Salis, P. (2024). IA e futuro: consigli per gli acquisti. *Indagine e Pratica Filosofica*, 2(1-Supplemento), 15-25. ISSN: 3035-0581.

DENOMINAZIONE ALTERNATIVA: *Philosophical Inquiry and Practice*

EN_ *AI and Future: Shopping Tips*

IA e futuro: consigli per gli acquisti

Pietro Salis
Università degli Studi di Cagliari
psalis@unica.it

RIASSUNTO: I *large language models* rappresentano una sfida significativa nell'ambito dell'Intelligenza Artificiale (IA) per il grande pubblico. Questa sfida ha generato un intenso dibattito pubblico, sollevando preoccupazioni e allarmi da diverse prospettive. Allo stesso tempo, l'IA è diventata un marchio ampiamente sfruttato nelle strategie di marketing, evidenziando un atteggiamento implicitamente positivo verso il suo utilizzo e sviluppo. Nel presente contributo cerco di delineare alcuni aspetti che potrebbero aiutare a superare questi atteggiamenti polarizzati e unidirezionali riguardo all'IA.

PAROLE-CHIAVE:
Intelligenza Artificiale. Intelligenza Artificiale Generativa. *Large Language Models*. Approccio Critico. Test di Turing.

AI and Future: Shopping Tips

ABSTRACT: Large language models are the focus of artificial intelligence's (AI) challenges for the general public. This sparked a broad public discussion, identifying several warning signs and concerns emanating from various directions. At the same time, AI is widely exploited in marketing strategies, underscoring an implicitly positive attitude towards its use and development. In this paper, I try to highlight some aspects that can be useful in unravelling such polarising and one-way attitudes.

KEYWORDS:
Artificial Intelligence. Generative Artificial Intelligence. *Large Language Models*. Critical Approach. Turing Test.

1. Introduzione

La nuova frontiera dell'intelligenza artificiale è sempre più al centro delle discussioni, tanto accademiche, quanto di interesse pubblico. Questo genera un contesto di ampio risalto mediatico, in cui talvolta i commentatori intervengono a sproposito e dove farsi un'idea chiara di ciò che viene descritto e valutato non è sempre semplice. Ciò accade anche perché – lo dico da non esperto – dei veri esperti di IA vengono raramente consultati e coinvolti attivamente in queste discussioni. Prima di finire per cadere nei luoghi comuni inevitabili di queste discussioni molto pervasive, sarebbe bene porre alcuni interrogativi centrali per provare a dare un inquadramento diverso al tema:

1) i cosiddetti *large language models* (LLM), la nuova versione cosiddetta generativa dell'IA, per intenderci ChatGPT e competitors (Bard, *et similia*) – sono *realmente* intelligenti?
2) la loro eventuale intelligenza è qualcosa che ci deve preoccupare?

L'IA è un argomento che suscita molte discussioni nei media, online e nelle conversazioni quotidiane, con opinioni che spaziano in ogni direzione. Giornalisti, politici, scienziati e filosofi sono sempre pronti a esprimere il loro punto di vista su questo tema. Per evitare di ripetere semplicemente i toni e i contenuti di questi discorsi, è legittimo interrogarci preliminarmente su alcune questioni fondamentali. È facile, infatti, sfruttare le emozioni e le paure delle persone su tali argomenti. Per prevenire questa deriva, è importante cercare di comprendere chiaramente cosa sono i LLM e cosa fanno esattamente.

2. Capacità dei LLM

Partiamo dalla prima domanda. I LLM sono modelli di intelligenza artificiale addestrati su grandi quantità di testo per generare risposte simili a quelle umane, ma la loro intelligenza è alquanto diversa da quella umana e si basa su pattern di dati[1] piuttosto che su comprensione e coscienza. Certamente i LLM sono in grado di leggere e consultare molti dati insieme in vista del compito consistente nel rispondere a uno specifico quesito, posto da noi (uno specifico prompt). Sono la frontiera cosiddetta generativa dell'IA per un semplice motivo: si tratta di generatori automatici di testo con alcune caratteristiche di rilievo. Ad esempio, hanno tempi di risposta notevoli ed hanno accesso ad ampie porzioni dei testi presenti nella rete.

