Digeat N.3 - 19 Settembre 2024

La fiducia nell’intelligenza artificiale: sfide etiche e prospettive future

Di Pasquale Dambrosio

Abstract

L’ intelligenza artificiale (IA) ha trasformato profondamente molti settori, portando con sé nuove opportunità, ma anche importanti sfide legali ed etiche. Se da un lato l’IA offre grandi opportunità, dall’altro solleva dubbi sulla sua conformità non solo ai diritti fondamentali, ma anche ai principi di equità e responsabilità. In questo contesto, è essenziale analizzare le questioni etiche legate all’IA, adottando un approccio che consideri non solo gli aspetti giuridici, ma anche morali ed etici. Questo articolo offre una riflessione critica sulle implicazioni etiche dell’IA, sottolineando l’importanza di un uso responsabile che tuteli i diritti fondamentali.

Indice

Una delle questioni principali riguardanti l’uso dell’intelligenza artificiale (d’ora in poi IA) è il rispetto del diritto alla trasparenza. Con l’uso sempre più diffuso di algoritmi complessi, soprattutto quelli basati sul *machine learning*, **cresce la preoccupazione per la mancanza di trasparenza**[1].

Il diritto alla trasparenza nei sistemi di IA

Gli algoritmi prendono decisioni attraverso processi interni che sono difficili da comprendere e verificare, sia dal punto di vista tecnico che per chi non è coinvolto direttamente.

> Questo fenomeno, noto come “opacità algoritmica”, solleva questioni etiche e legali e mette in discussione la trasparenza delle decisioni prese dai sistemi di IA. Quando le decisioni sono difficili da comprendere e da verificare, c’è il rischio di violare il diritto delle persone a sapere e comprendere le ragioni dietro tali decisioni.

Per garantire un livello adeguato di trasparenza, occorre affrontare una serie di questioni. Innanzitutto, è necessario sviluppare algoritmi che possano **spiegare chiaramente il loro funzionamento e le decisioni che prendono**. Inoltre, è essenziale **mantenere una documentazione dettagliata** riguardante il *design* e l’addestramento degli algoritmi e **comunicare in modo trasparente** come vengono utilizzati i dati. Successivamente, occorre **implementare meccanismi di controllo e *audit***

che permettano di verificare e correggere eventuali errori o abusi nel processo decisionale. Infine, è importante **garantire l'accesso alle informazioni** sui dati utilizzati e **valutare continuamente i *bias*** per assicurare che le decisioni siano eque e non discriminatorie[2].

Credo che affrontare questi aspetti contribuisca sicuramente a costruire sistemi di IA più responsabili e affidabili.

La responsabilità nell'utilizzo di sistemi di IA

Attribuire la responsabilità per le decisioni prese da sistemi di **IA** è sicuramente una delle sfide più complesse[3]. L'unicità di questi sistemi, che agiscono autonomamente basandosi su algoritmi avanzati e capacità di apprendimento continuo, **mette in crisi il tradizionale schema di responsabilità**, **solitamente fondato su un soggetto identificabile**. Infatti, nel caso di errori o danni causati da un sistema di IA, diventa difficile individuare con precisione il soggetto a cui imputare la responsabilità.

La crisi del modello tradizionale di responsabilità

Nel diritto tradizionale, la responsabilità è attribuita a un soggetto che, con la propria azione o omissione, ha violato una norma. Questo soggetto può essere il **programmatore** del sistema, il **proprietario** dello stesso, oppure l'**utente** finale che ne fa uso.

> Tuttavia, i sistemi di IA, in particolare quelli autonomi e basati sull'apprendimento automatico, agiscono in modo indipendente dalle intenzioni del loro creatore o utilizzatore e talvolta prendono decisioni non prevedibili, modificando il proprio comportamento in base a dati nuovi o inattesi.

Questo solleva una serie di domande: chi deve essere ritenuto responsabile nel caso in cui l'AI prenda una decisione errata o dannosa? Deve essere considerato responsabile il programmatore, che ha progettato l'algoritmo? O l'organizzazione che lo utilizza? E cosa dire dell'utente finale, che si affida alle decisioni dell'AI senza avere il controllo completo sul processo decisionale?

Le sfide dell'attribuzione di responsabilità

Una delle difficoltà legate all'IA è sicuramente quella di **stabilire una responsabilità tra il programmatore,** l'azienda che utilizza l'AI **e i danni eventualmente causati dal sistema**.

