AI Act e dintorni: le novità più rilevanti sulla corsa alla *governance*

27/11/2023

DIRITTO EUROPEO E DELLA CONCORRENZA, PRIVACY E CYBERSECURITY, PROSPETTIVE.

Jacopo Piemonte
Federico Aluigi

L'era moderna è testimone di una rivoluzione senza precedenti, dove l'intelligenza artificiale ("IA") ha assunto un ruolo sempre più preminente, permeando ogni aspetto della vita quotidiana.

La crescente espansione di detta tecnologia, con le sue potenzialità, pone di fronte a noi una serie di interrogativi, dubbi irrisolti e sfide etiche che richiamano l'attenzione di studiosi, *decision maker*, e della società nel suo complesso. Per di più, l'accelerazione esponenziale della capacità computazionale e l'accesso a enormi quantità di dati hanno catalizzato lo sviluppo dell'intelligenza artificiale, rendendola un motore trainante dell'innovazione.

Tuttavia, trattasi di uno sviluppo tanto rapido da sfidare i criteri di tempo nei quali la società civile è solita muovere il proprio progresso, in uno scenario dove – sul piano giuridico – provvedimenti, regolamentazioni e documenti di *soft law* si moltiplicano nel tentativo di "reggere il passo" rispetto al fulmineo progresso dell'IA.

In questo contesto, l'Unione Europea sta cercando di ergersi a primo attore nel regolamentare tale materia con il Regolamento di intelligenza artificiale (anche conosciuto come "***AI Act***")[1]. A riguardo, in data 6 dicembre 2023 si terrà il prossimo trilogo[2], nel quale si giocherà una partita fondamentale per capire se il testo definitivo di *AI Act* potrà essere licenziato già entro la fine del 2023 (si veda sotto, il punto 3).

[1] Per una panoramica sull'argomento, si veda il nostro precedente contributo al seguente LINK.
[2] Per ulteriori informazioni si veda "*Le tempistiche di approvazione dell'AI Act*", nel nostro precedente contributo al seguente LINK.

In attesa di tali sviluppi, inauguriamo con questo articolo una serie di contributi che avranno cadenza periodica ed in cui cercheremo di dare atto di tutte le novità in tema di intelligenza artificiale che si susseguiranno nel tempo. Iniziamo con le seguenti tre notizie in materia di IA che coprono l'arco temporale tra fine ottobre 2023 e la data odierna.

1) Il Codice di Condotta Internazionale per gli sviluppatori di IA

In data 30 ottobre 2023, è stato annunciato dai *leader* del G7 un accordo[3] sui Principi Guida Internazionali sull'Intelligenza Artificiale (IA) e sul Codice di Condotta per gli sviluppatori di IA nel contesto dell'"*Hiroshima AI Process*"[4].

In primo luogo, sono stati individuati undici principi-guida allo scopo di fornire indicazioni alle aziende sviluppatrici di sistemi di IA, nonché ad enti accademici, alla società civile e ai settori pubblico e privato coinvolti nello sviluppo dei sistemi avanzati. I principi annoverano diverse misure per valutare e mitigare i rischi durante lo sviluppo dei sistemi di IA. Richiedono inoltre la comunicazione trasparente delle capacità e delle limitazioni degli stessi, la condivisione di informazioni tra organizzazioni sviluppatrici promuovendo politiche di *governance* basate sul rischio e incentivi a investimenti in sicurezza e ricerca. Peraltro (ed è interessante da notare) per molti di tali ambiti viene ripreso fedelmente quanto già previsto dall'*AI Act*[5].

