ISSN 1826-3534

federalismi.it
RIVISTA DI DIRITTO PUBBLICO ITALIANO, COMPARATO, EUROPEO

24 GENNAIO 2024

L'Artificial Intelligence Act Europeo: alcune questioni di implementazione

di Claudio Novelli
Assegnista di ricerca in Filosofia del diritto
Alma mater studiorum - Università di Bologna

L'Artificial Intelligence Act Europeo: alcune questioni di implementazione*

di Claudio Novelli
Assegnista di ricerca in Filosofia del diritto
Alma mater studiorum - Università di Bologna

Abstract [It]: L'articolo esamina la proposta europea di regolamento sull'intelligenza artificiale, AI Act (AIA). In particolare, esamina il modello di analisi e valutazione del rischio dei sistemi di IA. L'articolo identifica tre potenziali problemi di implementazione del regolamento: (1) la predeterminazione dei livelli di rischio, (2) la genericità del giudizio di significatività del rischio e (3) l'indeterminatezza della valutazione sull'impatto dei diritti fondamentali. Il saggio suggerisce alcune soluzioni per affrontare questi tre problemi.

Title: The European Artificial Intelligence Act: some implementation issues
Abstract [En]: The article examines the European proposal for an artificial intelligence regulation, the AI Act (AIA). Specifically, it analyses the model for assessing and evaluating the risk of AI systems. The article identifies three issues in implementing the regulation: (1) the predetermination of risk levels, (2) the generality of the judgment on the significance of the risk, and (3) the indeterminacy of the evaluation on the impact on fundamental rights. The essay provides some solutions to address these three issues.

Parole chiave: AI Act, Intelligenza artificiale, Unione Europea, rischio, diritti fondamentali
Keywords: AI Act, Artificial Intelligence, European Union, risk, fundamental rights

1. Introduzione

Il 21 aprile 2021 la Commissione europea ha redatto la prima bozza dell'AI Act (AIA), un regolamento diretto ad armonizzare le regole per lo sviluppo, produzione, ed utilizzo dei sistemi di Intelligenza Artificiale (IA). Venerdì 8 dicembre 2023, dopo lunghe negoziazioni, il Parlamento e il Consiglio hanno raggiunto un accordo provvisorio sul testo definitivo nella fase di trilogo.[1]

L'AIA costituisce il primo quadro giuridico che disciplina la progettazione, lo sviluppo, e l'utilizzo dei sistemi di IA in maniera estensiva, ossia, non solo in relazione a profili specifici come la protezione dei dati o la responsabilità civile. Nei prossimi anni, chiunque vorrà produrre e/o commercializzare queste tecnologie nell'Unione Europea dovrà adeguarsi agli standard contenuti nell'AIA (e nei suoi atti implementativi). Mutatis mutandis, il regolamento sarà la Costituzione dell'UE sull'IA, con implicazioni

* Articolo sottoposto a referaggio.
[1] Per una sintesi dell'accordo: https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai.

decisive per la strategia europea sull'IA. Se l'UE riuscirà a essere un mercato attrattivo per l'IA, o se addirittura riuscirà a influenzare altri sistemi normativi con i propri standard (cd. Brussel effect), questo dipenderà in gran parte dall'AIA e dal modo dal modo in cui esso verrà implementato.

In questo saggio, mi occupo di un aspetto che sarà oggetto di dibattito nei mesi successivi all'approvazione finale del regolamento, nel corso dei quali si dovranno definire gli standard applicativi del regolamento: la valutazione del rischio dei sistemi di IA. Anche se è presto per trarre conclusioni sull'implementazione dell'AIA, il modello di rischio che emerge dall'AIA consente di fare delle ipotesi sulle implicazioni pratiche del regolamento. In particolare, uno degli aspetti chiave dell'AIA riguarda la categorizzazione dei sistemi di IA in quattro livelli di rischio: inaccettabile, alto, limitato e minimo. Il legislatore attribuisce gli oneri normativi in maniera proporzionale a questi livelli, con l'intento di prevenire o mitigare le conseguenze dannose dell'IA nella maniera più efficiente possibile. Le categorie di rischio sono definite sulla base della tecnologia di IA impiegata e su ambiti di applicazione generalmente individuati: e.g., educazione, giustizia, processi democratici e servizi essenziali. Non sono invece previste valutazioni del rischio basate su scenari concreti di applicazione; e quando sono accennate – solo in forma di revisione, dopo il testo compromissorio di giugno 2023 – non sono fornite metodologie generali per calcolare il rischio. Ciò è in parte dovuto alla natura dello strumento giuridico utilizzato, che deve essere abbastanza generale da poter consentire l'armonizzazione flessibile tra gli Stati Membri. Tuttavia, se il modello di rischio rimane così generico e non vengono introdotti elementi di maggiore granularità nell'analisi del rischio, si rischia di pregiudicare l'efficacia dell'AIA.

L'articolo individua tre debolezze dell'AIA: (a) la predeterminazione dei livelli di rischio; (b) il giudizio di significatività del rischio in sede di revisione; (c) la valutazione di impatto sui diritti fondamentali (FRIA). Proverò a fornire alcune raccomandazioni per affrontare i problemi innescati da queste debolezze. L'ambito di esecuzione delle raccomandazioni è quello degli atti delegati o di implementazione di cui si fa carico la Commissione. Pertanto, non si suggeriscono particolari modifiche del testo regolamentare.[2, 3] Invero, il recente accordo politico raggiunto nel corso del trilogo introduce ulteriori elementi critici come, ad esempio, il riferimento al "rischio sistemico" che qualificherebbe un sistema di AI di scopo generale (cd. General Purpose AI, GPAI) – inclusa l'AI generativa di testi o immagini (e.g, ChatGpt o Midjourney) – come ad alto rischio. Cosa si intenda per rischio sistemico non è ancora chiaro, vi è solo un riferimento alla potenza di calcolo usata per l'addestramento del sistema. Anche se non approfondirò questo innesto

[2] Ciò è particolarmente vero si si considerano i punti (b) e (c).

[3] Anche se nel corso del trilogo è stato raggiunto un accordo politico che cristallizza alcuni aspetti centrali dell'AIA, non esiste ancora un testo unico. Tuttavia, le tre versioni proposte da Commissione, Parlamento, e Consiglio sono in gran parte sovrapponibili, soprattutto per ciò che riguarda l'impostazione generale di analisi e valutazione del rischio.

recente all'AIA, molte delle osservazioni che farò per il giudizio di significatività del rischio sono applicabili anche al rischio sistemico dei sistemi GPAI.

L'articolo è così strutturato. Al paragrafo 2, sintetizzerò l'approccio di regolamentazione basato sul rischio adottato dal legislatore europeo nell'AIA. Al paragrafo 3, esporremo i punti di forza del regolamento. Al paragrafo 4, tre elementi critici nella implementazione dell'AIA: (a) la predeterminazione del rischio, (b) la misurazione della significatività del rischio, e (c) la valutazione di impatto sui diritti fondamentali. Il paragrafo 5 conclude l'articolo.

2. L'approccio al rischio dell'AIA

La maggior parte dei regolamenti basati sul rischio trova ispirazione nel principio di precauzione per il quale se un'attività può avere conseguenze gravi e incerte, allora bisogna adottare le precauzioni necessarie o evitarla del tutto. Pertanto, le misure di contenimento e prevenzione del rischio, anche quando introdotte per via legislativa, possono essere avviate anche in assenza di evidenze conclusive sui potenziali effetti dannosi.