[1] Un pattern di dati è un insieme di dati che mostra una struttura o una regolarità riconoscibile. In altre parole, è un modello o schema che emerge da un'analisi dei dati e che può essere utilizzato per identificare relazioni, comportamenti o tendenze generali all'interno di un particolare dataset.

Certamente possiamo essere d'accordo che non sono caratteristiche banali: per essere efficienti da questo punto di vista occorrono potenza di calcolo e accesso istantaneo a una grande mole di dati. Questo naturalmente ci aiuta e ci è utile per molti compiti, ci fa risparmiare tempo, e spiega abbastanza bene tutto il successo e il recente entusiasmo sociale e mediatico intorno all'IA. Pensate solo a quanto marketing ci bombarda negli ultimi anni: smartphone con l'IA, *cpu* specifiche per l'IA, e così via. Se non ci fosse un qualche entusiasmo, l'IA, è inutile dirlo, non venderebbe come di fatto vende (e non sarebbe protagonista degli inserti pubblicitari nel modo in cui lo è attualmente). Infine, come si sa, la potenza di calcolo dei nostri microprocessori è in crescita costante, anche solo in rapporto all'energia consumata o all'aumento delle frequenze operative, così come è in crescita anche la capienza e la prestanza dei server/data center che forniscono molti dati da base alle IA generative. Questo vuol dire che la potenza delle IA generative è inserita in una traiettoria di crescita continua per quanto riguarda le sue *performances*. Saper fare queste cose, e farle in questo modo, significa realmente essere intelligenti? Prima di rispondere è bene guardare anche il rovescio della medaglia: i LLM hanno anche dei difetti e dei limiti di tipo strutturale.

3. Difetti e limiti dei LLM

Va detto però, a dispetto di questo clima di entusiasmo, che le *performances* dei LLM non sono sempre impeccabili: a seconda della domanda che si fa, si ottengono spesso risposte deludenti. Chiedo al lettore di fare una prova, o se volete un piccolo esperimento, alla portata di chiunque: scegliete un argomento *X* di cui sapete tutto o comunque molto, insomma qualcosa che vi appassiona, e chiedete a ChatGPT una domanda su un aspetto di *X* che richiede un po' di competenza e discernimento: un minimo di idee chiare. Vedrete che la risposta di ChatGPT lascerà alquanto a desiderare e conterrà almeno un errore subito evidente ai vostri occhi (esperti). Cosa significa? Che certe competenze di discernimento – tra dati attendibili e non – non sono sempre *decidibili* e *discernibili* da un LLM.

Se ChatGPT sul tema *X* accedesse al 90% a dei dati e a delle informazioni inattendibili e fosse addestrato solo sulla base di questi testi e dati, al 90% fornirebbe risposte inattendibili[2]. Infatti, le loro *performances* si basano fondamentalmente su di una procedura simile all'auto-completamento: quali dati accessibili completano meglio un certo prompt/una certa domanda? Il lavoro sui dati e sulle informazioni sembra essere cioè più quantitativo che qualitativo, e dipende dal tipo esatto di dati e/o testi usati per

[2] Va precisato che i LLM sono addestrati solitamente su varie tipologie di fonti; quindi, questo scenario di dati totalmente inattendibili è più un esperimento mentale basato su di una possibilità logica che non una descrizione realistica.

l'addestramento – un addestramento con dati non pertinenti o troppo datati per un certo prompt genererà, in prospettiva, risposte deludenti[3]. Il che vuol dire che questi sistemi non hanno, al di là delle varie suggestioni agitate spesso superficialmente nel dibattito pubblico, una reale comprensione dei testi che generano e maneggiano[4].

Questi problemi sono alla base della lamentela per cui i LLM soffrirebbero in senso figurato di allucinazioni – cioè possono produrre risultati inesatti o inventati – e/o sarebbero particolarmente in difficoltà nel discernere il contesto esatto cui un certo prompt di ricerca si riferisce. Ancora, ad esempio, una delle difficoltà oggettivamente più limitanti di questi sistemi è la loro incapacità di distinguere il vero dal falso. Se uno dei LLM fosse addestrato con, e collegato a, dei data center che contengono solo informazioni false, produrrebbe il 100% di risposte false. Quindi i dati/testi utilizzati per l'addestramento di un LLM possono fare tutta la differenza del mondo.

