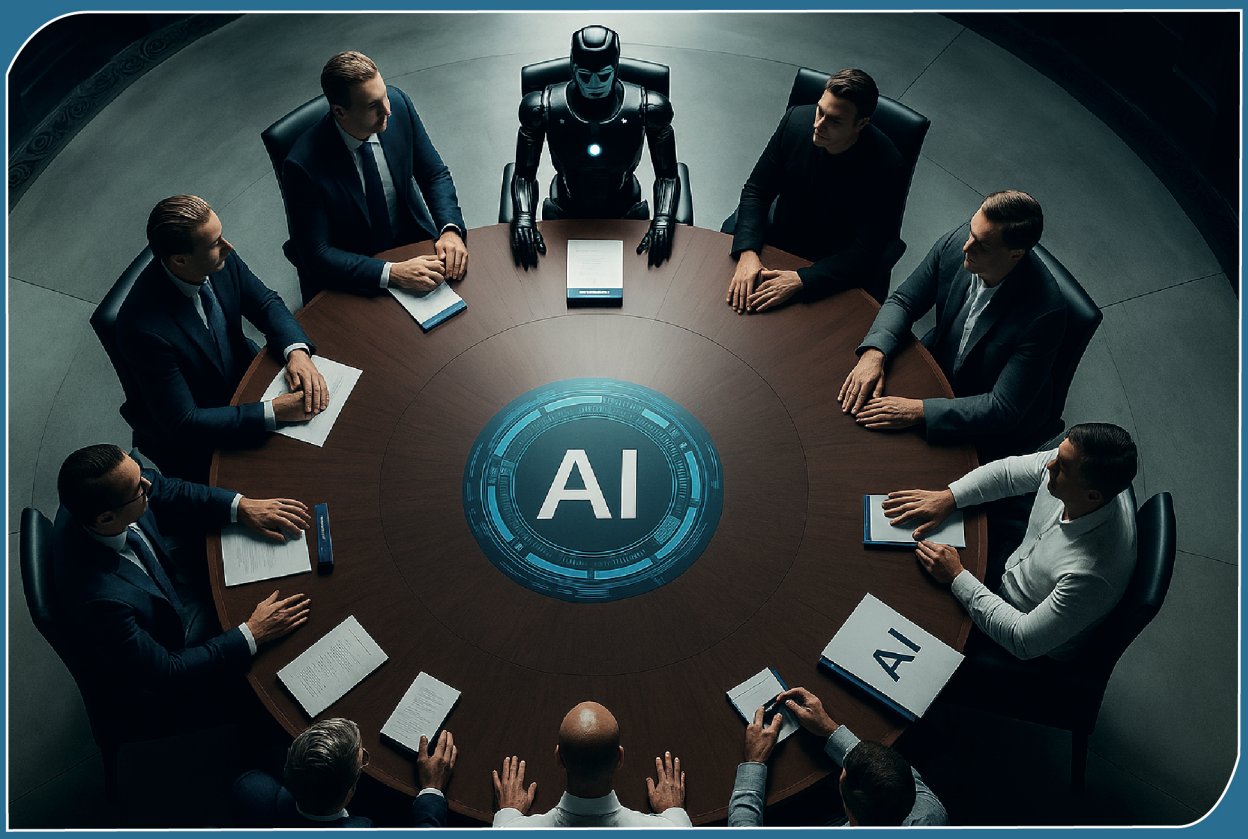
Consiglio Nazionale delle Ricerche
IRISS Istituto di Ricerca su Innovazione e Servizi per lo Sviluppo

Studi e Ricerche per l'Innovazione
Collana Consiglio Nazionale delle Ricerche
diretta da Paolo Landri
- 6 -

Dialoghi sull'intelligenza artificiale

Atti dei Seminari del CNR-IRISS (marzo-maggio 2024)

A cura di: Natale Rampazzo



edizioni
Consiglio Nazionale delle Ricerche

a cura di
Natale Rampazzo

Dialoghi sull'intelligenza artificiale

Atti dei Seminari del CNR-IRISS

(marzo-maggio 2024)

Roma
CNR Edizioni 2025

Coordinatrice e rapporti con la casa editrice
Maria Grazia Spronati
Progetto grafico, impaginazione ed editing
Antonio Marino
Copertina
Angela Petrillo

