AI Act, il regolamento europeo sull'intelligenza artificiale: punti di forza e punti di debolezza

Gaetana Natale*
Augusto M. Lazzè**

L'Europa è ormai giunta a definire un regolamento sull'intelligenza artificiale approvato dal Parlamento Europeo in data 13 marzo 2024. Ma che tipo di regolazione introduce: funzionale o strutturale, generica o per settori omogenei? A quali modelli culturali si ispirano le regole in esso introdotte? Possiamo parlare di *eteroregolazione*, *co-regolazione* o di *self-regulation* temperata da una *compliance* volontaria cd. Ai Pact? Introduce *principles* o *model-rules* o regole operative? Prevede solo obblighi o anche diritti tutelati da un efficace *enforcement*? È formalmente un regolamento, ma nella sostanza una direttiva per le clausole generali in esso contenute? È un regolamento sulle cd. certificazioni? Rappresenta un valido compromesso tra l'esigenza di non bloccare lo sviluppo tecnologico e nel contempo tutelare il cd. *human in the loop ex* art. 22 del GDPR? È una regolazione *future proof*? È volta alla realizzazione di un'armonizzazione o a creare una base di diritto uniforme? Come sono regolati i cd. "sistemi fondazionali" o i cd. "sistemi fondativi"? Il cd. FRIA, ossia il *Fundamental Right Impact Assessment of Generative Artificial Intelligence* può superare quella che Natalino Irti (1) definiva *atopia ed anomia*? Il nuovo regolamento definisce un nuovo concetto di *accountability* diverso da quello che abbiamo imparato a definire dal GDPR, nel senso che segue un approccio *top-down* e non *bottom up* (2)?

Proviamo a fare un'analisi critica delle principali norme in esso contenute per poter tracciare le coordinate di tale nuovo orizzonte di senso, cd. *Wertbegriff* dell'Intelligenza Artificiale che già permea e conforma silenziosamente le nostre vite.

Compiendo un'opera di normogenesi, così come richiesta da Betti (3),

---

(*) Avvocato dello Stato e Professore di Sistemi Giuridici Comparati.
(**) Dottore in Giurisprudenza.

(1) N. Irti, *Il diritto nell'età della tecnica*, Editoriale Scientifica, Napoli, 2007.

(2) Sul punto è stato sostenuto che, mentre il GDPR introduce un principio di *accountability* inteso nel senso di responsabilizzare gli operatori coinvolti nel trattamento dei dati personali, attribuendo agli stessi un ruolo decisivo nella valutazione del rischio (e realizzando così un sistema decentralizzato), l'AI Act introduce invece una classificazione dei sistemi di IA in base all'entità dei rischi (il cd. *risck-based approach*) connessi alla loro operatività. Un simile quadro è definito direttamente dal nuovo regolamento, garantendo così una valutazione del rischio centralizzata (appunto per questo l'approccio si definisce "*top-down*") e con la notoria forza cogente e immediatamente applicabile del regolamento (articolo 288 TFUE). Così, G. Resta, *Cosa c'è di "europeo" nella Proposta di Regolamento UE sull'intelligenza artificiale?*, in *Diritto dell'Informazione e dell'Informatica* (II), fasc. 2, 2022, pp. 323-342.

non si può non affermare che i principi sono espressione di valori (4): tra i valori e le regole il ponte sono i principi, intesi come metanorme (5) con funzione normogenetica (6) e normopoietica. Per Bobbio (7) i principi hanno 4 funzioni: interpretativa, integrativa, di direttiva e limitativa. Per Josef Esser (8) è importante la precomprensione e la scelta del metodo nel processo di individuazione del diritto. Seguendo tale metodologia proviamo ad entrare nel *core* di questa nuova regolazione, confrontandola poi con quella degli altri Paesi, ossia degli Stati Uniti e della Cina.

Il testo si compone di 113 articoli e tredici allegati (cd. tecnica degli annessi con una periodica revisione quinquennale). Il regolamento (che, in quanto tale, sarà effettivo senza bisogno che le singole nazioni lo recepiscano) è impostato su un'architettura di rischi, cd. *Risk Assessment* suddivisi in quattro categorie: 1) *unaceptable risk*, 2) *high risk*, 3) *limited risk* e 4) *low and minimal risk*: inaccettabili, alti, limitati, minimi. Maggiore è il rischio (9), maggiori le responsabilità e i paletti per chi sviluppa o adopera sistemi di intelligenza artificiale, fino alle applicazioni considerate troppo pericolose per essere autorizzate.

Gli unici casi che non ricadono sotto l'ombrello dell'AI Act sono le tecnologie adoperate per scopi militari (10) e per quelli di ricerca (11).

---

(3) E. Betti, *Interpretazione della legge e degli atti giuridici (Teoria generale e dogmatica)*, II ed., a cura di G. Crifò, Milano, 1971, pp. 112 e ss., ove l'Autore definisce la nomogenesi come il « (…) modo come in origine la norma fu pensata e come i tipi di interessi in giuoco furono valutati e coordinati (…)».

(4) C. Schmitt, *La tirannia dei valori*, III ed., presentazione a cura di G. Accame, Pellicani, Roma, 1996.

(5) Infatti, il termine "principio" è stato definito come «(…) il pensiero, l'idea germinale, il criterio di valutazione, di cui la norma costituisce la messa in opera, calata in una specifica formulazione. Esso fa riscontro al problema pratico risolto dalla norma: ne ispira la *ratio iuris* sotto l'aspetto teleologico, in quanto ne fornisce il criterio di soluzione», così E. Betti, *Interpretazione della legge e degli atti giuridici (Teoria generale e dogmatica)*, cit., p. 312.

(6) Circa la funzione normogenetica dei principi, v. F. Modugno, *Principi generali dell'ordinamento*, in *Enciclopedia giuridica*, vol. XXIV, 1991, pp. 4, 8 e ss.; J. Raz, *Legal Principles and the Limits of Law*, 81 Yale Journal Law 823 (1972), p. 841; S. Bartole, *Principi generali del diritto (diritto costituzionale)*, in *Enciclopedia del diritto*, vol. XXXV, 494, 1986, pp. 515 e 531.

(7) N. Bobbio, *Studi per una teoria generale del diritto*, edizione a cura di T. Greco, Giappichelli, Torino, 2012.

(8) J. Esser, *Precomprensione e scelta del metodo nel processo di individuazione del diritto*, traduzione a cura di S. Piatti, G. Zaccaria, Edizioni Scientifiche Italiane, Napoli, 1983.

(9) La cui nozione è fornita dall'articolo 3, paragrafo 1, punto 2 dell'AI Act come «(…) la combinazione della probabilità del verificarsi di un danno e la gravità del danno stesso». Bisogna inoltre notare come la definizione del livello di rischio sia operata in relazione all'ambito di applicazione del sistema di IA. Ciò è dovuto al fatto che il rischio stesso è parametrato in base all'impatto non sui valori di mercato, bensì sui diritti fondamentali dell'Unione. Così, G. Resta, *Cosa c'è di "europeo" nella Proposta di Regolamento UE sull'intelligenza artificiale?*, cit., p. 339.

(10) Così come risulta già dal *Considerando 24* dell'AI Act. Nell'ambito di quest'ultimo, il legislatore europeo ha in primo luogo escluso dall'ambito applicativo del Regolamento i sistemi di IA sviluppati, immessi sul mercato, messi in servizio o utilizzati con scopo militare, di difesa o sicurezza

Procediamo a considerare quattro macroaree, ossia: 1) gli usi proibiti; 2) le intelligenze artificiali ad alto rischio; 3) i sistemi di AI per uso generale e 4) il sistema degli uffici, innovazioni e controllo.

Gli usi proibiti.