I regolamenti basati sul rischio consistono generalmente di tre parti: la valutazione, la gestione e la comunicazione del rischio.[4] Anche l'AIA mostra questa impostazione, dedicando maggiore spazio alle misure di gestione del rischio, sottoforma di oneri normativi a carico dei soggetti che fanno parte della catena del valore dei sistemi di IA (principalmente deployers e providers). Analizzerò tutte queste tre parti dell'AIA, anche se le riflessioni critiche saranno perlopiù rivolte alla metodologia di analisi e valutazione del rischio, da cui le misure di gestione derivano. Queste ultime possono essere più facilmente riviste, oltre al fatto che per giudicarle con cognizione di causa dovremo attendere gli sviluppi del trilogo, gli atti di implementazione dell'AIA e la definizione degli standard di compliance. L'impianto di fondo, invece, è abbastanza stabile per poter inferire quali valutazioni politiche e regolative hanno influenzato il legislatore europeo nella stesura dell'AIA.

La concezione di rischio alla base dell'AIA è piuttosto tradizionale: è la probabilità che un pericolo si traduca in una perdita, una lesione o un danno. Oppure, definito in modo anche più generale, è rischio l'incertezza sulle possibili conseguenze avverse di un evento o attività per ciò che gli umani considerano importante [5].

[4] E. MILLSTONE et al., *Science in Trade Disputes Related to Potential Risk: Comparative Case Studies*, Siviglia, European Commission, 2004.

[5] M. FLORIN E M. T. BÜRKLER, a c. di, *Introduction to the IRGC Risk Governance Framework*, Losanna, EPFL, 2017. T. AVEN e O. RENN, *On risk defined as an event where the outcome is uncertain*, in *Journal of Risk Research* 12, n. 1, 2009, pp. 1–11. T. AVEN, O. RENN, e E. A. ROSA, *On the Ontological Status of the Concept of Risk*, in *Safety Science* 49, n. 8, pp. 1074–79.

Nell'AIA, le fonti di pericolo corrispondono ai sistemi di IA. Dopo un lungo dibattito su quale la definizione descrivesse meglio questi sistemi, l'AIA riprende la definizione fornita dall'OCSE: "machine-based system[s] designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments".[6]

Nell'AIA, Le entità esposte a queste fonti di pericolo sono qualificate in maniera assiologica: esse corrispondono a valori e diritti fondamentali dell'EU come i diritti umani fondamentali, la sicurezza, e le procedure democratiche (AIA, Articolo 1).[7] L'impostazione valoriale dell'AIA ha delle conseguenze significative per l'applicazione del regolamento, ma torneremo su questo profilo più avanti.

In questo contesto, l'AIA classifica i sistemi di IA secondo quattro livelli di rischio e associa ad ognuno di essi delle salvaguardie che ne compensino la pericolosità. Sono queste salvaguardie, sottoforma di obblighi a carico dei vari soggetti coinvolti nel ciclo di vita dei sistemi di IA, che rendono tollerabile il rischio di alcuni sistemi di IA. Al contrario, il rischio è giudicato intollerabile se non esistono salvaguardie in grado di controbilanciare il pericolo o se non sono realizzate dai soggetti tenuti a farlo (inadempienza). La classificazione del rischio dell'AIA può essere sintetizzata così:[8]

Categorie di Rischio	**Esempi Applicativi**	**Implicazioni Normative**
Rischio Inaccettabile (Titolo II)	IA per manipolazione psicologica, sfruttamento delle vulnerabilità, categorizzazione biometrica[9], social scoring, riconoscimento delle emozioni	Proibizione (salvo deroghe in casi eccezionali, e.g., pericolo terrorista per i sistemi di categorizzazione biometrica)

[6] OECD Recommendation of the Council on Artificial Intelligence.

[7] Quando non viene fatto esplicito riferimento ad una versione specifica dell'AIA, questo vuol dire che il riferimento è alla bozza della Commissione del 2021: Proposal for a Regulation of The European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM/2021/206 final).

[8] Alcuni cambiamenti sono stati successivamente proposti, ad esempio, nel testo compromissorio di giugno 2023. Ad esempio, praticamente tutti i sistemi di categorizzazione biometrica sono diventati di rischio inaccettabile.

[9] Questi sistemi sono proibiti tutte quelle volte che categorizzano sulla base di attributi sensibili o protetti (sesso, genere, età, etnia, orientamento politico ecc.).

Rischio Alto (Titolo III)	Sistemi di IA applicati a: componentistica di sicurezza, educazione, impiego, servizi pubblici e privati essenziali, gestione dell'immigrazione, amministrazione della giustizia, law enforcement	Sistema di gestione del rischio continuo, controllo dei dati, documentazione tecnica, tracciamento, valutazione di impatto sui diritti fondamentali (FRIA), monitoraggio post-marketing
Rischio Limitato[10] (Titolo IV)	Sistemi di IA che interagiscono con persone naturali (e.g., chatbots), e che creano o manipolano suoni, immagini, e video (e.g., deepfakes)	Obblighi di trasparenza
Rischio Minimo (Titolo IX)	Sistemi di IA per il filtraggio dello spam e per videogiochi	Codici di condotta (volontari)

Dalla tabella emerge chiaramente che gli oneri normativi sono distribuiti proporzionalmente: pericoli minori richiederanno cautele minori a carico, ad esempio, degli sviluppatori e dei fornitori di sistemi di IA. La maggior parte del costo di compliance ricade sui providers dei sistemi ad alto rischio (AIA, Articolo 3(2)). Numerosi oneri normativi ricadono anche su coloro che impiegano concretamente questi sistemi di IA (i "deployers"), come gli utenti professionali (e.g., un medico o una banca), (AIA, Articolo 3(4)).[11] Tra le altre cose, i deployers devono utilizzare il sistema secondo le istruzioni d'uso redatte dai fornitori, monitorare che lo stesso le rispetti durante l'intero ciclo d'uso, segnalare eventuali incidenti (AIA, Articolo 62), e svolgere una valutazione di impatto sulla protezione dei dati (AIA, Articolo 29(6)).

L'AIA prevede (dei massimi di) sanzioni amministrative per i providers e deployers che violino il contenuto del regolamento (AIA, Articolo 71). Nello specifico, in caso di inadempienza delle regole sui sistemi proibiti (Articolo 5), prevede una multa fino a un massimo di 40 milioni di euro; una multa fino a un massimo di 20 milioni di euro per inadempienza delle regole sulla gestione dei dati e gli obblighi di

[10] Questa categoria non è mutualmente escludente con quella dei sistemi ad alto rischio.
[11] Dopo il testo compromissorio del Parlamento EU utenti professionali e 'deployers' sono equiparati.

trasparenza; una multa fino a 10 milioni di euro per inadempienza con qualsiasi altro requisito o condizione.[12] Tuttavia, se il soggetto inadempiente è una società che nell'anno finanziario precedente ha avuto un fatturato mondiale annuo superiore a queste cifre, allora la multa consisterà in una percentuale del medesimo; rispettando l'ordine, del 7, 4, e 2 percento del fatturato. La misura concreta della multa deve essere stabilita considerando la natura, la gravità, e la durata dalla inadempienza, ma anche le dimensioni, il fatturato annuo, e l'eventuale condotta riparativa o virtuosa dell'operatore (Articolo 71(6)). L'individuazione di sanzioni di natura non-monetaria, che possono essere associate o sostituite dalle multe, resta agli Stati Membri, secondo principi di efficacia, proporzionalità, e deterrenza. Infine, alcune sanzioni amministrative possono essere imposte dal Garante europeo della protezione dei dati in caso di violazione del dettato del regolamento da parte delle istituzioni ed organi UE.