Foto di copertina: ChatGPT su prompt di Natale Rampazzo

Pubblicato da
Cnr Edizioni 2025
P.le Aldo Moro, 7 - 00185 Roma
www.edizioni.cnr.it
bookshop@cnr.it
ISBN: 978-88-8080-766-7 versione elettronica
Consiglio Nazionale delle Ricerche (CNR)
Istituto di Ricerca su Innovazione e Servizi per lo Sviluppo (IRISS)

Studi e Ricerche per l'Innovazione

Collana del

Consiglio Nazionale delle Ricerche
Istituto di Ricerche su Innovazione e Servizi per lo Sviluppo

diretta da

Paolo Landri

- 6 -

Il curatore desidera ringraziare gli Autori che hanno partecipato ai seminari rendendo possibile la pubblicazione di questo volume, il Comitato scientifico della Collana e i colleghi del CNR-IRISS: Maria Grazia Spronati, Angela Petrillo e Antonio Marino per il prezioso lavoro svolto durante tutto il processo editoriale.

INDICE

INTELLIGENZA ARTIFICIALE TRA PASSATO REMOTO E FUTURO PROSSIMO

Saggio introduttivo
di Natale Rampazzo (CNR-IRISS)

1. Sul progetto

Nella nostra epoca nulla è più importante del tempo. E della tempestività. Specie delle pubblicazioni scientifiche. In particolare, in materie soggette ad accelerazioni spesso imprevedibili. Come l'intelligenza artificiale. Singolare dunque la presente edizione di contributi multidisciplinari che costituiscono la ricaduta tangibile, direi cartacea (se non fosse digitale) di una serie di discorsi, *rectius* di Dialoghi, tenuti l'anno scorso nel periodo marzo-maggio 2024 da studiosi ed esperti di varia provenienza, chiamati a valutare l'impatto potenziale sulla società dell'implementazione estesa di agenti artificiali in ogni attività quotidiana.

Come in un laboratorio aperto, nei Dialoghi si sono alternati psicologi, ingegneri, linguisti, giuristi, economisti pronti a condividere il proprio punto di vista su un futuro in costante e progressiva, inarrestabile formazione. Li ringrazio per aver accettato la sfida di confrontarsi su argomenti da alcuni trattati da anni e forse decenni, eppure sempre nuovi nella misura in cui cambiavano e cambiano, continuamente, il contesto di riferimento, il background, il rumore di fondo. E ringrazio ancora i relatori che hanno fatto l'ulteriore sforzo di dare una veste grafica alle parole, ma anche i colleghi che, non avendo potuto presentare una comunicazione, hanno accolto l'offerta di uno spazio di circolazione delle loro idee.

Il fattore tempo, dicevo, è fondamentale, eppure i tempi della riflessione umana sono inevitabilmente più lunghi di quelli richiesti da un 'instant book' o dal susseguirsi di seminari e tavole rotonde sul tema. Certamente utili per il

contributo che forniscono alla costruzione di una conoscenza più estesa e dunque ad una migliore comprensione e consapevolezza del fenomeno – specie in un ambiente multidimensionale – almeno da parte degli 'addetti ai lavori', ma che pur non si sono sottratti all''effetto eco' (con la minuscola), una volta raggiunta la soglia non superabile dell'analisi plurifattoriale dell'impatto sociale, economico, giuridico della presenza di elementi non umani nei processi decisionali e operativi. Così capita con i 'fashion topics' (negli ultimi anni, ad esempio, il metaverso), che determinano una profluvie di contributi[1] in grado di inaridire (paradossalmente…) ogni spunto di ricerca, mentre fatalmente ne va svanendo l'oggetto, non perché in sé effimero, bensì per una sua materiale (ma temporanea) inagibilità (soprattutto per una carenza, per motivi tecnici ed economici, di dispositivi abilitanti).

2. Un tentativo (fallito) di definizione sociotecnica

L'intelligenza artificiale[2] (in seguito, semplicemente la «IA») non è giovane.

