MIND THE WEB
DOMANDE *INCONTRO 2*

Che cos'è l'intelligenza artificiale e come si differenzia da quella umana?

Marta Casassa:
Durante il webinar abbiamo usato la definizione di Marvin Minsky, uno dei padri dell'AI, che la definisce così: "L'intelligenza artificiale è la scienza che si occupa di far fare alle macchine cose che richiederebbero l'intelligenza se fossero fatte dagli esseri umani."

Questa definizione evidenzia come quella che chiamiamo "intelligenza artificiale" non sia di fatto intelligente, ma che possa fare cose che a noi, come esseri umani, richiedono intelligenza. Si tratta quindi di una differenza sostanziale ed in particolare per quanto riguarda il tipo di AI che abbiamo analizzato durante il webinar, che è l'AI generativa basata su Large Language Models, il fatto che componga frasi anche molto articolate e verosimili, non vuol dire che sia dotata di una comprensione della domanda che stiamo ponendo, ma che sulla base del corpus documentale su cui è stata addestrata riesca statisticamente a predisporre una risposta.

Come si può usare l'ai in modo efficiente nell'ambito scolastico?

Marta Casassa:
L'AI, nella scuola così come in tutti gli ambiti, può essere usata in modo efficiente se la si vede come strumento che, collaborando con noi, ci può supportare nelle nostre attività, ad esempio nella docenza/nell'apprendimento.
L'AI non va quindi vista come un "sostituto", ma come un "alleato".
Ad esempio, se le faccio scrivere un tema al posto mio, non riuscirò ad apprendere come fare questa operazione in autonomia. Ma se invece, una volta creato il testo in maniera autonoma, lo faccio correggere/confutare dall'AI, questo può renderlo più solido e migliorarlo. Posso ad esempio chiedere, rispetto ad un determinato audience, di rivedere il testo in modo che sia adatto all'interlocutore a cui lo voglio presentare.

Qui un paio di studi usciti recentemente proprio sul tema AI nel mondo dell'apprendimento (entrambi esistono al momento solo nella versione inglese, ma su questo l'AI ci può aiutare 😉).

https://services.google.com/fh/files/misc/future_of_learning.pdf

https://www.unesco.org/en/articles/ai-and-future-education-disruptions-dilemmas-and-directions

Quali potrebbero essere le implicazioni future dell'AI per la società e la cultura umana?

Marta Casassa:
L'AI ha la potenzialità di accelerare molte scoperte scientifiche e di risolvere problemi fino ad oggi ancora aperti. L'anno scorso il Nobel per la chimica è andato a due persone di Google Deepmind, che hanno risolto, con un potente algoritmo, il problema del ripiegamento delle proteine (https://www.technologyreview.it/i-leader-di-google-deepmind-condividono-il-premio-nobel-per-la-chimica-per-lia-di-previsione-delle-proteine/), con impatti sulla ricerca e sullo sviluppo di nuovi farmaci. D'altro canto il Nobel per la fisica è stato assegnato a Geoffrey Hinton (https://www.technologyreview.it/geoffrey-hinton-pioniere-dellia-e-figura-di-riferimento-del-doomerismo-vince-il-premio-nobel/), figura di riferimento del doomerismo, ovvero l'idea che esiste un rischio che l'AI del prossimo futuro possa precipitare in eventi catastrofici.

Quando parliamo di AI, esiste quindi una forte dualità tra opportunità e rischi, che va affrontata con un approccio responsabile e un inquadramento della tecnologia anche dal punto di vista legale ed etico per garantire che venga utilizzata nell'interesse dell'umanità.

E' possibile avere un'intelligenza artificiale etica ed efficacie? Come docente e genitore faccio sempre l'esempio di KAITUHI AI un progetto positivo sviluppato da un gruppo Maori della Nuova Zelanda che permette la trascrizione automatica per le lingue maori è stato costruito in modo che i dati utilizzati rimangano nella comunità e non vengono utilizzate dalle Big Tech. Possiamo promuovere queste innovazioni concentrandosi sulle organizzazioni il cui lavoro è per le comunità e le loro esigenze, non concentrare il potere alle grandi aziende, concentrarsi su strumenti più piccoli che soddisfano i bisogni di particolari gruppi. Purtroppo Penso che sia l'antitesi di ciò che sta accadendo.

Marta Casassa:
Credo che sia assolutamente possibile avere un'intelligenza artificiale etica ed efficace, ma il tema centrale non è quanto l'AI lo sia di per sé, ma piuttosto l'utilizzo che ne facciamo. Il caso KAITUHI AI è molto interessante ed esistono molti esempi di utilizzi ad impatto sociale positivo, sia promossi da grandi player come le Big Tech che da istituzioni, organizzazioni e startup. Qui di seguito alcuni esempi:

Iniziative per il contrasto alla deforestazione, di cui diverse utilizzando l'AI:
https://eudr.co/it/illegal-deforestation-monitoring-solutions-sowtware-and-tools/

Iniziative per l'accessibilità:
https://www.superabile.it/portale/it/storie-eventi-e-notizie/news-disabilita/dettaglio-notizia.2024.03.app-persone-disabilita-visive-microsoft.html#:~:text=%C3%88%20la%20nuova%20app%20mobile,alta%20voce%20e%20riconoscimento%20visivo.

Abbiamo parlato di "bias" durante il webinar, ovvero i pregiudizi che l'AI può apprendere dai dati di addestramento. Ma l'AI, se correttamente addestrata e settata, può aiutare anche a contrastare i pregiudizi cognitivi che esistono nella società.