In secondo luogo, è stato approfondito il contenuto di ciascuno degli undici principi elencati nel Codice di Condotta, dove sono dettagliatamente trascritte le indicazioni pratiche rivolte agli sviluppatori di IA. Di particolare rilievo, a nostro avviso, sono: *i)* il primo principio, che postula la necessità di adozione di adeguate misure per identificare, valutare e ridurre i rischi per tutto il ciclo di vita dell'IA[6]; *ii)* il terzo e il quarto principio, che riguardano la "*transparency"* dei modelli. Questi prevedono, in particolare, la pubblicazione ed il continuo aggiornamento di rapporti di trasparenza contenenti informazioni e istruzioni per l'uso e la documentazione tecnica, richiedendo garanzia di intelligibilità delle informazioni per gli utenti[7]. Viene contemplata inoltre la condivisione delle informazioni verso tutti gli attori della società civile (specie relativamente ai rapporti di valutazione delle informazioni sui rischi per la sicurezza e sulla pericolosità dei sistemi durante tutto il ciclo di vita dell'IA[8]); iii) l'undicesimo principio, che apre ai vasti settori della proprietà intellettuale e della privacy, sottolineando la necessità di un coordinamento con l'intelligenza

---

[3] Così l'introduzione del documento: *"… the International Code of Conduct for Organizations Developing Advanced AI*
Systems aims to promote safe, secure, and trustworthy AI worldwide and will provide voluntary guidance for actions by organizations developing the most advanced AI systems, including the most advanced foundation models and generative AI systems (henceforth "advanced AI systems") …".
Per ulteriori informazioni si veda il seguente LINK.

[4] L'*Hiroshima AI Process*, istituito in occasione del vertice del G7 del 19 maggio 2023, mira a promuovere a livello globale dei limiti per i sistemi avanzati di AI.

[5] Si veda l'Allegato IV della proposta di Regolamento.

[6] Si veda il Paragrafo 1 del Codice di Condotta, "*Take appropriate measures throughout the development of advanced AI systems, including prior to and throughout their deployment and placement on the market, to identify, evaluate, and mitigate risks across the AI lifecycle*".

[7] Si veda il Paragrafo 3 del Codice di Condotta, "*Publicly report advanced AI systems' capabilities, limitations and domains of appropriate and inappropriate use, to support ensuring sufficient transparency, thereby contributing to increase accountability*".

[8] Si veda il Paragrafo 4 del Codice di Condotta, "*Work towards responsible information sharing and reporting of incidents among organizations developing advanced AI systems including with industry, governments, civil society, and academia*".

artificiale che possa implementare salvaguardie in tali settori[9].

Il Codice di Condotta, sebbene vi si possa conformare su mera base volontaria, risulta di particolare importanza poiché, diversamente dall'*AI Act*, è immediatamente applicabile e con un potenziale campo di applicazione non europeo, ma globale. Peraltro, è probabile che detto Codice rivestirà un ruolo centrale nella regolamentazione della nuova tecnologia vista la massima attenzione, in questo momento storico, sulla IA ed i riflessi di quest'ultima sulle valutazioni di *accountability* operate dalle imprese.

2) **L'Ordine Esecutivo statunitense sull'uso sicuro, protetto e affidabile dell'IA**

In data 30 ottobre 2023, è stato emanato dal Presidente degli Stati Uniti d'America Joe Biden l'Ordine Esecutivo sull'uso sicuro, protetto e affidabile dell'Intelligenza Artificiale[10].

Tale documento segna una presa di posizione degli Stati Uniti sulla materia, ed in concomitanza con il Codice di Condotta di cui al punto 1) e l'*AI Act*, sembra delineare le linee guida globali e i principi di sviluppo nel campo dell'Intelligenza Artificiale. All'interno dell'Ordine Esecutivo si rinvengono diverse disposizioni dirette alle agenzie federali, da attuare entro il prossimo anno, su questioni che spaziano dalla sicurezza nazionale all'immigrazione, dal settore delle costruzioni all'assistenza sanitaria. Di seguito, i punti più rilevanti:

