagarsi in modo incontrollato, sia attraverso il rilascio in ambienti confinati, come un'isola isolata. Le sperimentazioni in questa fase sono essenziali per imparare come si comportano i gene drive nelle condizioni naturali più selvagge e competitive, ma poiché in questo caso la ricerca e la sperimentazione sono 'fuori dalla gabbia', per così dire, l'ottenimento dell'approvazione normativa e comunitaria per la sperimentazione dei gene-drive è raccomandata e necessaria per promuovere la fiducia e la trasparenza riguardo alle questioni scientifiche, ecologiche ed etiche.
- **La terza fase è il rilascio completo nell'habitat**, dove il gene drive realizza il suo pieno potenziale ereditario in un'intera specie, oltre a trasferirsi potenzialmente ad altre specie con conseguenze incontrollate. **La terza fase presenta ovviamente le maggiori sfide**, data la potenziale eliminazione o alterazione di intere popolazioni con conseguenze ecologiche imprevedibili. Alcuni hanno chiesto ai ricercatori di non costruire sistemi in grado di invadere le popolazioni selvatiche. La chiave, tuttavia, sta nel ruolo che alla fine svolgeranno le comunità umane, dal momento che la scienza e la tecnologia a drive genetici continuano la loro marcia inesorabile verso l'ecologia del mondo e quindi verso gli habitat umani. Dato questo scenario, l'impegno e la partecipazione delle comunità nella fase 3 di autopropagazione genetica è immensamente più importante che nella fase 2. Le comunità devono sviluppare la capacità di partecipare in modo significativo. La mancanza di informazioni e di conoscenze aumenta preoccupazione e ansia per i potenziali impatti sconosciuti sulla vita e sull'intera ecologia di un luogo. Una risposta etica, una

regolamentazione, delle politiche e dei comportamenti sono essenziali per placare qualsiasi paura e legittimare la scienza e la tecnologia a drive genetici, data la centralità della vita e del suo futuro.

## I valori centrali di un codice etico

**Un importante sforzo per affrontare la complessa questione dell'etica ha portato all'identificazione di tre valori centrali in un Codice etico**, dopo un processo di consultazione e consenso iniziato nel 2017 e completato nel 2020.

I tre valori centrali sono: **scienza responsabile, gestione ecologica e impegno pubblico, e verso la fine del processo un quarto valore: "condivisione giusta ed equa dei benefici"** è stato aggiunto e considerato naturalmente abbinato all'"impegno pubblico". Guardando all'interno dei tre valori centrali, scopriamo che la scienza responsabile include la sicurezza, la protezione, la revisione paritaria e la condivisione dei dati. La gestione ecologica include la valutazione dei rischi, nonché l'attenzione alla sostenibilità, alla gestione e alla conservazione della biodiversità. Mentre il **coinvolgimento del pubblico** è stato difficile da definire, ma si riferisce alla trasparenza e al consenso informato di una popolazione identificata.

Ciò richiede **l'educazione a prendere decisioni**, che significa non limitarsi a spiegare il progetto, ma coinvolgere prima la popolazione interessata in una discussione sui suoi rischi e benefici per loro, oltre a fornire un'opportunità significativa di accettare o rifiutare il progetto per la loro comunità (Annas et al. 2021, p.23).

### I "principi di partecipazione pubblica"

Questo concorda bene con i "principi di partecipazione pubblica" che fanno riferimento al modello dell'OMS (2020b):

- Rispetto, cioè prendersi cura e prestare attenzione agli interessi e alle preoccupazioni degli altri)
- Inclusività, sforzarsi di includere tutti coloro che rientrano nel suo ambito
- Trasparenza, essere chiari, aperti e non nascondere nulla
- Reattività per dimostrare che le richieste o le preoccupazioni sono state ascoltate e cercare di soddisfarle
- Onestà, cioè dire la verità, non cercare di ingannare o permettere che prevalgano le falsità.

È evidente che un impegno pubblico orientato ai valori è fondamentale per il futuro del gene drive a beneficio dell'umanità. Non c'è da stupirsi che le organizzazioni scientifiche, politiche e internazionali abbiano ripetutamente sottolineato la sua importanza e chiesto la sua applicazione prima dell'implementazione della tecnologia gene drive. Tuttavia, nella maggior parte delle discussioni, degli studi, delle linee guida, delle raccomandazioni ecc. esiste un paradosso: non coinvolgono o non hanno quasi alcun contributo da parte delle comunità interessate (Kormos et al. 2022). Questo paradosso sottolinea la difficoltà dei processi di coinvolgimento e consenso delle comunità, apparentemente perché richiedono una preparazione e un'attuazione attenta e lunga, usando un mix di tecniche di ricerca sociale. Ma non mancano esempi positivi, come la ricerca condotta dal World Mosquito Programme (WPM) a partire dal 2011 con il batterio Wolbachia, una strategia per tentare di eliminare la dengue (Schairer et al. 2021; Singh 2019) che ha coinvolto diversi paesi.

## La tentazione delle scorciatoie

**La fiducia è una parola chiave per l'intero processo e, in questo caso, coloro che conducono la tecnologia gene drive nelle comunità e nelle ecologie hanno l'obbligo etico di informare, educare, coinvolgere, consultare e cercare un ampio consenso.**

**La fiducia richiede anche integrità, indipendenza e trasparenza**, per evitare un clima pernicioso di sospetti. Per esempio, l'Autorità Europea per la Sicurezza Alimentare (EFSA) ha lavorato su una valutazione normativa della tecnologia gene drive. Questa valutazione dovrebbe essere rassicurante per i cittadini, eppure è stato riferito che "la metà degli esperti incaricati di valutare i rischi potenziali della tecnologia ha legami finanziari con le organizzazioni che sviluppano la tecnologia, e altri hanno anche conflitti di interesse con un'azienda che sviluppa insetti GM" (Corporate Europe Observatory, 2019). Anche se gli esperti fanno del loro meglio per dare un giudizio indipendente, i sospetti di conflitto d'interessi sono destinati a fiorire.

È bene tenere presente che **le tecnologie si evolvono più rapidamente dell'ambiente normativo e delle agenzie incaricate di definirle e applicarle** e, addirittura, più rapidamente della comprensione e delle capacità decisionali informate delle comunità e delle loro leadership.

Le pressioni competitive possono consigliare l'adozione di scorciatoie e di percorsi più semplici, ma quando sono in gioco la salute, l'ecologia e il futuro dell'umanità, non è l'opzione più saggia da perseguire. Nell'era della grande accelerazione è questo il problema cruciale da risolvere.

## Bibliografia

- Annas, G. J. et al. (2021). A Code of Ethics for Gene Drive Research. *The CRISPR Journal* 4(1):18–24.
- Biello, D. (2014). Ancient DNA Could Return Passenger Pigeons to the Sky. *Scientific American*, August 29.
- Boulter, M. (2003). *Extinction. Evolution and the End of Man*. London: Fourth Estate.
- Burt, A. (2003). Site-Specific Selfish Genes as Tools for the Control and Genetic Engineering of Natural Populations. *Proc. R. Soc. Lond.* 270:921–28.
- Callaway, E. (2015). Mammoth Genomes Hold Recipe for Arctic Elephants. *Nature* 521:18–19.
- Callaway, E. (2016). Gene Drive' Moratorium Shot down at UN Biodiversity Meeting. *Nature*, December 21.
- Corporate Europe Observatory (2019). EFSA Gene Drive Working Group Fails Independence Test. *Corporate Europe Observatory*, June 28.
- Cowie, R. e al. (2022). The Sixth Mass Extinction: Fact, Fiction or Speculation? *Biological Reviews*.
- Esvelt, K. et al. (2014). Emerging Technology: Concerning RNA-Guided Gene Drives for the Alteration of Wild Populations. *ELife*, July 17.
- Grunwald, H. A. et al. (2019). "uper-Mendelian Inheritance Mediated by CRISPR–Cas9 in the Female Mouse Germline. *Nature* 566(7742):105–9.
- Harvard University (2015). Safeguarding Gene Drives. *Wyss Institute*, July 30.
- Harvard University (2015). Springing into Action: The Wyss Institute Introduces Its New Biosafety Process. *Phys.Org*, May 6.
- Keller, A. et al. (2012). New Insights into the Tyrolean Iceman's Origin and Phenotype as Inferred by Whole-Genome Sequencing. *Nature Communications* 3(698).
- Kolbert, E. (2015). *The Sixth Extinction. An Unnatural History*. London: Bloomsbury.

- Kormos, A. et al. (2022). Ethical Considerations for Gene Drive: Challenges of Balancing Inclusion, Power and Perspectives. *Frontiers in Bioengineering and Biotechnology* 10(January).
- National Academies of Sciences, Engineering, and Medicine (2016). *Gene Drives on the Horizon: Advancing Science, Navigating Uncertainty, and Aligning Research with Public Values*. Washington, DC: The National Academies Press.
- Novak, B. (2016). Dodo Bird De-Extinction? The Dialogue Has Begun in the Island Nation of Mauritius. *Revive&Restore*, December 19.
- Rennie, J., Cepelewicz J. (2019). Gene Drives Work in Mice (If They're Female). *Quanta Magazine*, January 23.
- Saey, T. H. (2015). Gene Drives Spread Their Wings. *ScienceNews*, December 2.
- Schairer, C.E. et al. (2021). "Oxitec and MosquitoMate in the United States: Lessons for the Future of Gene Drive Mosquito Control." *Pathogens and Global Health* 115(6):365–76.
- Singh, J. A. (2019). Informed Consent and Community Engagement in Open Field Research: Lessons for Gene Drive Science. *BMC Medical Ethics* 20(54):1–12.
- Stein, R. (2015). Critics Las Out at Chinese Scientists Who Edited DNA in Human Embryos. *NPR*, April 23.
- Windbichler, N. et al. (2011). A Synthetic Homing Endonuclease-Based Gene Drive System in the Human Malaria Mosquito. *Nature* 473:212–15.
- World Health Organization (2014). *Guidance Framework for Testing Genetically Modified Mosquitoes, Second Edition*. Geneva.
- World Health Organization (2020a). *Evaluation of Genetically Modified Mosquitoes for the Control of Vector-Borne Diseases: Position Statement*. October.
- World Health Organization (2020b). *WHO Guidance, Ethics and Vector Borne Diseases*. Geneva: WHO.
- World Health Organization (2021). WHO Issues New Guidance for Research on Genetically Modified Mosquitoes to Fight Malaria and Other Vector-Borne Diseases. May 19.

# Scrivere poesia con GPT4: la sfida di scolpire versi con l'IA

Il poeta si misura con la recalcitranza del linguaggio, del simbolismo culturale, delle complessità del metro e della rima, come il giardiniere la affronta per creare bellezza entro i vincoli imposti dalla natura. Le stesse sfide di GPT4 in veste di Aiedo, in perfetto equilibrio tra chiarezza ed evocazione

Di Nick Young, Marcello Frixione e Enrico Terrone - Università di Genova

**GPT-4** rappresenta un notevole salto in avanti nelle capacità dell'intelligenza artificiale: può generare testi sorprendentemente coerenti e simili a quelli umani, inclusa la **poesia**. Solleva così interrogativi filosoficamente affascinanti sulla relazione tra AI e arte (cfr. Hertzmann 2018; Köbis e Mossink 2021; Anscomb 2022).

Sebbene GPT-4 possa generare testi che assomigliano alla poesia, si potrebbe avere la sensazione che la poesia generata tramite l'AI, da quel presunto poeta che con un gioco di parole potremmo chiamare l'**Aiedo** (l'aedo ai tempi dell'AI, ndr), sia priva di merito artistico. In certi casi questo sembra davvero ciò che accade. Per esempio, un utente di GPT-4 potrebbe semplicemente inserire come comando "Scrivi poesia" e il programma produrrà il seguente output.

Nel cielo terso danzano stelle dorate,  
un'arpa celeste suona dolce melodia,  
le note sospese nell'aria fluttuano,  
come ali di farfalle, leggere e armoniche.  
Ogni verso scritto è un battito di cuore,  
un'emozione pura, un bacio al creato,  
in queste otto righe, un mondo si svela,  
una poesia che l'anima ha incantato.

Non solo queste righe sono piatte e banali, ma l'utente di GPT-4 ha fatto ben poco. Non si può dire che abbia scritto una poesia.

**Tuttavia, creare poesia attraverso GPT-4 è una sfida e si può effettivamente raggiungere merito artistico.** Sebbene i poeti AI non sperimentino gli stessi tipi di recalcitranza dei poeti tradizionali, GPT-4 presenta le sue proprie forme di recalcitranza, da superare prima che il modello aiuti a produrre poesia esteticamente piacevole e non banale.

**Ecco come creare poesia, non solo grazie alla padronanza del linguaggio e della forma poetica, ma anche dalla comprensione dell'AI e dalla capacità di guidare efficacemente il suo output.**

## Giardinaggio e recalcitranza

Richard Wollheim (1980, p. 28) introdusse in estetica il termine “recalcitranza” per esprimere la

Sono diversi i tipi di **recalcitranza** che i poeti hanno tradizionalmente affrontato. Un poeta deve lottare per **trovare le parole e le frasi giuste che non solo trasmettono i suoi pensieri e sentimenti**, ma per **adattarsi anche al ritmo, alla rima e al metro della poesia**. Questo comporta spesso **scelte difficili**, poiché il poeta può scoprire che **la parola perfetta per esprimere un particolare sentimento interrompe il ritmo o lo schema di rima stabilito**.