3. I punti di forza delle regolamentazioni basate sul rischio

L'AIA ambisce a stabilire standard chiari e affidabili per la progettazione, l'implementazione e l'utilizzo dei sistemi IA, mirando a creare un mercato interno efficace. Un mercato interno ben funzionante dovrebbe essere sicuro per i consumatori, attrattivo per gli investitori e in grado di incentivare l'innovazione tecnologica. L'Unione Europea, con questa regolamentazione, potrebbe diventare un pioniere influente, avendo già stimolato paesi come Stati Uniti e Regno Unito a sviluppare proprie normative sull'IA.[13] Tuttavia, esiste il rischio che, in caso di mancata approvazione del testo normativo entro le elezioni del 2024 o comunque di inefficacia dell'infrastruttura regolamentare, l'UE possa perdere il vantaggio da first-muover che aveva acquisito muovendosi con anticipo rispetto ad altri sistemi giuridici. In tal caso, l'Unione Europea rischierebbe di perdere l'occasione di diventare un mercato attrattivo per la progettazione e lo sviluppo dell'intelligenza artificiale.

Per scongiurare questa ipotesi, è fondamentale che la regolamentazione stabilisca regole che mantengano elevati standard di sicurezza senza ostacolare la produzione, l'utilizzo e l'innovazione nell'IA. L'approccio basato sul rischio adottato dall'UE ha diversi punti di forza. Per esempio, consente di stabilire priorità e obiettivi in modo chiaro e trasparente, proteggendo i valori fondamentali dell'UE e favorendo lo sviluppo del mercato interno dell'IA. In questa ottica, maggiore è il rischio che un sistema IA possa compromettere questi valori, ad esempio compromettendo alcuni diritti umani fondamentali, più elevato sarà il livello di regolamentazione richiesto. Inoltre, priorità e obiettivi esplicitamente delineati e bilanciati permettono

[12] A queste multe vanno aggiunte quelle per fornitura di informazioni inesatte, incomplete o fuorvianti agli organismi notificati e alle autorità nazionali competenti. La multa può arrivare fino a 5 milioni di euro o, se il trasgressore è una società, fino all'1 % del suo fatturato totale mondiale dell'esercizio finanziario precedente, se superiore.
[13] https://www.ncsc.gov.uk/news/uk-develops-new-global-guidelines-ai-security.

non solo di visualizzare la direzione della governance, ma anche di rendere i policy-maker più responsabili per le proprie decisioni.

Questo tipo di regolamentazione consente anche di valutare l'allocazione dei costi sociali connessi all'IA, includendo sia i costi di adeguamento e certificazione sia quelli derivanti da malfunzionamenti o violazioni di altra natura. L'AIA impone così una sorta di tassa sulle esternalità negative delle IA ad alto rischio, concentrandola soprattutto sui deployers. Tuttavia, rimane da valutare l'efficienza di questa allocazione di costi e risorse.

Le normative basate sul rischio sono inoltre utili nella gestione dell'incertezza, fornendo risposte in termini sia qualitativi che quantitativi per le incertezze dovute alla varianza delle situazioni (incertezze aleatorie) e/o all'assenza di informazioni complete (incertezze epistemiche).[14] Tipicamente, possono fornire metodologie per calcolare la probabilità di eventi pericolosi e le loro potenziali conseguenze, anche al fine di implementare strategie di prevenzione e mitigazione del rischio.

In alcuni casi, la valutazione dell'incertezza può trarre vantaggio da un supporto quantitativo attraverso l'analisi costi-benefici (CBA).[15] Mentre un principio generico di gestione del rischio può accettare che i costi di mitigazione superino i benefici[16], purché non eccessivi, la CBA presuppone che un intervento sia giustificato solo se i costi sono inferiori o uguali ai benefici. Anche se la CBA non considera costi e benefici incerti, può fornire un contributo preliminare nella gestione del rischio: quantificare i costi e benefici noti.[17] Queste informazioni possono poi essere integrate con valutazioni di natura qualitative su ciò che è "ragionevolmente praticabile". Senza valutazioni qualitative, infatti, la CBA si rivela uno strumento limitato poiché esprime il valore delle cose con un singolo parametro numerico, solitamente i prezzi di mercato, mentre le normative come l'AIA riguardano un rischio legale, il cui bene esposto consiste in diritti e valori fondamentali (principi natura costituzionale e identitari dell'ordine giuridico dell'UE).

Infine, le regolamentazioni basate sul rischio offrono un intervento flessibile, adattandosi ai cambiamenti politici, tecnologici ed economici. L'AIA, ad esempio, prevede la possibilità di aggiornare l'elenco dei sistemi di IA ad alto rischio (Articoli 84-85, AIA). Tuttavia, attualmente, l'inclusione di nuove IA è limitata ai campi già definiti, suggerendo la necessità di criteri di categorizzazione del rischio diversi, ossia, più granulari e revisionabili nel tempo.

---

[14] T. AVEN, *Risk Assessment and Risk Management: Review of Recent Advances on Their Foundation*, in *European Journal of Operational Research*, 253, n. 1, 2016, pp. 1–13.

[15] S. FRENCH, T. BEDFORD, E E. ATHERTON, *Supporting ALARP decision making by cost benefit analysis and multiattribute utility theory*, in *Journal of Risk Research*, 8, n. 3, 2005, pp. 207–23.

[16] Un esempio potrebbe essere il principio 'As Low As Reasonably Practicable' (ALARP) – talvolta chiamato ALARA – che presente nella legislazione sulla sicurezza inglese e neozelandese.

[17] B. J. M. ALE, D. N. D. HARTFORD, E D. SLATER, *ALARP and CBA All in the Same Game*, in *Safety Science*, n.76, 2015, pp. 90–100.