Ciò in futuro potrà essere migliorato con lo sviluppo di algoritmi di *machine learning*[5] più precisi, che si possono addestrare meglio al discrimine dell'attendibilità delle fonti e magari anche a saper leggere meglio i segnali contestuali. In effetti, la scansione dei testi nella rete da parte delle IA generative funziona su basi probabilistiche[6] e sulla base della coerenza, e ciò non è sempre garanzia sufficiente di risposte pertinenti e soddisfacenti. In futuro si aspettano prestazioni migliori per quanto riguarda sia la velocità sia l'attendibilità dei dati, ma è totalmente plausibile ritenere che i LLM resteranno sempre incapaci di una comprensione effettiva dei testi generati.

4. Dubbi sulle capacità dei LLM

Quindi è forse legittimo asserire che i *large language models* non sono così intelligenti come spesso si sostiene nelle discussioni recenti. Un problema non da poco, in questo contesto, è che le definizioni di che cosa effettivamente conti come intelligenza hanno sempre lasciato a desiderare e non hanno mai messo d'accordo tutti. E forse questa è una prima difficoltà che viene sfruttata dalle discussioni più superficiali: senza un

[3] L'addestramento di un LLM opera grosso modo come segue: nel contesto di una massiccia esposizione a siti web, articoli, libri e così via, un LLM impara gradualmente a riconoscere schemi, strutture che chiameremmo *grammaticali* e specifici contesti semantici, cioè i ragionamenti tramite cui certe parole vengono utilizzate in modo sistematico.

[4] Si veda l'utile intervista ad Antonio Lieto, professore di informatica dell'Università di Salerno: https://www.smarknews.it/facciamo-chiarezza-su-chatgpt-limiti-potenzialita-e-falsi-miti-intervista-a-antonio-lieto-ricercatore-informatico/?print=print.

[5] Gli algoritmi di *machine learning* sono, in buona sostanza, insiemi di istruzioni che permettono ai computer di imparare dai dati a cui hanno accesso. In pratica, questi algoritmi analizzano vaste quantità di dati, identificano pattern, e utilizzano queste informazioni per fare previsioni o prendere decisioni senza essere esplicitamente programmati per ogni singolo compito. Questa assenza di programmazione esplicita è efficacemente espressa con l'espressione *apprendimento automatico*.

[6] Con che frequenza statistica, ad esempio, certi tipi di dati si legano ad altri nei testi cui si ha accesso.

criterio chiaro, tutto e nulla risulta essere intelligente. Nel caso dell'intelligenza artificiale, tuttavia, le cose forse sono un po' diverse e qualche indizio ci spinge ad essere meno pessimisti. Vediamo un po' perché.

I sostenitori dei *large language models* come punto di svolta nell'IA fanno invece leva sul fatto che, si dice da più parti, siamo forse vicini al superamento della soglia tracciata dal cosiddetto test di Turing. Questo è un tema che ci è utile introdurre in questo contesto. Che cosa è il test di Turing?

Il test di Turing – ideato dal celebre matematico britannico, e padre dell'informatica, Alan Turing – consiste, nel modo più diretto possibile, in una prova in cui le risposte di un computer ad un umano in un'interazione basata su domande e risposte scritte – di fronte a uno schermo e a una tastiera – dovrebbe permettere all'umano di capire se ha di fronte una persona o una macchina. Vale a dire, se si trova di fronte ad un'intelligenza naturale (una persona) o una invece artificiale (un'IA). Se l'umano *non sa decidere* o se direttamente risponde che sta parlando *con una persona* il test è da considerarsi superato: la macchina ha ingannato l'umano facendogli credere di essere una persona sulla base dell'intelligenza e pertinenza delle sue risposte. C'è una cosiddetta indistinguibilità nella risposta ottenuta, rispetto ad una riconoscibile come umana, che fa passare il test ad una macchina. Inutile aggiungere altro, questo test e la sua importanza sono oggi direttamente tirati in ballo dalla prepotente ascesa dei LLM. Vediamo come e perché con qualche esempio.