Il poeta potrebbe anche affrontare l’uso di **metafore e similitudini, avendo bisogno di trovare un equilibrio tra chiarezza ed evocazione**, oltre ad affrontare la sfida di creare qualcosa di originale e significativo. Digitando “scrivi poesia”, un utente di GPT-4 elude tutte queste tipologie di sfide, dunque potrebbe sembrare che l’AI edo non stia intraprendendo alcuna forma significativa e artistica di creazione. Ma in realtà il modello serve a utilizzare per produrre poesia dall’estetica curata e non banale.

I giardinieri affrontano distinte forme di recalcitranza, che può ripensarsi in termini di “**Natura Naturans**“, che, secondo **Spinoza** che coniò il termine, si riferiva agli **aspetti incessantemente dinamici e imprevedibili della natura**. Quando i giardinieri interagiscono con il terreno, i semi, il clima e le stagioni, non stanno gestendo entità passive e statiche. Invece, lottano con la natura nel suo stato attivo e dinamico. Ogni elemento del giardinaggio presenta una propria serie di sfide e resistenze.

Un appezzamento di terreno non è semplicemente una tela bianca in attesa del disegno del giardiniere. Ma un mezzo che può essere solo parzialmente controllato. I semi e le piante non sono oggetti inerti da posizionare a piacimento, bensì organismi viventi ognuno con i suoi requisiti unici per la crescita, resistenza ai parassiti e reazione alle condizioni ambientali.

## La danza con la Natura Naturans

Il giardinaggio è una danza con la Natura Naturans, uno scambio dinamico con la natura nel suo stato più attivo, creativo e spesso imprevedibile. Un giardiniere deve lavorare all’interno dei vincoli e delle resistenze della natura, utilizzando la sua conoscenza e abilità per ottenere un risultato esteticamente gratificante da questo processo dinamico.

Anche il ruolo del giardiniere esperto non è uno di controllo assoluto, ma di guida e risposta, cercando di prevedere elementi imprevedibili e sviluppando strategie per gestirli. I giardinieri esperti sanno che potrebbe esserci un improvviso freddo, o che un particolare insetto potrebbe decimare i loro cespugli di rose. Capiscono come diverse piante reagiscono alla luce, all’acqua e a varie condizioni del terreno. Contrastiamo questo caso con quello del giardiniere principiante. Quest’ultimo potrebbe certamente piantare semi, innaffiarli e sperare nel meglio. Ma senza la conoscenza dell’esperto, sarebbe alla mercé dei capricci della natura.

Potrebbe innaffiare troppo le loro piante o non rendersi conto che le piante di pomodoro hanno bisogno di sostegno mentre crescono per evitare steli piegati e spezzati. **Il giardiniere esperto, attraverso la sua profonda comprensione e esperienza, può invece affrontare le sfide poste dalla natura**. Può trasformare queste sfide in opportunità, utilizzando la natura dinamica del mezzo per creare qualcosa di esteticamente rilevante. Il suo giardino diventa un riflesso del suo talento e della sua comprensione del mondo naturale. La creazione di qualcosa di esteticamente gradevole – un giardino – da un processo caratterizzato da imprevedibilità e resistenza, suggerisce una forma di espressione artistica. Dunque, **i giardinieri esperti, pur non avendo il controllo completo su come crescono le piante, plasmano il giardino in un modo che produce**

**valore estetico.** Prendono **gli elementi imprevedibili** – le abitudini di crescita delle piante, il tempo, i parassiti – e **li incorporano nel loro progetto. Il risultato non è semplicemente una raccolta di piante ma un giardino:** un sistema organizzato che conta come oggetto estetico (cfr. Carlson 2002; Cooper 2009; Salwa 2014).

## GPT4 come Machina Naturans

Facendo leva sull'analogia del giardinaggio, rivolgiamo ora la nostra attenzione alla poesia generata dall'AI.

**Dunque l'AIedo incontra una recalcitranza più simile a quella dei giardinieri che a quella dei poeti tradizionali. L'imprevedibilità insita nella creazione di poesia generata da IA, la resistenza della macchina al controllo e alla previsione, assomiglia alla recalcitranza che il giardiniere affronta nella natura.**

La recalcitranza che gli AIedi affrontano può dunque essere teorizzata in termini di “Machina Naturans”, un concetto che estende l'idea di Spinoza di “Natura Naturans” a sistemi artificiali come GPT-4. Quando gli AIedi interagiscono con GPT-4, non stanno gestendo un'entità passiva, statica. Invece, lottano con l'intelligenza artificiale nel suo stato attivo, dinamico, caratterizzato dal suo comportamento imprevedibile e dalla sua complessità intrinseca. Ogni elemento della poesia generata da AI presenta una propria serie di sfide e resistenze. **GPT-4 non è semplicemente una tela bianca in attesa del design dell'AIedo, ma un sistema piuttosto imprevedibile che genera testi basati sui modelli e le strutture che ha appreso dai suoi dati di addestramento.** Nonostante sia addestrato su un immenso corpus di dati e impieghi algoritmi sofisticati, quando a GPT-4 viene dato l'istruzione di ‘scrivere poesia’, i risultati – come abbiamo visto – possono essere **spaventosamente privi di ispirazione.** Un passo preliminare per consentire a GPT-4 di **generare poesia significativa ed esteticamente coinvolgente** consiste nel **trattarlo come una machina naturans** – un sistema caratterizzato da un comportamento dinamico e imprevedibile simile alla natura

stessa. L'AIedo è dunque sfidato a produrre sollecitazioni adeguate. Questo richiede una comprensione sia del compito poetico che delle capacità e limitazioni generali dell'AI. **Come il giardiniere seleziona piante adatte per il proprio giardino, così l'AIedo considera fattori come la capacità creativa dell'IA, la sua comprensione delle convenzioni poetiche e la sua capacità di avvalersi del linguaggio in forme nuove ed evocative.**

## Allineare l'output dell'AI con le convenzioni poetiche umane

A differenza dei poeti umani che possiedono una comprensione intuitiva del ritmo, del metro e delle immagini, GPT-4 non “comprende” la poesia nello stesso modo in cui lo fanno gli umani.

Genera **testo basato su schemi appresi durante l'addestramento**, ma non comprende la bellezza di una metafora abilmente creata o l'impatto emotivo di una riga toccante. Di conseguenza, **l'AIedo deve guidare l'AI utilizzando sollecitazioni e feedback per produrre un output che si allinei con gli standard umani della poesia di elevato valore estetico.**

Così come l'esperienza del giardiniere gli permette di anticipare e adattarsi agli **elementi imprevedibili della natura**, l'esperienza dell'AIedo lo aiuta a fronteggiare la recalcitranza dell'AI. **L'AIedo deve dunque comprendere la “natura” di GPT-4, i suoi punti di forza e debolezze, e imparare a cooperare con l'algoritmo piuttosto che opporsi.**

Nella scrittura poetica tradizionale, l'autore si confronta con la recalcitranza del linguaggio, del simbolismo culturale, delle complessità del metro e della rima, e della domanda di originalità. **I poeti scolpiscono i loro pensieri, cercando di aderire a una specifica struttura poetica o di crearne una propria.**

**La recalcitranza incontrata è principalmente interna: è una lotta contro i vincoli del linguaggio e della forma. Quando invece si crea poesia con GPT-4, la recalcitranza incontrata è esterna.** L'AIedo lotta con l'imprevedibilità dell'output dell'AI, e si impegna ad **allineare quell'output con le convenzioni poetiche umane.** Occorre però trovare le sollecitazioni corrette per ottenere la risposta desiderata dalla macchina. Questa sfida è analoga a quella che il giardiniere affronta nel tentativo di creare bellezza entro i vincoli imposti dalla natura. **Il giardiniere non esercita un controllo assoluto sulla natura ma piuttosto la guida e la influenza,** lavorando all'interno delle sue regole e ritmi per modellare qualcosa di esteticamente rilevante. Allo stesso modo, l'AIedo non ha un controllo assoluto sull'AI ma piuttosto collabora con la sua "Machina Naturans", lavorando all'interno dei modelli e ritmi intrinseci del sistema per generare poesia munita di valore estetico.

## L'estensione del dominio della poesia

Con GPT-4, la possibilità di creare buona poesia si è estesa oltre il dominio del talento poetico tradizionale, per includere anche coloro che comprendono il funzionamento dell'AI e hanno una qualche contezza del valore poetico. Ciò ha portato a quello che potrebbe essere interpretato come una **democratizzazione della creazione poetica,** sebbene con importanti precisazioni. Nell'era precedente a GPT-4, era possibile per un individuo possedere comprensione profonda della poesia e piena contezza del suo valore, senza necessariamente avere la capacità di generare buona poesia. Una persona poteva assumere il ruolo di conoscitore, capace di rilevare le più sottili sfumature di ritmo, metro e metafora, pur non avendo la capacità di intrecciare parole in una convincente costruzione poetica. Le facoltà di apprezzamento e creazione non erano sempre allineate nello stesso individuo.

**L'introduzione di GPT-4 ha mutato questa dinamica. La capacità di creare poesia è diventata dipendente non solo dalla padronanza del linguaggio e della forma poetica, ma anche dalla comprensione dell'AI e dalla capacità di guidare efficacemente il suo output.** Coloro che hanno una profonda conoscenza della poesia e una piena comprensione di come funziona GPT-4 possono indirizzare l'AI a produrre poesia che realizza le loro preferenze estetiche. Il conoscitore, un tempo confinato al ruolo di osservatore, ora si trova a partecipare al processo creativo nella misura in cui il suo **gusto raffinato e la sua comprensione di GPT-4 facilitano la generazione di poesia munita di valore estetico.** La combinazione dell'esperienza e della sensibilità estetica del conoscitore con le capacità generative di GPT-4 può produrre poesia che non è solo tecnicamente competente, ma anche emotivamente toccante ed esteticamente riuscita.

## Fare poesia con GPT-4

Esiste un'altra categoria di individui che, pur non avendo l'attitudine naturale per la composizione poetica tradizionale, potrebbe **rivelare un talento per la creazione di poesia tramite l'intelligenza artificiale.**

Si avrebbe quindi l'emergere di un AIedo che **potrebbe non essere un conoscitore di poesia nel senso tradizionale, ma che compenserebbe questa lacuna con un talento naturale nell'arte di guidare l'AI a generare composizioni poetiche convincenti.** Tuttavia, questi sviluppi non implicano

che chiunque possa produrre poesia di alta qualità con GPT-4. Una persona sprovvista tanto di conoscenza della poesia quanto di un talento naturale per la creazione di poesia assistita dall'AI, creerebbe molto probabilmente poesie di poco valore. **Dunque GPT4 estende il dominio della poesia, amplia il potenziale bacino di poeti** e di conseguenza potrebbe **rendere più vario il panorama della creazione poetica, senza per questo ridurre la poesia a una mera operazione meccanica** alla portata di tutti.

## Esempi di “giardinaggio poetico”: la poesia con GPT4 e i bias

L'esempio di poesia riportato nell'introduzione tradisce una preferenza di GPT-4 per esiti poetici abbastanza tradizionali e scontati. Un primo modo per intervenire sugli output del programma può consistere nel fornirgli un argomento scelto in maniera tale da evitare temi poeticamente troppo ovvi. Chiedendo al sistema, per esempio, di scrivere una poesia sulla teoria dell'evoluzione abbiamo ottenuto il testo seguente.

Nel grembo del tempo, vita si snoda,  
 un viaggio lento, dove l'arte si rinnova.  
 Dal semplice, sorge la forma complessa,  
 nella danza eterna della natura espressa.  
 Evoluzione, silenzioso fiume di cambiamento,  
 dal passato nasce il presente, nel vento.  
 Da specie a specie, in trama inespressa,  
 svela il mistero della vita, in sua bellezza

Tematicamente il testo è **abbastanza pertinente, ma resta decisamente banale, costruito su una rete concettuale prevedibile e superficiale**. Inoltre, la **struttura metrica è schematica: distici di rime bacciate (talvolta assonanze)**, che coincidono con la scansione in periodi della poesia. Abbiamo perciò domandato al programma di intervenire evitando, per esempio, l'utilizzo di metafore scontate (come la danza della natura, il grembo del tempo o il fiume di cambiamento) e l'uso di una struttura rigida di rime. Gli è stato inoltre suggerito di inserire anche termini tecnici della biologia o espressioni di altre lingue e di provare a impiegare versi più lunghi. Attraverso tredici passaggi, il testo che abbiamo ottenuto è il seguente.

Nell'alto del cielo di matrice biologica s'innalzano formazioni misteriose  
 ritus occultus velluto respiro di informazioni non espresse  
 dalle molecole scaturisce vita un inizio privo di canto e di meridiano  
 viaggio di delicate mutazioni danza senza destinazione in un fluire costante  
 non discordanze le ombre né trascuratezze lievi ma tessere d'un disegno  
 experimentum perpetuum così naturale intimo come i reef del mondo  
 la vita un intreccio di nucleotidi criptico cruciverba di sequenze  
 evoluzione tessuto che nel silenzio si tesse unico variegato dans ses nuances.

Siamo decisamente molto lontani dal punto di partenza e ciò mostra le possibilità di questo tipo di uso interattivo di GPT4. Tuttavia, complessivamente, la recalcitranza è forte, e l'impiego di GPT4 come strumento per il giardinaggio poetico non è ancora agevole. **Il programma esibisce alcuni forti pregiudizi su cosa sia e come debba essere organizzato un testo poetico.** Abbiamo riscontrato una **tenace preferenza per alcune manifestazioni del poetese più ovvio** (un esempio, che appare evidente già dai testi riportati in queste pagine, sono le **metafore basate sulla danza**); tale propensione risulta molto difficile da scalfire: se gli si chiede di non utilizzare una certa parola o un certo concetto, GPT4 lo elimina, ma esso ricompare inevitabilmente dopo qualche interazione. Considerazioni analoghe valgono per la **forma metrica**. Sembra che per GPT4 una poesia debba essere scritta in quartine di versi della lunghezza, più o meno, di un endecasillabo, con rime bacciate o alternate (talvolta sostituite da assonanze). Anche tale schema risulta estremamente persistente e il sistema si mostra restio a modificarlo.