In primo luogo, viene richiesta trasparenza agli sviluppatori di IA, i quali dovranno condividere i risultati dei loro *test* ed altre informazioni critiche con il governo degli Stati Uniti. Contestualmente, vengono affidate una serie di competenze ad enti pre-identificati: il *National Institute of Standards and Technology* ("NIST"), in coordinamento con il *Secretary of Energy*, ed il *Secretary of Homeland Security,* sono investiti del ruolo di sviluppare gli standard e le linee guida per lo sviluppo di modelli di IA sicuri ed affidabili[11]. Peraltro, riguardo alla diffusione di notizie false e la necessità di prevenire condotte fraudolente e manipolazioni, spetterà invece al *Department of Commerce* esercitare una mirata ed approfondita guida tecnica allo scopo di garantire l'autenticità dei contenuti generati dall'Intelligenza Artificiale[12]. In particolare, tale ente dovrà verificare che tutti gli attori coinvolti nel processo garantiscano un'indicazione chiara dell'origine dei contenuti prodotti attraverso l'impiego di modelli e software di IA[13].

---

[9] Si veda il Paragrafo 11 del Codice di Condotta, "*Implement appropriate data input measures and protections for personal data and intellectual property".*
Sull'argomento, si veda il nostro precedente contributo al seguente LINK.
[10] Si veda l'Ordine Esecutivo al seguente LINK.
[11] Si veda il paragrafo 4.1. dell'Ordine Esecutivo, "*Developing Guidelines, Standards, and Best Practices for AI Safety and Security".*
[12] Si veda il paragrafo 4.5. dell'Ordine Esecutivo, "*Reducing the Risks Posed by Synthetic Content*".
[13] Di rilievo il punto *(a)* del paragrafo 4.5. per cui "… *Within 240 days of the date of this order, the Secretary of Commerce, in consultation with the heads of other relevant agencies as the Secretary of Commerce may deem appropriate, shall submit a report to the Director of OMB and the Assistant to the President for National Security Affairs identifying the existing standards, tools, methods, and practices, as well as the potential development of further science-backed standards and techniques, for:*
(i) authenticating content and tracking its provenance;
(ii) labeling synthetic content, such as using watermarking;
(iii) detecting synthetic content;
(iv) preventing generative AI from producing child sexual abuse material or producing non-consensual intimate imagery of real individuals (to include intimate digital depictions of the body or body parts of an identifiable individual);
(v) testing software used for the above purposes; and
(vi) auditing and maintaining synthetic content …".

In secondo luogo, all'interno di un più ampio riordino della materia relativa ai diritti civili in relazione all'IA[14], l'Executive Order pone enfasi sullo sviluppo patologico di *bias* cognitivi che possano, a causa dell'imprevedibilità dei sistemi di IA, creare discriminazioni in settori sensibili quali giustizia, sanità ed edilizia residenziale. Nell'intento, dunque, di evitare che l'origine geografica di un individuo o l'appartenenza ad una specifica etnia conducano a risultati iniqui nelle decisioni prese dagli algoritmi, viene disposta, nello sviluppo dei sistemi di IA, una riconsiderazione di numerosi criteri concernenti vari settori (per esempio, riguardanti la concessione di credito nel mercato immobiliare) al fine di individuare eventuali distorsioni o disparità a danno di minoranze o fasce "protette". Vengono dunque richieste particolari cautele nella gestione di procedure automatizzate in modo da minimizzare eventuali *bias*[15].

In terzo luogo, appare incisiva ed inedita la sezione del documento riguardante il supporto dei lavoratori[16], in relazione alla previsione di una ingente perdita di posti di lavoro in alcuni settori, successivamente all'implementazione pratica dei sistemi di IA nella società civile. Sul punto, si prevede la preparazione e presentazione di un rapporto circa l'impatto di dette tecnologie sul mercato lavorativo e l'elaborazione di *best practice* per i datori di lavoro che potrebbero essere utilizzate per mitigare i potenziali danni dell'IA sul benessere dei dipendenti e per massimizzarne i potenziali benefici.