4. Alcuni limiti dell'analisi del rischio nell'AIA e come affrontarli

Anche se la regolamentazione basata sul rischio è uno strumento potenzialmente efficace per governare l'incertezza e la dinamicità dello sviluppo tecnologico, il modello di valutazione del rischio dell'AIA mostra delle debolezze. Alcune di esse fanno emergere dei dubbi sulla efficacia dell'implementazione del regolamento. Nei prossimi sottoparagrafi ne individuo tre: (1) la predeterminazione del rischio, (2) il giudizio di significatività del rischio, e (3) la valutazione di impatto sui diritti fondamentali.[18]

4.1. La predeterminazione dei livelli di rischio

Il problema della predeterminazione dipende dal fatto che nella concezione di rischio dell'AIA prevale la componente assiologica; quindi, la probabilità che l'impiego di un sistema di IA danneggi valori e diritti fondamentali dell'UE (e.g., salute e sicurezza). La classificazione che ne deriva mostra una natura altrettanto valoriale, come conferma la lista dei sistemi a rischio inaccettabile e alto, oltre che l'attenzione all'impatto dell'IA sui diritti fondamentali (AIA, articolo 29a), sui cui tornerò a breve. Ad esempio, la categoria dei sistemi a rischio alto dipende da una lista di valori cardine dell'UE, rappresentanti come ambiti di adozione dei sistemi di IA: istruzione, lavoro, uguaglianza, diritto d'asilo, stato di diritto, democrazia, e altri. Ciò non sorprende, soprattutto perché lo stesso approccio caratterizza altri atti legislativi UE adottati nell'ambito del digitale.[19]

Tuttavia, spesso l'AIA predetermina il peso di questi valori senza sottoporre la categorizzazione del rischio a una valutazione più dettagliata. Questo vuol dire che la classificazione del rischio di un sistema di IA darà priorità ai valori giuridici coinvolti, e in particolar modo a quelli considerati gerarchicamente preordinati, rispetto agli scenari applicativi concreti. Ma la predeterminazione è una strategia solo subottimale, da cui deriva una visione statica dei pericoli dell'IA; ignora l'interazione dinamica tra fonti di pericolo, interazione sociotecnica, profili di vulnerabilità, e fattispecie esterne di rischio. Questi fattori di rischio, presi singolarmente, hanno un ruolo all'interno dell'AIA – e.g., vulnerabilità (Art. 5 (b)) – ma non vengono messi in relazione dinamica per valutare il rischio complessivo di un sistema di IA. Questo approccio, inoltre, restituisce un'immagine statica dei valori (giuridici) stessi – principalmente diritti fondamentali – che vengono trattati come standard tecnici a realizzazione binaria invece che come precetti di ottimizzazione (come accade tipicamente per i principi giuridici) [20].[21]

[18] I punti 2) e 3) sono stati introdotti attraverso emendamenti alla bozza iniziale e approvati il 14 giugno 2023.
[19] M. E. GONÇALVES, *The risk-based approach under the new EU data protection regulation: a critical perspective*, in *Journal of Risk Research*, n. 23, fasc. 2, 2020, pp. 139–52.
[20] R. ALEXY, *A Theory of Constitutional Rights*, Oxford University Press, Oxford, 2002.
[21] Invero, questo difetto è stato in parte corretto con l'introduzione di una valutazione di impatto sui diritti fondamentali (FRIA) che i deployer di sistemi ad alto rischio sono obbligati a fare. Vedremo che anche rispetto a questa valutazione, emergono significativi problemi.

La predeterminazione del rischio ha delle implicazioni negative per il successo e l'implementazione dell'AIA. Infatti, se la valutazione è statica, aumenta la probabilità che le categorie di rischio siano sovra- o sotto-inclusive.

Nella prima ipotesi, si porranno problemi di onerosità dell'adeguamento agli standard giuridici da parte di produttori e fornitori i cui sistemi di IA, pur operando in aree considerate sensibili dall'AIA, non pongano rischi elevati. Come abbiamo visto, il carico di obblighi e garanzie a carico di produttori e fornitori di sistemi ad alto rischio è molto più alto rispetto agli altri livelli di rischio. Dal momento che l'adeguamento alle norme presenta sempre un costo, sia in termini di organizzazione interna che di costo-opportunità, potenziali produttori e fornitori potrebbero essere disincentivati a investire nel mercato europeo.

Non è chiaro, ad esempio, perché tutti i sistemi impiegati in ambito educativo o di formazione professionale debbano adeguarsi all'intero pacchetto di obblighi previsti per i sistemi ad alto rischio. In questo ambito, i sistemi di IA vengono spesso organizzano in maniera automatizzata il materiale didattico e formativo. Il malfunzionamento di questi sistemi ha un impatto piuttosto limitato su i diritti e le vulnerabilità degli individui. Alcuni di questi problemi possono essere mitigati attraverso interventi tecnici mirati e non sembra ragionevole imporre ai fornitori di questi sistemi una valutazione di impatto sui diritti fondamentali ai fini della certificazione (come è attualmente richiesto).[22]

Nella seconda ipotesi, invece, il rischio di alcuni sistemi di IA potrebbe essere sottovalutato. Un esempio, a mio avviso rilevante, è quello dell'IA usata per i videogiochi che, secondo l'attuale disciplina, dovrebbe essere trattata sempre come a rischio minimo, quindi quasi senza oneri normativi. Tuttavia, in questo ambito, l'IA può porre rischi significativi, come rinforzare la dipendenza attraverso tecniche di engagement o manipolare il comportamento al fine di indurre ad acquisti online giocatori non maggiorenni.[23]

Sia in caso di sovra-inclusione che di sotto-inclusione, questi problemi possono pregiudicare il cd. Brussel Effect [24] e il successo generale dell'AIA e, con esso, una grossa parte della strategia europea sull'IA. Il vantaggio competitivo da first-mover che l'UE aveva guadagnato e che, si sperava, avrebbe tradotto la certezza giuridica in attrattività del mercato, si è già significativamente ridimensionato dopo i numerosi

[22] Qualcuno potrebbe obiettare che in tal caso il sistema non porrebbe un rischio considerato "significativo" e che quindi sarebbe possibile contestare la classificazione di alto livello per il sistema in considerazione. Tuttavia, questa la valutazione di significatività opera solo in seconda istanza, con modalità e tempistiche ancora incerte.

[23] Non sto suggerendo che tutti i sistemi usati per i videogiochi sia ad alto rischio – né che tutti quelli impiegati in ambito professionale non lo siano – ma che questi giudizi richiederebbero appunto un modello di valutazione del rischio (o delle misure di gestione) più granulare.

[24] A. BRADFORD, *The Brussels Effect*, in *Northwestern University Law Review*, n. 107, fasc. 1, 2013, pp. 1–68. A. BRADFORD, *The Brussels Effect: How the European Union Rules the World*, *Faculty Books*, 2020.

stalli del processo legislativo europeo.[25] Se alla meticolosità dell'accordo politico, si aggiunge una grossolana valutazione del rischio e allocazione dei costi, allora l'esperienza europea finirà per essere soltanto un esempio da non seguire (un Brussel Effect al contrario).

D'altro canto, la predeterminazione del livello di rischio può avere dei vantaggi in termini di semplicità procedurale e uniformità di applicazione dell'AIA. Sotto il primo profilo, l'associazione statica del livello di rischio di un sistema di IA al suo ambito generale di impiego facilita l'approvazione e il monitoraggio da parte dei fornitori e delle autorità nazionali di supervisione, rendendo le procedure più semplici, veloci ed economiche. Tale approccio permette inoltre di conoscere con anticipo i costi di adeguamento alla normativa e, indirettamente, per lo sviluppo e commercializzazione dei sistemi di IA. Al contrario, una procedura di valutazione del rischio più granulare, come quella che auspico, può rivelarsi super-erogatoria e fornire meno certezze, scatenando l'effetto opposto a quello che si desidera. Sotto il secondo profilo, un elemento critico di una valutazione non predeterminata del rischio consiste nella maggiore probabilità di applicare l'AIA in maniera diversificata e frammentata. Ciò danneggerebbe l'obiettivo dell'AIA di armonizzare le regole sull'IA tra gli Stati Membri. Allo stesso tempo, bisogna sottolineare che non c'è alcuna garanzia che l'AIA nella sua impostazione attuale favorisca un'applicazione uniforme. Anzi, restano molte regole suscettibili di interpretazioni soggettive e arbitrarie come, lo vedremo più avanti, quelle sulla valutazione di impatto dei diritti fondamentali (FRIA).