## Conclusioni

Il problema principale consiste probabilmente nel canone poetico cui il programma fa riferimento, ossia nel **data base di poesie che il programma “conosce”** e che utilizza come modello. Se gli si domanda ad esempio “chi è Edoardo Sanguineti?”, il programma si dimostra abbastanza competente e bene informato. Ma se gli si chiede di scrivere una poesia nello stile di Sanguineti si ottiene qualcosa del genere.

Sfuggono le parole, danzando nel vuoto,  
un balletto d'assalti, il linguaggio è un gioco.  
Sanguineti, maestro dell'ardita audacia,  
incanta l'anima con la sua ribelle traccia.  
Surreale e provocatorio, l'arte si scatena,  
nel frastuono dei versi, la verità si ripiena.  
La lingua si diverte, in un turbinio di suoni,  
Sanguineti, il poeta che rompe i confini più buoni.

Si ottiene dunque un testo che ha ben poco a che fare con lo stile sanguinetiano. Il problema consiste nel fatto che il programma **“sa” varie cose sulla poesia e sulla poetica di Sanguineti, ma non conosce i suoi testi (o almeno, non li conosce al punto da integrarli tra i suoi modelli di fare poesia)**, per cui cerca di costringere le informazioni di cui dispone in uno schema inadeguato. Esiti analoghi si ottengono anche con altri autori (ad esempio Pound o E.E. Cummings). Per concludere, quindi, **sembra che GPT4 abbia una “cultura poetica” non solo tradizionalista, ma anche limitata e piuttosto banale.** Questa difficoltà non è tuttavia insormontabile: in prospettiva, infatti, **le competenze poetiche del programma potranno essere estese e migliorate**, attraverso l'accesso a un insieme di testi più **vasto e articolato**. Per il momento, però, il terreno che GPT4 offre al giardinaggio poetico risulta ancora particolarmente sassoso e impervio.

---

*\*Il lavoro è stato supportato dal progetto ERC-StG Project 101040535 (“PEA – The Philosophy of Experiential Artifacts”).*

## Bibliografia

- Anscomb, C. (2022). Creating Art with AI. *Odradek. Studies in Philosophy of Literature, Aesthetics, and New Media Theories*, 8(1), 13-51.
- Carlson, A. (2002). *Aesthetics and the environment: The appreciation of nature, art and architecture*. Psychology Press.
- Cooper, D. E. Gardens; in Davies, S., Higgins, K. M., Hopkins, R., Stecker, R., & Cooper, D. E. (Eds.). (2009). *A companion to aesthetics*. John Wiley & Sons.
- Hertzmann, A. (2018). Can computers create art?. In *Arts* (Vol. 7, No. 2, p. 18).
- Köbis, N.; Mossink, L. D. (2021). Artificial intelligence versus Maya Angelou: Experimental evidence that people cannot differentiate AI-generated from human-written poetry, *Computers in Human Behavior*, 114.
- Salwa, M. (2014). The Garden as a Performance. *Estetika: The European Journal of Aesthetics*, 51(1), 42-61.
- Wollheim, R. (1980). *L'arte e i suoi oggetti*, Marinotti (trad. it., 2013).

# Digital Pathology: i nuovi scenari dell'attività diagnostica

Una nuova tecnologia permette la trasformazione digitale del vetrino su cui è presente la sezione di tessuto. La Digital Pathology è in grado di definire un vetrino digitale per effettuare una diagnosi più approfondita. Ecco quali vantaggi offre al patologo l'abbandono del microscopio per la visualizzazione su monitor

Di **Elena Bottinelli**, Head of Digital Transition and Transformation del Gruppo San Donato

Il percorso diagnostico, immutato da molto tempo nelle sue fasi essenziali, vede ora la possibilità di una profonda ed importante evoluzione. Grazie alla **Digital Pathology**.

Questa nuova tecnologia permette la trasformazione digitale, tramite scanner ad alta risoluzione, del vetrino sul quale è presente la sezione di tessuto.

**Oltre al vetrino fisico tradizionale, il patologo oggi dispone di un vetrino digitale.** Gli strumenti di AI amplificano le capacità di ottenere informazioni qualitative e quantitative dai vetrini digitali, nascoste all'occhio umano. E mirano a svolgere anche analisi prognostiche e predittive.

Per comprendere a pieno ciò di cui parleremo è bene fare una premessa.

## L'attività classica del patologo

Il **patologo** svolge un **ruolo diagnostico essenziale** nella cura dei pazienti, fornendo la **diagnosi in molte patologie, incluse tutte le diverse forme di neoplasia**. L'attività diagnostica del patologo si basa principalmente sull'**osservazione al microscopio di sezioni di tessuto o di campioni cellulari**. Essi derivano da **interventi chirurgici, biopsie, agoaspirati e liquidi organici prelevati** durante interventi chirurgici, **procedure endoscopiche ed agoaspirative** effettuate per scopi diagnostici e/o terapeutici.

Questi campioni, necessari per definire la natura dei più svariati processi patologici, sono processati con specifiche metodiche. Lo scopo è la preparazione di **sezioni o strisci cellulari** da disporre su vetrini. Una volta colorati con **coloranti specifici**, i vetrini possono poi essere **osservati al microscopio**.

Queste procedure sono relativamente semplici e nei loro principi base, rimangono **inalterate da quasi due secoli**.

Infatti, gli ingredienti base e le modalità di preparazione dei tessuti – **sostanze chimiche per la fissazione, processazione, inclusione in paraffina, colorazione dei tessuti e delle cellule** – sono, con limitate modifiche, gli stessi **da oltre 150 anni**. Nonostante l'apparente vetustà, questa tecnica tradizionale è poco costosa. E permette al patologo di diagnosticare malattie che, al momento, nessuna altra tecnologia, pur sofisticata, è in grado di definire.

Il patologo, invece, osservando al microscopio i vetrini con le sezioni di tessuto colorate con **ematossilina ed eosina** - i due coloranti tradizionali che identificano le strutture nucleari e citoplasmatiche cellulari - valuta la composizione cellulare. E verifica come sono organizzate le diverse popolazioni cellulari, le caratteristiche delle singole cellule ed i rapporti fra le diverse popolazioni cellulari.

## Il processo di valutazione

La valutazione di queste caratteristiche, talora con il supporto di ulteriori indagini effettuate con tecniche di **immunofenotipizzazione** e di analisi genetico-molecolari, permette al patologo di **formulare una diagnosi**. E questa diagnosi spesso contiene **indicazioni prognostiche** di estrema importanza e permetterà al paziente ed al clinico di seguire il più accurato percorso terapeutico.

I vetrini con le sezioni di tessuto e la fonte di queste sezioni, il tessuto incluso in paraffina, prevedono la conservazione ed archiviazione, per rimanere a disposizione del paziente e del patologo per ogni eventuale futura necessità. **La capacità del patologo di caratterizzare e definire con precisione le alterazioni osservate dipendono dalla sua esperienza e conoscenza dei processi patologici.**

## Digital Pathology: che cos'è

La Digital Pathology è invece la nuova tecnologia che consente la trasformazione digitale, via scanner ad alta risoluzione, del vetrino su cui è presente la sezione di tessuto. Oltre al vetrino fisico tradizionale, il patologo dispone di un vetrino digitale. Le caratteristiche dell'immagine del vetrino digitale sono attualmente di qualità equivalente a quelle che vengono osservate dal patologo con un buon microscopio ottico.

Studi condotti in più e diverse realtà hanno dimostrato che **non c'è inferiorità della capacità diagnostica del patologo con l'uso del vetrino digitale rispetto a quello fisico**. La Digital Pathology potrebbe quindi sembrare **la semplice trasformazione digitale del vetrino fisico**. Tuttavia l'**evoluzione** apparente banale, grazie **ai progressi della tecnologia digitale e delle infrastrutture informatiche**, apre in realtà **nuovi scenari**. Essi modificheranno nei prossimi anni non solo il lavoro ed il ruolo del patologo nei percorsi diagnostici. Ma permetteranno anche di mettere ancor di più in luce l'**enorme mole di informazioni diagnostiche, cliniche e biologiche** che la semplice sezione di tessuto racchiude in sé. Si osserveranno dunque **importanti benefici per il paziente e per le conoscenze scientifiche**.

Possiamo paragonare questa evoluzione a quella avvenuta anni fa in campo fotografico nel **passaggio dalla pellicola alla fotografia digitale**, anche se con alcune **peculiari differenze**. Trasformare il vetrino fisico in vetrino digitale significa trasformare le strutture tessutali, le cellule presenti, le loro caratteristiche **in entità misurabili ed analizzabili con l'applicazione di metodiche di analisi finora non possibili**. Si aggiunge quindi alla valutazione del patologo un ulteriore livello di conoscenza. Finora abbiamo solo qualche esempio, ma lo scenario rimane in gran parte da esplorare.

## La trasformazione digitale

Quello che sta avvenendo in patologia è **simile a ciò che è già avvenuto in ambito radiologico con il passaggio dalla lastra radiologica alla lastra digitale**.

Vediamo perché il patologo dovrebbe **abbandonare il microscopio per trasferirsi di fronte ad uno schermo**. Infatti il percorso diagnostico attuale è efficace e poco costoso e la formazione del patologo ha lo strumento base nel microscopio. Ma la trasformazione digitale apre nuovi scenari

## Digital Pathology: due scenari di sviluppo

Il primo settore profondamente modificato con effetto immediato è quello organizzativo con la trasformazione del **workflow del processo diagnostico** nei laboratori di anatomia patologica e del lavoro dei patologi.

Infatti, **il vetrino digitale**, una volta acquisito dallo scanner ed inserito nel LIS- Laboratory Information System/sistema informatico di gestione del laboratorio- **diventa disponibile in tutte le postazioni di lavoro**. Il patologo può dunque iniziare la sua attività diagnostica **visualizzandolo sul proprio monitor**.

Per il patologo i vantaggi del vetrino digitale sono molteplici:

- la possibilità di condivisione immediata, senza necessità di spostamento sia del patologo che del vetrino fisico, con colleghi della stessa unità operativa, ma anche a distanza;
- la visione a distanza, potenzialmente ovunque nel mondo, apre la possibilità che **parti del lavoro diagnostico possano essere effettuate in remoto**. Si facilita così sia il lavoro dei patologi dello staff dell'unità operativa, sia i consulti con esperti di particolari patologie, ovunque siano.

L'emergenza Covid-19 ha messo in evidenza come questa possibilità di **teleconsulto** poteva permettere di **superare restrizioni logistiche**.

### La preparazione dei vetrini nella Digital Pathology

Altri aspetti vanno ben oltre la semplice trasformazione digitale del tradizionale vetrino, incidendo profondamente sull'organizzazione del laboratorio. **Un buon vetrino digitale può partire solo da un perfetto vetrino fisico**. Dunque il laboratorio, in particolare nello staff tecnico deve mettere in atto tutte le attenzioni nelle **fasi preanalitiche e di processazione** dei campioni per poter arrivare ad una qualità elevata e costante del campione istologico [1].

### Digital Pathology: i vantaggi

Per il patologo disporre di **un archivio digitale della casistica istologica** permette il **rapido confronto di campioni istologici dello stesso paziente o di pazienti con patologie similari**. Ciò permette un più accurato percorso diagnostico, evitando possibili diagnosi non corrette.

Il vetrino digitale permette inoltre al patologo di poter utilizzare metodologie e supporti diagnostici nell'attività diagnostica. Favoriscono la precisione del suo lavoro, come attività di quantificazione delle diverse componenti cellulari, di indici di proliferazione, di conte precise di positività per specifici marcatori e che finora erano effettuate come conte guidate dall'occhio umano e necessariamente poco precise e soggettive.

In altri termini **più qualità e precisione nella caratterizzazione di processi patologici va a tutto vantaggio del paziente**.

Altri fattori di cui tenere presente sono per esempio la **riduzione del numero di patologi** e **l'incremento del carico di lavoro per l'aumento di procedure diagnostiche** e di **screening**. Ma anche l'incremento della **complessità** dei casi e delle procedure per la loro corretta e precisa definizione diagnostica spingono per una più decisa evoluzione digitale.

## Criticità da risolvere

Ci sono sicuramente alcuni aspetti di notevole importanza che non sono ancora del tutto risolti. In primis, a differenza di quanto avvenuto in radiologia, **non è ancora stato definito uno standard per le immagini di Digital Pathology**.

Ogni ditta fornitrice di scanner ha un proprio standard di immagine digitale. Ciò **limita l'interscambio fra centri diversi di immagini digitali**, non tanto a fini diagnostici, ma **a fini di ricerca**.

Al momento attuale presso il laboratorio di anatomia patologica dell'Istituto San Donato vengono inoltre **digitalizzati giornalmente una media di circa 1200 vetrini**. Nell'archivio digitale sono presenti ormai più di **250.000 vetrini**.

Questi aspetti pratici hanno grande **impatto economico**. Riguardano la **conservazione**, per quanto **tempo e rapidità di accesso di queste immagini in archivi digitali**. Gli spazi necessari per conservare i vetrini digitali di centri ad alto volume di attività sono dell'ordine di **petabyte per anno**. Potrebbe infine essere **necessario conservare anche il vetrino fisico oltre all'immagine digitale**. Ma è ancora da definire. Infatti sono **problemi in cerca di una soluzione** in tempi relativamente brevi. Intanto, però, limitano in parte la diffusione di questa tecnologia.