L'AI Act mette subito in chiaro quali sono gli impieghi vietati. Sono elencati all'articolo 5 (12). Tra questi vi sono sistemi che sfruttano tecno-

---

nazionale, salvo poi precisare che ove temporaneamente o permanentemente quei sistemi fossero adoperati per scopi differenti, sarebbero destinatari della disciplina dell'AI Act. Da ciò deriva il conseguente obbligo in capo al soggetto che utilizza il sistema di garantire idonea *compliance* ai requisiti richiesti dalla normativa europea.

(11) Il *Considerando 25* precisa, infatti, che l'AI Act «(…) dovrebbe sostenere l'innovazione, rispettare la libertà della scienza e non dovrebbe pregiudicare le attività di ricerca e sviluppo. È pertanto necessario escludere dal suo ambito di applicazione i sistemi e i modelli di IA specificamente sviluppati e messi in servizio al solo scopo di ricerca e sviluppo scientifici. È inoltre necessario garantire che il regolamento non incida altrimenti sulle attività scientifiche di ricerca e sviluppo relative ai sistemi o modelli di IA prima dell'immissione sul mercato o della messa in servizio (…)».

(12) Il cui paragrafo 1 dispone quanto segue: «Sono vietate le pratiche di IA seguenti: a) l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che utilizza tecniche subliminali che agiscono senza che una persona ne sia consapevole o tecniche volutamente manipolative o ingannevoli aventi lo scopo o l'effetto di distorcere materialmente il comportamento di una persona o di un gruppo di persone, pregiudicando in modo considerevole la sua capacità di prendere una decisione informata, inducendo pertanto una persona a prendere una decisione che non avrebbe altrimenti preso, in un modo che provochi o possa provocare a tale persona, a un'altra persona o a un gruppo di persone un danno significativo; b) l'immissione sul mercato, la messa in servizio o l'uso di un sistema di IA che sfrutta le vulnerabilità di una persona o di uno specifico gruppo di persone, dovute all'età, alla disabilità o a una specifica situazione sociale o economica, con l'obiettivo o l'effetto di distorcere materialmente il comportamento di tale persona o di una persona che appartiene a tale gruppo in un modo che provochi o possa ragionevolmente provocare a tale persona o a un'altra persona un danno significativo; c) l'immissione sul mercato, la messa in servizio o l'uso di sistemi di IA ai fini della valutazione o della classificazione delle persone fisiche o di gruppi di persone per un determinato periodo di tempo sulla base del loro comportamento sociale o di caratteristiche personali o della personalità note, inferite o previste, in cui il punteggio sociale così ottenuto comporti il verificarsi di uno o di entrambi i seguenti scenari: i) un trattamento pregiudizievole o sfavorevole di determinate persone fisiche o di interi gruppi di persone in contesti sociali che non sono collegati ai contesti in cui i dati sono stati originariamente generati o raccolti; ii) un trattamento pregiudizievole o sfavorevole di determinate persone fisiche o di gruppi di persone che sia ingiustificato o sproporzionato rispetto al loro comportamento sociale o alla sua gravità; d) l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di un sistema di IA per effettuare valutazioni del rischio relative a persone fisiche al fine di valutare o prevedere la probabilità che una persona fisica commetta un reato, unicamente sulla base della profilazione di una persona fisica o della valutazione dei tratti e delle caratteristiche della personalità; tale divieto non si applica ai sistemi di IA utilizzati a sostegno della valutazione umana del coinvolgimento di una persona in un'attività criminosa, che si basa già su fatti oggettivi e verificabili direttamente connessi a un'attività criminosa; e) l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di sistemi di IA che creano o ampliano le banche dati di riconoscimento facciale mediante *scraping* non mirato di immagini facciali da internet o da filmati di telecamere a circuito chiuso; f) l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di sistemi di IA per inferire le emozioni di una persona fisica nell'ambito del luogo di lavoro e degli istituti di istruzione, tranne laddove l'uso del sistema di IA sia destinato a essere messo in funzione o immesso sul mercato per motivi medici o di sicurezza; g) l'immissione sul mercato, la messa in servizio per tale finalità specifica o l'uso di sistemi di categorizzazione biometrica che classificano individualmente le persone fisiche sulla base dei loro dati biometrici per

logie subliminali per manipolare i comportamenti di una persona; quelli che abusano di persone vulnerabili e fragili; la categorizzazione biometrica che fa riferimento a dati personali sensibili, come il credo religioso, l'orientamento politico o sessuale; la pesca a strascico (il cd. *web scraping*) da internet di volti, come fece anni fa la contestata startup Clearview AI; il riconoscimento delle emozioni sul posto di lavoro o a scuola; i sistemi di punteggio o *social scoring*. Il testo vieta anche la polizia predittiva, ossia la tecnica investigativa che utilizza informazioni come tratti della personalità, nazionalità, situazione familiare o economica, per stabilire la probabilità di commissione di un reato.

Tuttavia, sono previste alcune eccezioni. Per esempio, il divieto di categorizzazione biometrica non vieta l'etichettatura o il filtro di dataset biometrici, legalmente acquisiti, per scopi di polizia. E sono ammessi sistemi di analisi del rischio che non facciano profilazione di individui, come quelli per smascherare transazioni sospette o per tracciare le rotte del narcotraffico, sulla base dello storico accumulato nei database.

E poi c'è il paragrafo h (13), uno dei più combattuti nelle negoziazioni tra Consiglio e Parlamento. Perché riguarda l'impiego di sistemi di riconoscimento facciale e biometrico in tempo reale. Applicazione proibita, perché, come si legge nelle premesse, può portare "*a risultati marcati da pregiudizi e provocare effetti discriminatori*", salvo in tre "*situazioni ampiamente elencante e ben definite*", nelle quali il ricorso al riconoscimento facciale "*è necessario per raggiungere un sostanziale pubblico interesse, la cui importanza supera i rischi*". E i tre casi sono quelli annunciati a dicembre: la ricerca di vittime di reati e di persone scomparse; minacce certe alla vita o alla sicurezza fisica delle persone o di attacco terroristico; localizzazione e identificazione dei presunti autori di una lista di 16 reati contenuti nell'allegato II (14).

---

trarre deduzioni o inferenze in merito a razza, opinioni politiche, appartenenza sindacale, convinzioni religiose o filosofiche, vita sessuale o orientamento sessuale; tale divieto non riguarda l'etichettatura o il filtraggio di set di dati biometrici acquisiti legalmente, come le immagini, sulla base di dati biometrici o della categorizzazione di dati biometrici nel settore delle attività di contrasto; h) (...)».

(13) «(…) h) l'uso di sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico a fini di attività di contrasto, a meno che e nella misura in cui tale uso sia strettamente necessario per uno dei seguenti obiettivi: i) la ricerca mirata di specifiche vittime di sottrazione, tratta di esseri umani o sfruttamento sessuale di esseri umani, nonché la ricerca di persone scomparse; ii) la prevenzione di una minaccia specifica, sostanziale e imminente per la vita o l'incolumità fisica delle persone fisiche o di una minaccia reale e attuale o reale e prevedibile di un attacco terroristico; iii) la localizzazione o l'identificazione di una persona sospettata di aver commesso un reato, ai fini dello svolgimento di un'indagine penale, dell'esercizio di un'azione penale o dell'esecuzione di una sanzione penale per i reati di cui all'allegato II, punibile nello Stato membro interessato con una pena o una misura di sicurezza privativa della libertà della durata massima di almeno quattro anni. La lettera h) del primo comma lascia impregiudicato l'articolo 9 del regolamento (UE) 2016/679 per quanto riguarda il trattamento dei dati biometrici a fini diversi dall'attività di contrasto».

