L'*AI Act* assume un volto. Storico accordo tra Parlamento e Consiglio per garantire sicurezza e sviluppo

15/12/2023 DIRITTO EUROPEO E DELLA CONCORRENZA, PRIVACY E CYBERSECURITY, SOCIETÀ, PROSPETTIVE.

Roberto A. Jacchia
Jacopo Piemonte
Federico Aluigi

In data 8 dicembre 2023, successivamente ad una "maratona negoziale" di 36 ore, è stato raggiunto un cruciale accordo politico tra Parlamento Europeo e Consiglio, riguardante la proposta di Regolamento volta a garantire la sicurezza e lo sviluppo dell'Intelligenza Artificiale in Europa ("*AI Act*")[1]. L'obiettivo – ed il compromesso - è garantire il rispetto dei diritti fondamentali e della democrazia, consentendo contemporaneamente lo sviluppo e la leadership europea del settore.

Le ultime battute del processo legislativo avevano visto, successivamente all'emersione di una serie di punti di discussione durante il trilogo del 24 ottobre 2023[2], la pubblicazione di un documento informale diffuso da Francia, Italia e Germania, avente ad oggetto un accordo "a tre" circa le modalità di regolamentazione dei modelli di fondazione[3], facenti parte del più vasto

[1] Per ulteriori informazioni sull'Intelligenza Artificiale e le sue implicazioni nel complesso, si veda il nostro precedente contributo al seguente LINK.
[2] Per ulteriori informazioni sulle questioni aperte dal trilogo del 24 ottobre, si veda il nostro precedente contributo al seguente LINK.
[3] Nell'ambito dell'Intelligenza Artificiale, con il termine "modello" si fa riferimento a un insieme strutturato di algoritmi e parametri che permettono di eseguire specifici compiti di apprendimento automatico. I modelli sono addestrati tramite l'analisi e l'elaborazione di dati, al fine di identificare e apprendere schemi o relazioni tra di essi.
Un *foundation model* rappresenta un tipo specifico e avanzato di modello di Intelligenza Artificiale. I modelli "generici" di Intelligenza Artificiale sono progettati e addestrati per svolgere compiti specifici

ramo della IA generativa (la categoria in cui rientra, per esempio, ChatGTP), nella prospettiva più permissiva di un'autoregolamentazione obbligatoria attraverso codici di condotta[4]. Si era temuto che quest'iniziativa potesse minare il raggiungimento di un accordo di massima tra gli Stati Membri sul testo dell'*AI Act* nel successivo trilogo fissato per il 6 dicembre 2023. Tale scenario è stato scongiurato.

Di seguito, riassumiamo il contenuto dell'accordo raggiunto tra Parlamento e Consiglio all'esito di tale ultima tornata.

Applicazioni vietate

Riconoscendo la potenziale minaccia derivante da alcune applicazioni dell'intelligenza artificiale ("**IA**") verso i diritti dei cittadini e la democrazia, i co-legislatori hanno concordato il **divieto** delle seguenti pratiche:

- sistemi di categorizzazione biometrica che si basino su caratteristiche "sensibili" delle persone (ad esempio, convinzioni politiche, religiose, filosofiche, orientamento sessuale e razza);
- *scraping*[5] indiscriminato di immagini facciali reperite su Internet o catturate da registrazioni di telecamere a circuito chiuso al fine creare *database* di riconoscimento facciale;
- sistemi di riconoscimento delle emozioni delle persone sui luoghi di lavoro e nelle istituzioni educative;
- sistemi di valutazione sociale basati su comportamenti o caratteristiche personali di individui (c.d. *social scoring*);
- sistemi di IA che manipolino il comportamento umano per eluderne il libero arbitrio;
- l'utilizzo dell'IA per sfruttare le vulnerabilità delle persone (a causa dell'età, disabilità, situazione sociale o economica).

Casi tassativi di utilizzo lecito della identificazione biometrica

I negoziatori hanno inoltre concordato una serie di strette eccezioni per l'impiego di sistemi di identificazione biometrica in luoghi ad accesso pubblico[6].

Il loro **utilizzo** dovrà essere subordinato a preventiva autorizzazione giudiziaria e sarà funzionale alla repressione di determinati reati previamente definiti. Il riconoscimento biometrico *ex post* (con utilizzo delle immagini "in differita") potrà poi essere utilizzato solamente per la ricerca mirata di individui condannati o sospettati di aver commesso reati pre-identificati come gravi; l'utilizzo del riconoscimento biometrico "in tempo reale" (su cui le discussioni sembra siano state effettivamente molto accese nel corso dei triloghi) dovrà invece soddisfare una serie di condizioni rigorose, con **limitazione** ai fini di:

- ricerche mirate di vittime di determinati reati gravi (ad esempio, sequestro, tratta, sfruttamento sessuale);
- prevenzione di una minaccia terroristica specifica e attuale;

---

e ben definiti e possono essere addestrati su set di dati di dimensioni variabili, a seconda del compito che dovranno svolgere.
I *foundation models*, invece, sono addestrati su enormi quantità di dati e con moltissimi parametri. Ciò permette loro di svolgere una varietà di compiti più ampia rispetto ai modelli tradizionali.