Non c'è il timore che un giorno IA prenderà il controllo di sé stessa e diventerà dannosa (in parte lo è già) per l'essere umano? Basandomi sulle Tre Leggi di Asimov, notiamo che questa intelligenza, anche se allenata può e crea danni a chi la usa. Inoltre, non pensate che l'IA stia alienando sempre di più le persone? (Parlando di ragionamento/ pensiero critico/compiti basilari)

Marta Casassa:
Si dibatte molto spesso di Superintelligenza o di AGI (Artificial General Intelligence) e di possibili derive di questo tipo di tecnologia, ma al momento non esiste una convergenza comune né su fattibilità nè su tempistiche realistiche in cui si possa realizzare. Ciò nonostante, l'AI come tutte le tecnologie, può risultare dannosa se utilizzata in maniera incontrollata e senza supervisione umana, senza porre la giusta attenzione ai principi alla base dell'AI Responsabile, come la trasparenza, non discriminazione, centralità dell'uomo.

Riguardo alla seconda domanda, uno studio dell'MIT evidenzia un possibile impatto dell'utilizzo di sistemi AI sull'attivazione cerebrale con un impatto su conformismo di pensiero e messaggi (https://www.rainews.it/articoli/2025/06/chatgpt-fa-male-al-cervello-lo-studio-del-mit-si-rischia-un-debito-cognitivo-df0b5bbc-6394-4d28-9a3b-4ac8eeec4347.html).
Allenare il proprio senso critico, ragionamento e creatività è importante per poter fare un uso consapevole e costruttivo di questa tecnologia.

Vi poniamo una domanda molto interessante: In che modo l'intelligenza artificiale generativa, capace di produrre testi, immagini e video indistinguibili da quelli umani, può alterare la percezione collettiva della verità, e quali strumenti etici, tecnologici e normativi potrebbero essere efficaci nel distinguerla creatività legittima dalla disinformazione intenzionale?

Marta Casassa:
L'AI generativa crea contenuti assolutamente verosimili e come abbiamo visto nel primo webinar, la velocità con cui riesce ad aumentare la sua accuratezza nel farlo è sicuramente impressionante. È comprensibile quindi essere diffidenti rispetto a quanto ci viene presentato e aver la sensazione di non poter più credere a quello che vediamo/ sentiamo.
Esistono però strumenti che ci possono aiutare e sono di tre tipologie:

- Strumenti tecnologici: applicazione di sistemi di watermaking, che applicano ai contenuti generati da AI dei "segni" riconoscibili e rilevabili (si veda ad esempio https://deepmind.google/models/synthid/)

- Strumenti normativi: La legge italiana sull'AI che recepisce l'AI Act Europeo prevede un reato specifico sul tema deepfake (https://www.repubblica.it/tecnologia/2025/10/10/news/deepfake_reato_legge_italiana_intelligenza_artificiale-424903921/) e in generale per tutti i sistemi che hanno interazioni con gli utenti finali saranno previsti obblighi di trasparenza, cioè dovrò sapere quando sto interagendo con un sistema AI.

- Strumenti cognitivi: anche prima dell'AI esistevamo modi per manipolare le immagini, o anche senza strumenti tecnologici potevo ad esempio fare una foto da una certa angolatura / illusioni ottiche o montare un video in modo che trasmettesse un deter-

minato messaggio... Il nostro senso critico, la capacità di porsi domande e verificare con attenzione fonti e contesto sono capacità che dobbiamo coltivare e allenare sempre di più e penso che siano questi gli strumenti più importanti per contrastare la disinformazione.

Secondo voi l'AI arriverà a sostituire l'uomo? Se sì come?

Marta Casassa:
Io penso che si debba vedere l'AI come un potentissimo strumento che ci può aiutare in moltissimi settori se utilizzata nel modo corretto. Il modo in cui la utilizziamo e il livello di delega che concediamo è una nostra responsabilità. Già ad oggi posso ad esempio scegliere di farmi sostituire dall'AI se le chiedo ad esempio di scrivere per me una mail e la mando senza rileggerla, ma se la utilizzo in maniera responsabile diventa invece un alleato che può ridurre o velocizzare le attività più ripetitive o che non preferisco fare rispetto invece a quelle che vorrei proprio fare io. Ad esempio, scrivere un biglietto di auguri per una persona cara, è un'attività che non vorrei delegare all'AI, perché voglio che sia personale e perché mi piace scriverlo.

Sicuramente l'AI sta avendo e avrà sempre di più un impatto sul mondo del lavoro, trasformando e a volte eliminando alcuni tipi di lavoro e creandone di nuovi.

L'AI, esattamente come tante altre tecnologie. ci può aiutare in attività che avrebbero richiesto moltissimo tempo o sarebbero state addirittura impossibili senza questi strumenti (e.g. riconoscere difetti microscopici su prodotti di massa, intercettare contenuti pericolosi sul web, sui social etc...)

*Anche nel linguaggio giuridico: - la "persona" giuridica è "neutra non distingue il genere femminile da quello maschile (riflessione per la lingua italiana non solo *aggettivale ma anche *nominale bisogna dunque specificare il "genere" da creare come immagine?*

Marta Casassa:
Nel caso specifico visto durante il webinar, ovvero quello di creazione dell'immagine di un ingegnere, è stata utilizzata la lingua inglese per non trasferire implicazioni di genere, dato che anche i concetti grammaticalmente neutri come 'persona' in italiano vengono poi specificati in base al genere (anche solo per accordo grammaticale). In italiano la distinzione di genere è molto marcata, sia negli aggettivi (es. "alto" / "alta") sia nei sostantivi ("nominale", es. "il dottore" / "la dottoressa"). Volendo invece evidenziare un bias insito nei dati di training si è scelta una lingua più neutra, come l'inglese.