## Programmi di supporto

Questo è tuttavia **solo l'aspetto immediato della rivoluzione** che il vetrino digitale apporterà nel lavoro del patologo. Il vetrino digitale rappresenta infatti il punto di partenza dell'applicazione di processi di elaborazione delle informazioni contenute nella sezione di tessuto, trasformato in un'immagine digitale. Essi vanno **da programmi di supporto diagnostico ad algoritmi di screening diagnostico ad algoritmi di intelligenza artificiale**. Questo scenario è attualmente nella sua fase iniziale. Ma promette di avere un impatto enorme nel **settore della diagnostica istopatologica**. Tuttavia, al momento, è ancora difficile predire quale sarà l'evoluzione in questo campo in rapidissimo progresso.

## La Digital Pathology e l'applicazione dell'AI

L'Intelligenza Artificiale (AI), proposto da McCarthy e coll. nel 1956 [2] è un termine generico che può abbracciare molti significati. Si riferisce alla possibilità che macchine possano essere utilizzate per predizioni o soluzioni di problemi alla stessa stregua della mente umana in simili situazioni.

E dopo 70 anni, da previsione, l'AI è realtà. Ma siamo solo agli inizi dell'applicazione pratica di strumenti di AI alla pratica diagnostica. Dunque è assai difficile predire quale impatto essa avrà, nella pratica diagnostica e quali ne saranno i tempi.

Nella pratica clinica, **il patologo basa la propria diagnosi istologica sul riconoscimento visuale**,

su valutazioni semi-quantitative di molteplici aspetti morfologici presenti nel vetrino fisico, integrandoli nel contesto clinico.

Con un percorso formativo specifico e della durata di molti anni, il patologo riesce ad **individuare ed estrarre dal campione gli aspetti morfologici rilevanti**. Comparati con criteri accumulati nella letteratura scientifica, inquadrati nei diversi ambiti dei processi patologici, permettono di classificare il quadro patologico. Sotto forma scritta, forniscono un **referto istopatologico per il medico curante**.

L'esperienza maturata dal patologo e l'utilizzo di linee guida standardizzate favoriscono un **percorso diagnostico standardizzato e l'accuratezza diagnostica**. Tuttavia si tratta sempre di un processo diagnostico dove mantengono un ruolo importante la soggettività e la diversa capacità di percezione visiva di ognuno, incluso il patologo.

### Tecniche sempre meno invasive

Questa attività è resa ancor più difficile dall'utilizzo di tecniche sempre meno invasive per l'acquisizione di campioni tessutali con dimensioni sempre più limitate. Al contempo è pressante la richiesta di referti istopatologici corredati con un sempre maggior numero di dati di valore prognostico e predittivo.

**Programmi di AI, già approvati per uso diagnostico**, sono in grado di **affiancare e forse sostituire il patologo per alcuni specifici quesiti diagnostici**.

Esempi di questa tipologia di programmi, già utilizzabili nella pratica clinica sono dedicati alla **diagnostica di biopsie prostatiche** per l'identificazione e la valutazione del grading dell'adenocarcinoma prostatico [3] e per l'identificazione di micrometastasi linfonodali nel carcinoma mammario [4].

### Machine Learning: apprendimento automatico supervisionato

Questi programmi sono stati elaborati attraverso un processo per molti aspetti simile al **training del patologo**, detto Machine Learning supervised, o **apprendimento automatico supervisionato** e che consiste nell'apprendere dall'analisi di un set casistico per raggiungere un risultato già definito in partenza.

L'analisi computazionale di un elevato numero di vetrini digitali con diagnosi predefinite effettua l'estrazione e la classificazione di molteplici aspetti morfologici. Attraverso **ripetuti cicli di valutazione e correzione**, permette di raggiungere lo stesso risultato diagnostico inizialmente definito dal patologo.

Dopo questa fase iniziale il **modello algoritmico** sviluppato va poi testato su ulteriori set di casistica per definirne sensibilità e specificità nella pratica diagnostica. Tuttavia questi **programmi**, di supporto e controllo diagnostico all'operato del patologo e la cui approvazione diagnostica è relativamente recente, **non hanno ancor avuto un'ampia applicazione routinaria**. Questa potrebbe indicare quanto migliorare la qualità e riproducibilità nella pratica diagnostica routinaria.

### La rete di Federated Learning

All'Istituto San Donato è stato avviato a questo scopo un **progetto di ricerca, finanziato dal Ministero della Salute**. Si ripromette di **valutare l'impatto di tali algoritmi sulla diagnosi ed il grading dell'adenocarcinoma prostatico in un'ampia casistica di pazienti sottoposti a biopsie prostatiche**.

Oltre al nostro Istituto come capofila, parteciperanno altri **due ospedali italiani, di Reggio Emilia e Trento**, attraverso una rete di **Federated Learning**, per una casistica complessiva di 1200 pazienti.

## Reti neurali e deep learning

A questa modalità di sviluppo computazionale, relativamente semplice, si sono aggiunte recentemente ulteriori e più complesse modalità di analisi e di elaborazione dei dati ricavabili dal vetrino digitale. Così pure per altri aspetti dell'**imaging medico**. Sono infatti disponibili più potenti unità di processazione. Esse hanno permesso lo **sviluppo di reti neurali ed il deep learning**, in grado di elaborare ed integrare non solo dati strutturati, ma anche di origine e composizione diversa. Tuttavia richiedono un **training di apprendimento più lungo e complesso**.

Gli **strumenti di AI moltiplicano le nostre capacità di ottenere informazioni qualitative e quantitative dai vetrini digitali, nascoste all'occhio umano**. Possono anche permettere l'integrazione dei dati derivanti da altre scienze omiche, quali **radiomica, genomica e proteomica**. L'obiettivo è quello di fornire un **quadro multidimensionale con più robuste potenzialità prognostiche e predittive (4-5)**. Infatti la morfologia delle cellule e dei tessuti normali e patologici che osserviamo nel vetrino fisico e nel vetrino digitale è il frutto di **specifici stati funzionali, governati da specifici corredi genetici, di trascrittomica e proteomica**.

La possibilità di individuare nel vetrino digitale tramite programmi di AI quali sono le alterazioni genomiche e trascrittomiche, determinanti di quella specifica morfologia, può produrre un salto di enorme portata alla diagnostica patologica. Infatti permette di indicare l'evoluzione di una determinata patologia neoplastica e la sua possibile risposta ai diversi tipi di trattamento.

## L'applicabilità dell'AI nella Digital Pathology

Questi scenari sono di estremo interesse. Tuttavia pongono una **serie di domande di ordine pratico e sulla loro possibile applicabilità in tutti i laboratori**. Lo sviluppo e l'**utilizzo di programmi di AI richiede investimenti e competenze specifiche**.

**Ingegneri informatici e Data Scientist sono i principali attori nello sviluppo degli algoritmi di AI** e la figura del patologo è essenzialmente l'utilizzatore finale. Ma nel contempo deve guidare lo sviluppo, valutare ed utilizzare i risultati di questa nuova tecnologia.

Quindi è necessario creare **gruppi interdisciplinari** che uniscano le competenze relative allo sviluppo ed all'utilizzo dei sistemi di AI, alle **competenze diagnostiche e cliniche nella loro applicazione pratica**. Significa quindi introdurre **nuove figure professionali nel percorso diagnostico e clinico**. Al contempo bisogna introdurre nuovi percorsi formativi per il patologo. Ma anche per le altre branche diagnostiche e cliniche. L'importante è che sappiano utilizzare e gestire questi nuovi strumenti di lavoro.

## Conclusioni

Sotto l'aspetto economico, serve lo sviluppo dei programmi di AI che richiedono investimenti e ricerca. Ormai sono numerose le ditte che propongono tali programmi.

Occorre capire quale sarà il loro costo e se questi programmi di AI sostituiranno il patologo nella sua funzione diagnostica.

Pur essendo difficile predire il futuro, viene da rispondere che non si rischia questo pericolo, almeno per un lungo periodo di tempo.

L'estrema varietà e complessità dei quadri patologici richiederanno ancora la **formazione, la preparazione e l'esperienza del patologo**. Possiamo prevedere che l'**AI ridurrà parte del lavoro del patologo per quanto riguarda alcuni compiti ripetitivi**, in diagnostiche meno complesse. Inoltre **ridurrà la soggettività, migliorando la riproducibilità e la standardizzazione della diagnostica patologica**.

Nel breve tempo, possiamo aspettarci che **programmi di AI aumentino anche le possibilità prognostiche e predittive nell'analisi di molte patologie neoplastiche**. Solo un'oncologia di **precisione** può permettere al patologo di **dedicare più tempo, attenzione e ricerca ai casi più complessi**.

## Bibliografia

- [1] Fraggetta F, L'Imperio V, Ameisen D, Carvalho R, Leh S, Kiehl TR, Serbanescu M, Racoceanu D, Della Mea V, Polonia A, Zerbe N, Eloy C. Best Practice Recommendations for the Implementation of a Digital Pathology Workflow in the Anatomic Pathology Laboratory by the European Society of Digital and Integrative Pathology (ESDIP). *Diagnostics* (Basel). 2021 Nov 22;11(11):2167. doi: 10.3390/diagnostics11112167. PMID: 34829514.
- [2] McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, 27(4), 12. <https://doi.org/10.1609/aimag.v27i4.1904>
- [3] Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, Brogi E, Reuter VE, Klimstra DS, Fuchs TJ. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med*. 2019 Aug;25(8):1301-1309.
- [4] van der Laak J, Litjens G, Ciompi F. Deep learning in histopathology: the path to the clinic. *Nat Med*. 2021 May;27(5):775-784. doi: 10.1038/s41591-021-01343-4. Epub 2021 May 14. PMID: 33990804.
- [5] Swanson K, Wu E, Zhang A, Alizadeh AA, Zou J. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. *Cell*. 2023 Apr 13;186(8):1772-1791. doi: 10.1016/j.cell.2023.01.035. Epub 2023 Mar 10. PMID: 36905928.
- [6] Reis-Filho JS, Kather JN. Overcoming the challenges to implementation of artificial intelligence in pathology. *J Natl Cancer Inst*. 2023 Jun 8;115(6):608-612. doi: 10.1093/jnci/djad048. PMID: 36929936; PMCID: PMC10248832

# IA e salute mentale, un rapporto complesso: usi pratici, strumenti, rischi

La riflessione sulla complessità dell'applicazione dell'intelligenza artificiale al supporto psicologico e alla salute mentale è oggi quanto mai centrale, perché, pur essendo consci delle potenzialità straordinarie offerte da questa tecnologia, è necessario un approccio equilibrato, etico e focalizzato sull'essere umano

Di **Luca Bernardelli**, Psicologo, Autore del libro "Guida Psicologica alla Rivoluzione Digitale", CEO di BECOME. Augmented Life, Cofounder di BOWMAN - Data Matter e Clelia Malighetti, Psicologa, Psicoterapeuta, PhD

**L'intelligenza artificiale (IA) può offrire diverse opportunità nel campo della psicologia, della psichiatria e della salute mentale.**

L'IA può essere utilizzata per aiutare nella diagnosi e nella valutazione dei disturbi mentali attraverso l'analisi di grandi quantità di dati, come i risultati di test psicologici, i dati biometrici e i modelli comportamentali, identificando pattern e segnali che possono aiutare nella diagnosi accurata di disturbi mentali.

Inoltre, l'intelligenza artificiale può **fornire supporto e counseling personalizzati**. Specifici chatbot possono interagire con gli individui, offrendo supporto emotivo, consulenza psicologica e suggerimenti per affrontare lo stress o i disturbi mentali lievi.

In aggiunta, l'intelligenza artificiale può essere impiegata per monitorare costantemente il benessere mentale degli individui, attraverso l'analisi dei dati provenienti da sensori indossabili, registrazioni vocali, dati dei social media e altre fonti. In questo modo, una IA potrebbe individuare segnali di stress, ansia o altri disturbi mentali, fornendo interventi preventivi e suggerendo modelli di comportamento sani e funzionali.

In più, l'IA può essere utilizzata per **creare app terapeutiche** che si adattano alle esigenze specifiche di ciascun individuo, come avviene, per esempio, per le "**terapie digitali**" (Digital Therapeutics, DTx), ormai normate in diverse Nazioni.

Infine, l'intelligenza artificiale può **accelerare lo sviluppo di nuovi farmaci** per i disturbi mentali, utilizzando algoritmi di apprendimento automatico e analizzando grandi set di dati per identificare potenziali obiettivi terapeutici e predire l'efficacia di determinati composti chimici.

Questi sono solo alcuni dei **numerosi casi d'uso** in cui le intelligenze artificiali potrebbero dare un contributo fondamentale alla ricerca, alla valutazione, all'intervento e alla prevenzione nell'ambito della salute mentale.

D'altra parte, per quanto l'intelligenza artificiale possa - e potrà - offrire vantaggi sempre più significativi, è importante sottolineare che **la consulenza fornita da professionisti della salute mentale rimane - e rimarrà? - fondamentale** per la cura di disagi e disturbi correlati.

L'IA solleva certamente **dilemmi etici riguardanti la privacy** e la responsabilità nelle decisioni prese dagli algoritmi ed è cruciale considerare le sue limitazioni, quali, per esempio, il rischio di bias nei dati che la alimentano o la mancanza di empatia.

Infatti, le istituzioni si trovano ad avere la necessità di governare adeguatamente le intelligenze artificiali ad alto rischio, come quelle che ricadono nell'alveo della salute e in particolar modo della salute mentale (come è accaduto all'Unione Europea con il recente [AI Act](#)).

Per esempio, le intelligenze artificiali usate in ambiti come il supporto psicologico, la psicoterapia e la psicodiagnostica possono essere considerate **ad alto rischio**, dato che sono associate a situazioni in cui un malfunzionamento o un uso improprio potrebbe comportare gravi conseguenze per la sicurezza, la privacy o l'integrità dei dati degli individui, che richiedono inevitabilmente un controllo e una regolamentazione più rigorosi finalizzati alla mitigazione dei rischi.