(14) Il quale, sotto la rubrica "Elenco di reati di cui all'articolo 5, paragrafo 1, lettera e), punto iii)", precisa che i reati richiamati dall'articolo menzionato nella rubrica sono da intendersi i seguenti:

L'elenco comprende: terrorismo; traffico di esseri umani; abusi sessuali su minori e pedopornografia; traffico di droghe e sostanze psicotrope; traffico illecito di armi, munizioni ed esplosivi; omicidio o gravi feriti; traffico di organi; traffico di materiale radioattivo e nucleare; sequestro di persona e ostaggi; crimini sotto la giurisdizione della Corte penale internazionale; dirottamento di aerei e navi; stupri; crimini ambientali; rapine organizzate e armate; sabotaggio; partecipazione a una organizzazione criminale coinvolta in uno o più crimini tra quelli elencati.

Il riconoscimento biometrico da remoto in tempo reale deve essere utilizzato "*solo per confermare l'identità*" della persona che è stata individuata come target, dopo aver bilanciato il rischio che si corre senza fare ricorso a questa tecnologia rispetto ai risultati consentiti dal suo impiego e per lo stretto necessario, "*nello spazio e nel tempo*". Per adottare questi strumenti, le forze di polizia devono prima fare un controllo sugli impatti sui diritti fondamentali dei cittadini e avere il *placet* di un giudice o di un ente indipendente. L'AI Act, tuttavia, garantisce una procedura d'urgenza. In questo caso si può attivare la sorveglianza biometrica e ci sono 24 ore di tempo per richiedere l'autorizzazione. Se manca tale autorizzazione, l'uso del riconoscimento facciale va bloccato immediatamente e tutti i dati devono essere cancellati.

I garanti nazionali dei dati personali e del mercato devono spedire ogni anno alla Commissione un rapporto sull'uso dei sistemi di riconoscimento biometrico in tempo reale, così come di eventuali usi proibiti. Ad ogni modo, gli Stati dell'Unione possono adottare leggi nazionali per ampliare il raggio d'azione della sorveglianza biometrica, nel rispetto dei paletti fissati dall'AI Act. Le stesse regole si applicano anche per il riconoscimento facciale usato *ex post*. In questo caso la finestra per ottenere l'autorizzazione in casi di urgenza è di 48 ore.

Le intelligenze artificiali ad alto rischio.

Sotto i sistemi vietati, si collocano quelli ad alto rischio, che pongono un significativo rischio per la salute, la sicurezza o i diritti fondamentali dei cittadini. A questa categoria è espressamente dedicato il Titolo III dell'AI Act, il quale si apre con l'articolo 6. Questa disposizione è preordinata, nei paragrafi 1 e 2, all'individuazione dei sistemi classificabili come IA "ad alto rischio"

«terrorismo, tratta di esseri umani, sfruttamento sessuale di minori e pornografia minorile, traffico illecito di stupefacenti o sostanze psicotrope, traffico illecito di armi, munizioni ed esplosivi, omicidio volontario, lesioni gravi, traffico illecito di organi e tessuti umani, traffico illecito di materie nucleari e radioattive, sequestro, detenzione illegale e presa di ostaggi, reati che rientrano nella competenza della Corte penale internazionale, illecita cattura di aeromobile o nave, violenza sessuale, reato ambientale, rapina organizzata o a mano armata, sabotaggio, partecipazione ad un'organizzazione criminale coinvolta in uno o più dei reati elencati sopra».

(15). Vi rientrano sistemi di identificazione e categorizzazione biometrica o per il riconoscimento delle emozioni; applicativi di sicurezza di infrastrutture critiche; *software* educativi o di formazione, per valutare i risultati di studio, per assegnare corsi o per controllare gli studenti durante gli esami. E poi vi sono gli algoritmi usati sul lavoro, per valutare *curriculum* o distribuire compiti e impieghi; quelli adoperati dalla pubblica amministrazione o da enti privati per distribuire sussidi, per classificare richieste di emergenza, per smascherare frodi finanziarie o per stabilire il grado di rischio quando si sottoscrive un'assicurazione.

Qui occorre fare il bilanciamento tra *tecne* ed *episteme*, affinché l'algoritmo "*if this, then that*" non si trasformi nel binomio servo-padrone di cui parla Hegel nella *Fenomenologia dello Spirito.*

Infine ricadono in questa categoria gli algoritmi usati dalle forze dell'ordine, dal potere giudiziario e dalle autorità di frontiera per valutare rischi, scoprire flussi di immigrazione illegale o stabilire pericoli sanitari, impedendo a una persona di varcare i confini dell'Unione. Se però l'algoritmo serve solo per svolgere una procedura limitata, per migliorare i risultati già ottenuti da un essere umano, per identificare deviazioni dagli usuali processi decisionali o per svolgere lavori preparatori di controllo, allora non può essere considerato ad alto rischio.

Entro 18 mesi dall'entrata in vigore del regolamento, la Commissione fornirà linee guida per applicare in pratica le norme sui sistemi ad alto rischio. Così come modificare la lista degli algoritmi che ricadono sotto questa categoria. Per farlo occorre stabilire gli scopi della tecnologia, l'estensione d'uso e di autonomia decisionale, natura e quantità di dati processati, abusi su gruppi di persone, così come la possibilità di correggere un errore o i benefici ottenuti. Un *database* conterrà l'elenco aggiornato dei sistemi ad alto rischio usati in Europa.

---

(15) «1. A prescindere dal fatto che sia immesso sul mercato o messo in servizio in modo indipendente rispetto ai prodotti di cui alle lettere a) e b), un sistema di IA è considerato ad alto rischio se sono soddisfatte entrambe le condizioni seguenti: a) il sistema di IA è destinato a essere utilizzato come componente di sicurezza di un prodotto, o il sistema di IA è esso stesso un prodotto, disciplinato dalla normativa di armonizzazione dell'Unione elencata nell'allegato I; b) il prodotto, il cui componente di sicurezza a norma della lettera a) è il sistema di IA, o il sistema di IA stesso in quanto prodotto, è soggetto a una valutazione della conformità da parte di terzi ai fini dell'immissione sul mercato o della messa in servizio di tale prodotto ai sensi della normativa di armonizzazione dell'Unione elencata nell'allegato I. 2. Oltre ai sistemi di IA ad alto rischio di cui al paragrafo 1, sono considerati ad alto rischio i sistemi di IA di cui all'allegato III». Dunque, mentre il primo paragrafo pone due condizioni (rinviando all'Allegato I ove sono riportate normative europee di armonizzazione) al verificarsi delle quali il sistema di IA è da ritenersi ad alto rischio, il secondo paragrafo contiene una clausola residuale. Infatti, dall'analisi dell'Allegato III (cui rimanda l'articolo 6, paragrafo 2) si evince la decisione del legislatore comunitario di applicare ai sistemi di IA la disciplina prevista per i sistemi ad alto rischio in base all'ambito applicativo dello specifico sistema. La *ratio* di una simile posizione si coglie in considerazione del fatto che è proprio il contesto di utilizzazione di una determinata tecnologia di IA a definire il grado di impatto potenzialmente pregiudizievole sui diritti fondamentali.

Sarà ad esempio ad alto rischio il nuovo sistema *Life2vec* di cui si sta parlando in questi ultimi giorni per predire e forse ritardare la morte? Come è noto i ricercatori di Copenaghen e della Northeastern University di Boston hanno sviluppato un algoritmo definito appunto *Life2vec*, utilizzando una vasta quantità di dati provenienti dall'anagrafe nazionale danese. Questi dati comprendono dettagli come istruzione, lavoro, stato di salute e altro ancora, trasformando ogni evento della vita in "parole" per l'algoritmo. Il risultato? Una previsione della mortalità prematura, con un'incredibile precisione del 79%, superando di gran lunga altri modelli predittivi.

Sarà da considerare ad alto rischio il sistema algoritmico definito *RAG*, ossia il *Retrieval Augmented Generation*, un sistema che può portare ChatGPT ad un livello superiore, permettendo agli algoritmi di "attingere" informazioni da fonti esterne, quasi se potessero consultare in tempo reale un'enciclopedia o un *database* esterni per fornire risposte sempre più dettagliate?

