Istruzione & Formazione News n. 21

Il Vaso di Pandora. Riflessioni sull'Intelligenza Artificiale (1)

Poco tempo fa, l'Unione Europea ha approvato la legge sull'intelligenza artificiale conosciuta come "AI Act", un evento che ha attirato una considerevole quantità di attenzione, in quanto si tratta della prima legge del genere al mondo. Ma cosa stabilisce, di preciso, questa legge? In breve, si tratta di un tentativo di regolare l'uso dell'IA, stabilendo quali applicazioni devono essere vietate (e le relative eccezioni) e quali devono essere considerate come "ad alto rischio", oltre che gli obblighi e requisiti a cui devono essere sottoposte queste ultime, in maniera tale da garantire il rispetto dei nostri diritti; inoltre, l'AI Act propone anche misure a favore dell'innovazione, promuovendo lo sviluppo di sistemi di IA più capaci, ma anche più sicuri. Vediamo quindi di esaminare i punti più importanti di questa legge.

Prima di iniziare, bisogna però chiarire alcune cose: innanzitutto, l'AI Act divide l'intelligenza artificiale[1] in multipli gruppi usando due misure, il rischio e le capacità. In termini di capacità, le IA possono essere "sistemi di IA", che svolgono un numero limitato di compiti, o "modelli/sistemi di IA per finalità generali", che possono svolgere un'ampia varietà di compiti; per quanto riguarda il rischio, invece, un sistema di IA può ricadere nelle categorie "vietato", "ad alto rischio" e "non ad alto rischio", mentre le IA per finalità generali possono essere "con rischio sistemico" o non. Inoltre, nel caso dei sistemi di IA il rischio dipende dall'ambito in cui l'intelligenza artificiale viene utilizzata più che dal compito che svolge; in altre parole, l'uso di un'IA potrebbe essere vietato in certe situazioni, ma accettabile in altre: ad esempio, le IA per il riconoscimento delle emozioni sono vietati in ambito scolastico e lavorativo, ma altrimenti ricadono nella categoria "ad alto rischio"[2].

Partendo da queste premesse, vediamo quindi di esaminare quali sono le applicazioni vietate:

1. Le IA che usano tecniche subliminali o "volutamente manipolative ed ingannevoli" per indurre le persone a prendere certe decisioni;
2. Le IA che sfruttano vulnerabilità come età, situazione economica o disabilità per indurre le persone a prendere certe decisioni;
3. Le IA utilizzate per creare sistemi di scoring che attribuiscono "punteggi sociali" alle persone valutandone comportamento e caratteristiche personali[3];
4. Le IA usate per la profilazione, ossia la pratica di valutare il rischio che una persona commetta crimini in base alla sua personalità e caratteristiche personali;
5. Le IA utilizzate per creare od ampliare banche dati per il riconoscimento facciale facendo uso di immagini ottenute su internet o provenienti da filmati di telecamere a circuito chiuso;
6. Come menzionato precedentemente, le IA capaci di riconoscere emozioni attraverso l'analisi di elementi come espressioni facciali, gesti e voce, qualora utilizzate in ambito scolastico o lavorativo;

7. Le IA che fanno uso uso di dati biometrici, ossia caratteristiche fisiche, comportamentali e fisiologiche (volto, forma del corpo, voce, postura ecc.) per dedurre informazioni come orientamento sessuale, religione od opinioni politiche;Le IA che fanno uso di dati biometrici per permettere l'identificazione in tempo reale delle persone all'interno di spazi pubblici.

Cosa sono, quindi, le IA "ad alto rischio"? In generale, l'AI Act stabilisce una serie di criteri che permettono di identificarle, ma propone anche diversi casi specifici, tra cui le eccezioni di cui sopra. Un'intelligenza artificiale è considerata ad alto rischio se "presenta un rischio significativo per salute, sicurezza o diritti fondamentali", una valutazione che dipende da una varietà di fattori, tra cui la finalità dell'IA, i dati che tratta, il suo livello di autonomia e la portata e misura dei danni potenziali; particolarmente interessante è il fatto che si fa riferimento anche ai benefici ottenibili dall'IA in questo calcolo, implicando che un certo livello di rischio è comunque accettabile, qualora i vantaggi siano sufficientemente alti.

In aggiunta, oltre ai sistemi biometrici e di riconoscimento emozioni non vietati, l'AI Act propone anche una serie di esempi specifici come IA ad alto rischio:

- IA usate in infrastrutture critiche, come gestione del traffico stradale o fornitura di acqua e gas;
- IA usate nell'ambito dell'istruzione e formazione allo scopo di valutare i risultati, monitorare il comportamento e determinare l'accesso o l'ammissione;
- IA usate nel settore dell'occupazione e della gestione dei lavoratori per selezionare ed assumere, per la gestione dei rapporti di lavoro e l'assegnazione degli incarichi, e per valutare e monitorare;
- IA usate per valutare l'ammissibilità ad accedere a servizi pubblici e privati, tra cui l'accesso al credito ed alle assicurazioni, all'assistenza sanitaria, e la classificazione delle chiamate di emergenza;
- IA usate nelle attività di contrasto (ossia per combattere la criminalità) o per gestire frontiere, richieste d'asilo e immigranti;
- IA usate nell'amministrazione della giustizia e nei processi democratici, qualora possano influenzare le decisioni delle persone esposte.

Si noti, inoltre, che, come nel caso delle applicazioni vietate, anche qui abbiamo delle eccezioni, casi in cui un'intelligenza artificiale è considerata non ad alto rischio, tipicamente quando sono usate per scopi puramente amministrativi.