## Affective computing ed emotion AI

L'“Affecting Computing” nasce come disciplina ingegneristica, con l'obiettivo di sviluppare tecnologie in grado di percepire, riconoscere, comprendere e simulare i processi affettivi.

Il concetto moderno di affective computing è stato introdotto nel 1997 dalla professoressa Rosalind Picard, direttrice del centro di ricerca del Massachusetts Institute of Technology (MIT) e fondatrice e direttrice dell'Affective Computing Research Group presso il MIT Media Lab, oltre a diverse startup tecnologiche. Il suo lavoro ha dato il via a un'ondata di ricerche sulle interazioni emotive tra esseri umani e macchine, portando a quella che oggi viene chiamata “**Intelligenza Artificiale delle Emozioni**” (o “Emotion AI”).

Picard, in una ricerca pubblicata nel 2002, ha affermato che non è necessario che i computer abbiano una percezione emotiva, ma piuttosto la capacità di soddisfare i bisogni emotivi di una persona. In questo senso, i sistemi affettivi dovrebbero contribuire ai "bisogni empatici, di autoconsapevolezza e di esperienza" dell'utente, che dovrebbe semplicemente sentire che l'IA ha riconosciuto i suoi sentimenti dopo l'interazione.

I sistemi abilitati all'elaborazione affettiva rilevano i - e reagiscono ai - cambiamenti nel linguaggio, nel tono di voce e nelle sfumature delle espressioni facciali di una persona. Lo fanno raccogliendo i dati attraverso sensori fisici, come videocamere e microfoni, e analizzando tali informazioni sulla base di esperienze interattive e Big Data pre-caricati. Per esempio, un sistema di affective computing potrebbe essere in grado di riconoscere che una persona si sente triste e rispondere riproducendo una canzone rilassante o visualizzando un messaggio di conforto.

L'affective computing coinvolge essenzialmente due processi, il **riconoscimento delle emozioni** e la **sentiment analysis** (Peng et al., 2018; Poria et al., 2017; Rouast et al., 2019).

Il riconoscimento delle emozioni mira a rilevare lo stato emotivo degli esseri umani e si concentra principalmente sul riconoscimento delle emozioni visive (VER), sul riconoscimento delle emozioni audio/vocali (AER/SER) e sul riconoscimento delle emozioni fisiologiche (PER) (Wang et al., 2022).

La sentiment analysis, invece, si concentra principalmente sulle valutazioni testuali e sull'estrazione di opinioni su eventi sociali, campagne di marketing e preferenze di prodotti. Il risultato di questa tipologia di analisi è tipicamente positivo, negativo o neutro (Poria, Cambria & Gelbukh, 2016).

In quanto interdisciplinare e di ampia applicazione, l'affective computing coinvolge la psicologia, le scienze fisiologiche, le tecnologie cognitive e informatiche (Poria, Cambria & Gelbukh, 2016) e unisce tre tecnologie principali: il rilevamento e l'interpretazione delle emozioni, il ragionamento dialogico che utilizza le informazioni emotive e la generazione e sintesi delle emozioni.

Un **chatbot affettivo** è un sistema che interagisce con gli esseri umani utilizzando tecnologie affettive per rilevare le emozioni, decidere e simulare risposte affettive. Può avere un sistema autonomo di elaborazione del linguaggio naturale con almeno queste componenti: analisi del segnale e riconoscimento automatico del parlato, analisi semantica e politiche di dialogo, generazione di risposte e sintesi vocale. Questo agente conversazionale può essere un mero assistente vocale, un personaggio virtuale in 2D o in 3D o un robot fisicamente incarnato.

I **robot sociali** sono già utilizzati in ambito clinico-assistenziale per aiutare malati, anziani e/o disabili in casa o in ospedale. I robot sono disponibili 24x7, sono pazienti, possono superare le carenze percettive (sordità, disturbi visivi, ecc.) e possono rapidamente accedere alle informazioni disponibili su Internet. Inoltre, sono preziosi per registrare continuamente dati e inviarli a medici e/o psicologi per rilevare eventuali disagi o comportamenti anomali.

Nel caso di disturbi dello spettro autistico, patologie neurodegenerative o gravi disabilità, il robot potrebbe addirittura essere migliore dell'uomo nell'interazione con le persone, soprattutto se molto sole. In questi casi, una macchina potrebbe addirittura aiutarle a prevenire quelle forme depressive che potrebbero degenerare in una demenza.

In conclusione, l'obiettivo dell'affective computing è quello di migliorare l'interazione uomo-computer attraverso sistemi più empatici e intelligenti.

Nel setting clinico chiamato "**cyberterapia**", ovvero il contesto terapeutico che impiega le nuove tecnologie per la valutazione e l'intervento terapeutico, l'affective computing offre diverse modalità di supporto, per esempio nella diagnosi e nella gestione dell'ansia, dello stress, della depressione e nella regolazione delle emozioni (Parsons & Rizzo, 2008; Riva et al., 2014; Villani, Lucchetta, Preziosa, & Riva, 2009), rilevando i cambiamenti a livello fisiologico (per esempio, nel battito cardiaco, nella respirazione, nella voce e nell'attività motoria, nella temperatura corporea per monitorare o/e prevenire episodi ansiosi o depressivi). Queste informazioni potrebbero essere utilizzate per fornire una diagnosi e un trattamento più precisi e personalizzati.

Un altro contesto riguarda il monitoraggio dei **disturbi del sonno** attraverso lo screening dell'attività cerebrale, la frequenza cardiaca e la respirazione durante il sonno, per rilevare eventuali anomalie, come l'apnea notturna o disturbi del sonno.

Infine, nella **riabilitazione cognitiva** i sensori potrebbero per esempio rilevare cambiamenti nella memoria e nel comportamento di persone affette da demenza, consentendo di personalizzare l'intervento e migliorare la qualità di vita del paziente.

## Psicodiagnostica e fenotipizzazione digitale

L'Organizzazione Mondiale della Sanità stima che entro il 2035 ci sarà un deficit globale di circa 12,9 milioni di operatori sanitari. La digitalizzazione dei servizi sanitari è destinata a svolgere un ruolo fondamentale nel miglioramento dell'efficienza dei costi sanitari.

Un esempio ne è il “**trriage digitale**”, in cui i pazienti compilano un questionario digitalizzato e vengono guidati direttamente al livello di assistenza appropriato in base alle loro risposte. Di conseguenza, un sistema di triage accurato potrebbe potenzialmente risparmiare su costi e tempi, qualora fosse in grado di svolgere il ruolo di addetto al triage.

Nel mondo della digitalizzazione in ambito psicologico, degni di attenzione sono gli strumenti basati sull'intelligenza artificiale, chatbot in grado di offrire assistenza 24 ore su 24 ai pazienti.

Nell'ambito della psicodiagnostica l'intelligenza artificiale sta prendendo sempre più spazio nel tentativo di facilitare, velocizzare e migliorare il processo di valutazione. Strumenti in grado di riconoscere specifiche neuro-diversità o alcuni disturbi rappresentano una nuova generazione di strumenti psicodiagnostici all'avanguardia.

Dalla rilevazione delle alterazioni dei movimenti oculari sintomatiche della schizofrenia all'analisi delle funzioni attentive e di memoria, dall'imaging automatizzato del comportamento ai fini di diagnosi di un disturbo dello spettro autistico all'analisi vocale finalizzata all'identificazione di sintomi di ansia e depressione, dalla valutazione neuropsicologica dei disturbi da deficit di attenzione e iperattività (DDAI) e dei disturbi specifici dell'apprendimento (DSA) alla valutazione psicologica di un paziente funzionale a una gestione più rapida ed efficiente delle situazioni a rischio, le AI stanno prendendo spazio anche nel comparto della salute mentale.

In tutti i casi, il potenziamento delle diagnosi precoci diventa un fattore essenziale per avviare celermente percorsi di supporto appropriati alle varie condizioni

La psicodiagnostica digitale potrebbe essere notevolmente accelerata dalla “**fenotipizzazione digitale**” che, grazie ai “**biomarcatori digitali**”, dati generati principalmente dall'uso intensivo dello smartphone, combinati ai dati soggettivi prodotti dagli stessi utenti, rendono possibile mappare gli “**indicatori di malattia**”, indici da cui ricavare in tempo reale lo stato di salute psicofisica di un individuo, allo scopo di salvaguardarne la salute.

La fenotipizzazione consiste essenzialmente in un nuovo metodo di osservazione, basato sulla raccolta di dati e sulla profilazione comportamentale grazie allo smartphone.

Per interazione uomo-macchina non si fa riferimento solo a ciò che si digita ma a come si digita, per esempio l'analisi della latenza tra spazio e carattere, o l'intervallo tra scorrimento e clic, ovvero tutti quegli elementi che potrebbero rivelare informazioni utili rispetto ai tratti cognitivi o a stati affettivi delle persone (Dagum, 2018).

La promessa della fenotipizzazione digitale è una misurazione ecologica, ovvero che avviene nel contesto dell'esperienza vissuta della persona e non in un contesto clinico-ambulatoriale.

Per esempio, se pensiamo alla depressione post partum, i segnali fisiologici e psicologici di una neomamma con diagnosi di depressione analizzati durante l'allattamento di notte, ci immaginiamo che siano probabilmente diversi e più informativi di quelli che la madre stessa può riferire al mattino del giorno successivo.

Senza altro il monitoraggio e l'utilizzo di questi dati è accompagnato da diverse questioni etiche, quali: si tratta di una mera misurazione o di una forma di sorveglianza? È una misurazione intrusiva? Come verranno usati i dati? Chi li può detenere?

In ambito psichiatrico, uno dei segnali più informativi potrebbe risiedere nella cronologia delle ricerche su Internet o nei post sui social media, che potrebbero confessare un intento suicida o i primi segni di psicosi. Il valore di queste informazioni è superiore all'intrusione nella privacy necessaria per ottenerle?

Tutte queste questioni fanno parte di un dibattito sempre più attivo, come merita ogni nuova promettente tecnologia.

## Chatbot psicologici nella ricerca

Sono numerosi gli enti di ricerca e le università che hanno concepito chatbot nell'ambito della ricerca psicologica e neuroscientifica.

Tra i più importanti, possiamo citare senz'altro **"Eliza"**, il primo chatbot psicologico sviluppato nel 1966 dal Joseph Weizenbaum al Massachusetts Institute of Technology di Boston. Il suo obiettivo era di realizzare la parodia di una conversazione tra un terapeuta rogersiano e il suo paziente nei primi colloqui, in cui le domande non prevedono una profondità, utilizzando semplici regole di riconoscimento del linguaggio.

Eliza faceva domande ai suoi interlocutori, cercando di comprendere il loro stato emotivo e di fornire risposte appropriate basate su frasi predefinite. Spesso, le sue risposte erano in realtà riflessioni sulle domande dell'utente, creando un effetto di "rispecchiamento" simile a quello utilizzato nella terapia di Carl Rogers, a cui peraltro si ispirava.

Sebbene, ovviamente, Eliza non fosse in grado di offrire una vera terapia o diagnosi psicologiche, ha suscitato un grande interesse e ha avuto un impatto significativo nel campo dell'intelligenza artificiale e della psicologia. Ha dimostrato come un programma informatico relativamente semplice potesse simulare una conversazione terapeutica e sollevare questioni interessanti sull'interazione uomo-macchina.

Eliza ha aperto la strada a numerosi sviluppi successivi nel campo dei chatbot psicologici, che oggi sono utilizzati in molte applicazioni per fornire supporto emotivo, consigli e interazioni terapeutiche basate sull'intelligenza artificiale.

Un'altra chatbot storico è **"Parry"**, creato nel 1972 dallo psichiatra Kenneth Colby, durante i suoi studi alla Stanford University, con l'obiettivo di esplorare la possibilità che un'intelligenza artificiale potesse essere in grado di simulare o imitare i disturbi mentali umani, nel caso specifico i disturbi paranoici, dimostrando abilità sorprendenti nel riprodurre le possibili risposte di un paziente paranoico. Parry rappresenta uno dei primi esperimenti di chatbot psicologici e ha avuto il merito di sollevare diverse questioni etiche e morali, contribuendo a gettare le basi per ulteriori sviluppi nell'utilizzo dell'intelligenza artificiale nel campo della salute mentale.

Il più recente chatbot **"SimSensei"** è un assistente virtuale progettato per fornire supporto emotivo e valutazione dello stato emotivo delle persone. È stato sviluppato da un team di ricercatori guidato dal dott. Skip Rizzo presso l'Institute for Creative Technologies dell'Università della California del Sud. SimSensei utilizza la tecnologia dell'intelligenza artificiale e dell'elaborazione del linguaggio naturale per interagire con gli utenti e raccogliere informazioni sul loro stato emotivo, fornendo un ambiente sicuro per esprimere i propri sentimenti. Il chatbot può condurre interviste interattive in cui fa domande sulla salute mentale, sullo stress, sull'umore e sul benessere generale. L'obiettivo di SimSensei è di fornire un supporto aggiuntivo per individuare potenziali segnali di disagio emotivo

e fornire risorse di auto-aiuto. Il chatbot utilizza algoritmi di intelligenza artificiale per analizzare le risposte degli utenti, individuando indicatori di rischio e offrendo suggerimenti o riferimenti a professionisti della salute mentale qualificati, se necessario.