> Esempio, un sistema di IA commette un errore nel giudizio su una decisione finanziaria, chi dovrebbe essere ritenuto responsabile per i danni subiti dal cliente? Il programmatore potrebbe sostenere di non aver previsto tutte le circostanze che hanno portato all'errore, e l'organizzazione potrebbe affermare di aver implementato correttamente il sistema secondo gli *standard* richiesti. L'utente finale, d'altra parte, potrebbe non avere le competenze per comprendere le dinamiche del sistema, e quindi non potrebbe essere considerato responsabile delle sue decisioni.

Inoltre, **l'opacità algoritmica**, ovverosia la difficoltà di comprendere i meccanismi interni attraverso cui un algoritmo giunge a determinate conclusioni, rende ancora più complesso stabilire chi abbia effettivamente il controllo del sistema e, quindi, la responsabilità delle sue decisioni.

Responsabilità oggettiva: una possibile soluzione?

Per far fronte a queste sfide, una delle possibili soluzioni è **l'adozione del principio di responsabilità oggettiva.**

> In altre parole, chi utilizza un sistema di IA dovrebbe essere ritenuto responsabile delle conseguenze negative generate dal sistema, indipendentemente dalla prevedibilità o meno degli errori. L'applicazione della responsabilità oggettiva ai sistemi di IA potrebbe fornire una soluzione più chiara e semplice al problema dell'attribuzione della responsabilità.

Ad esempio, un'organizzazione che utilizza un sistema di IA per **selezionare personale o gestire decisioni finanziarie** sarebbe automaticamente responsabile per eventuali errori o discriminazioni commesse dal sistema, indipendentemente dalla difficoltà di prevedere o spiegare tali errori.

L'adozione della responsabilità oggettiva pone, però, ulteriori questioni, in particolare in termini di **bilanciamento degli interessi**. Da un lato, garantisce **una maggiore tutela per gli individui danneggiati**, i quali possono ottenere un risarcimento senza dover dimostrare la colpa del responsabile. Dall'altro, però, potrebbe **scoraggiare l'innovazione tecnologica**, imponendo alle imprese oneri eccessivi, soprattutto quando non è possibile controllare pienamente i risultati di un sistema di IA.

La responsabilità distribuita

Luciano Floridi e **Mariarosaria Taddeo** [4] hanno introdotto la nozione di **"responsabilità distribuita uomo-macchina".** Questo concetto evidenzia come, nelle decisioni prese dai sistemi di IA, la responsabilità non possa essere attribuita esclusivamente a un individuo, ma debba essere condivisa tra i progettisti, gli utilizzatori e l'algoritmo stesso.

> La responsabilità distribuita implica, di conseguenza, che sia necessario ripensare il modello tradizionale di attribuzione della responsabilità, riconoscendo che l'interazione tra umano e macchina crea nuove forme di co-responsabilità.

L'idea di una responsabilità distribuita, offre una visione utile, ma a mio avviso dovremmo continuare a riflettere su come applicarla in pratica.

Il rischio di discriminazione nell'uso dell'IA

Gli algoritmi di IA, utilizzati in settori come l'occupazione, la giustizia, la salute e i servizi pubblici, prendono decisioni che **possono avere un impatto significativo sulla vita di ciascun individuo**. Questi algoritmi si basano su grandi quantità di dati, e rischiano di **perpetuare e amplificare pregiudizi sociali esistenti**[5].

Ad esempio, un algoritmo progettato per supportare decisioni di assunzione può essere addestrato su *dataset* che riflettono le pratiche storiche di discriminazione nel settore lavorativo. Se il *dataset* contiene prevalentemente dati relativi a candidati uomini, l'algoritmo potrebbe inconsciamente favorire i candidati maschi rispetto alle donne, perpetuando così una forma di **discriminazione di genere**. Allo stesso modo, un algoritmo utilizzato per la concessione di prestiti potrebbe sfavorire i membri di minoranze etniche se basato su dati che riflettono disparità storiche nell'accesso ai finanziamenti.

> Per garantire che l'utilizzo dell'IA non violi il diritto alla non discriminazione, è necessario promuovere l'equità algoritmica**.**

L’equità algoritmica, in sostanza, si riferisce allo sviluppo e all’implementazione di algoritmi che non solo evitano di riprodurre pregiudizi esistenti, ma che attivamente perseguono decisioni eque e imparziali.

La privacy e i diritti sulla protezione dei dati personali

Un altro aspetto critico dei sistemi di IA riguarda la **protezione dei dati personali**.

I sistemi di AI richiedono grandi quantità di dati per poter funzionare correttamente, e spesso si tratta di **dati sensibili**, come i dati sanitari, finanziari etc. L’uso indiscriminato di questi dati, soprattutto in mancanza di un chiaro consenso informato, come sappiamo, può portare a gravi **violazioni della privacy**.