[4] Per ulteriori informazioni sull'accordo, si veda il nostro precedente contributo al seguente LINK.

[5] Per *scraping* si intende la tecnica informatica di estrazione di dati e informazioni dalla rete.

[6] Per "sistema di identificazione biometrica" si intende un particolare tipo di sistema informatico che ha la funzionalità e lo scopo di identificare una persona sulla base di una o più caratteristiche fisiologiche e/o comportamentali (biometria), confrontandole con i dati, precedentemente acquisiti e presenti nel *database* del sistema, tramite degli algoritmi e dei sensori di acquisizione dei dati in *input*.

- localizzazione o identificazione di una persona sospettata di aver commesso uno dei reati gravi specificamente menzionati (ad esempio, terrorismo, tratta, sfruttamento sessuale, omicidio, rapimento, stupro, rapina a mano armata, partecipazione a un'organizzazione criminale).

Obblighi per i sistemi classificati "ad alto rischio"

Come abbiamo avuto modo di rappresentare in precedenti contributi sul tema[7], in base al c.d. *risk-based approach*, la proposta di Regolamento introduce una **classificazione in base al rischio su quattro livelli**, in funzione delle ripercussioni che potrebbe avere l'IA sulla sicurezza delle persone e sui diritti fondamentali.

Il livello di **rischio alto** si presenta come quello più oneroso in termini di *compliance*, comportando molteplici obblighi a cui i fornitori devono attenersi prima dell'immissione del sistema sul mercato europeo e durante il suo ciclo di vita. In tale categoria sono inclusi anche i sistemi suscettibili di arrecare danni alla salute, alla sicurezza, ai diritti fondamentali o all'ambiente.

Per i sistemi di IA classificati come ad alto rischio (in considerazione degli interessi che potrebbero coinvolgere), sono stati definiti nuovi obblighi (addizionali rispetto a quelli previamente previsti). I co-legislatori hanno incluso la necessità, tra gli altri requisiti, di prevedere una valutazione obbligatoria dell'impatto sui diritti fondamentali denominata *Fundamental-Rights Impact Assessment* ("FRIA") per certi sviluppatori di sistemi ad alto rischio. Dovranno, ad esempio, svolgere una FRIA gli sviluppatori di sistemi di intelligenza artificiale nei settori assicurativo e bancario (si pensi, ad esempio, a *software* che decidano in maniera "meccanica" e automatizzata a chi concedere un mutuo e a chi no). Tale obbligo sarà poi estendibile anche ai sistemi di IA utilizzati per influenzare l'esito di elezioni politiche e il comportamento degli elettori. Resta fermo il diritto dei cittadini di presentare reclami sui sistemi di IA ad alto rischio, nonché di ricevere spiegazioni sulle decisioni basate su detti sistemi qualora lamentino un pregiudizio ai loro diritti.

Limiti per i sistemi di Intelligenza Artificiale aventi scopi generali

Per tenere conto della vasta gamma di compiti che i sistemi di IA possono svolgere e della rapida espansione delle loro capacità, si è convenuto che i sistemi di intelligenza artificiale aventi **scopi generali** definiti come "*General-Purpose AI*" ("GPAI")[8] e i modelli su cui si basano dovranno attenersi ai requisiti di trasparenza inizialmente proposti dal Parlamento. Questi includono la necessità, per chi sviluppa tali sistemi, di redigere documentazione tecnica, di conformarsi alla legislazione sul diritto d'autore dell'Unione e di diffondere resoconti dettagliati sui dati utilizzati per l'addestramento degli algoritmi.

Inoltre, per i modelli GPAI ad alto impatto che comportano rischi sistemici, è passata la posizione del Parlamento che garantisce obblighi ancora più rigorosi: se questi modelli soddisfano determinati criteri, i loro sviluppatori dovranno effettuare valutazioni del modello e analisi volte a mitigare i rischi sistemici nonché condurre specifici *test* e segnalare alla Commissione incidenti gravi che dovessero occorrere a causa dell'utilizzo. Peraltro, l'accordo prevede che, fino a quando non saranno pubblicati *standard* armonizzati dell'Unione, i GPAI con rischi sistemici dovranno comunque basarsi su codici di condotta.

Misure a sostegno dell'innovazione e delle PMI

[7] Per ulteriori informazioni, si veda il nostro precedente contributo al seguente LINK.
[8] Per *"General-Purpose AI"* si intendono dei sistemi di IA che hanno un'ampia gamma di usi possibili, sia previsti che non previsti dagli sviluppatori. Possono essere applicati a molti compiti diversi in vari campi, spesso senza modifiche sostanziali e messe a punto.

Infine, l'accordo raggiunto garantisce che le imprese, in particolare le PMI, possano sviluppare soluzioni di IA in condizioni idonee a contenere il rischio che i *player* che controllano l'industria sviluppino una posizione dominante. A tal fine, l'accordo promuove i cosiddetti "*regulatory sandboxes*"[9] e i "*real-world-testing"*, che dovranno essere istituiti dalle autorità nazionali e che consentiranno a tutti gli operatori interessati di sviluppare e addestrare delle IA innovative prima della loro messa sul mercato.