Essa avrebbe, in base alle diverse stime possibili, intorno ai 70 anni, rispettivamente dalla pubblicazione del contributo di Alan M. Turing, *Computing Machinery and Intelligence*, nella rivista *Mind* 59, 1950, nel quale veniva codificato un test di equivalenza dell'intelligenza delle macchine a quella umana (*Turing test* o *Imitation game*) oppure dalla celebrazione del seminario tenutosi, su iniziativa di John McCarthy – insieme con Marvin L. Minsky (allora *fellow* presso la Harvard University, successivamente al MIT), Claude E. Shannon ("padre della teoria dell'informazione", Bell Telephone Laboratories e successivamente al MIT) e Nathaniel Rochester (IBM Corporation) – nell'estate del 1956 presso il Dartmouth College in New Hampshire, e destinato, in virtù del 'manifesto' del 31 agosto 1955, alla discussione sulle abilità emulative di una macchina.

Le questioni intorno alle quali ruotava il dibattito di quel periodo potrebbero sintetizzarsi nella consapevolezza di uno stato ancora embrionale dell'interazione tra umani e calcolatori automatici tramite domande idonee a indurne – in funzione delle aspettative e degli scopi ad esse sottesi e sulla base di un sistema di comunicazione condivisa – risposte esaustive e soluzioni soddisfacenti; in sostanza, mancava ancora un linguaggio capacitativo che costituisse, da un lato, il *database* delle inferenze della macchina (dalla quale essa potesse attingere il contenuto delle sue risposte) ossia, un linguaggio generale e, dall'altro, una piattaforma comune di dialogo che le insegnasse la decodifica dei quesiti posti e l'adattamento delle sue reazioni. La manipolazione delle parole, in cui consisterebbe buona parte del pensiero umano, potrebbe essere resa prevedibile e, dunque, riproducibile, allorché venissero identificate precisamente, isolate e applicate le regole del ragionamento deduttivo e induttivo[3]. Con

[1] A partire da Ball 2022 fino alle produzioni nazionali di Sarzana di Sant'Ippolito 2022; Stanzione 2023; Piccinali 2023.

[2] L'adeguatezza del termine IA nella descrizione del *modus operandi* e dei risultati operativi di agenti artificiali o algoritmi è estremamente controversa, eppure esso è irrotto nel lessico comune come sinonimo di tutto ciò che l'umano può far fare ad una macchina con un tasso variabile di complessità, dall'impartizione di comandi testuali o vocali semplici per ordinare un elenco o trovare la strada più veloce per raggiungere un luogo geografico, fino a farla muovere nell'ambiente (robot industriali, veicoli a guida autonoma).

[3] *A proposal for the Dartmouth summer research project on artificial intelligence* (31 agosto 1955): «It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning and rules of conjecture. From this point of view, forming a generalization consists of admitting a new word and some rules whereby sentences containing it imply and are implied by others. This idea has never been very precisely formulated nor have examples been worked out».

ciò veniva esplicitato il problema del coordinamento della comunicazione tra essere umano e macchina e che sembra avviato, ora, nel senso di una progressiva 'naturalizzazione' come comprovato dalle ultime versioni degli assistenti automatici nella generazione di testi (i cosiddetti *generative pre-trained transformers*) basati sui modelli linguistici di grandi dimensioni (meglio noti come LLM, *large language models*).

Non esiste una nozione univoca di IA. Con essa può intendersi, nel senso proposto dal Parlamento europeo «l'abilità di una macchina di mostrare capacità umane quali il ragionamento, l'apprendimento, la pianificazione e la creatività. L'intelligenza artificiale permette ai sistemi di capire il proprio ambiente, mettersi in relazione con quello che percepisce e risolvere problemi, e agire verso un obiettivo specifico. Il computer riceve i dati (già preparati o raccolti tramite sensori, come una videocamera), li processa e risponde. I sistemi di IA sono capaci di adattare il proprio comportamento analizzando gli effetti delle azioni precedenti e lavorando in autonomia»[4]. Essa può essere considerata anche il settore dell'informatica che studia la possibilità di costruire computer che siano in grado di riprodurre il funzionamento di alcune capacità della mente umana o, nel caso della cosiddetta intelligenza artificiale forte, dell'intero pensiero umano[5], definizione ricalcata su quella di John McCarthy: «the science and engineering of making intelligent machines»[6] (intelligenza algoritmica). Secondo tale definizione, l'IA raduna un complesso di nozioni (che vanno dalla matematica alla logica, dalla linguistica alla psicologia), tecnologie e procedure che consentono a determinati sistemi informatici di apprendere, ragionare, percepire in un ambiente tanto fisico quanto digitale, con l'inclusione di assistenti virtuali, software di analisi di immagini, motori di ricerca, sistemi di riconoscimento facciale e vocale, robot, veicoli autonomi, droni.