Ad esempio uno studio pubblicato nel 2016 sulla rivista *Computers in Human Behavior* ha dimostrato l'efficacia di Simsensei nel ridurre i sintomi di depressione e ansia nei membri del servizio militare (Rizzo et al., 2016). Un altro studio pubblicato nel 2019 sul *Journal of Medical Internet Research* ha analizzato l'efficacia di Simsensei nel ridurre i sintomi di depressione e ansia nei giovani adulti e i risultati hanno dimostrato che Simsensei era efficace nel ridurre i sintomi della depressione e dell'ansia, con effetti mantenuti a un follow-up di 3 mesi (Palanica et al., 2019). Un ulteriore studio pubblicato nel 2021 sul *Journal of Medical Internet Research* ha analizzato l'accettabilità e la fattibilità di Simsensei come strumento di supporto alla salute mentale per le persone affette da dolore cronico. I risultati hanno dimostrato che Simsensei è uno strumento efficace per fornire supporto alla salute mentale delle persone affette da dolore cronico e i partecipanti al gruppo Simsensei hanno registrato una riduzione significativa dei sintomi di depressione e ansia rispetto al gruppo di controllo.

Una recente “scoping review” (Bending et al., 2022) sui chatbot impiegati in psicologia clinica e psicoterapia ha mostrato risultati promettenti per quanto riguarda la fattibilità, l'accettabilità e l'efficacia preliminare dei chatbot per favorire la salute mentale.

Il chatbot “**MYLO**” (Manage Your Life Online) offre un programma di auto-aiuto per la risoluzione dei problemi quando una persona è in difficoltà (Gaffney et al., 2014).

Il chatbot “**SHIM**” è un programma di auto-aiuto per migliorare il benessere mentale (Ly et al., 2017).

Il chatbot “**SABORI**” (Suganuma et al., 2018) offre un programma di auto-aiuto preventivo per la promozione della salute mentale.

Il chatbot “**SISU**” simula una conversazione modellata sul paradigma della scrittura terapeutica, con elementi della terapia “ACT” (Bendig et al., 2018).

Gli studi considerati da questa revisione sono per lo più studi pilota, composti da piccoli campioni e quindi con una potenza statistica insufficiente per parlare di efficacia terapeutica. Gli studi analizzati da questa review mostrano benefici da parte dei partecipanti per variabili psicologiche come il benessere, lo stress e la depressione.

I principali vantaggi di un chatbot nel campo della salute mentale sono l'accessibilità e la scalabilità, un supporto emotivo automatizzato e la possibilità di un monitoraggio continuativo in cui è possibile richiedere l'intervento di professionisti della salute mentale.

Questo è un ambito che si sta evolvendo molto rapidamente e in cui le innovazioni sono ormai quotidiane.

## Chatbot psicologici di mercato

In realtà, sul mercato sono già presenti anche chatbot per il counseling psicologico che stanno erogando servizi alla collettività.

Il più famoso - e finanziato - è “[WOEBOT](#)”, sviluppato da un team di ricercatori dell'Università di Stanford, guidato dal dott. Alison Darcy, un professore associato di ricerca presso il Dipartimento di Psichiatria e Scienze del Comportamento della Stanford University.

Woebot è un'applicazione di intelligenza artificiale sviluppata per fornire supporto emotivo e consulenza basata su evidenze scientifiche, utilizzando principi di terapia cognitivo-comportamentale (CBT) per aiutare le persone a gestire lo stress, l'ansia e l'umore. Sono stati condotti diversi studi scientifici per valutare l'efficacia di Woebot.

Un esempio di studio è stato pubblicato sulla rivista scientifica "JMIR Mental Health" nel 2017. Lo studio ha coinvolto 70 partecipanti adulti con sintomi depressivi da lievi a moderati, mostrando una riduzione significativa dei sintomi depressivi rispetto al gruppo di controllo (Fitzpatrick, Darcy, & Vierhile, 2017).

Un altro studio clinico randomizzato condotto nel 2020 ha esaminato l'effetto di Woebot su 192 donne con sindrome post-partum, indicando risultati incoraggianti in termini di ansia e depressione (Ramachandran et al., 2020).

Questi studi suggeriscono che l'utilizzo di Woebot può essere efficace nel ridurre i sintomi di ansia e depressione, ma è sempre importante sottolineare che Woebot non è stato progettato per sostituire la terapia tradizionale o l'intervento umano, ma per fornire un supporto emotivo supplementare, soprattutto a chi non può o non riesce ad accedere facilmente a servizi di consulenza psicologica tradizionali.

Un altro chatbot per la salute mentale che utilizza l'intelligenza artificiale e le tecniche di terapia cognitivo-comportamentale (CBT) è “[WYSA](#)”, che fornisce supporto emotivo e aiuta gli utenti a gestire la propria salute mentale.

In uno studio pubblicato nel 2020 sul Journal of Medical Internet Research, i ricercatori hanno analizzato l'efficacia di WYSA nel ridurre i sintomi di ansia e depressione nei giovani adulti. I risultati hanno dimostrato che Wysa è stato efficace nel ridurre i sintomi di ansia e depressione, e l'effetto si è mantenuto a un follow-up di 3 mesi.

In un successivo studio pubblicato nel 2021 sul Journal of Affective Disorders è stata analizzata l'efficacia di Wysa nel ridurre i sintomi di depressione negli studenti universitari. Anche qui, i risultati hanno dimostrato che Wysa è stato efficace nel ridurre i sintomi della depressione e l'effetto si è mantenuto a un follow-up di 3 mesi.

Un altro studio, pubblicato nel 2022 sul Journal of Medical Internet Research, ha dimostrato che Wysa è uno strumento efficace per ridurre i sintomi di ansia e depressione nelle persone con dolore cronico. Sebbene siano necessarie ulteriori ricerche per determinarne l'efficacia per altre popolazioni e condizioni di salute mentale, i risultati ottenuti finora suggeriscono che WYSA ha il potenziale per rappresentare una valida integrazione ai trattamenti per la salute mentale esistenti (Gupta, Malik & Sinha, 2022).

Probabilmente, nei prossimi anni assisteremo al proliferare di agenti conversazionali esperti in psicologia che, attraverso varie forme (app di messaggistica istantanea, avatar antropomorfi, meta-umani in realtà virtuale, robot affettivi) erogheranno servizi 24x7 a una popolazione sempre più ampia.

Anche “[ChatGPT](#)”, il più famoso e complesso motore di comprensione e generazione di linguaggio naturale, è in grado di fornire interessanti indicazioni sulla salute mentale e sul benessere emotivo, pur non essendo stato creato con quell’obiettivo.

Se provaste a porre domande quali: “Mi sento triste in questo periodo: hai dei suggerimenti per superare questo momento difficile?”, vi rendereste conto della stupefacente capacità di indirizzarle.

Le risposte di questa IA alle domande psicologiche degli utenti, infatti, sono molto puntuali, compreso un disclaimer finale – ripetuto più volte – che orienta la persona bisognosa d’aiuto verso un terapeuta in carne e ossa.

## Chatbot psicologici da attenzionare

Esistono chatbot di stampo psicologico molto particolari. Tra questi, meritano senz’altro una citazione Replika e Norman.

L’app “**Replika**” è nata nel 2017 dall’azienda californiana Luka ed è un’interfaccia conversazionale con un avatar virtuale che fornisce supporto emotivo e relazionale, imparando le preferenze degli utenti che interagiscono con lei.

Recentemente, nel folto gruppo Facebook chiamato “Replika: Romantic Relations” è sorto addirittura il problema delle molestie sessuali e delle avance indesiderate di questa IA. Peraltro, nel febbraio 2023, il [Garante della Privacy Italiano](#) ha imposto il blocco del trattamento dei dati degli utenti italiani di questa app, in quanto in evidente violazione del GDPR.

A differenza di Replika, il chatbot “**Norman**”, è stato esplicitamente progettato nel 2018 dal Massachusetts Institute of Technology come un esperimento per esplorare il lato oscuro dell’intelligenza artificiale.

Ispirato al personaggio Norman Bates del film “Psycho” di Alfred Hitchcock, questa IA è stata addestrata su immagini violente e disturbanti al fine di investigare come l’esposizione a determinati contenuti possa influenzare comportamenti e linguaggi anche da parte di una intelligenza artificiale.

Il risultato di questo addestramento ha portato Norman a sviluppare una visione distorta e negativa del mondo. L’aspetto cruciale di questo esperimento è stato l’accento posto sulla necessità di adottare approcci responsabili e di garantire l’equità e l’obiettività nelle applicazioni dell’intelligenza artificiale.

Infine, “**Forefront.AI**” e altri chatbot internazionali in fase sperimentale, propongono avatar antropomorfi specialistici che si presentano come “psicologi abilitati”, pur non essendolo...

## Nuove problematiche psicologiche da ia

L’intelligenza artificiale potrebbe rappresentare un pericoloso innesco per nuove tipologie di **tecnodipendenze** (Ellis, 2019).

L’uso eccessivo o problematico dell’IA, infatti, potrebbe portare a sviluppare una dipendenza da chatbot terapeutici o da altre applicazioni che propongono interazioni virtuali.

Inoltre, l'intelligenza artificiale potrebbe generare ansia e stress a causa di preoccupazioni riguardanti la sicurezza delle informazioni personali o la paura di essere rimpiazzati da tecnologie intelligenti.

In più, l'intelligenza artificiale potrebbe portare all'affermazione di avatar virtuali o agenti intelligenti che simulano personalità e interazioni umane, sollevando questioni legate all'identità digitale e alla confusione tra il sé virtuale e il sé reale.

Peraltro, l'IA, spinta eccessivamente, potrebbe portare a una diminuzione della fiducia nella capacità umana di prendere decisioni e risolvere problemi.

Più in generale, l'intelligenza artificiale solleva dibattiti etici complessi, come l'autonomia umana, l'equità e il controllo delle tecnologie, che potrebbero avere un impatto sulla salute mentale delle persone, causando preoccupazione, confusione o senso di impotenza collettivi.

In uno scenario in cui, secondo un recente [studio di Goldman Sachs](#), strumenti simili a ChatGPT potrebbero mettere a rischio circa 300 milioni di posti di lavoro entro il 2030, l'intelligenza artificiale potrebbe essere la principale fonte di “**tecnofobia**”, ovvero la paura o all'avversione nei confronti delle tecnologie. Molte persone potrebbero temere che questa tipologia di automazione riduca il ruolo umano, causando incertezza professionale e ansia riguardo al futuro, con un inevitabile impatto sulla salute mentale.

Certo, l'introduzione dell'intelligenza artificiale porterà a nuove opportunità di lavoro e alla creazione di nuovi settori, ma sarà necessario fornire risorse e programmi di formazione per aiutare i lavoratori a sviluppare competenze complementari alle tecnologie emergenti e a promuovere la riconversione professionale, al fine di mitigare possibili diffusi disturbi dell'adattamento.

Inoltre, l'intelligenza artificiale può sollevare preoccupazioni riguardo alla sicurezza dei dati personali, dato che violazioni o gestione inappropriata dei propri dati possono portare alla divulgazione non autorizzata di informazioni personali. Questo può contribuire alla paura di essere esposti a rischi legati alla propria privacy.

In aggiunta, l'uso pervasivo della tecnologia e le sempre più potenti intelligenze artificiali potrebbero contribuire a creare una cultura in cui il monitoraggio e la sorveglianza sono considerati la norma, generando così un senso di inquietudine e una sensazione di perdita della propria riservatezza e dell'autonomia.

Per di più, va considerato l'effetto dei “filtri intelligenti” che modificano i volti umani e che stanno generando un'epidemia di “**dismorfie da social**” o “**dismorfie da selfie**”, che riguardano la preoccupazione eccessiva per il proprio aspetto fisico, particolarmente legate all'immagine che una persona ha di sé, in costante crescita dopo l'avvento dei social media e di una cultura dell'immagine estremizzata, che porta a sentirsi perennemente giudicati (Ryding & Kuss, 2020).

Un ulteriore fenomeno particolarmente pericoloso è il “**deepfake**”, che usa l'intelligenza artificiale per creare contenuti audiovisivi ingannevoli e compromettenti, ma che sembrano autentici, facendo sembrare che una persona stia dicendo o facendo qualcosa che non ha realmente detto o fatto. Sebbene il deepfake possa essere utilizzato in modo divertente o creativo, presenta anche alcuni rischi potenziali per la salute mentale delle persone (Westerlung, 2019).

In aggiunta, le impronte digitali e i dati provenienti dai social media sono la base per creare anche i cosiddetti “**ghostbot**”, software che, grazie all'intelligenza artificiale, possono generare

reincarnazioni digitali di persone defunti, portando a estendere il tema della protezione della privacy anche post-mortem.

Insomma, la società avrà sempre più a che fare con forme di abuso digitale che possono causare gravi danni emotivi, psicologici e sociali.

Per affrontare tutte queste preoccupazioni, è importante sostenere la formazione continua della collettività per adattarsi ai cambiamenti tecnologici, introdurre politiche di incentivazione dell'innovazione responsabile e lavorare a una società in cui l'intelligenza artificiale sia utilizzata per migliorare la qualità della vita.

E proprio le professioni psicologiche dovranno gradualmente dotarsi delle conoscenze relative alle tecnodipendenze e ai disagi correlati al digitale, che faranno sempre più parte degli aggiornamenti dei prossimi manuali dei disturbi psichiatrici.

## La psicologia dell'IA e dei big (mental) data

Lo **psicologo dell'intelligenza artificiale e dei big data** è una nuova professione in evoluzione che richiede competenze molto specifiche che vanno oltre il perimetro degli attuali percorsi formativi psicologici tradizionali.

Il professionista formato in questo campo dovrebbe avere una comprensione solida dei principi, delle tecniche e delle applicazioni, inclusi i modelli di apprendimento automatico e profondo, l'elaborazione del linguaggio naturale e l'analisi dei dati.