Questi supposti vantaggi della predeterminazione del rischio andrebbero ridimensionati. Il profilo della semplicità/costo andrebbe considerato anche alla luce della perdita del costo-opportunità dato da categorie di rischio più efficienti nel bilanciare sicurezza e innovazione. Quindi, mentre nel breve periodo è probabile che valutazioni granulari e contestuali appesantiscano le procedure, nel lungo periodo, misure per la gestione del rischio ritagliate sugli scenari concreti di applicazione potrebbero risultare in oneri normativi più leggeri e, nella migliore delle ipotesi, anche in una maggiore tutela dei valori esposti. Per arginare i problemi di breve periodo, bisogna favorire la cooperazione tra istituzioni dell'UE centrali e autorità nazionali di supervisione, in modo che, anche nel contesto di una valutazione del rischio context-specific, vi siano alcuni fattori di rischio costanti, magari individuati proprio all'interno di quelle aree di applicazione generali previste dall'AIA (ad esempio, all'Annesso III).

Per quanto riguarda il supposto vantaggio di un'applicazione uniforme dell'AIA, è vero che una maggiore imprevedibilità e discrezionalità potrebbe emergere con l'aumento delle variabili (e delle interazioni) da includere nella valutazione di rischio, specie se considerate in situazioni specifiche (gli scenari) e a livello territoriale. Tuttavia, l'individuazione di parametri comuni, per quanto contestuali, tramite gli atti delegati

[25] Il più recente, e potenzialmente in grado di far ritardare di molto l'AIA, è quello sui modelli fondativi: https://www.euractiv.com/section/artificial-intelligence/news/eus-ai-act-negotiations-hit-the-brakes-over-foundation-models/.

o di implementazione dell'AIA, può mitigare questo pericolo di frammentazione. D'altronde, il legislatore europeo stesso evoca la possibilità di una valutazione del rischio più specifica, consentendo ai fornitori di riesaminare la classificazione predefinita e basata su ampie aree di applicazione attraverso un giudizio suppletivo di "significatività del rischio" (e.g., AIA, Recital 32).[26]

Per mitigare la predeterminazione del rischio presente nell'AIA si può: (a) introdurre una categorizzazione più granulare del rischio dei sistemi di IA, basata su (simulazione di) scenari concreti di applicazione; (b) prevedere più livelli di rischio rispetto ai quattro attualmente previsti, anche nella forma di sottolivelli; (c) differenziare maggiormente gli oneri normativi a carico di produttori, fornitori e utenti dei sistemi di IA, anche se fanno parte della stessa categoria di rischio. Tutte queste soluzioni richiederebbero modifiche sostanziali al testo dell'AIA che, considerata la fase legislativa in cui ci troviamo, sono molto improbabili.

Va precisato che molti dei problemi che derivano dalla predeterminazione sono contenuti dal giudizio di significatività introdotto nel testo compromissorio del Parlamento. Approfondirò questo aspetto nel prossimo sottopragrafo.

Altre soluzioni sono auspicabili come, ad esempio, intervenire sugli atti di implementazione dell'AIA, magari introducendo una metodologia comune per calcolare il rischio nei casi concreti. Il ventaglio di strumenti che possono essere sperimentate in questa sede è ampio. Si possono introdurre deroghe, eccezioni, clausole a scadenza, oppure revisioni post-implementazione. Attualmente strumenti simili sono previsti limitatamente all'aggiornamento della lista dei sistemi ad alto rischio (Artt. 84 e 85, AIA). Inoltre, facilitare il funzionamento delle regulatory sandboxes potrebbe permettere ai produttori di IA di sperimentare senza doversi adeguare alle condizioni normalmente richieste. Certo, questi sono più dei rimedi che delle vere e proprie soluzioni, dal momento che dovremmo sempre confrontarci con le categorie di rischio rigide attualmente previste dall'AIA.

4.2. Il giudizio di significatività del rischio

Una delle innovazioni più significative del testo compromissorio approvato dal Parlamento il 14 giugno 2023 è il giudizio di significatività del rischio.

"As regards stand-alone AI systems, […] it is appropriate to classify them as high-risk if, in the light of their intended purpose, they pose a significant risk of harm to the health and safety or the fundamental rights of persons and […] to the environment. Such significant risk of harm should be identified by assessing on the one hand the effect of such risk with respect to its level of severity, intensity, probability

[26] Versione redatta dal Parlamento europeo "Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)."

of occurrence and duration combined altogether and on the other hand whether the risk can affect an individual, a plurality of persons or a particular group of persons. Such combination could for instance result in a high severity but low probability to affect a natural person, or a high probability to affect a group of persons with a low intensity over a long period of time, depending on the context […]" (Recital 32, AIA)

Dalla combinazione dei Recital 32, 32a, e Articolo 6(2), si evince il ruolo di tale giudizio: i fornitori di sistemi ad alto rischio possono contestare e potenzialmente ribaltare tale classificazione provando che il loro sistema di IA, pur operando negli ambiti pericolosi elencati all'Allegato III (e.g., educazione, impiego, salute, processi democratici), non comporterà effettivi rischi per i valori giuridici dell'UE, i diritti fondamentali delle persone o l'ambiente. Se desidera declassare il livello di rischio del proprio sistema prima di lanciarlo sul mercato il fornitore deve redigere un riassunto delle informazioni e delle ragioni per le quali il proprio sistema non rappresenta un rischio significativo (Recital 32a) e presentarlo all'autorità nazionale di sorveglianza. Quest'ultima potrà contrastare entro tre mesi[27] le ragioni del fornitore che a sua volta potrà appellare la decisione dell'autorità.

Come anticipato, il giudizio suppletivo di significatività mitiga la predeterminazione del rischio. Pertanto, questo giudizio va nella direzione auspicata nel precedente sottoparagrafo, fornendo un criterio di revisione della classificazione basato sulle circostanze dello scenario di applicazione di un sistema di IA. Tuttavia, il legislatore europeo ha definito la nozione significatività in modo piuttosto generico. In particolare, emergono due problemi: l'ambito del giudizio di significatività e la sua misurazione.

Per quanto riguarda l'ambito di azione del giudizio di significatività, il recital 32 consente la revisione del livello di rischio esclusivamente per i sistemi di IA ad alto rischio. Non si coglie bene il senso di questa limitazione. Sarebbe più utile avere criteri di revisione per ogni categoria di rischio, con l'obiettivo di promuovere misure di gestione del rischio più efficaci e flessibili. Anche i sistemi con rischio inaccettabile potrebbero essere rivalutati sulla base del rischio effettivo. Infatti, l'evoluzione delle conoscenze potrebbe rendere alcuni sistemi di IA meno rischiosi nel tempo grazie a soluzioni tecniche di mitigazione dei pericoli.