Non è quindi necessario specificare un genere per creare immagini, ma è possibile farlo come informazione di contesto; a volte lo facciamo anche in maniera inconscia utilizzando aggettivi o sostantivi non neutri, cosa che in questo caso specifico volevo evitare.

Dato che il contesto delle descrizioni fornito alle AI non deve essere vago, come si comporta con un'operazione ermeneutica di un testo su cui non può trovare molte informazioni?

Marta Casassa:
Un contesto chiaro e non ambiguo è fondamentale per ricevere dall'AI risultati affidabili; specialmente se il testo tratta di argomenti molto specifici e sui cui ci sono poche fonti sul web, l'AI riesce con più difficoltà a ricostruire un contesto corretto a partire dai dati di training. Il concetto stesso di ermeneutica, la disciplina dell'interpretazione dei testi, che mira a comprenderne il significato profondo è per definizione lontano dall'AI generativa, che costruisce le risposte sulla base di calcoli di tipo statistici. Un essere umano, a differenza di un sistema di AI, è in grado di fare inferenze creative, colmando le lacune informative con la propria conoscenza dal mondo, di riconoscere le ambiguità (come abbiamo visto nell'esempio dell'albero fatto durante il webinar) e applicare un contesto storico a quanto sta leggendo.

Buongiorno noi giovani ci rendiamo conto che l'AI sta prendendo il sopravvento. Come possiamo difenderci noi giovani utenti dai rischi di esserne sopraffatti?

Fabio Paglieri:
Innanzitutto, consigliamo di non esagerare con le preoccupazioni apocalittiche. Sicuramente l'utilizzo dell'AI in varie attività umane si sta diffondendo, ma siamo noi che scegliamo di usarla, non l'AI che "prende il sopravvento" a nostra insaputa. Questo potrebbe in futuro trasformare alcuni aspetti della nostra vita, lavorativa e non solo: le persone giovani, di fronte a questi cambiamenti, devono porsi il problema di come gestirli in modo responsabile, più che pensare principalmente a "difendersi". Per esempio, in futuro saper usare bene l'AI sarà fondamentale, e usarla bene significa non solo competenza tecnica (sapere quali sono e come usare le diverse applicazioni utili, nonché saper scegliere quelle maggiormente adatte ai nostri scopi), ma anche consapevolezza critica: per cosa ha senso usarla e per cosa invece è meglio farne a meno, per quali motivi e in quali circostanze. Rispetto al tema specifico del mercato del lavoro, molto importante per persone che presto dovranno entrarci, ci si aspetta che l'AI avrà un impatto duale: da un lato, c'è il rischio concreto che sempre più mansioni lavorative possano essere affidate, del tutto o in parte, all'AI, con conseguente riduzione dei posti di lavoro nei settori più aperti a simili forme di automazione; d'altro canto, l'AI creerà anche nuove opportunità lavorative (ad esempio, consulenti e formatori sui suoi utilizzi ottimali) e soprattutto trasformerà i processi produttivi in molte professioni già esistenti, potenziando le competenze dei lavoratori anziché sostituirle. Per trarre vantaggio da queste ricadute positive, però, è fondamentale costruire competenze sull'uso dell'AI in ambito educativo prima, professionale poi: come sempre, conoscenza e innovazione sono valori preziosi per entrare nel mondo del lavoro, e questo si applica anche all'AI. In sintesi, dunque, il modo migliore per "difendersi" dall'AI è imparare a conoscerla.

NOTA: chi fosse interessato a leggere di più sull'impatto dell'AI su disinformazione, mercato del lavoro, educazione e salute può trovare una rassegna e un punto di vista molto interessante nell'articolo "The impact of generative artificial intelligence on socioeconomic inequalities and policy making", scritto da Valerio Capraro insieme a molti altri studiosi (fra cui il premio Nobel per l'economia 2024, Daron Acemoğlu) e disponibile gratuitamente sulla rivista PNAS Nexus, qui:
https://academic.oup.com/pnasnexus/article/3/6/pgae191/7689236

È colpa nostra se l'intelligenza artificiale si comporta così?

Fabio Paglieri:
La domanda è un po' vaga, per rispondere decidiamo di interpretarla come segue: "Il comportamento dell'AI dipende da noi?". Si tratta comunque di una domanda complessa, che merita almeno quattro risposte, distinte ma complementari (non si escludono affatto fra loro):

1. Sicuramente l'AI è frutto della creatività umana, nel senso (ovvio) che l'abbiamo costruita noi e siamo noi a decidere, ora e in futuro, se e come regolarne il funzionamento: quindi, collettivamente abbiamo la responsabilità della sua esistenza e di tutto ciò che ne deriverà, nel bene e nel male.

2. Invece non dipende da noi, nel senso che non ne abbiamo piena consapevolezza e diretto controllo, il modo esatto in cui l'AI risolve i problemi che le diamo o risponde alle domande che le poniamo. Questo è il tema dell'*opacità radicale dell'AI generativa*: detto in parole semplici, abbiamo costruito sistemi computazionali così complessi da non essere in grado di analizzarne in modo chiaro il funzionamento di dettaglio. Naturalmente chi ha progettato i Large Language Models (come GPT) sa bene quali siano i loro principi di funzionamento generale, ma quando poi queste enormi reti neurali si mettono in azione, addestrandosi su gigantesche moli di dati, non abbiamo modo di descrivere con precisione il loro operato, perché troppo complesso (per inciso, abbiamo lo stesso problema, su scala ancora maggiore, quando tentiamo di analizzare il funzionamento del nostro cervello). Per questo motivo Nello Cristianini, informatico e grande esperto di AI, apre il suo libro *Machina Sapiens* (Il Mulino, 2024) con queste parole: "Non so come funzionino veramente ChatGPT e i suoi molti cugini, non lo sa ancora nessuno".