A differenza delle applicazioni vietate, le IA ad alto rischio possono essere utilizzate, ma sono sottoposte ad un "sistema di gestione dei rischi", in sostanza una sorta di esame che deve essere effettuato prima dell'immissione sul mercato del sistema di IA, e che deve essere ripetuto ed aggiornato periodicamente; in parole povere, il fornitore deve esaminare i possibili rischi che il sistema potrebbe causare, inclusi quelli che potrebbero sorgere a causa di un uso improprio, ed adottare misure per limitare o gestire questi rischi. Inoltre a ciò, deve anche essere rispettata una serie di obblighi, alcuni dei quali si applicano al fornitore, altri al deployer (utilizzatore), che garantiscano che l'IA sia quanto più possibile sicura: i dati devono essere completi e rappresentativi

per evitare la creazione di bias, il fornitore deve disporre di una documentazione tecnica che ne spieghi lo sviluppo e il funzionamento e fornire un manuale di istruzioni al deployer, e l'IA deve essere sviluppata in modo da permettere la sorveglianza umana e garantire un elevato livello di robustezza rispetto ad errori e guasti, cybersicurezza inclusa.

Per quanto riguarda le IA non ad alto rischio, esse sono solo brevemente menzionate, in quanto non affette dall'AI Act; in generale sono identificate come intelligenze artificiali che non influenzano l'esito del processo decisionale, ad esempio nel caso in cui siano usate per svolgere compiti altamente limitati, oppure per migliorare il risultato di un'attività umana. Per quanto riguarda obblighi e limitazioni, i fornitori di IA non ad alto rischio "non sono tenuti a rispettare i requisiti stabiliti per i sistemi di IA ad alto rischio", ma sono incoraggiati ad adottarli volontariamente, e creare codici di condotta, facendo riferimento a sette "principi etici non vincolanti"[5].

La situazione per i modelli di IA per finalità generali è leggermente differente, imponendo alcuni obblighi (prevalentemente legati al fornire documentazione tecnica e informazioni su capacità e limiti) che valgono indipendentemente dal livello di rischio, mentre i modelli che presentano un rischio sistemico, definiti come sistemi che hanno "capacità di impatto elevato" o "un impatto significativo sul mercato interno" sono anche soggetti ad ulteriori obblighi riguardanti l'individuazione ed attenuazione dei rischi e il garantire un elevato livello di cybersicurezza; quest'ultimo punto è particolarmente importante, in quanto l'AI Act inserisce la possibilità che qualcuno sfrutti vulnerabilità del sistema tra i più importanti rischi per questa tipologia di IA, insieme all'uso improprio, ai problemi di controllo, e ai modelli autoreplicanti. Bisognerebbe infine menzionare il caso delle IA generative, ossia capaci di creare o modificare immagini, video e audio (e quindi categorizzate come sistemi per finalità generali), che hanno attirato molta attenzione negli ultimi mesi per via del fenomeno dei "deepfake": in virtù della difficoltà nel distinguere questi contenuti "AI-generated" da quelli autentici l'AI Act impone un obbligo di marcatura a fornitori e deployer, richiedendo che i primi trovino un modo per marcare i contenuti in un modo che permetta di rilevarli come prodotti dall'IA, mentre i secondi dovrebbero segnalare che si tratta di contenuti AI-generated qualora vengano pubblicati o condivisi.

Nel suo insieme, l'AI Act è sorprendentemente completo ed efficace, specialmente se si considera che è la prima legge sull'uso dell'intelligenza artificiale al mondo; ciò non vuol dire che sia perfetta, e un'attenta analisi rivelerà certamente alcune mancanze o imprecisioni, ma l'attenzione dedicata ad esaminare le possibili applicazioni dell'IA è considerevole, come dimostra la quantità di eccezioni che vengono evidenziate. Soprattutto, l'AI Act riesce a bilanciare la necessità di fornire indicazioni precise e complete con quella di rimanere adattabile ed aperto a cambiamenti ed aggiornamenti, praticamente un obbligo in un campo in continua evoluzione, risultando così in indicazioni che, sebbene a volte un po' vaghe, risultano efficaci e versatili. Altrettanto importante è il modo in cui l'UE ha deciso di gestire il problema del rischio: viene ampiamente riconosciuto che l'intelligenza artificiale comporta dei rischi (al punto che i sistemi di IA non ad alto rischio sono menzionati solo brevemente, specialmente in confronto a quanto spazio è dedicato ai sistemi ad alto rischio), ma il numero di applicazioni vietate è piuttosto ridotto, e gli obblighi e le limitazioni per i sistemi ad alto rischio sono, per la maggior parte, piuttosto ragionevoli; tutto ciò, unitamente alla dichiarata intenzione di supportare l'innovazione e lo sviluppo, risulta in un regolamento che riconosce chiaramente le capacità dell'intelligenza artificiale, sia in positivo che in negativo, e che

cerca di proseguire in maniera tale da minimizzare i rischi, ma non in maniera tale da nuocere alla ricerca.

(A cura di Manfredi Negro, neolaureato in Scienze Cognitive dell'Università degli Studi di Milano)

1 Di cui fornisce una definizione basata su autonomia e capacità inferenziale come elementi portanti che permettono di distinguerla dai più semplici algoritmi.

2 Ad esempio in ambito medico, dove la capacità di riconoscere emozioni può essere estremamente utile.

3 In maniera molto simile al sistema cinese del "credito sociale".

4 Si pensi al riconoscimento facciale usato, ad esempio, dai cellulari.

5 In ordine, intervento e sorveglianza umani, robustezza tecnica e sicurezza, vita privata e governance dei dati, trasparenza, non discriminazione ed equità e diversità, benessere sociale ed ambientale, e responsabilità.

Milano, 27.03.2024