Di fronte a tale tecnologia siamo oltre il test di Turing e la legge di Moore e oltre quello che ha immaginato Geoffrey Hinton, padre dell'AI, psicologo cognitivo ed informatico che ha lasciato il suo ruolo in Google per poter parlare liberamente dei rischi dell'AI. Si ricorda che l'impatto di Hinton su AI è dovuto, soprattutto, al suo lavoro sulle cd. *backpropagation* o *retropropagazione*, la base del *deep learning* (16): una tecnica di apprendimento che aiuta le reti neurali a migliorare le loro previsioni attraverso il *sistema neurale convoluzionale.*

Chi sviluppa sistemi di AI ad alto rischio è tenuto a stabilire sistemi di controllo, gestire in modo trasparente i dati (17), chiarendo l'origine delle in-

---

(16) Espressione coniata da A.L. Samuel, *Some studies in Machine Learning using the game of Checkers*, 3 IBM Journal of Research and Development 210 (1959), la quale indica «(…) un campo di ricerca appartenente alla famiglia del *machine learning* (*omissis*) e dell'intelligenza artificiale (*omissis*), finalizzato alla creazione di un algoritmo di apprendimento capace di risolvere problemi complessi sulla base di informazioni catalogate e rielaborate in un ordine consequenziale di nozioni. Mediante l'impiego di calcoli matematici e informatici, definiti "reti neurali artificiali", l'apprendimento profondo, che può seguire diversi modelli, si basa sulla raccolta, analisi e selezione di diversi dati al fine di raggiungere una determinata conclusione, al pari di quanto accade all'interno di un cervello biologico quando si debba trovare una soluzione ad un problema. Come un cervello umano, il *deep learning*, inoltre, sviluppa nuovi processi di apprendimento e ragionamento utilizzando e combinando in modo nuovo le maggiori informazioni acquisite», L. Torchia, *Lo stato digitale*, il Mulino, Bologna, 2023, p. 186.

(17) È infatti imposto ai fornitori di sistemi di IA ad alto rischio di porre in essere un "sistema di gestione della qualità" (*ex* articolo 17, AI Act) al fine di garantire la conformità al Regolamento. Oggetto di questo sistema di gestione sono anche gli aspetti concernenti i dati, a norma del paragrafo 1, lettera f dello stesso articolo 17. La menzionata disposizione sancisce infatti quanto segue: «I fornitori di sistemi di IA ad alto rischio istituiscono un sistema di gestione della qualità che garantisce la conformità al presente regolamento. Tale sistema è documentato in modo sistematico e ordinato sotto forma di politiche, procedure e istruzioni scritte e comprende almeno i seguenti aspetti: (…) f) i sistemi e le procedure per la gestione dei dati, compresa l'acquisizione, la raccolta, l'analisi, l'etichettatura, l'archiviazione, la filtrazione, l'estrazione, l'aggregazione, la conservazione dei dati e qualsiasi altra operazione riguardante i dati effettuata prima e ai fini dell'immissione sul mercato o della messa in servizio di sistemi di IA ad alto rischio; (…)».

formazioni usate e mantenendole aggiornate, e registrare in automatico i log, da conservare per tutta la vita commerciale dell'algoritmo (18), per poter risalire a eventuali situazioni di rischio e indagare sulle origini (compiendo il cd. *reverse engineering*). Sono centrali i concetti di *security* (19), *safety* (20), *transparency* (21) e *explanibility* (22). Inoltre devono essere forniti i documenti tecnici (da conservare per 10 anni) (23), in versione *light* per *startup* e piccole e medie imprese. Gli sviluppatori di sistemi ad alto rischio dovranno comunicare il livello di accuratezza dell'AI, compresa una serie di metriche stabilita dalla Commissione, robustezza e sicurezza informatica (24). Il tutto sotto il controllo di un essere umano, il quale, in caso di pericolo imminente, può bloccare l'intelligenza artificiale attraverso un "*bottone di stop o una procedura simile, che consente al sistema di bloccarsi in modo sicuro*" (25). Gli sviluppatori sono tenuti a istituire un sistema di verifica della qualità, a sottoporsi alle analisi di conformità, applicare il marchio CE, che identifica un prodotto autorizzato nell'Unione, comunicare eventuali incidenti alle autorità. Anche importatori o distributori sono tenuti a conservare i documenti sulla sicurezza dell'AI che hanno venduto. E a sottoporsi a più controlli, se modificano l'algoritmo al punto da farlo ricadere nella categoria ad alto rischio. È previsto anche un sistema di monitoraggio dopo l'immissione di un sistema sul mercato, dal quale sono escluse le forze dell'ordine.

Primo problema: accanto a norme cd. prudenziali volte a definire il cd.

---

(18) *ex* articolo 12, paragrafo 1 del Regolamento in questione.

(19) L'articolo 15 al paragrafo 1 sancisce infatti che i sistemi ad alto rischio devono essere sviluppati in modo tale da presentare un idoneo grado di accuratezza, robustezza e cibersicurezza.

(20) A tal fine, è previsto dall'articolo 14 che sia garantita la supervisione umana sull'attività svolta dai sistemi ad alto rischio. Il paragrafo 2 dello stesso articolo prevede che «[l]a sorveglianza umana mira a prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali che possono emergere quando un sistema di IA ad alto rischio è utilizzato conformemente alla sua finalità prevista o in condizioni di uso improprio ragionevolmente prevedibile, in particolare qualora tali rischi persistano nonostante l'applicazione di altri requisiti di cui alla presente sezione».

(21) In specifico riferimento alla categoria dei sistemi di IA ad alto rischio, l'articolo 13 postula la necessità di progettazione e sviluppo degli stessi in modo trasparente, essendo questa una condizione necessaria e indefettibile affinché gli utenti possano fare degli *output* prodotti dai sistemi un utilizzo consapevole. Dunque, il paragrafo 2 dello stesso articolo prevede che «[i] sistemi di IA ad alto rischio sono accompagnati da istruzioni per l'uso in un formato digitale o non digitale appropriato, che comprendono informazioni concise, complete, corrette e chiare che siano pertinenti, accessibili e comprensibili per i *deployer*», demandando al seguente paragrafo 3 l'individuazione del contenuto minimo delle istruzioni per l'uso.

(22) Elemento funzionale alla trasparenza, come evidenziato dal *Considerando 27* nella parte in cui enuncia che «(…) [c]on "trasparenza" si intende che i sistemi di IA sono sviluppati e utilizzati in modo da consentire un'adeguata tracciabilità e spiegabilità, rendendo gli esseri umani consapevoli del fatto di comunicare o interagire con un sistema di IA e informando debitamente i *deployer* delle capacità e dei limiti di tale sistema di IA e le persone interessate dei loro diritti (…)».

(23) A norma dell'articolo 18 dell'AI Act.

(24) A norma del paragrafo 2 dell'articolo 15.

(25) Articolo 14, paragrafo 4, lettera e).

pre-emptive remedy (rimedio preventivo ed ingiunzione dinamica) le certificazioni sono rimesse alle stesse società di sviluppatori. Siamo sicuri che in tal modo i soprarichiamati concetti di *safety* e *security* saranno salvaguardati o prevarranno le logiche del profitto? Per ora possiamo solo dire ai posteri l'ardua sentenza se è vero che dovrà prevalere la cd. "*teoria della competenza specifica del rischio*" in una logica Gaussiana e nella logica dell'insegnamento di Lessig (26) della Nuova Scuola di Chicago che vede una distinzione tra diritto, norme eterogiuridiche, mercato e architettura. In tale quadro problematica sarà la definizione del dato sintetico secondo la metodologia GANS (*Generative Adversarial Network*).