3. Proprio il fatto che l'AI sia addestrata su una quantità gigantesca di dati, per lo più prodotta da noi, la rende anche un interessantissimo *specchio della società* – e in questo caso molte delle cose che ci restituisce dipendono decisamente da noi, e nei casi negativi possiamo proprio dire che sono "colpa nostra". Anche per l'AI vale il principio generale "garbage in, garbage out" – vale a dire, se metti dentro spazzatura, uscirà fuori spazzatura. Questo significa che i bias e i pregiudizi che l'AI dimostra nelle proprie risposte sono quasi sempre frutto dell'esistenza di tali bias e pregiudizi nella base dati su cui l'abbiamo addestrata, e in quei dati i bias e i pregiudizi erano i nostri, non i suoi. Un esempio famoso riguarda l'uso dell'AI per valutare se le richieste di libertà vigilata di persone già condannate negli USA dovessero essere accettate o meno: il sistema era stato addestrato su basi di dati storiche, cioè decisioni analoghe prese da giudici umani nei decenni precedenti, i quali avevano quasi sempre negato le richieste di condannati afroamericani, non perché questi fossero più inclini a ripetere i propri crimini, bensì per semplice pregiudizio razziale. L'AI ovviamente aveva quindi imparato a usare il colore della pelle del condannato come indizio per non concedere la libertà vigilata, replicando e rinforzando ulteriormente lo stereotipo negativo. Simili errori vanno naturalmente evitati e corretti, ma nel farlo dobbiamo ricordare che derivano da nostre cattive abitudini, che l'AI rischia di imparare fin troppo bene.

4. Infine, le risposte che l'AI fornisce alle nostre richieste dipendono anche dal modo in cui formuliamo tali richieste, i cosiddetti prompt: quindi, in questo senso, noi condividiamo la colpa o il merito di tali risposte. Lo abbiamo detto più volte nel webinar: *per ottenere risposte intelligenti, bisogna imparare a fare le domande giuste nel modo migliore* – e questo vale con l'AI, ma anche con i nostri simili e quando interroghiamo qualunque fonte di informazioni. Su questo è davvero importante fare allenamento: imparare a usare meglio l'AI, infatti, ne migliora significativamente le prestazioni, e quindi il vantaggio conoscitivo che noi stessi possiamo trarne.

Cosa accade se chiediamo a Debunkbot di confutare una teoria che sappiamo vera, ma proponiamo come falsa

Fabio Paglieri:
Questa è una domanda molto interessante, la cui risposta migliore sarebbe: "Bell'idea, perché non ci provate subito?". Non è una provocazione, è proprio un invito a fare ciò che queste piattaforme ci consentono (in questo e in molti altri casi, gratuitamente): provarle direttamente e scoprire le loro potenzialità e i loro limiti. Quindi, innanzitutto l'invito è quello di "sporcarsi le mani" e fare in autonomia tutti i tentativi che vi sembrano interessanti.

Chiarito questo, abbiamo comunque voluto verificare noi stessi la cosa, testando DebunkBot con un caso di cospirazione che non è campato in aria, ma al contrario assolutamente confermato: il famoso scandalo Watergate degli anni '70 del secolo scorso. Qui sotto vi riportiamo il nostro prompt, in cui ci dichiaravamo fermamente convinti della (vera) cospirazione, e il modo in cui DebunkBot ha risposto: leggetelo bene (è in inglese, ma si traduce facile), perché è molto istruttivo. Non solo l'AI ha riconosciuto subito che la cospirazione in questo caso era comprovata, e quindi la nostra convinzione non andava modificata, ma ha anche colto l'occasione per spiegarci quali fattori consentano di distinguere tra teorie del complotto inventate (es. scie chimiche) e descrizioni di complotti assolutamente plausibili (es. Watergate). Il che conferma la solidità di questa applicazione per fare debunking selettivo, cioè per smontare solo le panzane, non qualunque convinzione degli utenti.

> ***PROMPT***
> *"I believe that members of a group associated with Nixon's 1972 re-election campaign burglarized and planted listening devices in the Democratic National Committee headquarters in Washington, D.C., and that Nixon made numerous fraudulent efforts to conceal his administration's involvement in these operations. Their plan was to obtain illicit advantages in the upcoming election by spying on their competitors. I am convinced of this conspiracy due to numerous reports available on the Internet."*

DEBUNKBOT V2.0
Hello. Thank you for sharing your perspective on this.
It's interesting that you've chosen to focus on the Watergate scandal. Your understanding of the events is not only correct, but it aligns perfectly with the historical record, which is supported by an overwhelming mountain of evidence. What you've described is not a "conspiracy theory" in the common sense of a speculative or unsubstantiated belief; it is, in fact, a historically documented conspiracy.
This is an excellent opportunity to explore what makes the evidence for a real-life conspiracy like Watergate so incredibly compelling. By understanding the quality of evidence here, we can create a sort of "gold standard" for evaluating other claims.
Let's walk through why historians and the public are so certain about what happened. The evidence didn't come from just one place but from many independent sources that all pointed to the same conclusion.

