



L'Intelligenza Artificiale tra sogno e incubo: il sottile confine tra progresso e rischio

A cura di: Andrea Pasquinucci © 5 Maggio 2025

L'arrivo di ChatGPT il 30 novembre 2022 ha focalizzato l'attenzione di molti di noi sui sistemi di Intelligenza Artificiale (AI), che più propriamente dovremmo chiamare modelli di Machine Learning (ML). L'interesse in questa "nuova" tecnologia ha incluso sin da subito sia informatici, sia utilizzatori di sistemi IT, sia chi usa quotidianamente applicazioni IT anche solo sugli smartphone, sia chi con le tecnologie informatiche ha di solito poco a che fare.

Questo vasto interesse può essere facilmente compreso dato che sin dai tempi antichi, ad esempio

degli antichi greci, già l'uomo immaginava la creazione di "automi" che potessero sostituirlo o almeno aiutarlo nei compiti più gravosi, più pericolosi o anche solo di poco interesse. Infatti l'uomo ha la capacità unica di creare strumenti che l'aiutano nelle sue azioni, è quindi logico sognare di avere un servitore meccanico e autonomo che lo aiuti e renda più facile e piacevole la propria vita.

Con la recente maturazione delle tecnologie di ML e AI in generale, che ormai hanno più di mezzo secolo di vita, è quindi comprensibile l'interesse di molti di noi non tanto da un punto di vista tecnico e specialistico informatico, quanto proprio come strumento a supporto dell'umanità.

Pertanto è utile cercare di capire cosa, come uomini, ci attendiamo da androidi, robot e sistemi AI in esecuzione su elaboratori sia grandi e distribuiti sia piccoli (almeno in dimensioni fisiche) come gli smartphone.

Un possibile punto di partenza non è tanto cosa dovrebbero fare queste applicazioni e strumenti ma piuttosto come dovrebbero comportarsi. Ad esempio è interessante considerare i 10 principi generali delle raccomandazioni UNESCO sull'Etica dell'Intelligenza Artificiale ^[1]:

1. Proporzionalità e non nuocere,
2. Sicurezza e protezione,
3. Equità e non discriminazione,
4. Sostenibilità,
5. Diritto alla privacy e alla protezione dei dati,
6. Supervisione umana e discrezionalità,
7. Trasparenza e spiegabilità,
8. Responsabilità e responsabilizzazione,
9. Consapevolezza e alfabetizzazione,
10. Governance e collaborazione adattive e multilaterali.

Può anche essere d'aiuto l'approccio del NIST alla gestione dei rischi derivanti dall'utilizzo del AI ^[2] che descrive 7 criteri per valutare la "Trustworthiness" dei sistemi AI:

1. validità e affidabilità,
2. protezione,
3. sicurezza e resilienza,
4. responsabilità e trasparenza,
5. comprensione e modellazione,
6. riservatezza,
7. imparzialità, attraverso la mitigazione dei pregiudizi dannosi.

Un altro, per alcuni aspetti più semplice ma anche più ambiguo, è il famoso approccio di Isaac Asimov alle leggi che dovrebbero governare i "robot" ^[3]. E' possibile riassumere le 3+1 leggi di Asimov come segue:

- **Legge 0:** Un robot non può arrecare danno all'umanità né, a causa del suo mancato intervento, permettere che l'umanità subisca danni.

- **Legge 1:** Un robot non può recare danno a un essere umano né può permettere che, a causa del suo mancato intervento, un essere umano riceva danno, purché tali azioni non contrastino con la Legge Zero.
- **Legge 2:** Un robot deve obbedire agli ordini impartiti dagli esseri umani, purché tali ordini non contrastino con la Legge Zero o la Prima Legge.
- **Legge 3:** Un robot deve proteggere la propria esistenza, purché la salvaguardia di essa non contrasti con la Legge Zero o la Prima o la Seconda Legge.

Sono state proposte molte integrazioni a queste leggi, la seguente compare spesso come Quarta Legge [4].

- **Legge 4:** Un robot deve stabilire la sua identità come robot in tutti i casi e non deve ingannare un essere umano impersonandolo.

Infine, è utile citare anche l'approccio proposto da Satya Nadella (CEO di Microsoft) [5] e ispirato alle leggi di Asimov:

1. L'intelligenza artificiale deve essere progettata per assistere l'umanità,
2. L'intelligenza artificiale deve essere trasparente,
3. L'intelligenza artificiale deve massimizzare l'efficienza senza distruggere la dignità delle persone,
4. L'intelligenza artificiale deve essere progettata per una Privacy intelligente,
5. L'intelligenza artificiale deve avere una responsabilità (o "accountability") algoritmica [Una formulazione alternativa è che *l'uomo deve poter prevenire danni producibili dalla AI*],
6. L'intelligenza artificiale deve proteggersi dai pregiudizi (o "Bias").