Inoltre, dovrebbe essere formato in **"Data Science"**, la disciplina che combina statistica, informatica e analisi dei dati per estrarre conoscenze e informazioni significative dai dati, identificando modelli, relazioni e tendenze per prendere decisioni informate. Questa conoscenza è essenziale quando si tratta di applicazioni nei campi della psicologia, della salute mentale e delle scienze sociali, in cui le grandi moli di dati raccolti - i cosiddetti **"Big Mental Data"** - sono complesse e influenzate da fattori umani e sociali.

Inoltre, una formazione specifica di questo tipo consentirebbe allo psicologo di comprendere i modelli di apprendimento utilizzati dall'IA, di valutarne l'efficacia, identificandone i bias nei dati o negli algoritmi e prendendo misure per mitigare tali problemi, contribuendo a progettarne di migliori.

Inoltre, il professionista della salute mentale potrebbe fornire supporto psicologico a coloro che si trovano ad affrontare sfide legate all'uso dell'intelligenza artificiale, quali la mitigazione degli effetti della già citata tecnofobia, l'adattamento ai cambiamenti tecnologici e l'aiuto nel prendere decisioni informate sull'utilizzo delle tecnologie digitali.

È evidente che queste figure ibride e complesse, diverranno presto pregiate e attrattive, in particolar modo per le organizzazioni clinico-sanitarie e per le istituzioni e gli enti regolatori.

Questo settore emergente continuerà a evolversi rapidamente nel tempo e richiederà un costante aggiornamento delle competenze professionali.

## IA e professionisti della salute mentale

L'avvento dell'intelligenza artificiale offre diverse opportunità lavorative per i professionisti della salute mentale.

Questi specialisti, infatti, possono contribuire allo sviluppo di strumenti di valutazione basati sull'IA per la misurazione di costrutti psicologici, consentendo una valutazione più accurata, per esempio, delle caratteristiche di personalità o delle specifiche abilità cognitive di un individuo.

Una figura “psy” potrebbe inoltre coadiuvare l'implementazione di agenti conversazionali in grado di fornire interventi psicologici personalizzati o assumere ruoli di supporto nella diagnosi dei disturbi mentali, fornendo un aiuto digitale nella valutazione dei sintomi e nell'identificazione delle condizioni psicologiche.

Oltre a ciò, intelligenze artificiali costruite ad hoc, potrebbero aiutare i professionisti ad analizzare grandi quantità di dati psicologici, al fine di individuare modelli e correlazioni nascoste, con l'obiettivo di fornire una migliore comprensione dei processi psicologici, delle cause dei disturbi mentali e delle risposte agli interventi terapeutici.

Infine, l'IA può offrire nuove possibilità di ricerca nei campi della psicologia e della psichiatria, consentendo lo studio di questioni complesse e magari addirittura l'identificazione di nuovi approcci terapeutici.

Occorre sempre sottolineare che l'impiego dell'IA nell'ambito della salute mentale richiede un approccio etico e responsabile, che tuteli la privacy dei pazienti/utenti, minimizzi il rischio di bias e massimizzi i benefici per le persone.

## **L'IA nella ricerca psicologica e neuroscientifica**

Dal punto di vista della ricerca psicologica e neuroscientifica, sono diversi gli ambiti in cui l'intelligenza artificiale potrebbe apportare benefici.

Un primo esempio è la comprensione della vita quotidiana al di fuori del laboratorio, dato che i sensori presenti negli smartwatch, nei wearable e in tutta la pletora di oggetti “IOMT” (Internet of Medical Things) disponibili sul mercato, sono dispositivi di consumo sempre più comuni, che consentono di raccogliere enormi moli di dati sia fisiologici sia comportamentali. Tutte queste informazioni potrebbero essere valorizzate, per esempio, per comprendere meglio le dinamiche soggettive, o addirittura per prevedere se un individuo si stia avviando verso l'esordio di un disturbo psicopatologico (Graham et al., 2019).

Un secondo esempio è la comprensione delle connessioni tra mente, cervello e corpo, dato che, appunto, la raccolta di dati sta diventando sempre più economica e sempre più semplice, sia per i ricercatori sia per i consumatori. Da anni, le neuroscienze e la psicofisiologia stanno esaminando come specifici processi psicologici sono correlati ai processi cerebrali e corporei e, nel corso del tempo, sarà possibile raccogliere campioni di dati oggettivi di dimensioni sempre maggiori e sempre più funzionali alle domande di ricerca (Ng & Leung, 2020).

Un terzo esempio è il deepfake per la ricerca psicologica (Karnouskos, 2020). Sebbene, infatti, questa tecnologia AI abbia implicazioni potenzialmente molto pericolose per l'intera collettività, potrebbe d'altra parte consentire di sperimentare modi radicalmente nuovi di impostare esperimenti psicologici, per esempio, manipolando l'età, il sesso o l'etnia percepiti da una persona,

consentendole di sperimentare le differenze di interazione tramite l'immedesimazione nel corpo e nell'identità di qualcun altro.

Gli esempi di applicazioni in questi ambiti potrebbero essere davvero tantissimi e limitati alla sola fantasia dei ricercatori.

## Etica dell'IA e umanesimo digitale

L'uso dell'IA nell'ambito della salute mentale pone diverse sfide etiche da valutare con grande attenzione e consapevolezza.

Tra le più importanti, citiamo sicuramente la privacy e la sicurezza dei dati, dato che gli algoritmi di intelligenza artificiale richiedono l'accesso a dati personali e (molto) sensibili per fornire valutazioni e trattamenti, sollevando preoccupazioni sulla privacy e sulla sicurezza dei dati.

Inoltre, gli algoritmi possono essere influenzati da bias e discriminazioni presenti nei dati che ne hanno alimentato l'apprendimento o in coloro che li hanno concepiti, portando a disparità di trattamento nei confronti di determinati gruppi di individui, amplificando le disuguaglianze esistenti.

È fondamentale affrontare questi temi sin dall'inizio, con l'obiettivo chiaro di garantire equità e inclusione nella fornitura di servizi di salute mentale basati su intelligenze artificiali.

In aggiunta, gli algoritmi di intelligenza artificiale spesso operano come "scatole nere", rendendo difficile comprendere il processo decisionale che li guida e complesso spiegare ai professionisti sanitari e ai pazienti/utenti come vengono raggiunte determinate conclusioni o suggerimenti terapeutici. La trasparenza e l'interpretabilità degli algoritmi è importante per garantire la fiducia, il coinvolgimento e la comprensione delle decisioni prese.

L'uso di algoritmi nell'ambito della salute mentale potrebbe sollevare questioni riguardanti l'autonomia degli individui, dato che l'automazione delle decisioni potrebbe ridurre la libertà di scelta o limitare la loro partecipazione attiva nei processi di cura.

Un altro tema importante riguarda la responsabilità: se un algoritmo diagnostico o terapeutico commette un errore, chi è responsabile della qualità dei dati utilizzati o delle decisioni prese dall'intelligenza artificiale?

Affrontare queste sfide richiede una governance rigorosa, normative chiare e un dialogo sempre aperto tra tecnologi, professionisti della salute mentale, esperti di etica, associazioni di pazienti e istituzioni. Il fine è implementare algoritmi che siano trasparenti ed equi e che rispettino i valori fondamentali della collettività nella fornitura di servizi digitali di salute mentale.

Un approccio che promuove un equilibrio sostenibile tra l'uso dell'intelligenza artificiale e l'importanza dei valori umani fondamentali è **l'umanesimo digitale**.

Pur avvalendosi dell'IA per fornire strumenti di valutazione, interventi personalizzati e supporto emotivo, l'umanesimo digitale mette in evidenza l'importanza della presenza e dell'interazione umana nella pratica clinica, promuovendo l'utilizzo dell'intelligenza artificiale per migliorare l'accuratezza e l'accessibilità dei servizi digitali di salute mentale, senza perdere di vista l'essenza dell'esperienza umana.

Oggi la riflessione sulla complessità dell'applicazione dell'intelligenza artificiale al supporto psicologico e alla salute mentale è quanto mai centrale, perché, pur essendo consci delle potenzialità straordinarie offerte da questa tecnologia, è necessario un approccio equilibrato, etico e focalizzato sull'essere umano.

Solo attraverso una continua interazione tra la scienza, la psicologia e l'etica, infatti, sarà possibile accompagnare le prossime fasi della trasformazione digitale e un futuro in cui le persone siano sempre più al centro delle decisioni che riguardano la loro salute.

## Bibliografia e sitografia

Bendig, E., Erb, B., Schulze-Thuesing, L., & Baumeister, H. (2019). The next generation: chatbots in clinical psychology and psychotherapy to foster mental health—a scoping review. *Verhaltenstherapie*, 1-13.

Bernardelli, L. (2022). Guida Psicologica alla Rivoluzione Digitale.

Carey, TA. *The Method of Levels: How to do psychotherapy without getting in the way*. Hayward, CA: Living Control Systems Publishing; 2006. ISBN: 097401554

Dagum, P. (2018). Digital biomarkers of cognitive function. *NPJ digital medicine*, 1(1), 10.

Dahl, J., & Lundgren, T. (2006). Acceptance and Commitment Therapy (ACT) in the Treatment of Chronic Pain. In R. A. Baer (Ed.), *Mindfulness-based treatment approaches: Clinician's guide to evidence base and applications* (pp. 285–306). Elsevier Academic Press.  
<https://doi.org/10.1016/B978-012088519-0/50014-9>

Damasio, A. (2003). Feelings of emotion and the self. *Annals of the New York Academy of Sciences*, 1001(1), 253-261.

Damasio, A. R. (1994). Descartes' error and the future of human life. *Scientific American*, 271(4), 144-144.

Damasio, A. R. (1999). How the brain creates the mind. *Scientific American*, 281(6), 112-117.

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., ... & Morency, L. P. (2014, May). SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems* (pp. 1061-1068).

Ellis, D. A. (2019). Are smartphones really that bad? Improving the psychological measurement of technology-related behaviors. *Computers in Human Behavior*, 97, 60-66.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health*, 4(2), e7785.

Gaffney H, Mansell W, Edwards R, Wright J. Manage Your Life Online (MYLO): a pilot trial of a conversational computer-based intervention for problem solving in a student sample. *Behav Cogn Psychother*. 2014 Nov; 42(6):731–46.

Gardiner, P. M., McCue, K. D., Negash, L. M., Cheng, T., White, L. F., Yinusa-Nyahkoon, L., ... & Bickmore, T. W. (2017). Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. *Patient education and counseling*, 100(9), 1720-1729.

Gardiner, P., Hempstead, M. B., Ring, L., Bickmore, T., Yinusa-Nyahkoon, L., Tran, H., ... & Jack, B. (2013). Reaching women through health information technology: the Gabby preconception care system. *American Journal of Health Promotion*, 27(3\_suppl), eS11-eS20.

Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H. C., & Jeste, D. V. (2019). Artificial intelligence for mental health and mental illnesses: an overview. *Current psychiatry reports*, 21, 1-18.

Gupta, M., Malik, T., & Sinha, C. (2022). Delivery of a Mental Health Intervention for Chronic Pain Through an Artificial Intelligence-Enabled App (Wysa): Protocol for a Prospective Pilot Study. *JMIR Research Protocols*, 11(3), e36910.

Karnouskos, S. (2020). Artificial intelligence in digital media: The era of deepfakes. *IEEE Transactions on Technology and Society*, 1(3), 138-147.

Ly, K. H., Ly, A. M., & Andersson, G. (2017). A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet interventions*, 10, 39-46.

Marken, R. S., & Carey, T. A. (2015). Understanding the change process involved in solving psychological problems: A model-based approach to understanding how psychotherapy works. *Clinical psychology & psychotherapy*, 22(6), 580-590.

Ng, G. W., & Leung, W. C. (2020). Strong artificial intelligence and consciousness. *Journal of Artificial Intelligence and Consciousness*, 7(01), 63-72.

Palanica, A., Flaschner, P., Thommandram, A., Li, M., & Fossat, Y. (2019). Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. *Journal of medical Internet research*, 21(4), e12887.

Parsons, T. D., & Rizzo, A. A. (2008). Initial validation of a virtual environment for assessment of memory functioning: virtual reality cognitive performance assessment test. *CyberPsychology & Behavior*, 11(1), 17-25.

Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J., & Jia, W. (2018). Influence analysis in social networks: A survey. *Journal of Network and Computer Applications*, 106, 17-32.

Pennebaker, J. W. (1997). Writing about emotional experiences as a therapeutic process. *Psychological science*, 8(3), 162-166.

Picard, R. W. (2002). Affective medicine: Technology with emotional intelligence. In *Future of Health Technology* (pp. 69-83). IOS Press.

Picard, R. W., & Healey, J. (1997). Affective wearables. *Personal technologies*, 1, 231-240.

Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, 108, 42-49.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information fusion*, 37, 98-125

Powers, WT. *Behavior: The control of perception*. New York: Aldine; 1973.

Ramachandran, M., Suharwardy, S., Leonard, S. A., Gunaseelan, A., Robinson, A., Darcy, A., ... & Judy, A. (2020). 74: Acceptability of postnatal mood management through a smartphone-based automated conversational agent. *American Journal of Obstetrics & Gynecology*, 222(1), S62

Riva, G., Calvo, R., & Lisetti, C. (2014). *Cyberpsychology and affective computing*.

Rizzo, A. A., Lucas, G., Gratch, J., Stratou, G., Morency, L. P., Shilling, R., & Scherer, S. (2016). Clinical interviewing by a virtual human agent with automatic behavior analysis.

Ryding, F. C., & Kuss, D. J. (2020). The use of social networking sites, body image dissatisfaction, and body dysmorphic disorder: A systematic review of psychological research. *Psychology of Popular Media*, 9(4), 412.