I sistemi di AI per uso generale.

Il testo regola i sistemi di AI per uso generale (i cd. *general purpose AI models*, o modelli GPAI) (27), in grado di svolgere compiti diversi (come creare un testo o un'immagine) (28) e allenati attraverso un'enorme mole di dati (i cd. *Big Data*) (29) non categorizzati, come GPT-4, alla base del potente chatbot ChatGPT, o LaMDA, dietro Google Bard. Gli sviluppatori devono assicurarsi che i contenuti siano marcati in un sistema leggibile da una macchina e siano riconoscibili come generati da un'AI (30). Un utente deve sapere se

(26) L. Lessig, *The New Chicago School*, 27 The Journal of Legal Studies 661 (1998).

(27) Nozione di modello GPAI che è definita dall'articolo 3, paragrafo 1, punto 63 come «(…) un modello di IA, anche laddove tale modello di IA sia addestrato con grandi quantità di dati utilizzando l'autosupervisione su larga scala, che sia caratterizzato da una generalità significativa e sia in grado di svolgere con competenza un'ampia gamma di compiti distinti, indipendentemente dalle modalità con cui il modello è immesso sul mercato, e che può essere integrato in una varietà di sistemi o applicazioni a valle, ad eccezione dei modelli di IA utilizzati per attività di ricerca, sviluppo o prototipazione prima di essere immessi sul mercato».

(28) La cd. Intelligenza Artificiale Generativa o, nella denominazione anglosassone, *Generative Artificial Intelligence.*

(29) Espressione che indica una particolare confluenza di volume, velocità e varietà di informazioni, così P. Zikopoulos, *Harness the Power of Big Data: The IBM Big Data Platform*, McGraw Hill Professional, 2012, p. 61. In relazione al rapporto tra IA e *big data*, è stato sostenuto che maggiore è la velocità, l'ampiezza e la varietà dei dati immessi nel sistema, maggiore saranno le capacità e la qualità del sistema. In tal senso, A. Guadamuz, *A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence inputs and outputs*, 2 GRUR International (di prossima pubblicazione, 2024), disponibile su SSRN: *https://ssrn.com/abstract=4371204* o *http://dx.doi.org/10.2139/ssrn.4371204*, p. 3. Questa conclusione è confermata dal fatto che all'impiego di una più vasta scala di dati corrisponde una riduzione della cd. "*loss function*" (la funzione di perdita), la quale determina l'entità di errore del modello di IA (così J. Drexl, R.M. Hilty, F. Beneke, L. Desaunettes, M. Finck, J. Globocnik, B. Gonzalez Otero, J. Hoffmann, L. Hollander, D. Kim, H. Richter, S. Scheuerer, P.R. Slowinski, J. Thonemann (gruppo di ricerca per la regolazione dell'economia digitale del *Max Plank Institute for Innovation and Competition*), *Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective*, versione 1.0, 2019, disponibile su *https://ssrn.com/abstract=3465577*, p. 7).

(30) La scelta del sistema del cd. *watermarking* è stata condivisa anche da Cina e USA. Con particolare riferimento alla Cina, è stato precisato che sono previste tre forme di *watermark*: esplicito, implicito e per apposizione di metadati. Il primo consiste nell'indicazione della provenienza macchinica del

sta interagendo con un chatbot. E i contenuti deepfake devono essere etichettati come tali (attraverso sistemi come il *watermarking*, la filigrana digitale applicata a foto o video) (31). Previsioni che, tuttavia, non è detto siano sufficienti a impedire la diffusione di *fake news* (32). Unica eccezione: l'impiego di questi sistemi per perseguire reati.

Il regolamento fissa una soglia per identificare i sistemi ad alto impatto, che hanno maggiori effetti sulla popolazione e perciò devono rispettare obblighi più stringenti. L'articolo 51 dell'AI Act, sotto la rubrica "Classificazione del modelli di IA per finalità generali come modelli per finalità generali con rischio sistemico", definisce al paragrafo 1 i criteri di qualificazione dei mo-

---

contenuto in un formato leggibile agli esseri umani. Il secondo si realizza tramite l'apposizione di "etichette" impercettibili all'occhio umano, sulle quali risulti quantomeno il nome del *provider* del sistema di IA utilizzato per produrre il contenuto. L'ultimo tipo di *watermark* è imposto solo nei casi in cui il contenuto sia salvato in forma di *file*. In tal senso, v. P. Henderson, *Should the United States or the European Union follow China lead and require watermarks for generative AI?*, in *Georgetown Journal of International Affairs*, 24 maggio 2023, disponibile presso *https://gjia.georgetown.edu/2023/05/24/should-the-united-states-or-the-european-union-follow-chinas-lead-and-require-watermarks-for-generative-ai/.*

(31) Una simile previsione trova giustificazione nel *Considerando 133*, il quale enuncia quanto segue: «[d]iversi sistemi di IA possono generare grandi quantità di contenuti sintetici, che per gli esseri umani è divenuto sempre più difficile distinguere dai contenuti autentici e generati da esseri umani. L'ampia disponibilità e l'aumento delle capacità di tali sistemi hanno un impatto significativo sull'integrità e sulla fiducia nell'ecosistema dell'informazione, aumentando i nuovi rischi di cattiva informazione e manipolazione su vasta scala, frode, impersonificazione e inganno dei consumatori. Alla luce di tali impatti, della rapida evoluzione tecnologica e della necessità di nuovi metodi e tecniche per risalire all'origine delle informazioni, è opportuno imporre ai fornitori di tali sistemi di integrare soluzioni tecniche che consentano agli *output* di essere marcati in un formato leggibile meccanicamente e di essere rilevabili come generati o manipolati da un sistema di IA e non da esseri umani. Tali tecniche e metodi dovrebbero essere sufficientemente affidabili, interoperabili, efficaci e solidi nella misura in cui ciò sia tecnicamente possibile, tenendo conto delle tecniche disponibili o di una combinazione di tali tecniche, quali filigrane, identificazioni di metadati, metodi crittografici per dimostrare la provenienza e l'autenticità dei contenuti, metodi di registrazione, impronte digitali o altre tecniche, a seconda dei casi. Nell'attuare tale obbligo, i fornitori dovrebbero tenere conto anche delle specificità e dei limiti dei diversi tipi di contenuti e dei pertinenti sviluppi tecnologici e di mercato nel settore, come rispecchia lo stato dell'arte generalmente riconosciuto. Tali tecniche e metodi possono essere attuati a livello di sistema o a livello di modello, compresi i modelli di IA per finalità generali che generano contenuti, facilitando in tal modo l'adempimento di tale obbligo da parte del fornitore a valle del sistema di IA. Per continuare a essere proporzionato, è opportuno prevedere che tale obbligo di marcatura non debba riguardare i sistemi di IA che svolgono principalmente una funzione di assistenza per l'*editing standard* o i sistemi di IA che non modificano in modo sostanziale i dati di *input* forniti dal *deployer* o la rispettiva semantica».

(32) Circa i limiti presentati dal sistema di *watermarking*, sono stati evidenziati i seguenti profili problematici: il fatto che l'implementazione tecnica di queste filigrane sia rimessa integralmente a carico dei fornitori dei sistemi di IA generativa, dovendo dunque affrontare questi ultimi le difficoltà esecutive (tra le quali rientra il dato per il quale l'apposizione di un *watermark* tramite una determinata tecnologia possa rendere il marchio non leggibile tramite un altro sistema); la sussistenza di una percentuale di errore nel rilevamento del marchio apposto (determinando dunque la possibilità di falsi positivi nel processo di controllo circa l'origine macchinica del contenuto); dubbi circa la robustezza di questo sistema, dovuti alla possibilità di una manipolazione, alterazione o rimozione dei marchi apposti tramite il *watermarking.* In tal senso, Parlamento Europeo, *Generative AI and watermarking*, del 13 dicembre 2023, *https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/757583/EPRS_BRI(2023)757583_EN.pdf*, p. 3.

delli di IA con scopi generali portatori di rischio sistemico (33), nelle lettere a) e b). La lettera a) fa riferimento all'elevata capacità di impatto valutata in base a strumenti tecnici adeguati e a indicatori e parametri di riferimento (34). La lettera b) riconosce alla Commissione il potere di adottare una decisione, *ex officio* o su qualificata segnalazione dello *Scientific Panel*, con la quale assoggetta uno specifico modello alla relativa disciplina poiché si ritiene che abbia capacità o un impatto equiparabile a quello di cui alla lettera a) (35).