Al contrario, un sistema di IA classificato come a basso rischio potrebbe nel tempo diventare pericoloso, richiedendo una rivalutazione della significatività del rischio. L'IA nei videogiochi è un esempio: pur sembrando innocua, può diventare rischiosa con l'aumento dell'autonomia o con l'interazione con altre tecnologie come la realtà aumentata o estesa.[28] In questo caso, anche se l'AIA prevede la possibilità di aggiornare, attraverso atti delegati della Commissione, la lista dei sistemi ad alto rischio per le IA che

[27] Salvo poter intervenire oltre questo periodo per rischi di carattere nazionale (come avviene per gli altri sistemi di IA).
[28] D. REINERS et al., T*he Combination of Artificial Intelligence and Extended Reality: A Systematic Review*, in *Frontiers in Virtual Reality*, 2, 2021.

presentano rischi comparabili o superiori a quelli già presenti – alla luce della severità, probabilità e impatto sui diritti fondamentali (e.g., Articolo 7) – non è chiaro in che modo questa valutazione dovrà essere fatta.

Quest'ultimo aspetto ci consente di affrontare il secondo profilo problematico, ossia, la misurazione della significatività del rischio. Il Recital 32 definisce un rischio come significativo sulla base della combinazione di alcuni elementi: da un lato, la gravità, l'intensità, la probabilità del verificarsi dell'evento, la sua durata e, dall'altro, il rischio di danneggiare individui o collettività. Sebbene queste variabili identifichino alcuni dei fattori più rilevanti nell'ontologia del rischio, altri di pari importanza sembrano trascurati o, quando inclusi, considerati in modo isolato. Ciò incide sull'accuratezza del calcolo della magnitudine di rischio. Per esempio, per valutare se e come un sistema di categorizzazione biometrica causerà danni, non è sufficiente stimare la probabilità di esiti indesiderati basandosi solo sulla frequenza con cui certe fonti di pericolo generano danni. È necessario combinare questa probabilità con altri fattori, come l'esistenza di contromisure in grado di contrastare il verificarsi del danno. Così, ad esempio, rischi connessi alla provenienza e gestione dei dati non possono essere calcolati indipendentemente dall'esistenza di salvaguardie tecniche e/o giuridiche (GDPR) già presenti e vincolanti. L'interazione tra tutti questi elementi può far aumentare o diminuire la probabilità complessiva dell'evento dannoso e di riflesso la magnitudine di rischio (ossia, la probabilità dell'evento sommata alle sue conseguenze negative). Pertanto, l'implementazione del giudizio di significatività dell'AIA richiede di sviluppare un modello che consideri molteplici fattori di rischio e, soprattutto, che li metta in relazione dinamica tra di loro.

Ma come far sì che queste considerazioni vengano integrate nell'AIA o nei suoi atti di implementazione? A mio avviso, appare più semplice intervenire sugli atti di implementazione dell'AIA, ossia quegli atti normativi non legislativi in cui vengono fissati standard specifici di conformità, piuttosto che sull'architettura generale del regolamento.

Dal punto di vista operativo, sarebbe a mio avviso utile sviluppare, negli atti di implementazione, una matrice di rischio che combini i due elementi fondamentali della magnitudine di rischio: la probabilità che l'evento si verifichi e le conseguenze negative.[29] Per poter calcolare questi due elementi, serve però una tassonomia chiara e dettagliata dei fattori di rischio e un quadro delle possibili interazioni tra i medesimi.

Un approccio interessante è offerto da alcuni modelli di analisi del rischio sviluppati nei policy reports dell'Intergovernmental Panel on Climate Change (IPCC). Come indicato in queste analisi, e nella relativa letteratura, il rischio complessivo di un fenomeno va valutato tenendo conto non solo delle fonti di

[29] H. NI, A. CHEN, e N. CHEN, *Some Extensions on Risk Matrix Approach*, in *Safety Science*, n. 48, fasc. 10, 2010, pp. 1269–78.

pericolo, della severità delle conseguenze o della probabilità, ma anche della natura e quantità dei beni esposti, della loro vulnerabilità e delle strategie di mitigazione e prevenzione del rischio [30]. L'AIA non ignora questi fattori, come la vulnerabilità menzionata all'Articolo 5. Il problema è che li tratta come variabili indipendenti e realizzazione binario, che esistono o non esistono, e non considera la loro interazione come un evento rilevante ai fini del calcolo della significatività del rischio.

A questo scopo, già in precedenti lavori abbiamo proposto di integrare il modello per la gestione da rischio di cambiamento climatico con l'AIA [31]. La tassonomia del rischio da IA distinguerebbe così quattro principali determinanti del rischio (hazard, exposure, vulnerability, e response), singole componenti di queste determinanti (i cd. driver), e le fattispecie di rischio esterne dal rischio osservato ma che ne influenzano la magnitudine.

Questi fattori, poi, andrebbero messi in relazione secondo almeno tre modalità di interazione: (1) interazioni aggregate, per cui i fattori di rischio si manifestano in modo indipendente, ma la loro presenza combinata aumenta il rischio complessivo. Ad esempio, in un sistema di intelligenza artificiale (IA) per la diagnostica medica, l'opacità del funzionamento del sistema e una scarsa rappresentatività delle immagini usate per l'addestramento sono due rischi distinti che, insieme, incrementano il rischio totale; (2) interazioni cumulative, per cui i fattori di rischio interagiscono in modo specifico, producendo un impatto maggiore quando combinati. Per esempio, in un sistema di IA per la valutazione del credito, la limitata rappresentatività dei dati può combinarsi con il rischio di overfitting o con pregiudizi nei dati, aggravando il rischio complessivo; (3) interazione a cascata, per cui un fattore di rischio può innescarne altri, creando una reazione a catena.[32] Così, ad esempio, l'opacità di un sistema di IA generativa può portare a rischi di imprevedibilità, non governabilità, violazione della privacy o del diritto d'autore.[33] Questi rischi emergono come conseguenza della caratteristica iniziale, ma non sono direttamente collegati ad essa.

La letteratura sull'analisi e valutazione del rischio, indica ulteriori fattori da considerare: la temporalità, ad esempio distinguendo tra effetti a breve e lungo termine oppure modulando nel tempo l'intensità o probabilità dell'evento di rischio (e le misure di contenimento);[34] i tempi di recupero, che sono inclusi

[30] N. SIMPSON et al., *A Framework for Complex Climate Change Risk Assessment*, in *One Earth*, n. 4, fasc. 4, 2021, pp. 489–501. A. AYANLADE et al., *Complex Climate Change Risk and Emerging Directions for Vulnerability Research in Africa*, in *Climate Risk Management*, n. 40, 2023. H. PÖRTNER et al., a c. di, *Summary for policymakers*, in *Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge, Cambridge University Press, 2022. N. P. SIMPSON et al., *Adaptation to Compound Climate Risks: A Systematic Global Stocktake*, in *Science* 26, fasc. 2, 2023.

[31] C. NOVELLI et al., *Taking AI Risks Seriously: A Proposal for the AI Act*, in SSRN Scholarly Paper, 2023. C. NOVELLI et al., *How to Evaluate the Risks of Artificial Intelligence: A Proportionality-Based, Risk Model for the AI Act*, in SSRN Scholarly Paper 2023.