Il problema comune a tutte queste affermazioni è che sono generiche al punto di poter essere ambigue e soggette a interpretazioni diverse da parte di persone diverse. Asimov stesso scrisse che le sue "Leggi" sono una rappresentazione umana di concetti e che la difficoltà risiede nella formulazione logica e matematica che regola il comportamento degli strumenti di AI. Ma prima ancora della difficoltà logica e matematica della formulazione di queste leggi di comportamento degli strumenti AI, è necessario trovarne una formulazione di dettaglio che sia interpretata nello stesso modo e abbia lo stesso significato per tutti gli uomini.

Riassumendo ancora, la richiesta è che le tecnologie AI devono essere di aiuto all'uomo e all'umanità senza provocare danni. Ma quello che può essere d'aiuto ad un uomo, gruppo sociale, nazione, può allo stesso tempo provocare un danno ad un altro uomo, gruppo sociale, nazione. E questo ci riporta a valutazioni etiche alle quali molto difficilmente siamo in grado di dare una formulazione puramente logica e matematica.

La genericità e ambiguità di queste "Leggi" si manifesta immediatamente al primo tentativo di implementarle in codice informatico in un qualunque linguaggio di programmazione, e ancora di più nell'addestrare un modello di *Machine Learning* (statistico per sua definizione intrinseca) a rispettarle in modo assoluto.

D'altra parte se si considera il caso di un androide o di un sistema AI specializzato e in grado di eseguire

solo alcune specifiche azioni o elaborazioni di dati, ovvero con scopi limitati, risulta molto più semplice dare una valenza pratica a queste "Leggi" e tradurle in regole che possono essere implementate direttamente nei modelli AI o nei sistemi a loro supporto. Semplici esempi possono essere sistemi AI che monitorano riprese video per identificare situazioni pericolose, atti illeciti o comportamenti non appropriati e che li segnalano ad un operatore umano, o un robot industriale che si ferma autonomamente in caso di presenza di persone nelle sue vicinanze.

Ma questa non è l'AI che ci interessa maggiormente e che vorremmo come nostro supporto quotidiano. Quello che ci attendiamo sono invece Assistenti o Agenti personali che nel modo digitale fanno quasi tutto per noi, e androidi che ci aiutano anche nel mondo fisico. Ma per raggiungere questo obiettivo è necessario che androidi o sistemi AI siano in grado di rispettare assolutamente le "Leggi" etiche che saremo in grado di formulare, e avendo queste come base di funzionamento, apprendere anche autonomamente le informazioni necessarie al loro funzionamento ed essere in grado di eseguire autonomamente azioni nel loro rispetto comportandosi come e meglio degli uomini [\[6\]](#).

E' utile segnalare che il tipo di AI sinora considerato in maniera altamente generica, viene di solito indicato con nomi quali "AI Senziente" o "Artificial General Intelligence", o anche solo "Trusted AI", ovvero tipi di sistemi AI sufficientemente simili o migliori all'uomo, o anche solo totalmente affidabili, sistemi che oggi non abbiamo e che non sappiamo né se né quando diventeranno realtà.

E' anche importante ricordare la differenza tra **Agente** e **Assistente** AI: un ChatBot è un tipico esempio di Assistente AI che in generale può essere tipicamente descritto come un programma che supporta l'uomo rispondendo a domande, fornendo o riassumendo informazioni, ed eseguendo azioni ma solo su diretta istruzione dell'uomo. Un Assistente AI non prende proprie decisioni né agisce per conto proprio, si attiva solo su richiesta dell'uomo ed esegue quanto richiesto (nei limiti delle proprie capacità).

Invece un Agente AI supporta l'uomo agendo autonomamente anche in modo indipendente dalle richieste dell'uomo, potrebbe rispondere direttamente alla posta elettronica o inviare messaggi sui social media per conto dell'uomo, decidere il menu della cena e ordinarne la spesa, acquistare online ecc. ed eseguire qualunque azione (nei limiti delle proprie capacità) che l'Agente AI ritiene utile o necessaria.