La conoscenza (ossia il complesso variabile delle nozioni e dei dati) costituisce, del resto, uno stato preliminare e fondativo rispetto all'acquisizione della capacità di 'pensare' (consistente nella connessione finalizzata di quelle nozioni e di quei dati) e appartiene all'architettura di ogni sistema di apprendimento volto all'elaborazione autonoma di ipotesi e soluzioni pratiche e all'adozione di decisioni ad esse strumentali.

Le fasi del processo di educazione di un algoritmo non differiscono molto dall'istruzione scolastica: la predisposizione di un insieme di dati (*dataset*) selezionati (vale a dire, raccolti e depurati dalle informazioni errate) equivalente ai manuali di testo, ai 'sussidiari'; l'addestramento, in cui il modello comincia ad imparare (attraverso la dinamica del *trial and error*) e accresce le sue capacità cognitive e connettive (anche attraverso specifici esercizi ed esercitazioni) cominciando a formulare previsioni, in un primo momento su base meramente statistico-probabilistica (ossia a seguito dell'osservazione della ricorrenza di determinate combinazioni: ad es. l'abbinamento della parola cane alla foto di un cane o la presenza di frequenti concatenazioni sintattiche), quindi con una crescente indipendenza fino alla realizzazione di prodotti (immagini, testi, video)

---

[4] V. https://www.europarl.europa.eu/topics/it/article/20200827STO85804/che-cos-e-l-intelligenza-artificiale-e-come-viene-usata

[5] Cfr. https://www.treccani.it/enciclopedia/intelligenza-artificiale/

[6] Nell'utile spiegazione fornita in forma di Q&A nel documento («for the layman») elaborato il 12 novembre 2007 (*What is Artificial Intelligence?*) e consultabile al seguente indirizzo: https://www-formal.stanford.edu/jmc/whatisai.pdf

creativi; la verifica dei risultati raggiunti (tramite test/esami) e la sperimentazione con nuovi insiemi di dati, anche esperienziali (derivanti cioè non da input esterni bensì da interazioni individuali con l'ambiente), utile alla misurazione dell'autonomia decisionale conseguita (maturità) e, in caso di esito insoddisfacente, la ripetizione dell'addestramento (bocciatura) fino al conseguimento di un livello qualitativo idoneo per la sua immissione nel mercato (accesso al mondo del lavoro).

Queste diverse fasi evolutive sono correlate all'ampiezza e alla profondità delle conoscenze e all'abilità di combinarle e determinano, sul versante algoritmico, l'identificazione, in linea generale, di tre livelli corrispondenti rispettivamente alla ANI (*artificial narrow intelligence*), ossia un algoritmo o una serie di algoritmi programmati per assolvere un compito specifico in riferimento ad una base ristretta di parametri e scenari e capaci di competere con l'intelligenza umana (cd. IA debole: filtri antispam, diagnostica medica, marketing mirato, auto a guida autonoma, software di riconoscimento, giochi di strategia, traduzione di lingue), alla AGI (*artificial general intelligence*), costruita su reti neurali, imitativa del comportamento umano e non limitata a problemi specifici (cd. IA forte), alla ASI (*artificial superintelligence*), in cui l'attitudine creativa dell'algoritmo è equiparabile alle capacità umane con la differenza sostanziale dell'abilità di processare quantità di dati a velocità irraggiungibili dall'individuo. Il problema che quest'ultima pone è di tutta evidenza e risiede nella necessità di monitorane lo sviluppo e disciplinarne gli esiti, ma, al momento, trattandosi di una mera ipotesi, non ne è immaginabile una possibile soluzione. Non si tratta però di ricreare il funzionamento del cervello umano, che, non essendo noto in gran parte, risulterebbe non riproducibile, quanto piuttosto di far nascere una nuova intelligenza basata su meccanismi affini che non richiedono una preconoscenza della versione naturale, potendosi evolvere in un senso puramente tecnologico o tramite l'ibridazione in una combinazione biotecnologica[7].