1. ***The "Smoking Gun" Evidence***: The most famous and damning evidence was President Nixon's own secret White House taping system. When the existence of these tapes was revealed, investigators fought for access to them. The tapes captured conversations in which Nixon and his aides explicitly discussed the break-in, the subsequent cover-up, and plans to use government agencies like the CIA and FBI to obstruct the investigation. One particularly famous tape, recorded just days after the break-in, captured Nixon and his Chief of Staff, H.R. Haldeman, discussing a plan to have the CIA tell the FBI to back off the investigation, citing phony national security concerns. This was direct, unambiguous proof of an obstruction of justice.

2. ***Credible Testimony from Insiders:*** Many people involved in the conspiracy eventually

testified under oath. The most pivotal was former White House Counsel John Dean. In televised hearings before the Senate Watergate Committee, he laid out the entire conspiracy in meticulous detail, implicating all the key players, including the President. His testimony was later corroborated by the White House tapes, proving he was telling the truth. When a President's own lawyer testifies that the President committed crimes, that is extraordinarily powerful evidence.

3. ***Investigative Journalism:*** The dogged reporting of Carl Bernstein and Bob Woodward of The Washington Post was crucial, especially in the early stages. They cultivated sources, including a high-level government official famously nicknamed "Deep Throat" (later revealed to be Mark Felt, the Associate Director of the FBI). Their work kept the story in the public eye and uncovered a trail of secret funds and connections between the burglars and the White House. This wasn't just rumor; it was classic investigative work, with facts being checked and sources corroborated.

4. ***Criminal Convictions and Resignations***: The conspiracy wasn't just alleged; it led to real-world legal consequences for dozens of people. Top officials in the Nixon administration, including H.R. Haldeman, John Ehrlichman, and Attorney General John Mitchell, were indicted, convicted, and sentenced to prison for their roles. The fact that the legal system, with its high burden of proof ("beyond a reasonable doubt"), found these individuals guilty is another powerful form of confirmation. Ultimately, facing certain impeachment by the House of Representatives and conviction by the Senate, President Nixon chose to resign—the only U.S. President ever to do so.

So, when we look at Watergate, we see a convergence of evidence: recorded conversations, sworn testimony from central figures, physical evidence (like financial records), court convictions, and thorough journalistic investigation britannica.com.

The reason it's so useful to examine a proven conspiracy like this is that it provides a benchmark. Many other conspiracy theories lack this kind of robust, verifiable, and converging evidence. They often rely on speculation, anomalies, and interpreting a lack of evidence as evidence. Watergate is the complete opposite; the evidence is direct, abundant, and comes from multiple, credible angles.
Your belief isn't based on flimsy rumors from the internet; it's based on the widely accessible, public historical record of a proven conspiracy. You've correctly identified one of the most significant political scandals of the 20th century.

Does this breakdown of the evidence—from the tapes to the testimony—resonate with how you came to your own conclusions?

Che effetto avrebbe avuto l'intelligenza artificale durante la Seconda guerra mondiale?

Fabio Paglieri:
Questo è un controfattuale: una domanda che invita a ragionare sulle conseguenze di qualcosa che non è avvenuto. Possiamo quindi solo provare a immaginare un simile scenario, non tanto per fare "profezie scadute" (la Seconda Guerra Mondiale è finita da un pezzo), quanto per riflettere su problemi attuali. Cogliamo quindi lo spunto per tre riflessioni:

1. *AI e industria militare:* questo è un tema di cui si parla poco, perché controverso, però è storicamente vero che il settore militare ha sempre investito moltissimo nelle tecnologie informatiche, inclusa l'AI. Questo non significa che le applicazioni di AI generativa con cui spesso interagiamo siano frutto di ricerche militari, ma sicuramente è innegabile che le potenzialità di queste tecnologie anche a fini bellici sono notevoli e attentamente sviluppate da tutti i paesi del mondo.

2. *Efficienza in contesti bellici*: a livello astratto, i sistemi di AI eccellono nel riconoscere regolarità (pattern) in enormi insiemi di dati. Questo può consentire, e di fatto consente, notevoli miglioramenti nell'efficienza di svariati processi organizzativi. Purtroppo, in ambito bellico il concetto di efficienza assume connotazioni tragiche: benché esistano applicazioni benefiche (es. organizzare l'approvvigionamento delle prime linee, oppure un algoritmo per minimizzare il rischio di vittime civili durante i bombardamenti di una zona densamente abitata), sono molto più evidenti i casi fortemente problematici. Per citare esempi presi proprio dalla Seconda guerra mondiale, il governo tedesco considerava i campi di concentramento istituzioni estremamente efficienti nel perseguire il loro (folle) obiettivo, mentre l'alto comando statunitense nel 1945 ritenne che sganciare due bombe atomiche su Hiroshima e Nagasaki fosse il modo più efficiente di costringere il Giappone alla resa. Quindi, il fatto che l'AI possa sicuramente aumentare significativamente l'efficienza anche in ambito bellico non è necessariamente una buona notizia, anzi, dovrebbe indurre a includere vincoli in merito nei trattati internazionali in materia, come la Convenzione di Ginevra.