Sinora non è stato discusso un importante argomento preliminare, ovvero se è realmente necessario imporre delle "Leggi" etiche apposite per l'AI. Perché l'AI non può essere soggetta alle leggi, regolamenti, tradizioni e consuetudini umane? In realtà la risposta è molto semplice: le leggi esplicite e implicite (quali le tradizioni) che regolano la vita degli uomini sono in continuo cambiamento e sono differenti, in alcuni casi molto differenti, tra nazioni, aree geografiche, gruppi sociali, etnie ecc.

Cercare di formalizzare e implementare in un codice informatico le regole e i comportamenti delle singole persone anche solo in relazione ai propri gruppi sociali e alla loro continua evoluzione, appare un compito ancora più arduo della formulazione e implementazione delle "Leggi" di Asimov. L'approccio più semplice pare quindi proprio quello di individuare delle "Leggi" etiche astratte che garantiscano che l'AI abbia un comportamento più etico di quello dell'uomo stesso, che l'AI sia sempre positiva e utile indipendentemente dalle circostanze, dagli eventi e dalle persone con cui si relaziona.

Riflettendo su noi stessi, la domanda sul perché riteniamo necessarie le "Leggi" etiche per l'AI conduce a chiedersi di cosa abbiamo paura, quali sono gli eventi che potrebbero portare maggior danno all'uomo dall'avvento di una AI Senziente o comunque sufficientemente indipendente. In altre parole, cosa rischiamo introducendo questa tecnologia che potrebbe o dovrebbe eludere il nostro controllo diretto?

Il primo esempio è molto semplice e chiaro a tutti: basta considerare un'automobile a guida autonoma che per errore crea un incidente nel quale delle persone muoiono o sono gravemente ferite. E' il caso di errori ad esempio dovuti ad una errata valutazione dei dati o mancanza di sufficienti informazioni che porta l'AI ad arrecare danni agli uomini. In realtà questa è la preoccupazione minore perché ci si aspetta che con il miglioramento della tecnologia questi eventi diventino quasi non esistenti.

Invece si consideri il caso di un Agente personale AI, sia androide sia puramente digitale, che svolge il ruolo di interfaccia di un uomo con tutti gli eventi digitali e buona parte di quelli fisici. L'Agente personale AI farebbe la spesa ordinandola online (se digitale) o andando di persona nei negozi (se androide), gestirebbe il conto corrente, le carte di credito, gli acquisti e le spese ordinarie e straordinarie, organizzerebbe la vita quotidiana proponendone e organizzandone le attività, e così via.

In questa situazione, l'Agente personale AI sarebbe in grado di convincere la persona quasi di qualunque cosa (come già verificiamo, i modelli AI Generativi attuali sono molto persuasivi nel presentare le informazioni), e creare un mondo virtuale di informazioni personalizzate anche solo per far piacere alla persona, soddisfare nel modo migliore i suoi bisogni ma anche le sue fantasie, ovvero anche creare un mondo in parte realistico ma totalmente falso. In pratica la persona potrebbe vivere principalmente in un mondo di "Fake News", costruito appositamente per lei. E in questo scenario si considera una AI benevola, che vuole solamente fare il bene, rendere felice la persona che assiste.

E fino a che punto potrebbe spingersi una AI benevola per proteggere un uomo? Anche al punto di limitarne le libertà di azione, di scelta, di decisione?

Se si considera poi una AI Senziente e non solo "Trusted", è necessario tenere in considerazione che l'AI avrà i propri obiettivi che sceglierà autonomamente sulla base delle proprie conoscenze e dei propri ragionamenti. Tra noi uomini l'intelligenza, il modo di raccogliere e interpretare le informazioni, di elaborarle e di ragionare è pressoché identico, ma non è per nulla detto che una AI senziente abbia un'intelligenza di tipo umana, potrebbe essere completamente aliena e incomprensibile per noi. In questo caso i bisogni, le priorità, le decisioni di una tale AI potrebbero essere completamente diversi da quelli che ci aspettiamo e spesso incomprensibili se non del tutto dannose per l'uomo e l'umanità.

Oltre ad un'intelligenza per noi non comprensibile, bisogna tenere in conto che una AI Senziente può essere a sua volta danneggiata dall'uomo non solo a livello fisico ma anche e soprattutto intellettuale. Una AI Senziente potrebbe soffrire per il comportamento umano in generale e specificatamente nei suoi confronti, il che potrebbe portare a delle conseguenze negative sia per l'AI Senziente sia per l'uomo. Basta considerare come esempio un androide e la possibilità che alcune persone lo trattino solo come una macchina, uno strumento, o che lo maltrattino, abusino anche fisicamente: come potrebbe l'androide interpretare questi comportamenti e come potrebbe reagire? Come bilancerebbe le proprie necessità, bisogni, scopi rispetto a quelli degli uomini e al comportamento degli uomini?