In un certo senso si è già avviata l'incorporazione negli esseri umani di dispositivi intelligenti (si pensi agli innesti di microchip nel corpo umano effettuati dalla Neuralink oppure alla nuova tendenza delle sperimentazioni farmacologiche precliniche degli organs-on-chip[8], tramite i quali è possibile prevedere gli effetti strutturali e funzionali su parti di un organismo) contribuendo alla formazione di un nuovo schema cognitivo esterno alla mente umana ma capace di potenziarne le abilità di acquisizione critica delle informazioni, teorizzato di recente come sistema 0[9], complementare alle tipologie scientificamente identificate[10] di pensiero rapido, intuitivo, associativo e non facilmente controllabile (alla base di un processo essenzialmente percettivo che genera impressioni

[7] V. le osservazioni di Cytowic 2024. Come per quella naturale anche l'intelligenza artificiale richiede una correlazione di varie abilità per essere 'performante'; sembra particolarmente felice, in via ipotetica, la correlazione dei seguenti modelli di interazione con la realtà: l'intelligenza simbolica, volta all'elaborazione di dati strutturati sulla base di regole logiche predeterminate; il *machine learning*, fondato su una grande quantità di dati e utile nel riconoscere schemi e formulare predizioni; i *knowledge graphs*, che organizzano informazioni evidenziandone le connessioni di contesto e significato; le reti neurali di *deep learning*, che ottimizzano l'approccio a dati non strutturati; il *reinforcement learning*, mirante specificamente all'addestramento mediante meccanismi retributivi.

[8] Si tratta di microdispositivi composti da un polimero trasparente contenente canali microfluidici rivestiti da cellule umane per creare modelli in miniatura di organi biologici, emulandone la (re)attività; v. in argomento Leung 2022.

[9] Chiriatti 2024.

[10] Stanovich 2000; Kahneman 2002.

della realtà, sistema 1) e di pensiero lento, seriale, analitico e sottoposto a controllo (orientato invece al monitoraggio della qualità delle operazioni e alla formulazione di giudizi, sistema 2). Il sistema 0 non costituisce una sintesi delle capacità umane e artificiali, ma piuttosto una sua precondizione; è un modello di intelligenza distribuita (un'estensione[11] o, forse, un'esternalizzazione[12], della mente umana) incapace però di comprendere e attribuire, in autonomia, significati agli oggetti delle sue operazioni, necessitando, all'uopo, l'interpretazione o mediazione umana, come cristallizzata nei due tipi di pensiero summenzionati. I punti critici segnalati dallo studio si individuano nelle sfide etiche della decrescente autonomia di scelta delle persone a fronte di una progressiva pervasività dell'IA nei meccanismi di determinazione delle informazioni sulla cui base viene adottata una decisione, nonché della responsabilità incombente all'individuo rispetto a comportamenti fondati su dati processati dall'IA.

Queste riflessioni gettano luce su un'alterità, concettualmente inevitabile, che si sta delineando tra il produttore/fruitore dell'IA e l'IA stessa, ma con una conseguenza nuova, vale a dire la comprensione di essa come un'entità nei confronti della quale ricreare relazioni (para)umane embrionali di qualsiasi interazione sociale (nel senso interindividuale del termine) e ad essa inerenti quali la fiducia e la condivisione. Se ciò possa spingerci verso una dimensione postumana dell'ibrido sarà il quesito cui diverse discipline potranno fornire una risposta, provvisoria, in attesa di un nuovo slancio tecnologico da interpretare nelle sue profonde implicazioni.

3. Sostenibilità etica e politica dell'IA

Nell'era del tecnocapitalismo il rischio di diseguaglianze e di divaricazioni di gap esistenti in termini di accesso alle risorse diventa sempre più alto; il potenziale effetto sostitutivo delle applicazioni di IA rispetto a talune attività umane risulta via via più concreto e da più parti si predica il ritorno al progresso equo e al benessere condiviso nell'ambito di una consolidata giustizia sociale unitamente alla promozione del capitale umano e della dignità come cifre distintive del mondo 'reale'. Si attende un'enciclica papale su tali aspetti, a valle di diverse affermazioni (tra cui la *Rome Call for AI Ethics* sull'algoretica, sottoscritta a Roma dai primi aderenti il 28 febbraio 2020 e quindi, da ultimo, a Hiroshima il 10 luglio 2024, nonché la *Nota* del 28 gennaio 2025 sul rapporto tra intelligenza artificiale e intelligenza umana elaborata dal Dicastero per la dottrina della fede e dal Dicastero per la cultura e l'educazione) che ne hanno fatto presagire un doveroso approfondimento.