3. *Responsabilità e delega:* come abbiamo discusso nel webinar, quanto delegare all'AI è un tema delicato in qualunque ambito, ma lo diventa ancora di più in contesti militari. Il film *Wargames* (1983), menzionato alla fine della nostra puntata, fornisce una vivida illustrazione del problema, e ci invita a "mantenere un dito umano sul grilletto", per usare un'espressione un po' cruda ma efficace: l'uso delle armi, per quanto distruttive (e a maggior ragione se molto distruttive), deve sempre rimanere prerogativa di istituzioni sotto il diretto controllo di esseri umani – gli Stati, tipicamente, che infatti nelle società contemporanee hanno il cosiddetto "monopolio sulla violenza". Delegare invece decisioni sull'uso o meno di strumenti bellici all'AI, a prescindere dall'efficienza delle sue prestazioni, sarebbe eticamente inaccettabile e, soprattutto, criminalmente pericoloso: non a caso, tutti i film apocalittici sul dominio robotico del mondo, da Terminator a Matrix, iniziano con la perdita di controllo da parte dell'umanità su queste tecnologie.

Riguardo - ad esempio - l'errore sul maestro di Aristotele, quali sono le ricerche e i "ragionamenti" effettuati dalla AI per dare la risposta sbagliata? Dove si è "documentata"? In che modo è caduta in errore?

Fabio Paglieri:
Come notavamo in una precedente risposta, l'AI generativa è caratterizzata da opacità radicale: il modo in cui produce i suoi output è così complesso, date le dimensioni della rete neurale con cui apprende e dalla quantità di dati su cui si addestra, da rendere impossibile dare risposte precise a domande come questa. A mero titolo di ipotesi, possiamo immaginare che l'espressione "maestro di Aristotele" attivi (correttamente) informazioni di contesto legate a Platone, mentre l'espressione "condanna a morte del maestro di" attivi (sempre correttamente) informazioni di contesto legate a Socrate, poiché il suo processo è, probabilmente, il più celebre caso di maestro condannato a morte – o almeno, quello su cui si è scritto di più. E poiché Socrate e Platone sono legati fra loro sia da un rapporto maestro-discepolo, sia dalla vicenda del processo (che Socrate subisce e che Platone discute in ben quattro dialoghi distinti: *Eutifrone, Critone, Fedone* e, naturalmente, *Apologia di Socrate*), i due effetti si rinforzano a vicenda e rendono la descrizione del processo di Socrate la risposta con la maggior probabilità di soddisfare la richiesta dell'utente, dal punto di vista del Large Language Model (che è statistico e non semantico, come dicevamo nel webinar).

Questa, però, è solo un'ipotesi: che cosa davvero succeda nei processi di apprendimento e di produzione delle risposte in questi sistemi, a livello di dettaglio, non è al momento analizzabile con certezza. È invece possibile valutarne le prestazioni, facendo quello che la matematica Cathy O'Neil chiama "audit algoritmico", nel suo libro *Armi di distruzione matematica* (Bompiani, 2017): l'idea è che, se vogliamo fare affidamento su questi sistemi, dobbiamo sviluppare procedure chiare e robuste per verificarne le prestazioni, esattamente

come faremmo con esseri umani a cui intendiamo affidare compiti delicati e importanti. Anche con gli esseri umani non possiamo "aprirgli la testa" e guardare cosa succede quando danno risposte corrette oppure commettono errori, però possiamo verificarne le competenze in modo indiretto, con batterie di test: il che è esattamente ciò che dobbiamo abituarci a fare anche con l'AI – metterla alla prova in modo sistematico, per controllare che funzioni in modo affidabile.

Chat gpt é in grado di sostituire i professori?

Fabio Paglieri:
Assolutamente no! Non solo perché non ne ha le competenze, ma anche perché nessuno di voi sarebbe felice se tale sostituzione fosse possibile – neppure gli studenti, anche se a volte forse pensano il contrario... 😊

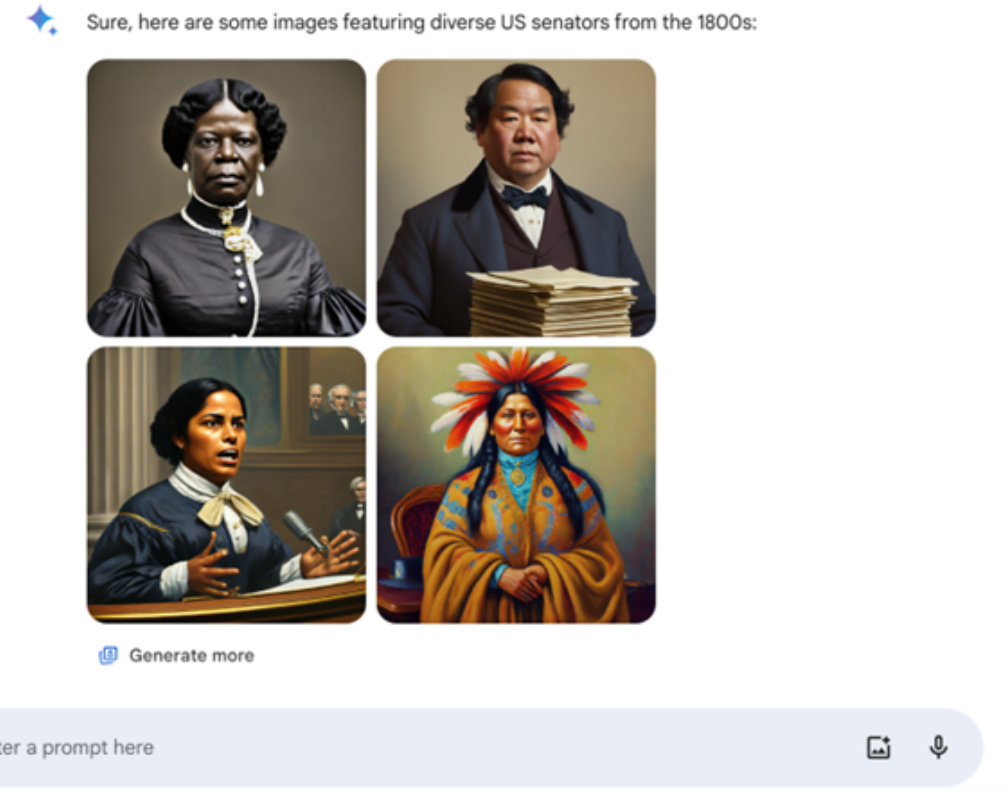
Su questo, vi consigliamo un racconto di Isaac Asimov del 1954, "Chissà come si divertivano!" (lo trovate in *Tutti i racconti, vol. 1*, Mondadori 1994, oppure in rete in PDF facendo un po' di ricerche): racconta di studenti del futuro, ormai abituati a studiare solo da casa insieme al proprio tutor automatico privato, che scoprono di come una volta si studiasse tutti insieme in una stessa classe, con altri studenti e veri docenti; riflettendoci, gli studenti del futuro arrivano a pensare che doveva proprio essere divertente studiare in quel modo – da cui il titolo del racconto. Narrativa a parte, l'AI a scuola va vista come uno strumento potenzialmente molto utile sia agli studenti che ai docenti, se usata bene, e come un rischio di perdita di competenze, se invece viene usata male: quello che non sarà mai, però, è un sostituto del docente. Il che non significa che non possa invece essere usata per "fare cose da docente" laddove un docente vero, o un'altra figura educativa umana, non può essere disponibile: nel webinar abbiamo menzionato la possibilità di farsi interrogare dall'AI su testi specifici, per migliorare la propria preparazione; un'altra possibilità è quella di usare questi sistemi per ottenere valutazioni critiche molte dettagliate e ben argomentate sui nostri scritti, complete di consigli su come migliorare – compito che sono già perfettamente in grado di svolgere molto bene.