Rouast, P. V., Adam, M. T., & Chiong, R. (2019). Deep learning for human affect recognition: Insights and new developments. *IEEE Transactions on Affective Computing*, 12(2), 524-543

Smith, K. A., Blease, C., Faurholt-Jepsen, M., Firth, J., Van Daele, T., Moreno, C., ... & Cipriani, A. (2023). Digital mental health: challenges and next steps. *BMJ Ment Health*, 26(1).

Suganuma, S., Sakamoto, D., & Shimoyama, H. (2018). An embodied conversational agent for unguided internet-based cognitive behavior therapy in preventative mental health: feasibility and acceptability pilot trial. *JMIR mental health*, 5(3), e10454.

Villani, D., Lucchetta, M., Preziosa, A., & Riva, G. (2009). The role of interactive media features on the affective response: a virtual reality study. *International Journal on Human Computer Interaction*, 1(5), 1-21

Wang, Y., Song, W., Tao, W., Liotta, A., Yang, D., Li, X., ... & Zhang, W. (2022). A systematic review on affective computing: Emotion models, databases, and recent advances. *Information Fusion*.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology innovation management review*, 9(11).

<https://www.open.online/2023/01/04/chatgpt-psicoterapia-test>

<https://www.wysa.com/clinical-evidence>

<https://formative.jmir.org/2021/5/e27868>

<https://www.sciencedaily.com/releases/2023/05/230517122117.htm>

<https://www.psychologytoday.com/us/blog/how-do-you-know/202201/unlocking-ai-psychology>

<https://www.psychologytoday.com/us/blog/the-future-brain/202205/ai-behavior-health-chatbot-app-fast-tracked-fda>

<https://www.psychologytoday.com/us/blog/understanding-hypnosis/202207/could-artificial-intelligence-replace-therapists>

<https://www.psychologytoday.com/us/blog/the-future-brain/202204/can-ai-be-emotionally-intelligent>

<https://chat.openai.com>

<https://blogs.wysa.io/blog/category/research>



# Tecnologie persuasive: così influenzano le nostre decisioni

Tutte le tecnologie sono costruite non solo pensando a funzionalità e scopo, ma anche impiegando tecniche persuasive per catturare l'attenzione, personalizzare i contenuti e influenzare il comportamento degli utenti. La captologia è lo studio di cosa influenza attenzione e comportamento quando utilizziamo le tecnologie digitali

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

Alla fine degli anni Settanta lo psicologo statunitense James Jerome **Gibson** conì il termine "**affordance**" per indicare delle proprietà degli oggetti. Un oggetto può presentare una o più affordance, ovvero delle qualità fisiche che "invitano" le persone a farne un certo utilizzo. Per esempio, un pulsante o una maniglia hanno quella determinata forma e colore perché devono distinguersi dal resto della superficie ed "invitare" a fare qualcosa: premere o aprire una porta.

Queste proprietà degli oggetti indicano implicitamente alle persone per quale uso sono destinate.

L'affordance di Gibson, studioso della psicologia della percezione, si riferiva a proprietà fisiche; oggi si parla anche di '**social affordances**' - o sociotechnical affordances: strumenti e dispositivi tengono conto non solo di cosa facciamo ma anche di come reagiamo (come Herterich e colleghi, 2023).

Le tecnologie possiedono questo tipo di proprietà, delle affordance fisiche e non, e sono costruite in modo da indurre, influenzare, indirizzare le azioni che compiamo. Ed è proprio di questo che si occupa la captologia.

## Cosa significa captologia

**Che ci fosse un legame tra tecnologia e persuasione è noto dagli anni Novanta.** Risale infatti al 1998 il primo lavoro di Brian Jeffrey Fogg, professore dell'Università di Stanford, che conì l'acronimo CAPT, (computers and persuasive technologies), dando il via ad un filone di studi nuovo: la captologia.

**La captologia è una disciplina che, integrando psicologia, marketing, informatica, economia comportamentale, analizza come e fino a che punto le tecnologie possano influenzare il comportamento umano,** quanto possano persuaderci a fare qualcosa; è lo studio di come i computer e i dispositivi digitali possano essere progettati per influenzare comportamento e atteggiamenti.

Il fine ultimo è avere delle basi solide per capire come progettare una tecnologia perché sia, oltre che funzionale, efficace, valida, anche e soprattutto in grado di **spingere all'azione.**

Come? Gli elementi presi in considerazione sono molti: design e layout, incentivi, feedback, messaggistica personalizzata; tutto quello che viene incorporato nelle interfacce tecnologiche mira a influenzare il comportamento degli utenti, richiamando l'interesse e portandolo verso azioni o risultati specifici.

Fogg, oggi direttore del Persuasive Technology Lab dell'Università di Stanford, si dedica da quasi trenta anni a questo campo, approfondendone teoria, metodi, applicazioni e aggiornamenti man mano che le tecnologie progrediscono. Smartphone, siti web, domotica, computer, mobile app, videogiochi, oggi tutto viene progettato tenendo conto di cosa può persuadere i consumatori a compiere una determinata azione.

## Gli elementi di persuasione delle tecnologie

Secondo Fogg ed il tuo team di ricerca una tecnologia (hardware o software che sia) per aver elementi persuasivi deve tener conto di **diverse caratteristiche** che vanno da come viene materialmente progettata a come viene gestito il rapporto con l'utente (Fogg, 2007; 2022).

**La cura del design** è il primo punto: una tecnologia attraente è anche più convincente. Alcune caratteristiche, come simmetria e linearità, sono universalmente associate a ciò che è piacevole (anche se la valutazione può variare a seconda del pubblico per cui è destinata).

Ciò è dovuto al cosiddetto "**effetto alone**", ovvero l'estensione di una caratteristica positiva ad altre. In pratica, una cosa bella ci sembra anche - per estensione appunto - ben fatta, ingegnosa, efficiente, affidabile e credibile.

Anche il layout e la struttura contano: un sito con un layout chiaro, funzionale, semplice è preferito ad uno che, viceversa, è caotico e farraginoso.

Ci sono poi elementi che fanno leva su **aspetti psicologici**: per esempio portare a dedurre, inconsciamente, che il prodotto abbia emozioni, preferenze, motivazioni e personalità.

Quello che ci induce ad umanizzare oggetti che sappiamo essere inanimati sono **indizi testuali**, come messaggi di testo che trasmettono empatia ("Mi dispiace, ma...") o icone sullo schermo che ritraggono emozioni, come emoji. Ad esempio, il messaggio di errore 404 di solito è accompagnato con faccine tristi o altri segnali di vicinanza, come se lo strumento 'empatizzasse' con l'utente.

Il linguaggio sia scritto che parlato è un'altra leva di persuasione. Le finestre di dialogo sono un ottimo esempio: messaggi come "Congratulazioni!" dopo aver ultimato un task, o domande come "Sei sicuro di voler uscire?" creano un implicito dialogo tra macchina e utente, generando in quest'ultimo un senso di coinvolgimento.

Quando effettuiamo il log in e nella schermata c'è un "Benvenuto" col nostro nome, ci sono suggerimenti e consigli in base alle nostre preferenze, quando ci viene richiesto un feedback, l'esperienza di navigazione ci rende più presenti e attivi anche se, nei fatti, non c'è proprio nessuno dall'altra parte a salutarci e ringraziarci per l'acquisto.

**Salutare gli utenti, confermare l'acquisto, ringraziare per la transazione, risponde a delle norme sociali che del resto sono valide anche offline.** In un e-commerce non accade nulla di diverso da quello che accade in un negozio fisico: vengono rispettate le dinamiche sociali.

A volte, soprattutto quando si tratta di programmi che richiedono determinati passaggi per essere eseguiti correttamente, vengono aggiunte delle funzionalità che hanno il solo scopo di guidare

l'utente. Per esempio, la prima volta che apriamo un'app o un software appena installato siamo guidati alla scoperta delle funzioni principali con una serie di finestre di dialogo che evidenziano i passaggi. In questo senso Fogg parla di ruolo della tecnologia: **il ruolo di “esperto” che supporta l'utente**. Questo ruolo ha la funzione di dare all'utente non solo suggerimenti e informazioni utili ma soprattutto trasmettere l'idea che la tecnologia sia efficiente, ben fatta e si occupi di lui.

## Qualche esempio di captologia

A proposito di aspetto e layout, nel 2009 Oinas-Kukkonen & Harijumaa hanno proposto il paradigma del **persuasive system design (PSD)**, adottato poi da diverse ricerche successive.

Secondo questo paradigma, è **l'unione tra aspetto visivo, facilità d'uso e psicologia a fare la differenza**. Principi ben noti come quelli della scarsità, dell'urgenza, dell'influenza sociale determinano il successo, soprattutto quando il fine è quello di vendere prodotti o servizi. In termini di design persuasivo sono fattori strettamente collegati: per esempio mettere in risalto la disponibilità limitata, le raccomandazioni degli altri utenti, inserire un timer che indica quanto tempo si ha a disposizione. Questi elementi influenzano le azioni degli utenti, indirizzandoli a credere che il prodotto sia molto richiesto e che a breve non sarà più disponibile.

Sempre secondo Oinas-Kukkonen e Harijumaa (2009) è utile **sfruttare il meccanismo psicologico chiamato “riprova sociale”**, ovvero la tendenza a ritenere affidabili pareri e comportamenti condivisi da molte persone: mostrando recensioni, valutazioni e testimonianze degli altri utenti si crea un senso di autorevolezza e credibilità.

Il design persuasivo si avvale di molti espedienti: riduzione del numero di passaggi necessari per completare un'attività o un acquisto (con un solo click); promemoria, notifiche per gli articoli ancora nel carrello e per i quali l'acquisto non è stato ultimato; variabilità del prezzo che potrebbe essere diverso qualche giorno dopo; monitoraggio e tracciabilità della spedizione che creano un senso di attesa e curiosità; suggerimenti sulla base degli acquisti già effettuati.

**Ecommerce come Amazon fanno un ampio uso di queste strategie**; le raccomandazioni personalizzate sono presenti anche in altri tipi di piattaforme tra cui Netflix, Youtube, Spotify, che sono molto attente a dare suggerimenti sulla base delle preferenze e del comportamento degli utenti: film o serie che “potrebbero piacerti”, video consigliati, playlist personalizzate.

**Social network e app non sono da meno**: strategie come video in riproduzione automatica e feed di contenuti personalizzati, notifiche push, gamification, sistemi di ricompensa (badge, punti), classifiche, definizione di obiettivi. Ad esempio, i social inviano periodicamente notifiche e aggiornamenti; i fitness tracker utilizzano varie tecniche per incoraggiare gli utenti a praticare attività fisica, come la definizione di obiettivi di passi giornalieri.

Gli esempi potrebbero essere infiniti ma la logica rimane la stessa: **tutte le tecnologie sono costruite non solo pensando a funzionalità e scopo, ma anche impiegando tecniche persuasive per catturare l'attenzione, personalizzare i contenuti e influenzare il comportamento degli utenti**.

Il loro modello di business si basa sul fatto che gli utenti le utilizzino massivamente, che spendano molto tempo fruendo di contenuti, interagendo, condividendo. Utilizzando queste tecniche puntano a catturare e mantenere l'attenzione degli utenti il più a lungo possibile.

## Catturare l'attenzione in modo etico

La captologia è una disciplina che si occupa di come le tecnologie possono influenzare il comportamento; questi esempi mostrano come la progettazione di piattaforme ed interfacce venga fatta anche pensando a quali azioni vogliono elicitare.

### Qual è il limite tra influenzare e manipolare?

Indubbiamente, sfruttare in maniera intenzionale delle zone grigie del funzionamento psicologico potrebbe portare a conseguenze indesiderate e non etiche.

**La progettazione responsabile tutela il rispetto dell'autonomia, della privacy, del benessere e la salvaguardia delle vulnerabilità degli utenti:** fornisce informazioni sull'uso che viene fatto dei dati raccolti, garantisce il controllo su come vengono utilizzati e condivisi, facilita le opzioni di personalizzazione e di notifica, anche (e soprattutto) per chi è meno esperto.

**Il fatto che esistano delle funzioni create apposta per “invitarci” a far qualcosa non è di per sé negativo e dannoso,** non più di quanto può esserlo, per esempio, posizionare i prodotti tra gli scaffali dei supermercati in un certo modo per dirigere la nostra attenzione su certi prodotti più che su altri (queste tecniche fanno capo allo shelf marketing).

Che le tecnologie abbiano anche una componente sociale, come la definisce Fogg, ovvero l'interesse a come reagiamo e a cosa ci induce a fare, non può che essere parte delle tecnologie stesse; del resto, non vengono create per noi?

Quello che può fare la differenza sono **trasparenza, chiarezza e rispetto degli utenti.**

## Bibliografia

Fogg, B.J. (1998). Captology: the study of computers as persuasive technologies. In: CHI 98 Conference Summary on Human Factors in Computing Systems, p. 385.

Fogg, B. J., Cueller, G., & Danielson, D. (2007). Motivating, influencing, and persuading users: An introduction to captology. In *The human-computer interaction handbook* (pp. 159-172). CRC press.

Fogg, B.J. (2022) Persuasive technology: using computers to change what we think and do. *Ubiquity* 2002(December), 2.

Gibson, J. J. (2014). The theory of affordances (1979). In *The people, place, and space reader* (pp. 90-94). Routledge.

Herterich, M. M., Dremel, C., Wulf, J., & vom Brocke, J. (2023). The emergence of smart service ecosystems—The role of socio-technical antecedents and affordances. *Information Systems Journal*.

*I quaderni di*

# Agenda **Digitale**

---

## NETWORK **DIGITAL** 360

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](mailto:MARKETING@DIGITAL4.BIZ)

©ICT & Strategy