Il successivo paragrafo 2 prevede invece una presunzione di classificazione come *general purpose AI model with systemic risk* ove il modello raggiunga o superi una determinata soglia ivi fissata. Il valore, come dichiarato a dicembre, è un potere di calcolo pari a 10^25 FLOPs (*floating point operations per second*, un'unità di misura della capacità computazionale) (36). Al momento, solo GPT-4 di OpenAI, Gemini di Google e qualche modello cinese rispetterebbero questa caratteristica. Ma dovranno essere gli sviluppatori a comunicarlo alla Commissione, che per adesso non si esprime sui modelli già nei radar dell'AI Act e potrà intervenire se viene a sapere che un sistema ad alto impatto non si è dichiarato tale. Da Bruxelles fanno sapere che la soglia potrà essere modificata in futuro, per meglio rispondere alle evoluzioni di mercato (37).

Le AI ad alto impatto dovranno applicare *ex ante* delle regole su sicurezza informatica, trasparenza dei processi di addestramento e condivisione della documentazione tecnica prima di arrivare sul mercato (38). Al di sotto si collocano

---

(33) La nozione di "rischio sistemico" è definita invece dall'articolo 3, paragrafo 1, punto 65 come un rischio «(…) specifico per le capacità di impatto elevato dei modelli di IA per finalità generali, avente un impatto significativo sul mercato dell'Unione a causa della sua portata o di effetti negativi effettivi o ragionevolmente prevedibili sulla salute pubblica, la sicurezza, i diritti fondamentali o la società nel suo complesso, che può propagarsi su larga scala lungo l'intera catena del valore; (…)».

(34) «1. Un modello di IA per finalità generali è classificato come modello di IA per finalità generali con rischio sistemico se soddisfa uno dei seguenti requisiti: a) presenta capacità di impatto elevato valutate sulla base di strumenti tecnici e metodologie adeguati, compresi indicatori e parametri di riferimento; (…)».

(35) «(…) b) sulla base di una decisione della Commissione, *ex officio* o a seguito di una segnalazione qualificata del gruppo di esperti scientifici, presenta capacità o un impatto equivalenti a quelli di cui alla lettera a), tenendo conto dei criteri di cui all'allegato XIII».

(36) «2. Si presume che un modello di IA per finalità generali abbia capacità di impatto elevato a norma del paragrafo 1, lettera a), quando l'importo cumulativo del calcolo utilizzato per il suo addestramento misurato in FLOP è superiore a 10^25».

(37) Infatti, il paragrafo 3 dell'articolo 51 prevede che la Commissione possa adottare «(…) atti delegati a norma dell'articolo 97 per modificare le soglie di cui ai paragrafi 2 e 3, nonché per integrare parametri di riferimento e indicatori alla luce degli sviluppi tecnologici in evoluzione, quali miglioramenti algoritmici o una maggiore efficienza dell'*hardware*, ove necessario, affinché tali soglie riflettano lo stato dell'arte».

(38) A norma dell'articolo 55, paragrafo 1 il quale dispone ulteriori obblighi, oltre a quelli previsti dal precedente articolo 53 per tutti i modelli GPAI. Inoltre, il paragrafo 2 dello stesso articolo 55 prevede la possibilità, riconosciuta ai *provider* dei modelli GPAI portatori di rischio sistemico, di fare affidamento su codici di condotta (rimandando al seguente articolo 56), adeguandosi ai quali si può dimostrare la *compliance* alla disciplina in questione.

tutti gli altri *foundational models.* Tra cui le due startup *made in Europe*: la francese Mistral e la tedesca Aleph Alpha. In questo caso l'AI Act scatta quando gli sviluppatori commercializzano i propri prodotti. E sono esclusi i modelli offerti con licenza *open source* a norma del *Considerando 104*, a meno che non siano qualificabili come modelli GPAI portatori di rischi sistemici.

Il sistema degli uffici, innovazioni e controllo.

L'AI Act delega una serie di controlli alle autorità locali, che entro due anni dall'entrata in vigore dovranno istituire almeno una *sandbox* regolatoria (o spazio di sperimentazione normativa per l'IA) (39) a livello nazionale (40). Ossia uno schema che consente di effettuare test in sicurezza, in deroga alla legge, per non soffocare l'innovazione a causa dei troppi obblighi da rispettare e sostenere l'addestramento di algoritmi, anche con test condotti nel mondo reale.

La Commissione si doterà di un Consiglio dell'AI, dove siede un esponente per ogni Stato dell'Unione. Il Garante europeo dei dati personali è invitato come osservatore, così come l'Ufficio dell'AI collocato sotto la Direzione generale Connect. Il Consiglio sarà strutturato in due sotto-gruppi, uno dedicato alla sorveglianza del mercato e uno alle notifiche delle autorità, e assisterà la Commissione nell'implementazione del regolamento, nell'identificazione di tendenze tecnologiche da monitorare e nella modifica delle norme. A sua volta il Consiglio dell'AI sarà affiancato da un forum di consulenti tecnici, mentre la Commissione potrà avvalersi di un comitato indipendente di scienziati ed esperti, sulla falsariga del gruppo di tecnici del clima che aiuta l'Onu nella regia delle politiche ambientali.

L'AI Act stabilisce le modalità per segnalare un incidente occorso a un sistema ad alto rischio, entro due giorni da quando avviene o se ne ha contezza. La violazione degli obblighi di *compliance* posti dall'AI Act determinerà l'applicazione di sanzioni pecuniarie. L'articolo 99 del regolamento, sotto la rubrica "Sanzioni", prevede al primo paragrafo che gli Stati membri dispongano

(39) Definita dall'articolo 3, paragrafo 1, punto 55 come «(…) un quadro controllato istituito da un'autorità competente che offre ai fornitori o potenziali fornitori di sistemi di IA la possibilità di sviluppare, addestrare, convalidare e provare, se del caso in condizioni reali, un sistema di IA innovativo, conformemente a un piano dello spazio di sperimentazione per un periodo di tempo limitato sotto supervisione regolamentare».

(40) Infatti, il Capo VI dell'AI Act (intitolato "Misure a sostegno dell'innovazione") si apre con l'articolo 57 il quale, sotto la rubrica "Spazi di sperimentazione normativa per l'IA" sancisce quanto segue: «1. Gli Stati membri provvedono affinché le loro autorità competenti istituiscano almeno uno spazio di sperimentazione normativa per l'IA a livello nazionale, che sia operativo entro ... [24 mesi dalla data di entrata in vigore del presente regolamento]. Tale spazio di sperimentazione può essere inoltre istituito congiuntamente con le autorità competenti di uno o più Stati membri. La Commissione può fornire assistenza tecnica, consulenza e strumenti per l'istituzione e il funzionamento degli spazi di sperimentazione normativa per l'IA. L'obbligo di cui al primo comma può essere soddisfatto anche partecipando a uno spazio di sperimentazione esistente nella misura in cui tale partecipazione fornisca un livello equivalente di copertura nazionale per gli Stati membri partecipanti».

un quadro normativo sanzionatorio che miri a realizzare un *enforcement* della disciplina dell'AI Act (41). Chi non si adegua all'AI Act rischia multe fino a 35 milioni di euro o al 7% del fatturato globale nel caso degli usi proibiti (ex articolo 99, paragrafo 3). Se a finire sotto la scure sono i sistemi ad alto rischio o quelli di uso generale, si arriva fino a un massimo di 15 milioni o del 3% del fatturato globale in caso di mancata ottemperanza alle regole (a norma dell'articolo 99, paragrafo 4). Altrimenti, se si contestano informazioni scorrette, la sanzione raggiunge un tetto di 7,5 milioni di euro o dell'1% del fatturato globale (secondo quanto disposto dall'articolo 99, paragrafo 5). L'articolo 99, al paragrafo 6 precisa inoltre che, nel caso in cui ad essere assoggettate a sanzione siano piccole o medie imprese o *start-up*, il valore da prendere in considerazione ai fini dell'irrogazione sia quello più basso tra i due indicati per ciascuna delle tre fasce appena richiamate (42).

