Rischio inaccettabile - pratiche AI vietate - ex art. 5 - allegato II

L'AI Act è il Regolamento dell'Unione Europea per lo sviluppo, l'introduzione nel mercato dell'UE e l'uso di prodotti e servizi di intelligenza artificiale. Particolare attenzione è stata rivolta alla gestione dei rischi per la salute, sicurezza e diritti fondamentali.

L'articolo 5 elenca tutte le pratiche di AI vietate, tutti quei sistemi che presenterebbero rischi inaccettabili. Vengono specificate 8 categorie, che rappresentano un elenco chiuso in cui alcuni sistemi sono proibiti con alcune eccezioni per scopi di applicazione della legge.

Prima categoria - A

Sono vietati quei sistemi che utilizzano **tecniche subliminali al di là della conoscenza delle persone** oppure **tecniche manipolative per andare a distorcere la loro capacità di prendere una decisione** informata, portandole ad assumere una decisione che non avrebbero preso, producendo di conseguenza un danno significativo.
Maggiori informazioni per questa categoria vengono fornite dal **Considerando 29**.
Sono infatti vietati i sistemi quei sistemi che hanno **l'obiettivo o l'effetto di alterare il comportamento umano** portando danni alla salute fisica, psicologica o sugli interessi finanziari.
Rientrano anche i **sistemi che fanno uso di stimoli audio, immagini, video che le persone non possono percepire**, questi compromettono l'autonomia e il libero arbitrio delle stesse perché al di là della percezione umana. E per questo si può pensare a interfacce macchina-cervello oppure realtà virtuale che presentano un grado di controllo più elevato sugli stimoli.
Un esempio sono quindi **le pratiche commerciali scorrette**, in cui vi sono quelle azioni sleali, aggressive che portano il consumatore a effettuare una scelta "commerciale" che non avrebbe preso.
Vi sono poi i **sistemi che possono sfruttare le vulnerabilità**, età, disabilità o una determinata situazione economica/ sociale.
Le uniche **eccezion**i riguardanti l'uso lecito di questi sistemi sono nel **contesto di trattamenti medici, come trattamento psicologico, riabilitazione fisica**. Usabili, quando sono eseguite in conformità alla legge e agli standard medici e con il consenso degli individui.

Seconda Categoria - B

Sono vietati quei sistemi che **sfruttano una qualunque vulnerabilità di una persona fisica a causa dell'età, disabilità, situazione sociale, economica** per andare ad **alterare il suo comportamento** in modo che causi o probabilmente causi un danno significativo.
Anche qui si fa riferimento al Considerando 29.
La **differenza rispetto alla prima** tipologia di sistemi qui **riguarda il target di persone protette dalla norma**. Si parla di **soggetti vulnerabili, dove queste vengono sfruttate dal sistema**. E tra queste vi è **l'età** (con riferimento ai minori, giovani, anziani), oppure una **minorazione fisica, mentale o intellettuale**, ma anche per **situazioni sociali o economiche**.

Terza Categoria - C

Vengono vietati quei sistemi che **valutano o classificano le persone** fisiche in un determinato periodo **in base al loro comportamento sociale** o caratteristiche personali **inferite tramite il social scoring**, per:

- Un **trattamento dannoso in contesti sociali non collegati a quelli in cui i dati sono stati generati**.
- Un **trattamento sfavorevole di persone ingiustificato in base al loro comportamento** sociale.

Questi sistemi che producono punteggi da parte di attori pubblici o privati possono portare all'esclusione o a effetti discriminatori rispetto certi gruppi/ persone. Un meccanismo di questo tipo, ma su piccola scala, è la **valutazione scolastica e i test di valutazione**.
Ampliando sugli aspetti della vita della persona si possono assumere tratti molto diversi.
Nel campo delle **attività di contrasto, il citizen scoring porta a perdere autonomia**, impoverisce il principio di non discriminazione e non può essere conforme ai diritti fondamentali.

Quarta Categoria - D

Rientrano i **sistemi che effettuano valutazioni del rischio di persone con lo scopo di valutare o prevedere il rischio che una persona commetta un reato** o lo reiteri.
Si basa sulla profilazione di una persona, sulla valutazione dei tratti, delle caratteristiche della personalità. (Polizia predittiva)
Questo **non viene applicato quando si supporta la valutazione umana** del coinvolgimento di una persona, ma che **si basa su fatti oggettivi e verificabili** direttamente connessi.
Le **persone dell'UE dovrebbero essere sempre giudicate in base al loro comportamento effettivo** e **mai in base al comportamento previsto dall'AI** basato su fattori come nazionalità, residenza, #figli. Vi sono poi **due categorie** legate all'oggetto di predizione: **place o person based**.

- Person based sono dedicate a **identificare i soggetti coinvolti in attività criminali** e operano sulla base di liste di persone ritenute a rischio.
- Place based vogliono **identificare le aree in cui è più probabile che si verifichino reati** e dislocare la polizia a seconda del risultato.