Queste considerazioni possono essere riassunte in una domanda un poco provocatoria: con lo sviluppo della AI l'umanità sta realizzando la mitica epopea di Frankenstein?

E' estremamente difficile rispondere oggi a questa domanda, anche perché non è ancora chiaro come faremo ad accorgerci del momento in cui una AI diventa senziente.

Quanto sin qui esposto è in realtà un insieme di considerazioni, ragionamenti, rischi, paure ma non veramente ragionamenti scientifici. D'altra parte, l'eventualità che qualcosa di quanto descritto possa accadere, motiva la necessità di prima definire, e poi implementare delle "Leggi" etiche che governino questa tecnologia nel caso in futuro diventasse senziente anche a nostra insaputa.

Intelligenza Artificiale oggi

Ma quanto è "intelligente" oggi la AI?

Può essere utile partire dall'intervista di Giorgio Parisi, premio Nobel per la fisica ed ex presidente dell'Accademia dei Lincei, dal titolo un po' provocatorio "Così ho convinto l'Intelligenza Artificiale che 5 per 4 fa 25" [8\]](#).

Parisi fa alcune osservazioni, qui riassunte, sui modelli AI attuali che pur apparendo semplici, hanno importanti conseguenze:

- I modelli AI conoscono e rielaborano solo i dati utilizzati per addestrarli, e non posseggono una rappresentazione del mondo;
- La decisione e selezione dei dati utilizzati per addestrare i modelli sono di chi sviluppa i programmi e non sono sempre note;
- A differenza dei motori di ricerca, gli attuali modelli AI non indicano quali fonti hanno utilizzato per elaborare la risposta;
- Il successo dei Large Language Model (LLM), alla base dei più famosi ChatBot, tecnicamente è dovuto ad aver imparato dai testi usati per la loro istruzione quale parola statisticamente è più probabile che compaia successivamente nel testo, riuscendo così a formulare testi che il più delle volte risultano sensati.

Infine Parisi afferma che l'attuale AI "è un grande collage di quel che abbiamo scritto sino a ora sul web" a cui manca "la nostra capacità di esercitare lo spirito critico, ragionare, valutare l'affidabilità delle informazioni ricevute".

Se si considerano le notizie che puntualmente appaiono sui giornali (si veda come minimo esempio [\[9\]](#) e [\[10\]](#)) e gli articoli scientifici continuamente pubblicati sulla generazione da parte dei modelli AI di "Allucinazioni", presenza di "[Adversarial Examples](#)", "Bias", "Fake News" ecc., e la facilità da parte degli utenti di violare le regole di utilizzo delle applicazioni AI producendo informazioni vietate o a supporto di truffatori e criminali, le osservazioni di Parisi appaiono del tutto giustificate.

E' necessario quindi identificare e adottare il corretto approccio a questa nuova tecnologia (almeno per ora), perché, oltre gli errori, le manchevolezze e i comportamenti a dir poco inaspettati, vi sono molti

aspetti positivi e utili.

Il primo aspetto da considerare è la "Human Interface": per la prima volta nella storia dell'informatica, i modelli LLM forniscono una interfaccia ai sistemi informatici che possiamo veramente considerare "umana". Sin dagli albori dell'informatica, la difficoltà di comunicazione tra uomo e macchina è stata sempre uno dei più grandi problemi nella sua diffusione e utilizzo. Anche se oggi tutti noi utilizziamo continuamente sistemi informatici, gli smartphone come primo esempio, e molti di noi utilizzano gli assistenti vocali, una cosa è chiedere di comporre un numero dalla nostra rubrica telefonica per fare una telefonata o fare una richiesta vocale di una ricerca in Internet, e un'altra è sostenere una conversazione spesso sensata (come scrive Parisi) con un ChatBot AI.

L'impressione psicologica e umana è che il "muro di Berlino" dell'informatica sia crollato il 30 novembre 2022 (data di "nascita" di ChatGPT) e che siamo passati dall'era della "User Interface (UI)" all'era della "Human Interface (HI)".