Gli usi di ChatGPT dove gli si chiedono ad esempio consigli psicologici ed esistenziali[7] – usi tratti da episodi che sono stati riportati frequentemente dalla stampa e dalla cronaca in tempi recenti – assumono spesso una dimensione per certi versi analoga a quella del test di Turing. In che senso? L'analogia è alquanto semplice. In queste interazioni, l'utente dell'IA generativa ha l'impressione che vi sia qualcuno dall'altra parte dello schermo a rispondere e a dare un feedback, proprio come farebbe una persona via chat; e pare anche talvolta che l'*interlocutore generativo* sia in grado di comprendere l'utente, si direbbe[8]. Ma le analogie non finiscono qua. Di recente, a questo proposito, dei colleghi del mio dipartimento presso l'Università di Cagliari hanno dimostrato come ChatGPT è di fatto in grado di svolgere compiti piuttosto complessi anche per le persone, come ad esempio superare il test di ingresso in medicina all'Università la Sapienza di Roma (Giunti et al., 2024). La notizia di questo esperimento ha suscitato lo scorso anno un po' di clamore ed è stata ripresa anche dal

[7] Questa dinamica sembra coinvolgere in maniera maggioritaria i giovani e in particolare gli adolescenti (È bene segnalare che le persone dovrebbero rivolgersi a un professionista e non a ChatGPT). Si veda ad esempio i reportages sul tema apparsi su *Open* e su *Wired*:
https://www.open.online/2023/01/04/chatgpt-psicoterapia-test/
https://www.wired.it/article/chatgpt-intelligenza-artificiale-psicoterapia-rischi-vantaggi-usi/

[8] Questo effetto, in linea di principio, si potrebbe spiegare nei termini della proiezione empatica inconsapevole da parte dello stesso utente.

Corriere della Sera[9]. Se questo non equivale a passare direttamente il test di Turing, dobbiamo dire che poco ci manca. Questo testimonia indubbiamente un notevole progresso dell'IA e indubbie capacità di elaborazione in termini di linguaggio e ragionamento. Siamo di fronte a una performance che ci dà la cifra della capacità attuale dell'IA generativa. I LLM sanno già fare qualcosa che la maggior parte di noi non sarebbe in grado di fare.

Cosa comporta tutto ciò? Come dobbiamo intendere questo successo e questo traguardo? Se stiamo a sentire alcuni discorsi generalisti si dice spesso che siamo all'alba di una nuova era o sull'orlo della catastrofe. L'IA diviene automaticamente il nuovo vaso di pandora e ogni miseria viene a turno evocata come conseguenza ineludibile. E ci si divide così sempre e ancora in quelle categorie bene inquadrate a suo tempo da Umberto Eco nei termini di *apocalittici* e *integrati*[10]: come gli *apocalittici* di Eco erano gli intellettuali che guardavano i media e la cultura di massa con distacco e disprezzo – come fossero alle prese con l'imminente avvento di una nuova barbare – oggi molti (intellettuali e non) vedono nell'IA generativa il punto di passaggio verso il post-umano, la tecno-barbarie, il declino o l'estinzione dell'umanità (per citare solo gli slogan più coloriti). Uno degli eventi più recenti e rilevanti, in questo contesto, è senz'altro la presentazione del documentario, piuttosto critico nei confronti dello sviluppo dell'IA generativa, *The End of Humanity*[11]. Come stanno le cose?

5. Singolarità e punto di svolta

Vuol dire forse che siamo al punto di svolta e che sta per accadere ciò che viene preconizzato dagli apocalittici commentatori? Stiamo forse vicini alla cosiddetta *singolarità* in cui l'IA diviene autonoma e accede alla possibilità di avere un rapporto da pari a pari con noi? Questa è sicuramente fantascienza, ma se la prospettiva si presenta comunque e in modo sistematico, ciò vuol dire che si tratta di tematiche che

[9] Si veda il seguente link: https://www.corriere.it/scuola/secondaria/23_marzo_02/chatgpt-passa-test-medicina-ma-fa-fatica-col-ragionamento-logico-1a4f14f8-b909-11ed-b194-e55eac395b06.shtml?refresh_ce.