L'Ordine Esecutivo si mostra infine particolarmente variegato nei contenuti, toccando i temi della giustizia penale in relazione all'IA; la cooperazione internazionale mirata all'espansione di accordi bilaterali e multilaterali nel settore; la previsione di azioni a tutela dei consumatori, dei pazienti clinici e degli studenti.

3) L'accordo tra Italia, Francia e Germania sulle IA generativa

In data 19 novembre 2023, Italia, Francia e Germania hanno diffuso un documento informale circa le modalità di regolamentazione dei modelli di fondazione, facenti parte del più vasto ramo della IA generativa (la categoria in cui rientra, per esempio, ChatGTP).

Il tema era già stato ampiamente oggetto di discussione fra gli Stati dell'Unione Europea, durante le intense negoziazioni per la definizione del testo dell'*AI Act*. Nel corso dell'ultimo trilogo in ottobre si era tentato, in particolare, di adottare un approccio basato sul rischio all'interno del settore dell'IA generativa. Tale tecnologia verrebbe, in questo caso, assoggettata a limiti e controlli già a livello legislativo[17].

Allontanandosi da detta prospettiva, l'accordo in esame individua i rischi dell'IA nella sua stessa applicazione pratica piuttosto che nella tecnologia per sé. Prospetta di conseguenza l'opportunità di un'autoregolamentazione obbligatoria dei modelli di fondazione attraverso codici di condotta, seguendo i principi definiti dal G7 nell'accordo di cui al punto 1). Con tale approccio spetterebbe, in capo agli sviluppatori di IA, l'obbligo di redigere "*model cards*" includenti le informazioni pertinenti sulle capacità e sui limiti dei modelli da sviluppare, basate sulle *best practice*

[14] Si veda la Sezione 7 dell'Ordine Esecutivo, "*Advancing Equity and Civil Rights".*
[15] Si veda, *ex multis*, il punto (b) del paragrafo 7.3. per cui "… *To address discrimination and biases against protected groups in housing markets and consumer financial markets, the Director of the Federal Housing Finance Agency and the Director of the Consumer Financial Protection Bureau are encouraged to consider using their authorities, as they deem appropriate, to require their respective regulated entities, where possible, to use appropriate methodologies including AI tools to ensure compliance with Federal law and:*
(i) evaluate their underwriting models for bias or disparities affecting protected groups; and
(ii) evaluate automated collateral-valuation and appraisal processes in ways that minimize bias …".
[16] Si veda la Sezione 6 dell'Ordine Esecutivo, "*Supporting Workers*".
[17] Sul punto, si veda "*Più regolamentazione per l'IA generativa*", nel nostro precedente contributo al seguente LINK.

all'interno della comunità degli sviluppatori. In più, viene proposto che un organo di *governance* dell'IA possa aiutare a sviluppare linee guida e verificare l'applicazione delle *model cards*, fornendo un accesso semplice alla segnalazione di qualsiasi violazione del codice di condotta. Infine, risulterebbe cruciale il punto del documento informale per cui non dovrebbero essere applicate sanzioni alla "prima infrazione" ed il regime prescelto prevederebbe invece conseguenze sanzionatorie solo a seguito di violazioni sistematiche dei codici di condotta e di un'analisi "adeguata" relativamente alle carenze individuate nell'autoregolamentazione.

Si sottolinea come l'accordo tra Italia, Francia e Germania sopra riepilogato intervenga in un momento cruciale delle negoziazioni interistituzionali per la finalizzazione dell'AI Act. Non si può escludere che questa iniziativa "trilaterale" possa minare la "fluidità" dei triloghi tra gli Stati Membri, condizionandone i futuri esiti e allungando i tempi di emanazione dell'*AI Act.*

Ulteriori aggiornamenti sul punto verranno riportati nei prossimi contributi che pubblicheremo sul tema nell'ambito di questa rubrica

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