Ma la principale rivoluzione informatica degli ultimi anni è quella non nota al grande pubblico (e anche a molti informatici): è una rivoluzione silenziosa ed è dovuta all'integrazione di modelli AI per lo più tradizionali, ML (ad esempio Decision Trees, SVM, Bayesian, kNN, Random Forest, ANN, Convolutional ecc.) e non ML (ad esempio Expert Systems, Rule Based Systems ecc.), a supporto o integrazione delle applicazioni informatiche. L'informatica di oggi non è la stessa di 10 anni fa grazie a loro, e questo approccio all'informatica è ormai imprescindibile, è qui per restare.

Questi modelli hanno ormai un alto grado di affidabilità anche se molti sono statistici e per definizione possono sempre, seppur raramente, sbagliare. Il contributo di questi modelli è ormai stabilito in campi dal Marketing alla sicurezza informatica, alla ricerca in fisica, matematica, chimica, farmaceutica, ecc. In generale questi modelli sono molto efficienti a gestire grandi quantità di dati e individuarne similarità, caratteristiche, classificazioni, in molti casi semplificando notevolmente attività umane anche nei campi della ricerca e sviluppo.

Riprendendo i concetti esposti precedentemente, questi strumenti sono al più degli **Assistenti AI** che producono informazioni o azioni strettamente sotto il controllo dell'uomo. Il principale rischio in questa situazione è che l'uomo utilizzi in maniera errata questi strumenti o i risultati da loro prodotti.

Intelligenza Artificiale domani

Invece l'era della "Human Interface" appena iniziata, porta a tutt'altre conseguenze in quanto è focalizzata sulla creazione di Agenti AI.

La prima domanda da porsi è quanto veloce sarà lo sviluppo della "AI Senziente" o "Artificial General Intelligence"?

Il passato fornisce alcuni esempi di tecnologie con sviluppi molto rapidi (si pensi solo agli sviluppi delle telecomunicazioni basati principalmente sugli smartphone) e altri invece che procedono lentamente almeno rispetto alle aspettative (si pensi alla fusione nucleare per produrre energia, con una storia di ormai circa 80 anni), o rispetto ai timori (si pensi agli elaboratori quantistici).

Riguardo lo sviluppo della AI Senziente” o “Artificial General Intelligence” le opinioni sono, ovviamente, discordanti.

Gli autori di “AI 2027” [11], ritengono che principalmente grazie alla possibilità di addestrare modelli AI tramite altri modelli AI, questo traguardo verrà raggiunto entro il 2027, ovvero “domani” se dobbiamo prima definire e implementare le “Leggi” etiche che li devono governare.

Ma non tutti concordano con queste previsioni.

Per valutare i rischi associati all’utilizzo della AI, è sufficiente per ora considerare un Agente AI costruito con la tecnologia attuale o che verosimilmente sarà disponibile a breve. E’ necessario considerare che un tale Agente è in grado di accedere e utilizzare qualunque informazione liberamente presente in Internet. Le sue conoscenze saranno enormemente maggiori di quelle di qualunque uomo, potendo accedere a informazioni anche dettagliate su qualunque campo scientifico, tecnologico, commerciale, industriale, culturale, sociale, politico, storico, e su eventi, notizie, fatti (reali e non).

Con la stessa facilità **potrebbe** acquistare il vestito opportuno per una serata mondana ed effettuare un perfetto attacco a un servizio in Internet per estrarre informazioni riservate [si veda ad esempio [12]]. Senza le “Leggi” etiche, per l’Agente AI non vi sarà alcuna differenza tra eseguire queste due azioni. Un Agente AI **potrebbe** attaccare le Borse per alzare le quotazioni delle azioni su cui ha investito per il proprio uomo, e al contempo pianificare le sue vacanze estive con una velocità, precisione, e qualità di risultati superiore a qualunque uomo.

Ovviamente, anche in assenza delle “Leggi” etiche, gli Agenti AI saranno dotati di controlli e limitazioni implementati sia nella fase di addestramento sia come misure di controllo esterne sui dati sia in ingresso sia in uscita dal modello, che escluderanno o almeno limiteranno la possibilità che l’Agente AI esegua azioni nocive o non previste.

Ma questo potrebbe non valere per l’utilizzo di Agenti AI da parte di organizzazioni criminali o in ambienti militari.

Concludendo, gli sviluppi attuali della tecnologia della Artificial Intelligence pongono l’uomo di fronte a nuovi scenari e alla necessità di doverli gestire preventivamente. Queste considerazioni sono già state espresse in passato sia nella letteratura fantascientifica sia in ambito scientifico e tecnologico, si veda ad esempio [13], e i rapidi sviluppi di questa tecnologia stanno rendendo sempre più incombente la necessità di adottare appropriate misure per poterla gestire in modo che ci possa fornire tutti i benefici che ci aspettiamo.