[32] N. SIMPSON et al., *A Framework for Complex Climate Change Risk...cit.*

[33] C. NOVELLI et al., *How to Evaluate...cit.*

[34] E. ZIO, *The Future of Risk Assessment*, in *Reliability Engineering & System Safety*, 177, 2018, pp. 176–90.

nella cd. resilienza (in un certo senso, l'opposto della vulnerabilità); [35]elementi ecologici, come le interazioni con tipologie di rischio esterne che non originano dall'uso di uno specifico sistema di IA ma che ne possono comunque influenzarne la magnitudine complessiva; rischi secondari, come quelli generati dalla regolazione stessa (cd. rischi ancillari); l'incertezza e i rischi residuali, ossia rischi non prevedibili a causa della scarsità di informazioni.[36] E si potrebbe andare ancora avanti, ad esempio, integrando concezioni più specifiche di probabilità (e.g., bayesiana o frequentista).

Non è questa la sede per esplorare le metodologie di valutazione del rischio da IA (e come farlo nel contesto dell'AIA europeo). Quello che però bisogna ribadire è che se il giudizio di significatività rimane generico, diventa indistinguibile dal modo in cui vengono inizialmente classificati (dal mio punto di vista, predeterminati) i sistemi di IA a monte, rendendo vano lo sforzo di introdurre un criterio flessibile di revisione.

Al contempo, intervenire sugli atti di implementazione dell'AIA redatti dalla Commissione, come si propone in questo articolo, crea ulteriori problemi. Ad esempio, la valutazione del rischio – e quindi delle relative misure di gestione – potrebbe essere parzialmente rimossa dal dibattito parlamentare, e gestita attraverso un'elaborazione normativa non democratici. Ciò non è insolito, dato che il Parlamento raramente interviene nei parametri tecnici e che l'UE prevede comunque meccanismi di rappresentanza democratica nella preparazione degli atti di implementazione. Così, prima di adottare un atto di implementazione, la Commissione Europea consulta un comitato rappresentativo di ogni Stato Membro, un processo noto come "comitologia".[37] Questo comitato supervisiona la Commissione durante l'adozione di atti di esecuzione. Inoltre, il programma della Commissione per una migliore regolamentazione consente ai cittadini e alle parti interessate di fornire riscontri su bozze di atti esecutivi per quattro settimane prima del voto del comitato.[38]

Nonostante questi meccanismi, il trasferimento di un considerevole potere decisionale agli apparati tecnici e alle autorità nazionali di supervisione e controllo potrebbe causare un'applicazione frammentata dell'AIA, o addirittura in contraddizione con l'intento del regolamento. Infatti, queste entità avrebbero il potere di decidere indirettamente se un sistema rientra o meno in una fascia di rischio e se la categorizzazione prevista dal regolamento debba essere stravolta. Pertanto, una soluzione a questi problemi deve essere ricercata nella governance dell'AIA, come insieme di regole che definiscono le

[35] S. THEKDI E T. AVEN, *Think Risk: A Practical Guide to Actively Managing Risk*, 1ª ed., London: Routledge, 2023.
[36] T. AVEN, *Ibid.*; S. THEKDI E T. AVEN, *Ibid.*
[37] https://commission.europa.eu/law/law-making-process/adopting-eu-law/implementing-and-delegated-acts/comitology_en
[38] C. F. BERGSTRÖM, *Comitology: Delegation of Powers in the European Union and the Committee System*, Oxford, Oxford University Press, 2005. A. BALLMANN, D. EPSTEIN, e S. O'HALLORAN, *Delegation, Comitology, and the Separation of Powers in the European Union*, in *International Organization*, n. 56, fasc. 3, 2002, pp. 551–74.

competenze e l'interazione tra le istituzioni sovranazionali e nazionali. Le autorità di sorveglianza nazionali, le cui competenze sono solo parzialmente indefinite nell'AIA, devono operare in modo coordinato. Potrebbero farlo attraverso un consorzio o sotto la guida di un'autorità centrale che assicuri una corretta coordinazione tra le autorità nazionali e che l'implementazione sia coerente con il mandato del Parlamento. Questo diminuirebbe il rischio di arbitrarietà e dispersione nell'attuazione basata su scenari concreti dell'AIA. Nell'accordo politico raggiungo il 9 dicembre 2023, viene confermata la presenza di un AI Office all'interno della Commissione Europea chiamato a coordinare l'applicazione del regolamento. Questo ufficio sarà inoltre affiancato da un gruppo di esperti indipendenti con il compito di segnalare i pericoli da rischi sistemici dei modelli "general-purpose" e contribuendo a classificare e testare questi modelli.

A questo proposito, è importante segnalare che a seguito della fase di trilogo, il legislatore europeo ha stabilito un parametro per valutare la significatività del rischio per la general-purpose AI (GPAI) e i modelli fondativi, che include la IA cd. Generative, come ChatGPT o Bard. A differenza dei sistemi ad alto rischio, per i quali la significatività del rischio viene valutata sulla base di una serie di fattori, per la GPAI il parametro unico è la potenza computazionale. Così, si riterrà che i modelli addestrati utilizzando più di 10^25 operazioni al secondo – tecnicamente, FLoating point Operations Per Second (FLOPS) – presentino rischi "sistemici", con maggiori esigenze di controllo. L'uso dei FLOP come indicatore si basa sull'ipotesi che una maggiore potenza di calcolo porti a modelli più complessi con implicazioni sociali più ampie. Tuttavia, questo parametro è piuttosto parziale per le ragioni che abbiamo esposto finora.

4.3. La valutazione di impatto sui diritti fondamentali (FRIA)

Il Parlamento Europeo, nella bozza iniziale dell'AI Act, ha inserito il Fundamental Rights Impact Assessment (FRIA). Questa valutazione, prevista dall'articolo 29(a) dell'AI Act, è obbligatoria per i deployers di sistemi di IA ad alto rischio prima del loro impiego.[39] Uno degli aspetti positivi del FRIA è che introduce uno standard di conformità qualitativo che va oltre la mera conformità ai requisiti tecnici. L'obiettivo è effettuare una valutazione prognostica per identificare eventuali pregiudizi ai diritti fondamentali.

L'ambito del FRIA è molto ampio: i deployers devono analizzare non solo l'utilizzo previsto del sistema IA, ma anche la sua portata temporale e geografica, gli impatti sui diritti fondamentali, le conseguenze sulle comunità emarginate, le ripercussioni ambientali e le implicazioni per la governance pubblica

[39] Il FRIA è uno degli aspetti su cui il trilogo fatica di più a trovare un accordo, con il Consiglio Europeo che spinge per limitarne l'applicazione soltanto alle autorità pubbliche perché "private companies will have to comply with similar obligations under the upcoming Due Diligence Directive". Fonte: https://www.euractiv.com/section/artificial-intelligence/news/ai-act-eu-countries-mull-options-on-fundamental-rights-sustainability-workplace-use/ .

(Articolo 29a, AIA).[40] Questa valutazione deve essere eseguita all'inizio dell'uso del sistema o basata su valutazioni esistenti, a meno che non vengano invalidate da usi successivi.

In base ai risultati del FRIA, i deployers sono tenuti a sviluppare piani per mitigare gli impatti negativi sui diritti fondamentali. Se non è possibile formulare un piano adeguato, devono cessare la distribuzione del sistema IA e informare i fornitori di IA e le autorità nazionali. Inoltre, durante la valutazione, è richiesto di coinvolgere le parti interessate, come le agenzie di protezione dei consumatori e le autorità di protezione dei dati, offrendo un periodo di sei settimane per i loro contributi.