[10] La dinamica descritta da Eco (1964) è rilevante perché sembra essere indipendente dal tema centrale di riferimento, e quindi si configura più come un tratto psicologico pervasivo che come un qualcosa di strettamente legato ai fatti: un'analoga polarizzazione degli atteggiamenti nel discorso pubblico ha accompagnato qualsiasi *problema planetario* – tale o presunto – sufficientemente minaccioso da apparire come l'inizio della fine. Si pensi ad esempio alla proliferazione di armamenti nucleari durante la guerra fredda, al buco nell'ozono, all'AIDS, all'avvento della rete Internet, al *millenium bug*, ecc. Indipendentemente da come poi si svolgessero i fatti, vi è sempre stata ampia rappresentanza di coloro che in questi fenomeni hanno visto la svolta fatale per l'umanità.

[11] Il documentario, proiettato in anteprima il 16 luglio 2024 a Vienna, è descritto online al seguente link: https://gsis.at/2024/06/18/artificial-intelligence-the-end-of-humanity/.

intercettano ansie e paure reali – o forse che qualcosa non va bene nel nostro immaginario (dove magari la fantascienza ha più peso della realtà).

Le ansie e le paure del pubblico, che nel caso dell'IA più che in altri si accompagno ad entusiasmo e ad un uso massiccio nel marketing, sono da un lato da mettere in conto e dall'altro raramente basate su di una comprensione sufficientemente approfondita del fenomeno. Andiamo però a vedere come stanno i dati reali riguardo a tali sentimenti e preoccupazioni.

Se osserviamo i dati sulla *percezione del pubblico* in Italia prevale un sentimento di preoccupazione, come da poco ha raccontato Livio Gigliuto dell'Istituto Piepoli che è stato recentemente ospite all'Università di Cagliari per un evento su comunicazione, nuovi media e IA. Quindi una preoccupazione del pubblico è effettivamente rilevabile in termini fattuali[12]. Tutto ciò, però, a dispetto del clima di entusiasmo sfruttato dal marketing di cui parlavamo prima. La percezione di paura va dunque mitigata dalla presenza di un contesto di sentimenti e aspettative di tipo positivo, dove l'IA generativa diviene un gadget che ha rilievo non marginale nelle dinamiche di vendita. Visto che i fatti, sia sociali che economici, sono testimoni di entrambe le tendenze, è bene porsi la domanda nel modo più diretto. Quando si parla della preoccupazione nei confronti dello sviluppo dell'IA generativa, si tratta di una preoccupazione seria e giustificata o di una percezione in qualche modo ingannevole? Dobbiamo correre ai ripari? Questa percezione negativa – che, come detto, è rilevata nei fatti – è legittima? Quali sono le preoccupazioni reali, distinte da quelle più pretestuosamente catastrofiste?

6. Ragioni contro l'eccessiva preoccupazione

Io devo confessare di essere piuttosto scettico riguardo alle preoccupazioni di tendenza millenarista. La mia impressione è che il test di Turing, oggigiorno, stia divenendo sempre più qualcosa di *sopravvalutato* come criterio per stabilire il reale livello di intelligenza e capacità cognitive dell'IA[13]. Intendo con ciò che si tratta di un test concepito grosso modo negli anni '50, con una prospettiva che allora pareva fantascientifica e che oggi, in un contesto radicalmente diverso riguardo allo sviluppo dell'IA, ci pare invece molto meno sorprendente.

Un'IA a cui attribuisco intelligenza dovrebbe saper fare qualcosa di più che generare testo sulla base di una mole gigantesca di testi letti istantaneamente e con ciò convincermi che dall'altra parte dello schermo o del *device* che utilizzo ci sia una persona

[12] Tuttavia, sarebbe bene indagare se c'è un qualche eventuale nesso tra questi sentimenti da un lato e la qualità del discorso pubblico sull'IA – e la precisione e la chiarezza di quanto viene affermato – dall'altro.
[13] Si legga questo mio giudizio anche in relazione diretta con ciò che dice Antonio Lieto su LLM e test di Turing nell'intervista citata.

e non una macchina. Dovrebbe, ad esempio, saper e poter *prendere l'iniziativa* e *fare qualcosa nel mondo per conto suo*, anche indipendentemente dalle nostre volontà. E nulla del genere sembra essere, francamente, all'ordine del giorno. I LLM non sono ancora in grado di attivare processi che interagiscono a tal punto con le nostre vite. Questo è sufficiente per escludere molti scenari di presa di consapevolezza e ribellione delle macchine come nel film *2001: Odissea nello spazio*, di Stanley Kubrick, o in *Terminator*, di James Cameron, dal novero delle opzioni plausibili. Significa che se escludiamo gli scenari apocalittici rimaniamo senza preoccupazioni più serie?