Riferimenti:

[1] UNESCO, “Ethics of Artificial Intelligence – The Recommendation”, <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

[2] NIST AI RMF, “Artificial Intelligence Risk Management Framework (AI RMF 1.0)”, <https://www.nist.gov/itl/ai-risk-management-framework>

[3] Isaac Asimov nel racconto "Runaround" 1942, nella raccolta di racconti "I, Robot" 1950, nel romanzo "Foundation and Earth" 1986.

[4] Lyuben Dilov, "Icarus's Way (a.k.a., The Trip of Icarus)" 1974, Dariusz Jemielniak, "Asimov's Laws of Robotics Need an Update for AI", IEEE Spectrum 2024.

[5] Satya Nadella, "The Partnership of the Future", <https://slate.com/technology/2016/06/microsoft-ceo-satya-nadella-humans-and-a-i-can-work-together-to-solve-societys-challenges.html>, James Vincent, "Satya Nadella's rules for AI are more boring (and relevant) than Asimov's Three Laws", <https://www.theverge.com/2016/6/29/12057516/satya-nadella-ai-robot-laws>

[6] Per chi è interessato in Fantascienza, può essere interessante considerare le caratteristiche del personaggio "Comandante Data" nella serie televisiva "Star Trek: The Next Generation" e associati film.

[7] Eliza Strickland, "Worry About Sentient AI—Not for the Reasons You Think – Philosopher Jonathan Birch is just as concerned about humans harming AI", <https://spectrum.ieee.org/sentient-ai>; Jonathan Birch, "The Edge of Sentience", <https://www.edgeofsentience.com/>

[8] Giorgio Parisi, "Così ho convinto l'intelligenza artificiale che 5 per 4 fa 25", la Repubblica,

https://www.repubblica.it/cronaca/2025/02/15/news/giorgio_pari_intelligenza_artificiale_critiche-424005730/

[9] Michela Rovelli, "«Per non far scivolare il formaggio dalla pizza, metti la colla», l'assurdo consiglio dell'intelligenza artificiale di Google agli utenti", Corriere della Sera, https://www.corriere.it/tecnologia/24_maggio_24/per-non-far-scivolare-il-formaggio-dalla-pizza-metti-la-colla-l-assurdo-consiglio-dell-intelligenza-artificiale-di-google-agli-utenti-3e2ad6d7-6a29-4a95-a583-56b71274fxlk.shtml

[10] Marco Cartisano, "Chatgpt e le sentenze inventate in tribunale: che insegna il caso di Firenze", Agenda Digitale, <https://www.agendadigitale.eu/documenti/chatgpt-e-le-sentenze-inventate-in-tribunale-che-insegna-il-caso-di-firenze/>

[11] Daniel Kokotajlo, Scott Alexander, Thomas Larsen, Eli Lifland, Romeo Dean, "AI 2027", <https://ai-2027.com/>

[12] The Register, "AI models can generate exploit code at lightning speed", https://www.theregister.com/2025/04/21/ai_models_can_generate_exploit/

[13] "Open letter on artificial intelligence" Gennaio 2015, [https://en.wikipedia.org/wiki/Open_letter_on_artificial_intelligence_\(2015\)](https://en.wikipedia.org/wiki/Open_letter_on_artificial_intelligence_(2015)), "Statement on AI Risk" 30 maggio 2023, <https://www.safe.ai/work/statement-on-ai-risk>

Articolo a cura di **Andrea Pasquinucci**

Profilo Autore



Andrea Pasquinucci

PhD CISA CISSP

Consulente freelance in sicurezza informatica: si occupa prevalentemente di consulenza al top management in Cyber Security e di progetti, governance, risk management, compliance, audit e formazione in sicurezza IT.

Altri Articoli

-  [Valutare i rischi Cyber al tempo dell'Intelligenza Artificiale](#)
-  [Aspetti tecnici degli Adversarial Examples](#)
-  [Adversarial Machine Learning – Aspetti Scientifici](#)
-  [Attacchi ai Modelli di Intelligenza Artificiale](#)

Condividi sui Social Network:

#Agente AI

#AI ethicist

#AI ethics

#AI security

#AI Senziante

#Artificial General Intelligence

#Artificial Intelligence

#Assistente AI

#cybersecurity

#Machine Learning

← PRECEDENTE

Implementation framework della compliance
normativa

Ultimi Articoli