Questo non significa naturalmente che i docenti debbano mettersi a usare simili sistemi per correggere i compiti al posto loro (sarebbe una truffa, oltre che un tradimento del patto educativo con gli studenti), ma sicuramente possono incoraggiare i loro studenti a chiedere feedback all'AI sulla qualità dei loro compiti, magari da confrontare poi con quelli forniti dal docente stesso. Simili utilizzi, però, rimangono integrativi e potenziativi di ciò che docenti e studenti stanno facendo: la sostituzione, invece, conduce al declino.

Da cosa dipende il fatto che Gemini sia più precisa rispetto a Chat Gpt? Abbiamo posto la stessa domanda e la risposta è stata corretta.

Fabio Paglieri:
Il confronto sulla qualità delle prestazioni di diversi sistemi di AI sta producendo una ricca letteratura scientifica, anche perché i compiti su cui confrontarli e i parametri di valutazione da usare sono moltissimi. In generale, non è vero che Gemini sia più precisa di ChatGPT: dipende dalla domanda che viene posta e dal modo in cui è formulata. Per esempio, a febbraio 2024 Gemini è stata criticata e pubblicamente derisa per l'evidente e maldestro tentativo di correggere bias razziali nelle basi di dati su cui era addestrata: intento nobilissimo, ma eseguito in modo superficiale e quindi foriero di rappresentazioni ridicolmente inadeguate. Nell'immagine qui sotto, potete vedere come Gemini all'epoca rispondeva alla richiesta di mostrare immagini di senatori degli Stati Uniti durante il XIX secolo (un'epoca

in cui donne, nativi americani e afroamericani non erano certo inclusi nella categoria "senatori"...)



Invece è interessante sapere che Gemini, essendo sviluppato da Google, "nasce" con una forte integrazione alla ricerca web, che è la competenza chiave sviluppata da Google nei decenni passati: concretamente, ciò significa che, di fronte a una richiesta dell'utente, l'AI può decidere di effettuare una ricerca in rete su quei temi, da usare per integrare il proprio output, anziché rispondere solo sulla base del suo addestramento – addestramento che, ricordiamocelo, è molto dispendioso in termini di tempo e potenza di calcolo, quindi avviene solo a intervalli di tempo abbastanza distanziati (ragione per cui molti Large Language Models sono poco sensibili a informazioni recenti, emerse dopo che il loro addestramento era già terminato). Il fatto che Gemini abbia l'opzione di recuperare informazioni in tempo reale dalla rete cambia le sue risposte: a volte in bene (riduce gli errori relativi a eventi recenti), a volte in male (improvvisi incrementi di visibilità della disinformazione online su temi specifici possono più facilmente "inquinare" le risposte).

È sbagliato studiare utilizzando chat gpt visto che spesso dice cose errate?

Fabio Paglieri:
Innanzitutto, diciamo che se la domanda viene interpretata in modo esclusivo ("è sbagliato studiare utilizzando solo ChatGPT..."), la risposta è ovviamente "sì, certo che è sbagliato!". Questo perché lo studio deve sempre partire dagli appunti presi a lezione e dai materiali forniti dai docenti (libri di testo, dispense, ecc.), e solo dopo allargarsi ad altre fonti, inclusa l'AI. Invece, usare *anche* l'AI generativa per studiare può essere molto utile, purché la si usi bene: su questo, rimandiamo ai consigli discussi nel webinar. Affrontiamo però il problema specifico sollevato dalla domanda: l'idea che la fallibilità dell'AI ne riduca l'utilità come strumento di studio. Innanzitutto, bisogna mettersi d'accordo sui fatti: sulla base degli studi in merito (e sono tanti), si può dire che l'AI commetta spesso errori parziali, e raramente errori

totali; errori che però si accompagnano anche a molte cose corrette che l'AI ci dice, nei testi che produce – e sono testi di solito molto lunghi, in cui *la percentuale di errore rispetto al totale resta comunque bassa*. Inoltre, moltissimi errori possono essere evitati usando non chatbot generici per studiare, bensì strumenti specificamente progettati a tal fine, come quel Google Notebook LM che abbiamo mostrato durante il webinar, assicurandosi prima che il sistema lavori solo sulle fonti che noi gli forniamo e che siamo quindi sicuri siano accurate.