7. Ragioni contro il troppo rilassamento

Vuol dire che tutto va bene e che non dobbiamo preoccuparci? Vuol dire veramente che tutte le preoccupazioni sono derubricabili a forme di sindrome dell'apocalisse? Anche qua, naturalmente, secondo il mio modesto parere non dobbiamo esagerare nell'altro senso. Certamente si sono registrati episodi di *robot autonomi* che hanno causato morti e lesioni nelle fabbriche e nelle catene di montaggio, così come negli stabilimenti di logistica ecc. Questi sono fatti, di chiara gravità, e che hanno imposto alle aziende di correre ai ripari con cambiamenti e ripensamenti di regolamenti e normative interne in materia. Ci sono anche previsioni economiche che parlano di perdite, nel prossimo futuro, di milioni di posti lavoro dovuti all'utilizzo dell'IA – che si dice rimpiazzerà molte tipologie di impiego.

Queste fosche previsioni implicherebbero alti costi sociali e per questo vengono spesso sbandierate nel discorso pubblico come un vero e proprio spauracchio. Questi aspetti invitano a una certa prudenza e non vanno sottovalutati. Queste sono sicuramente preoccupazioni più serie di molte altre. Altre proiezioni allarmanti riguardano il consumo di energia che servirà in futuro per alimentare microprocessori sempre più potenti e server/data center sempre più capienti – e quindi l'impatto energetico e l'impronta di carbonio dell'IA. Si tratta di sfide aperte che non conviene a nessuno ignorare o snobbare. Ma ciò, nonostante tutto, è sempre insufficiente per abbandonarsi all'abbraccio apocalittico.

7.1. Guida autonoma e sfide dell'IA

A questo proposito, un caso emblematico di utilizzo controverso dell'IA è quello della cosiddetta *guida autonoma*, visto che vi sono stati diversi casi conclamati di *omicidio involontario* perpetrato da veicoli con guida autonoma nelle fasi di test, specialmente

negli Stati Uniti[14]. Questi casi sono stati certamente terribili, e andrebbe fatto il possibile per evitarli. L'IA alla guida (con la complice negligenza dell'assistente umano al volante) è stata responsabile di alcuni incidenti fatali. Ma il fatto che sistemi di guida autonoma, che ricordiamo sono programmati esplicitamente per salvaguardare la vita delle persone, falliscano è un chiaro indizio fattuale che l'IA non è poi così intelligente. Queste morti non sono dovute cioè all'intelligenza dell'IA ma spesso a delle incapacità di leggere il contesto in cui si muovono i veicoli e i pedoni. E, paradossalmente, devo confessare che mi fa più paura se la mia vita, o quella delle persone care, è alla mercé di un sistema poco intelligente e poco efficace nello svolgere il proprio compito. Per la questione della guida autonoma gradiremmo avere un sistema affidabile e che funziona, e quindi ben più intelligente di come è configurato allo stato attuale.

La questione dell'impatto dell'IA sull'occupazione nel futuro è senza dubbio, come detto, più seria e preoccupante. Tuttavia, anche qua non serve a nessuno assumere un atteggiamento che si potrebbe definire di *disperazione preventiva*. Pare ad esempio probabile che l'IA sia destinata a coprire lavori più ripetitivi, (in molti casi) meno creativi e meno qualificati[15]. Non si tratta necessariamente di un male. Questo può essere uno stimolo ulteriore ad intraprendere carriere creative e imprenditoriali, e ciò potrebbe rivelarsi anche positivo in prospettiva. E in ogni caso, non abbiamo le basi per fare una predizione realistica su di un fenomeno così complesso come le trasformazioni dei mestieri sul lungo periodo. Quindi: bene stare attenti a ciò che accade e alle dinamiche in corso, ma non fasciamoci la testa.