Questa disciplina riuscirà a dimostrarsi a prova di futuro? L'approccio regolatorio europeo sarà idoneo a far fronte ai futuri sviluppi, ovvero manca qualcosa, a questo fine, in tale regolamento? L'Europa rischia di diventare un gigante regolatorio e un nano tecnologico? La tecnica da seguire, forse, era quella del *Foresight* oltre alla prevista *sandbox*?

Ebbene, nel panorama globale contemporaneo, in cui il progresso tecnologico avanza con una velocità sempre crescente, la "fluidità" e la "dinamicità" dell'innovazione prendono in contropiede il diritto (43). Gli operatori giuridici sono infatti abituati oramai a ragionare secondo approcci reattivi all'insorgere di problematiche che via via si pongono. Tuttavia, questo tipo di metodologia può oggi comportare rischi di inefficienza del loro intervento (44). Il *Foresight*

---

(41) «1. Nel rispetto dei termini e delle condizioni di cui al presente regolamento, gli Stati membri stabiliscono le regole relative alle sanzioni e alle altre misure di esecuzione, che possono includere anche avvertimenti e misure non pecuniarie, applicabili in caso di violazione del presente regolamento da parte degli operatori, e adottano tutte le misure necessarie per garantirne un'attuazione corretta ed efficace, tenendo conto degli orientamenti emanati dalla Commissione a norma dell'articolo 96. Le sanzioni previste sono effettive, proporzionate e dissuasive. Esse tengono conto degli interessi delle PMI, comprese le *start-up*, e della loro sostenibilità economica».

(42) «6. Nel caso delle PMI, comprese le *start-up*, ciascuna sanzione pecuniaria di cui al presente articolo è pari al massimo alle percentuali o all'importo di cui ai paragrafi 3, 4 e 5, se inferiore». Si tratta evidentemente di una previsione tesa a non impattare in misura sproporzionata su aziende che stanno cercando di farsi strada nel settore in questione, al fine di non minarne eccessivamente la concorrenza.

(43) In specifico riferimento all'IA, si pensi alle numerose richieste di sospendere l'attività di ricerca e di sviluppo dei sistemi di intelligenza artificiale avanzate lo scorso anno (in particolare quelle di Samuel Altman, Elon Musk o del Future of Life Institute). Si tratta di tre posizioni accomunate dalla paura della rapidità dell'evoluzione tecnologica che caratterizza la cd. "quarta rivoluzione industriale".

(44) A. Cataleta, S. Leucci, G. Rizzo, G. Vacagio, *Privacy, anticipiamo il futuro: un nuovo approccio è necessario*, *Agenda Digitale*, 20 settembre 2023, disponibile su *https://www.agendadigitale.eu/sicurezza/privacy/i-futuri-della-privacy-perche-serve-un-approccio-anticipante-alle-nuove-tecnologie/*.

si presenta dunque come uno strumento di primaria importanza nel ragionamento su cui si fondano le decisioni di politica legislativa, essendo una disciplina che propone l'adozione di un approccio proattivo teso a prevedere la maggiore varietà possibile di scenari futuri, in modo tale da comprendere anticipatamente come farvi fronte (45). Questo metodo, implicando la proiezione del maggior numero possibile di accadimenti ipotetici futuri, è divenuto una tecnica sempre più multidisciplinare, in particolar modo dalla fine dello scorso secolo (46). Queste considerazioni preliminari sono già utili per riflettere circa l'opportunità, sfortunatamente persa dal legislatore sovranazionale, di adottare il *foresight* come metodo regolatorio in ambito di IA.

Tuttavia, la consapevolezza, da un lato, dei limiti propri dei "canonici" approcci regolatori e, dall'altro, dei vantaggi che il *foresight* può comportare, ha raggiunto alcune Istituzioni, le quali hanno già iniziato a riconoscere a questa disciplina il rilievo che merita. Si intende fare specifico riferimento a tre autorità, due nazionali e una sovranazionale: la *Commission Nationale de l'Informatique et des Libertès* (o CNIL, autorità nazionale francese), la *Information Commissioner's Office* (o ICO, autorità inglese) e l'*European Data Protection Supervisor* (o EDPS, autorità europea) (47). Ciò che accomuna questi soggetti è l'adozione del *Foresight* come metodo di osservare al futuro, per comprendere quali siano le migliori strategie regolatorie da attuare nel presente.

La centralità del *Foresight* è stata evidenziata dall'EDPS, il quale ha sostenuto la centralità di questo metodo in base ad alcune constatazioni in rela-

---

(45) *ibidem*, ove viene inoltre precisato che il *Foresight* è una tecnica che «(…) non usa il futuro come obiettivo da raggiugere, ma piuttosto come un costrutto "usa e getta", il cui solo scopo è di far accedere ad una comprensione più ampia delle decisioni che sono importanti nel presente». Nello stesso articolo viene inoltre accennata l'origine del metodo del *Foresight*. Si tratta di un approccio, un metodo ideato e sviluppato a partire dalla metà dello scorso secolo. La principale motivazione che comportò la nascita di questo modo di osservare all'avvenire fu la necessità di mantenere la pace. In quest'ottica, si percepì che "progettazione" e "pianificazione" sono metodi «(…) utili, ma limitati di pensare al futuro, perché mirano a conoscere un singolo futuro». Si avvertì dunque il bisogno di pensare al futuro come insieme di pluralità di possibili scenari, i quali non devono essere conosciuti, bensì usati in ottica proattiva, di anticipazione nell'adozione delle decisioni strategiche nel presente. In questa prospettiva assume dunque rilevanza primaria l'individuazione delle esigenze future, in tal senso v. D. Mietzner, G. Reger, *Advantages and Disadvantages of Scenario Approaches for Strategic Foresight*, *1 International Journal Technology Intelligence and Planning 220* (2005), disponibile su *SSRN: https://ssrn.com/abstract=1736110*, p. 235.

(46) A. Cataleta, S. Leucci, G. Rizzo, G. Vacagio, *Privacy, anticipiamo il futuro: un nuovo approccio è necessario*, cit., ove si sottolinea che il *foresight* ha finito per inglobare elementi di sociologia, psicologia, economia e scienze ambientali. Una simile evoluzione della disciplina del *foresight* è giustificata, secondo i richiamati Autori, dal fatto che le sfide cui ci si trova oggi a dover far fronte (quali, ad esempio, il cambiamento climatico, le disuguaglianze sociali ed economiche e l'innovazione tecnologica) pongono l'esigenza di adottare un approccio interdisciplinare, essendo divenuto oramai obsoleto qualsiasi tentativo di intervento fondato di ragionamenti e valutazioni effettuate a compartimenti stagni.