Prossimi passi e Calendario dell'IA

Intanto, in data 12 dicembre la Commissione Europea ha pubblicato una versione aggiornata delle sue Q/R sull'*AI Act,* in seguito all'accordo provvisorio raggiunto tra Parlamento Europeo e Consiglio[10].

I punti salienti del documento riguardano:

i) l'introduzione, nell'esplicazione delle categorie di rischio, di uno "*Specific transparency risk"*, il quale configurerebbe un obbligo di trasparenza orizzontale applicabile a tutti i sistemi di IA[11];

ii) l'ambito di applicazione del Regolamento, che si dirige anche ai soggetti stabiliti al di fuori dell'Unione, a condizione che il sistema di IA sia immesso sul mercato unionale o che il suo utilizzo riguardi persone che si trovano nell'Unione. Gli importatori di sistemi di IA dovranno inoltre garantire che il fornitore extraeuropeo abbia già eseguito l'appropriata procedura di valutazione della conformità, che il sistema rechi una marcatura CE e che sia accompagnato dalla documentazione e dalle istruzioni per l'uso[12];

iii) l'approfondimento sul *Fundamental-Rights Impact Assessment*, attraverso una descrizione dei contenuti della valutazione a cui l'implementatore dei sistemi ad alto rischio sarà sottoposto[13].

L'accordo politico sarà ora soggetto all'approvazione formale del Parlamento Europeo e del Consiglio, che potrebbe protrarsi ancora per vari mesi nel corso del 2024. Occorreranno, infatti, incontri tecnici per raffinare il testo della proposta. Del pari, sarà necessario ottenere l'approvazione dei rappresentanti dei Governi e quella in seduta plenaria del Parlamento. A questo punto, il testo sarà pubblicato sulla Gazzetta ufficiale ed entrerà in vigore 20 giorni dopo la data di pubblicazione.

È previsto che l'*'AI Act* divenga applicabile **due anni** dopo la sua entrata in vigore, fatta eccezione per alcune disposizioni specifiche: i divieti (di cui abbiamo discusso più sopra) si applicheranno già dopo **sei mesi** dalla pubblicazione del testo mentre le norme sull'IA per scopi generali (pure oggetto di

[9] Come riporta la Commissione Europea, *"… The regulatory sandbox is a way to connect innovators and regulators and provide a controlled environment for them to cooperate. Such a collaboration between regulators and innovators should facilitates the development, testing and validation of innovative AI systems with a view to ensuring compliance with the requirements of the AI Regulation …".*
Per ulteriori informazioni, si veda il seguente LINK.
[10] Si veda il documento "*Questions & Answers"* al seguente LINK.
[11] Si veda il documento "*Questions & Answers"*, sotto la domanda "*What are the risk categories?".*
[12] Si veda il documento "*Questions & Answers*", sotto la domanda "*To whom does the AI Act apply?".*
[13] Si veda il documento "*Questions & Answers*", sotto la domanda "*What is a fundamental rights impact assessment? Who has to conduct such an assessment, and when?".*

approfondimento) saranno già applicabili dopo **dodici mesi**.

Infine, la Commissione Europea ha annunciato che nel periodo transitorio prima che il Regolamento diventi generalmente applicabile, verrà lanciato un "patto sull'intelligenza artificiale", attraverso la promozione di incontri tra sviluppatori di IA provenienti dall'Europa e da tutto il mondo per tentare di ottenere l'impegno congiunto ad aderire su base volontaria agli obblighi-chiave dell'*AI Act* anche prima della sua entrata in vigore[14].

[14] Per ulteriori informazioni, si veda il seguente LINK.

Roberto A. Jacchia

PARTNER

r.jacchia@dejalex.com
+39 02 72554.1
Via San Paolo 7
20121 - Milano

Jacopo Piemonte

ASSOCIATE

j.piemonte@dejalex.com
+39 02 72554.1
Via San Paolo 7
20121 – Milano
+32 (0)26455670
Chaussée de La Hulpe 187
1170 - Bruxelles

Federico Aluigi

ASSOCIATE

f.aluigi@dejalex.com
+32 (0)26455670
Chaussée de La Hulpe 187
1170 - Bruxelles

MILANO
Via San Paolo, 7 · 20121 Milano, Italia
T. +39 02 72554.1 · F. +39 02 72554.400
milan@dejalex.com

ROMA
Via Vincenzo Bellini, 24 · 00198 Roma, Italia
T. +39 06 809154.1 · F. +39 06 809154.44
rome@dejalex.com

BRUXELLES
Chaussée de La Hulpe 187 · 1170 Bruxelles, Belgique
T. +32 (0)26455670 · F. +32 (0)27420138
brussels@dejalex.com

MOSCOW
Potapovsky Lane, 5, build. 2, 4th floor, office 401/12/9 · 101000, Moscow, Russia
T. +7 495 792 54 92 · F. +7 495 792 54 93
moscow@dejalex.com