Un discorso analogo si può fare a riguardo del consumo energetico legato all'uso e allo sviluppo dell'IA. Sicuramente, vista la situazione dell'ambiente e la sfida posta dal cambiamento climatico, si tratta di questioni da tenere sotto controllo. Tuttavia, nessuno può escludere che un progresso in termini di approvvigionamento energetico e gestione dei consumi cambi drasticamente la situazione.

7.2. Segnali di speranza: AI ACT

Nel frattempo, però, segnali di speranza arrivano su altri fronti. Di recente è stato approvato l'*AI Act* (AIA) dall'Unione Europea[16] e si tratta di un primo importante passo in avanti, che vede l'Europa in prima linea a livello mondiale – il documento costituisce in assoluto la prima regolamentazione dell'IA da un punto di vista

[14] Si vedano in proposito, e a titolo d'esempio, le seguenti storie: https://www.wired.com/story/uber-self-driving-car-fatal-crash/; https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281

[15] A parte forse una qualche forma di creatività artistica che l'IA generativa sta manifestando nel campo dell'arte visiva e in quella musicale.

[16] Il testo finale dell'*AI ACT* è disponibile online al seguente link: https://artificialintelligenceact.eu/the-act/

istituzionale. Questo deve essere sicuramente motivo d'orgoglio per l'Unione Europea. Si tratta di un segnale decisamente positivo, per cui valutazioni nel merito e analisi nel dettaglio sono forse premature.

Ciò testimonia tuttavia della consapevolezza e volontà di provare a governare i processi e di non farsi governare da essi. Questo è sicuramente l'atteggiamento giusto. Aspettiamo l'arrivo di altre legislazioni simili, in altri paesi, per provare a fare una valutazione dell'AIA dalla prospettiva vantaggiosa di una comparazione generale. Naturalmente, tutto ciò richiederebbe approfondimenti e per questi occorrerebbero dei giuristi o degli esperti di filosofia della tecnologia – che avrebbero le competenze adeguate, a differenza mia – per cui su questo punto specifico mi fermo qua[17].

8. Conclusioni

Quindi, alla fine dei conti dobbiamo essere entusiasti o pessimisti rispetto alle nuove frontiere dell'IA generativa? Entrambi gli atteggiamenti, come è emerso, sono eccessivi e dettati da bias, emozioni e molto probabilmente da quella tendenza che gli economisti chiamano 'timore della perdita' (*fear of loss*)[18].

Se proviamo a vedere lucidamente sia i rischi reali che le opportunità la bilancia risulta molto meno drammatica in entrambi i sensi. Ma dipenderà anche dal nostro atteggiamento, e disperarci e/o esaltarci aprioristicamente, seguendo le solite tendenze polarizzanti non ci aiuterà. Un *moderato ottimismo vigile* – e preferibilmente ben informato – è probabilmente meglio di una qualsiasi 'disperazione preventiva' o di un semplicistico *ottimismo esagerato e miope*.

RIFERIMENTI BIBLIOGRAFICI

Eco, U. (1964). *Apocalittici e integrati: comunicazioni di massa e teorie della cultura di massa*, Milano: Bompiani.

Floridi, L. (2021). The European Legislation on AI: a Brief Analysis of its Philosophical Approach. *Philosophy & Technology*, *34*(2), 215-222.

[17] Per un primo tentativo di una valutazione, competente e di prospettiva, segnalo comunque Floridi (2021).

[18] In generale si tratta di quella tendenza comportamentale per cui si preferisce sistematicamente evitare una perdita piuttosto che acquisire guadagni. In proposito, si veda il classico Kahneman e Twersky (1979).

Giunti, M. Garavaglia, F.G., Giuntini, R., Pinna, S., & Sergioli, G. (2024). ChatGPT as a Prospective Undergraduate and Medical School Student. *PLoSONE 19*(10), e0308157. https://doi.org/10.1371/journal.pone.0308157

Kahneman, D., & Twersky, A. (1979). Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, *47*(2), 263-291.