Il FRIA, pur non influenzando direttamente la classificazione del rischio, include al suo interno specifiche misurazioni del rischio, come la valutazione dell'impatto sui gruppi marginalizzati (lettera f) dell'articolo 29). Da un lato, quindi, il FRIA può essere funzionale a risolvere alcuni difetti del modello statico di rischio dell'AIA, che tratta i valori giuridici come standard tecnici, portando a risultati predeterminati nel test di bilanciamento dei valori e degli interessi della comunità (Smuha et al. 2021).

Dall'altro lato, affinché la valutazione di impatto diritti fondamentali sia funzionale ad una analisi più granulare del rischio – e quindi generi misure di gestione né troppo rigide né troppo permissive – è necessario che venga fatta con un metodo chiaro. Metodo che, per il momento, non sono state oggetto del dibattito nelle istituzioni europee.

Arriviamo quindi ai due principali problemi del FRIA. Il primo riguarda chi conduce la valutazione: capire se un sistema di IA impatta, e come impatta, i diritti fondamentali è un'operazione di bilanciamento. Come noto, i diritti soggettivi possono essere visti qualitativamente come principi e, pertanto, vanno realizzati nella misura più ampia possibile senza mai derogarli o azzerarne l'efficacia (come accade per le regole giuridiche).[41] Ad ogni modo, un'operazione del genere si traduce per le aziende costi e risorse significativi. Il bilanciamento dei diritti, operazione complessa persino per i giuristi qualificati delle Corti, spingerebbe i deployers ad assumere consulenti esperti legali o esternalizzare questa valutazione. Anche assumendo che tutto ciò abbia un costo ragionevole per i deployers, questi ultimi si farebbero carico in definitiva di decisioni particolarmente sensibili. Non si può pensare a un bilanciamento del genere senza linee guida sulla esecuzione del FRIA e senza coordinazione con altre valutazioni di impatto che andranno eseguite in contemporanea, come il Data Protection Impact Assessments (DPIA).[42]

In assenza di linee guida, il FRIA sembra simile a una forma di auto-regolamentazione, e questo potrebbe portare a pratiche devianti tra i deployer. Infatti, l'autoregolamentazione incentiva comportamenti

[40] Versione del regolamento proposta dal Parlamento Europeo.

[41] R. ALEXY, *On the Structure of Legal Principles*, in *Ratio Juris*, n. 13, fasc. 3, 2000, pp. 294–304. R. ALEXY, *A Theory of Constitutional Rights…cit.*; R. ALEXY, *On Balancing and Subsumption. A Structural Comparison*, *Ratio Juris*, n. 16, fasc. 4, 2003, pp. 433–49.

[42] K. Demetzou, *Data Protection Impact Assessment: A tool for accountability and the unclarified concept of 'high risk' in the General Data Protection Regulation*, in *Computer Law & Security Review*, n. 35, fasc. 6, 2019.

"banditeschi": alcuni deployer potrebbero dedicarsi a un'approfondita valutazione FRIA, investendo tempo e risorse per assicurare la conformità e la protezione dei diritti fondamentali. Altri, al contrario, potrebbero limitarsi all'essenziale, risparmiando tempo e risorse e, forti di una posizione o una reputazione più consolidata, compromettere la tutela dei diritti fondamentali. Questa variabilità nell'applicazione del FRIA potrebbe causare incertezza nell'industria dell'IA, lasciando le aziende in dubbio sul rispetto dei requisiti normativi e sull'accettazione dei loro prodotti e servizi sul mercato, aumentando il rischio di sanzioni post-commercializzazione. Questa situazione renderebbe il mercato europeo più instabile e meno attraente per gli investimenti.

Il secondo problema è strettamente legato al primo e riguarda come garantire che le valutazioni FRIA siano omogenee, considerando che vengono eseguite individualmente dai deployers. È essenziale che queste valutazioni seguano un approccio uniforme nella ponderazione dei diritti coinvolti. La mancanza di armonizzazione porta di nuovo all'incertezza normativa e alla necessità di regolamenti supplementari, spesso al di fuori del dibattito parlamentare, rendendo il mercato europeo meno attraente.

Come risolvere questi due problemi? Ancora una volta, si suggerisce di farlo in fase di implementazione dell'AIA. Escludendo le soluzioni più radicali – come quella di eliminare del tutto il FRIA – una soluzione intermedia è quella di continuare a far ricadere il FRIA sui deployers, ma secondo un modello quanto più possibile standardizzato.[43] Ciò non significherebbe standardizzare anche le scelte di bilanciamento dei diritti coinvolti dall'utilizzo di un sistema di IA – che rimarrebbero decisioni strategiche e soggettive dei deployers – ma assicurerebbe una metodologia più chiara e comune. In questo modo, la checklist esistente nell'articolo 29a dell'AI Act dovrebbe essere trasformata in un processo facilmente interiorizzabile ed eseguibile dai deployer. Alcune proposte del genere sono già state fatte in letteratura, ma soltanto per la valutazione d'impatto sui diritti fondamentali prevista del DPIA.[44] Le due valutazioni non sono molto diverse – al punto che qualcuno ha proposto di combinarle – ma le cornici normative in cui si inseriscono e le diverse implicazioni pratiche richiedono riflessioni distinte.

[43] C. NOVELLI, G. GOVERNATORI, E A. ROTOLO, *Automating Business Process Compliance for the EU AI Act*, in *Legal Knowledge and Information Systems*, 2023, pp. 125–30.

[44] H. JANSSEN, M. S. A. LEE, e J. SINGH, *Practical fundamental rights impact assessments*, in *International Journal of Law and Information Technology,* n. 30, fasc. 2, 2022, pp. 200–232. H. JANSSEN, *Detecting New Approaches for a Fundamental Rights Impact Assessment to Automated Decision-Making,* in SSRN Scholarly Paper, 2020. F. BIEKER et al., *A Process for Data Protection Impact Assessment Under the European General Data Protection Regulation*, in *Privacy Technologies and Policy*, a c. di S. SCHIFFNER et al., Lecture Notes in Computer Science, Cham, Springer International Publishing, 2016, pp. 21–37. J. COLES, S. FAILY, e D. KI-ARIES, *Tool-Supporting Data Protection Impact Assessments with CAIRIS*, in *2018 IEEE 5th International Workshop on Evolving Security & Privacy Requirements Engineering (ESPRE)*, 2018, pp. 21–27.

5. Conclusioni

L'articolo ha esaminato tre criticità principali associate all'AI Act (AIA), tutte legate all'analisi e gestione del rischio.

Per superare la rigidità della predeterminazione del rischio, l'articolo suggerisce di introdurre una categorizzazione del rischio per l'IA basata sulla simulazione di scenari applicativi. Inoltre, propone di diversificare gli oneri normativi per produttori, fornitori e utenti di sistemi IA, anche all'interno della stessa categoria di rischio, per un approccio più equilibrato e adatto al contesto.

Riguardo alla valutazione di significatività del rischio, si raccomanda lo sviluppo di una metodologia più granulare. L'articolo cita modelli di rischio adottati nei report dell'IPCC e nella letteratura sul rischio climatico, suggerendo di trarre spunto da questi approcci per una valutazione più affidabile e coerente.

L'articolo consiglia, infine, di standardizzare la valutazione FRIA, rendendola un processo strutturato e uniforme per tutti i deployers di sistemi IA ad alto rischio, garantendo così una valutazione più bilanciata dei diritti fondamentali.