(47) *ibidem.*

zione al contesto normativo europeo: in primo luogo il Regolamento (UE) 2018/1725 (48) richiede espressamente che l'EDPS monitori gli sviluppi pertinenti all'impatto sui dati personali, con specifico riferimento alle nuove tecnologie dell'informazione e della comunicazione (49); inoltre, il monitoraggio della tecnologia e il *Foresight* sono palesemente collegati al ruolo di autorità di controllo svolta dallo stesso EDPS (50). Lo stesso EDPS ha precisato che lo scopo strategico che persegue è quello di anticipare il più possibile i futuri sviluppi del progresso tecnologico, e a tal fine il *Foresight* rappresenta uno strumento assolutamente indispensabile (51). L'ICO ha dichiarato l'intenzione di adottare il metodo del *Foresight* in relazione a 65 tecnologie emergenti, selezionate sulla base di un coefficiente di priorità che esprima una classificazione delle stesse in base a "*probability*", "*scale*", e "*associated harms and benefits*" in riferimento alla normativa sulla *privacy* (52). La CNIL ha istituito al proprio interno un apposito comitato (il cd. *Foresight Committee*), composto di 21 esperti con profili differenti al fine di implementare e arricchire il dibattito circa l'etica digitale (53). La giustificazione a fondamento della costituzione di un simile organo è stata individuata nella necessità di mostrare una maggiore apertura ad un lavoro coordinato con il mondo dell'innovazione tecnologica (54).

L'utilità del *foresight* in relazione alle nuove tecnologie si coglie dunque se si prova a riflettere sulla considerazione per la quale la proiezione e la comprensione dei possibili scenari futuri sono fattori imprescindibili, qualora si volessero individuare le modalità migliori per perseguire gli scopi che ci si

(48) Regolamento (UE) 2018/1725 del Parlamento Europeo e del Consiglio del 23 ottobre 2018 sulla tutela delle persone fisiche in relazione al trattamento dei dati personali da parte delle istituzioni, degli organi e degli organismi dell'Unione e sulla libera circolazione di tali dati, e che abroga il Regolamento (CE) n. 45/2001 e la Decisione n. 1247/2002/CE.

(49) X. Lareo, *Continous improvement process*, in *European Data Protection Supervisor*, Techsonar 2023-2024 Report, novembre 2023, su *https://www.edps.europa.eu/data-protection/our-work/publications/reports/2023-12-04-techsonar-report-2023-2024_en*, p. 1.

(50) *ibidem.*

(51) «*The aim is to anticipate as far as possible future technology trends and the privacy and data protection challenges posed by new technologies*», *ibidem.*

(52) Information Commissioner's Office, *Helping people understand how new technologies interact with the UK's data protection framework. Tech Horizon Report*, dicembre 2022, disponibile su *https://ico.org.uk/media/about-the-ico/documents/4023338/ico-future-tech-report-20221214.pdf*, p. 9. È stato sostenuto che, nel documento appena richiamato, «(…) l'ICO ha dato forma ad alcuni scenari futuri nel mondo di alcune tecnologie particolarmente invasive per analizzarne meglio le evoluzioni e gli impatti sulla protezione dei dati», così A. Cataleta, S. Leucci, G. Rizzo, G. Vacagio, *Privacy, anticipiamo il futuro: un nuovo approccio è necessario*, cit.

(53) Commission nationale de l'informatique et des libertès, *Data footprint and freedoms. Exploring the overlaps between data protection freedoms and the environment*, IP Reports Innovation and Foresight n. 9, giugno 2023, disponibile su *https://linc.cnil.fr/sites/linc/files/2023-09/cnil_ip9_data_footprint_and_freedoms.pdf*, p. 65.

(54) *ibidem.*

prefigge. Una simile conclusione è da ritenersi senz'altro applicabile anche all'intelligenza artificiale (55). Infatti, è stato sottolineato che, nella seppur inesorabile impossibilità di prevedere con certezza come la tecnologia si evolverà, resta ferma comunque la possibilità di informarsi sin da ora, al fine di prepararsi ed essere pronti ai possibili sviluppi futuri (56). A parere di chi scrive, un simile atteggiamento dovrebbe ritenersi doveroso nell'ambito della regolamentazione dell'IA, quantomeno in considerazione dei possibili risvolti che una simile tecnologia potrà avere (57).

In conclusione, si può sin da ora affermare che il legislatore sovranazionale, accanto alla previsione di principi etici (58) e delle cd. *regulatory sandboxes*, avrebbe dovuto attuare con maggiore sicurezza un approccio teso alla previsione e all'anticipazione (caratteristiche proprie del *foresight*) degli scenari futuri di medio e lungo periodo, onde evitare di adottare una regolamentazione che rischia di presentarsi già nel breve termine obsoleta.

In tutta questa situazione resta comunque aperta (e auspicabile) la possibilità che le autorità costituite a livello europeo dal regolamento e quelle nazionali che dovranno essere autonomamente formate dai singoli Stati membri vadano "in controtendenza", se così si può dire, rispetto all'AI Act sotto questo punto di vista, seguendo l'esempio dell'EDPS (il quale, come si è avuto modo di dare atto, ha già riconosciuto l'essenzialità di adottare un approccio di *foresight* non solo in relazione alla tutela dei dati personali, bensì anche in riferimento all'intelligenza artificiale). L'auspicio che si verifichi un simile sviluppo risulta ancor più plausibile ove si consideri l'obiettivo principe che si intende perseguire attraverso la regolamentazione dell'intelligenza artificiale: il mantenimento della centralità dell'essere umano, della sua dignità e della sua libertà di autodeterminazione. È assolutamente cruciale, in questo preciso momento storico, porre delle limitazioni, degli argini allo sviluppo

---

(55) È stato infatti precisato che «[t]*he rapidly evolving technological landscape requires us to anticipate new technological challenges, to be able to influence their evolution and use. A clear signal in this direction is the increasing pace of deployment of artificial intelligence in everyday life and machine learning applications, which requires a more proactive and anticipatory attitude towards an appropriate and effective governance of technology.* (…) *the need to become proactive in our relationship with technology has become increasingly compelling and has led us to start our foresight journey*», W. WIEWIÒROWSKI, *Readiness and adaptability in an evolving technology landscape*, in EUROPEAN DATA PROTECTION SUPERVISOR, *Techsonar 2023-2024 Report*, cit.

(56) *ibidem*, ove si sottolinea inoltre che il metodo di *foresight* adottato dall'EDPS si coniuga con un approccio "*risk-based*" (lo stesso adottato dall'AI Act), rivolgendo maggiori attenzioni alle tecnologie che impattano maggiormente sui diritti di *privacy* e di protezione dei dati degli individui.

(57) Si pensi al grado di autonomia con il quale i contemporanei sistemi di IA sono capaci di operare e migliorarsi in continuazione, attraverso le tecniche di *deep learning*. Una simile autosufficienza nello sviluppo delle macchine comporta il rischio che il loro controllo possa sfuggire all'essere umano, determinando la possibilità che si verifichino scenari potenzialmente pregiudizievoli nei riguardi dei diritti fondamentali.

(58) Tra i quali rientrano affidabilità, trasparenza, robustezza, *cybersecurity*, *data governance*, *privacy* e *accountability*.

tecnologico, senza però frenarlo. Occorre indirizzarlo verso un ideale di "umanesimo tecnologico" (59), nel quale sia ben chiaro il ruolo della macchina come strumento posto a servizio dell'essere umano, senza sostituirsi a quest'ultimo (60). Considerando anche i molteplici campi di applicazione dell'intelligenza artificiale, il *foresight*, tramite la sua interdisciplinarità, risulta a maggior ragione una strada più che valida da seguire nel tentativo di regolamentare adeguatamente il fenomeno in questione già da oggi, nell'ottica di ciò che sarà.

(59) Così come proposto in E. Battelli, *Necessità di un umanesimo tecnologico*, in *Diritto di Famiglia e delle Persone*, 3, 1096, 2022.

(60) E. Battelli, *Necessità di un umanesimo tecnologico*, cit., pp. 1107 e ss.