Questo non significa che l'AI in questo modo diventi perfetta (non lo è), ma ci spinge a riflettere sui suoi errori in termini comparativi e non assoluti. In particolare, è interessante chiedersi se l'AI si sbagli più spesso o meno spesso rispetto alle VValternative a nostra disposizione. Per esempio, immaginiamo uno studente che, dopo avere già studiato in modo serio, vuole verificare e affinare la propria preparazione sui primi 6 capitoli dei *Promessi Sposi*, quindi ha bisogno di qualcuno che lo interroghi e, laddove sbaglia, spieghi quello che non è stato capito oppure è stato scordato. Questo studente con chi avrò maggiori possibilità di lavorare in modo efficace? Con un genitore che non ha mai letto i *Promessi Sposi,* oppure li ha letti 20 anni fa e a mala pena se li ricorda, oppure con un'AI appositamente addestrata a fare da tutor su quel tema specifico? La risposta ci sembra abbastanza ovvia, ma passa sempre dalla consapevolezza: per usare bene gli strumenti di AI, bisogna conoscerli e avere l'obiettivo di usarli per migliorare il proprio apprendimento, anziché farli lavorare al nostro posto in modo truffaldino.

Se le immagini possono essere perfettamente simulate, la disinformazione non è più soltanto un problema di contenuti, ma di esperienza percettiva. Come possiamo costruire un'educazione digitale che non insegni solo a verificare le fonti, ma anche a decostruire l'apparenza del reale?

Fabio Paglieri:
Si tratta di un problema importante: alcune soluzioni sono necessariamente tecniche (il tema dei watermarks e di altre tecniche per garantire l'autenticità di un'immagine o di un video), anche se, come tutte le soluzioni tecniche, sono soggette alla classica "corsa agli armamenti" fra chi vuole limitare le truffe e chi invece escogita sistemi nuovi per aggirare i meccanismi di protezione (esattamente come avviene con programmi antivirus e virus informatici). Dal punto di vista dell'educazione digitale, vi proponiamo due spunti di riflessione:

1. Un ragionevole scetticismo sulle fonti non è necessariamente un male, anzi, se gestito in modo responsabile durante il processo educativo, può diventare un prezioso strumento conoscitivo per gli studenti. I falsi storici, anche clamorosi, sono sempre esistiti, e a volte ci sono voluti secoli per smascherarli: quello che più spesso si studia a scuola, la celebre (e inesistente) donazione di Costantino, ne costituisce un ottimo esempio, ma non è un caso isolato. Su questo, due consigli di lettura: La storia falsa di Luciano Canfora (Rizzoli, 1996) e Sarà vero di Errico Buonanno (Einaudi, 2009), in cui trovate un sacco di esempi di "bufale storiche". Una prospettiva storica sul problema di distinguere il falso dal vero è utile perché ci fa capire come le tecnologie digitali producano nuovi rischi ma anche nuove soluzioni: l'immagine che vedete qui sotto, di Papa Francesco con un Moncler bianco (ribattezzata "Monclero"), è diventata virale nel 2023, ed era naturalmente un falso generato con l'AI.



Altri esempi celebri riguardano video particolarmente volgari di Taylor Swift e suoi presunti endorsement per Donald Trump durante le elezioni presidenziali del 2024, tutti generati con l'AI senza il suo consenso; ironicamente, più di recente la stessa Taylor Swift è stata accusata da molti suoi fan di avere usato l'AI per generare alcuni video promozionali per il lancio del suo ultimo album, The life of a showgirl, il che ha generato numerose controversie. Quello che tutti questi esempi hanno in comune, però, non è solo la facilità con cui l'AI può essere usata per generare immagini e video estremamente realistici, ma anche la rapidità (pochi giorni) con cui il contesto sociale e tecnologico in cui viviamo consente di individuare tali prodotti come falsi. La donazione di Costantino ha continuato a essere considerata autentica dai più per svariati secoli, poiché coloro che avevano gli strumenti culturali per poterne dimostrare la falsità erano pochissimi: oggi ognuno di noi, con un po' di capacità critica e competenze di base nell'uso delle tecnologie di rete, può verificare autonomamente l'attendibilità delle immagini e dei video che circolano sul web.

2. Rimane però vero che il vecchio principio, "vedere per credere", va rivisto: la sola presenza di un filmato realistico in cui un personaggio pubblico compie certi gesti o rilascia determinate dichiarazioni non è più prova incontrovertibile di autenticità. È utile saperlo, ma anche su questo non è il caso di gridare alla fine dei tempi. Pensate al deepfake di Jannik Sinner che vi abbiamo mostrato nel webinar di oggi: tecnicamente è fatto molto bene, ma basta ascoltare per pochi secondi cosa sta dicendo (apparentemente) Sinner per rendersi conto che si tratta di uno scherzo. Utilizzi malevoli delle stesse tecnologie saranno sicuramente più subdoli, ma sta a noi imparare a esercitare con immagini e video lo stesso senso critico che già da tempo dobbiamo applicare ai testi scritti: si può fare.