Non vi rientrano i sistemi che tendono a prevedere probabilità di localizzare oggetti.
Ad esempio per intercettare carichi di droga e altri illeciti.

Quinta Categoria - E

Vengono vietati tutti quei **sistemi che creano o espandono database di riconoscimento facciale** con **l'estrazione non mirata di immagini da internet o da filmati di telecamere a circuito chiuso**.
Qui si fa uso dello **scraping**. Quella tecnica che **usa un programma per estrarre dati dall'output di un altro**. Ha diverse finalità. Addestramento di sistemi AI, analisi delle tendenze di un mercato, esame dei prezzi di prodotti, oppure estrarre contenuti da siti internet. È poi diverso dai crawler. In quanto i crawler indicizzano dei contenuti online, mentre lo scraper scarica determinati contenuti.
Questi sistemi **devono essere vietati perché aumentano la sensazione di sorveglianza** di massa e questo **porta a violazioni dei diritti fondamentali**, come quello della **privacy**.

Sesta Categoria - F

È vietato immettere sul mercato **sistemi che inferiscono le emozioni di una persona** nel contesto **del posto di lavoro e istituti scolastici**, a meno che il sistema sia per **motivi medici o di sicurezza**. Non rientrano nel campo delle emozioni gli stati fisici come dolore, fatica, quindi quei sistemi che rilevano lo stato di affaticamento, o espressioni, gesti evidenti. Sono **vietati perché le espressioni delle emozioni cambiano molto in base alle culture**, alle situazioni. Questi sistemi **hanno un'affidabilità limitata** e una **mancanza di specificità**.
Per fare questo, **utilizzano dati biometrici, micro espressioni del volto** e **possono risultare intrusivi rispetto ai diritti e libertà** delle persone interessate.
Un esempio è la fascia frontale di Boston, per quantificare l'attenzione degli studenti in base all'attività cerebrale.

Settima Categoria - G

Riguarda i **sistemi di categorizzazione biometrica che classificano le persone sulla base dei loro dati biometrici per dedurre o inferire informazioni** su razza, opinioni politiche, convinzioni religiose, filosofiche.
Non vi sono quei sistemi in cui i dati vengono acquisiti legalmente come immagini nelle **attività di contrasto**.
È un divieto "limitato" nel senso che sussiste solo se usato per deduzioni in merito a aspetti sensibili della vita di una persona.
Altri sistemi che non vengono compresi sono quelli che classificano le caratteristiche del viso o corpo nei sistemi di vendita online. Come non vi rientrano i filtri dei video/ foto nei social network, in quanto feature accessorie.

Ottava Categoria - H

Sistemi di identificazione biometrica remota in tempo reale in spazi pubblici a fini di contrasto, **a meno che** sia necessario per:

- Ricerca mirata di determinate vittime di rapimento, quindi **ricerca di persone scomparse**.
- **Prevenzione di una minaccia specifica** che provocherebbe danni alle persone.
- **Localizzazione/ identificazione di una persona sospettata di un reato** per lo svolgimento di un'indagine o azione penale per reati **all'allegato 2**.

Con **identificazione biometrica** si intende il **riconoscimento automatizzato delle feature umane** per andare a **determinare l'identità di una persona confrontando i dati con quelli salvati** nella banca dati.
Mentre con dati biometrici si parla delle caratteristiche fisiche, fisiologiche di una persona. Dati che possono essere usati sono: volto, movimento occhi, voce. La **verifica biometrica** è invece il **controllo dell'identità di persone e dei loro dati con i dati forniti in precedenza**.
L'identificazione è remota in quanto non richiede il coinvolgimento attivo del soggetto. **L'identificazione può anche avvenire a posteriori**. In questi i **dati sono stati rilevati prima che il sistema sia usato** e il **confronto e l'identificazione avvengono successivamente**. A differenza di quella in tempo reale in cui le due fasi si fanno in contemporanea.

I reati su cui è possibile usare questi sistemi sono elencati nell'allegato 2. Questi sono ispirati a quei reati per cui vige il mandato di arresto europeo. Tra questi:

- Per prevenire atti di terrorismo,
- Per rilevare vittime o trafficanti,
- Per mappare reti o tracciare flussi finanziari,
- Per identificare potenziali autori di omicidi,
- Per riconoscere contenuti illeciti,
- Per rilevare schemi di corruzione,
- Riconoscimento furti, rapine o vandalismi,
- Frodi, Falsificazioni oppure Cyber crimini.

Inoltre questi sistemi sono usabili solo per confermare l'identità della persona interessata, tenendo conto:

- Della natura della situazione che da luogo al possibile uso
- Delle conseguenze del sistema per i diritti e libertà delle persone interessate.

L'identificazione biometrica remota in tempo reale in luoghi pubblici è soggetta ad autorizzazione preventiva, la cui decisione è vincolante per lo stato in cui ha luogo l'uso. Se l'autorizzazione è negata, l'uso è interrotto con effetto immediato e tutti i dati verranno cancellati.