Il *deep learning* e l’opacità nel trattamento dei dati personali

I sistemi di IA che includono algoritmi di *machine learning* basati su architetture multistrato di reti neurali (il cosiddetto *deep-learning*), con la loro capacità di analizzare enormi quantità di dati, hanno sicuramente portato grandi progressi in molti settori. Tuttavia, **l’opacità che accompagna questo tipo di tecnologia pone dei limiti critici alla trasparenza** che ci si aspetta quando si trattano dati personali.

> Uno dei tanti problemi risiede nella natura stessa del *deep learning*. Gli algoritmi che ne fanno parte si basano su strutture così intricate da rendere difficile, se non impossibile, capire come si arrivi a una determinata decisione o risultato. Quando questi sistemi sono applicati all’analisi dei dati personali, c’è una barriera invisibile tra ciò che viene immesso nel sistema e ciò che ne esce. Chi è soggetto a una decisione presa da un algoritmo di *deep learning* non può sapere esattamente quali criteri siano stati utilizzati. L’opacità non riguarda solo la complessità tecnica, ma anche la difficoltà per i non esperti di comprendere come i propri dati vengano utilizzati.

Il risultato è che, per quanto i sistemi di *deep learning* possano essere “potenti”, la loro mancanza di trasparenza può **minare la fiducia che le persone ripongono nell’uso di queste tecnologie**. Aspetti poco chiari, non trasparenti e opachi. A maggior ragione quando i dati, le decisioni prese o consigliate all’operatore umano dal sistema AI hanno a che fare con la salute, la giustizia e la vita.

L’etica delle macchine: sfide e riflessioni personali

L’etica per le macchine o ***machine ethics*** esplora **come le macchine autonome possano prendere decisioni etiche e rilevanti**. Muller afferma che “l’etica per le macchine si occupa di garantire che il comportamento delle macchine nei confronti degli utenti umani, e forse anche di altre macchine, sia eticamente accettabile”.[6]

Michael Anderson e **Susan Leigh Anderson** [7]hanno, invece, proposto l’inserimento di regole etiche negli algoritmi, un’idea che considero interessante.

> Tuttavia, credo che l’etica applicata alle macchine debba essere vista come un processo dinamico piuttosto che come un insieme rigido di norme. Ciò detto, la complessità delle decisioni morali non deve impedire il tentativo di codificarle. Piuttosto, dovremmo sviluppare approcci che permettano ai sistemi di adattarsi e rispondere alle diverse situazioni in modo etico e responsabile.

Vedo la *machine ethics* non come un obiettivo irraggiungibile, ma come una sfida stimolante che può portare a innovazioni significative.

Conclusioni

L'IA rappresenta una delle **innovazioni più potenti e trasformative della nostra epoca**, **ma porta con sé sfide etiche e giuridiche** di grande portata. Sicuramente il progresso tecnologico non può essere disgiunto da una responsabilità etica e morale.

In questo senso, l'AI Act [8]è un passo importante. La strada per **bilanciare innovazione e tutela dei diritti** è complessa e richiede un impegno costante da parte di tutti gli attori coinvolti. **Credo che il futuro dell'IA debba essere costruito sulla fiducia, la collaborazione e una visione etica condivisa**. Solo affrontando insieme questi temi possiamo garantire che la tecnologia migliori davvero la nostra vita, senza compromettere la dignità umana o i valori fondamentali della nostra società[9].

L'IA è uno strumento incredibile, e come tale dovremmo gestirlo con saggezza, affinché serva l'uomo e non lo domini.

Bibliografia

[1] Gianfranco Basti (2024), *"La sfida etica dell'Intelligenza Artficiale e il ruolo della filosofia", Pontificia Università Lateranense*

[2] Jobin, A., Ienca, M., & Vayena, E. (2019). *"The Global Landscape of AI Ethics Guidelines"*. *Nature Machine Intelligence,* 1(9), 389-399.

[3] Floridi, L. (2019). *"Establishing the Rules for Building Trustworthy AI". Nature Machine Intelligence,* 1(6), 261-262*.*

[4] Floridi, L., & Taddeo, M. (2016). "*What Is Data Ethics?". Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2083).

[5] Binns, R. (2018). *"Fairness in Machine Learning: Lessons from Political Philosophy". Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency,* 149-159.

[6] C. Muller, *Ethics of Artficial Intelligence and Robotics*, in E.N. Zalta (ed.), "Stanford Encyclopedia of Philosophy", (June 1, 2021)

[7] Anderson, M., & Anderson, S. L. (2007), *"Machine Ethics: Creating an Ethical Intelligent Agent"*

[8] European Commission. (2021). *Proposal for a Regulation Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act)*

[9] Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). *"Artificial Intelligence and the "Good Society":* The US, EU, and UK Approach". *Science and Engineering Ethics, 24*(2), 505-528