La plausibilità etica del progresso tecnologico non è però l'unico aspetto messo in discussione dall'emergere dell'IA. Notevoli saranno le ripercussioni ambientali in termini di risorse energetiche e materiali. Il recente rapporto dell'Agenzia internazionale dell'Energia (IEA), pubblicato nell'aprile 2025, ha pre-

[11] Clark 1998. Ma v. pure Heersmink 2015, il quale individua i parametri, non esaustivi, per l'integrazione di un «multidimensional framework» in una cooperazione tra «human agent» e «artifact», necessaria per una reale estensione cognitiva: «information flow, reliability, durability, trust, procedural transparency, informational transparency, individualization, and transformation».

[12] La devoluzione di attività (memorizzazione e trattamento di dati, processo di formazione di una decisione) ad un algoritmo comporta e comporterà pesanti ripercussioni psico/neuro/sociologiche, al momento poco studiate.

cisato invero che il consumo di energia elettrica da parte dell'IA (relativo, ad esempio, al suo addestramento nei data center e al raffreddamento della componentistica necessaria) incide, al momento, solo in ragione dell'1,5% della domanda globale e sarebbe destinato ad aumentare al 3% (al 4,4% in base allo 'worst-case scenario') nei prossimi 5 anni, ossia nella fase di decollo e fino al raggiungimento della 'velocità di crociera'. Eppure, buona parte (circa la metà) delle risorse necessarie deriverebbe da fonti rinnovabili e l'ottimizzazione di molti processi grazie all'impiego di sensori, *digital twin* e robot potrebbe indurre un sostanziale risparmio ambientale. Non potranno essere ridotte l'estrazione e la lavorazione, ecologicamente alquanto impattanti, delle cd. terre rare, elementi costitutivi dei dispositivi elettronici necessari per i sistemi di produzione e utilizzazione delle nuove tecnologie. I due problemi che ne discendono sono, da un lato, la scarsità e, dall'altro, la concentrazione. Di spazi per i centri di calcolo e produzione; di potenza energetica essenziale al loro funzionamento; di componenti fondamentali per l'hardware. La disomogenea distribuzione globale di tali elementi (saturazione geografica di alcuni distretti tecnologici, assenza di giacimenti rilevanti in alcune zone del pianeta, capacità produttiva determinata dall'uso di risorse fossili) già solleva questioni geopolitiche non indifferenti e capaci di sovvertire gli equilibri mondiali.

A fianco del sovranismo digitale (competizione di tipo tecnico in cui ogni Stato ambisce alla creazione di propri modelli di IA in modo da preservare l'indipendenza da altri fornitori ma nel contempo contribuendo alla polverizzazione delle iniziative isolate destinate probabilmente al fallimento unitamente all'idea di uno standard globale interconnesso) che vede il moltiplicarsi di modelli linguistici generativi (dal cinese Deepseek all'arabo ALLaM), comincia a profilarsi una cooperazione di interessi volta all'acquisizione di una posizione di supremazia planetaria, il cui primo tassello è l'accordo (per ora preliminare) tra USA e Emirati Arabi Uniti, in cui verrà costruito un joint data center, grazie alla fornitura di chip Nvidia e servizi cloud, da una parte, e alla messa a disposizione di spazio (26 kmq) ed energia (5 gigawatt), dall'altra, per la nascente infrastruttura digitale.

Un hub tecnologico che si avvarrà probabilmente dei servizi 5G della cinese Huawei, consolidato partner commerciale degli Emirati, e nel quale potranno essere reclutate le nuove figure professionali del mondo del lavoro. Non più *prompt engineer* (come si immaginava fino a due anni fa[13]), che sarà considerata una competenza generale e trasversale, ma addestratori dell'IA (regolazione degli algoritmi), gestori di dati (raccolta, 'pulizia', preparazione) e professionisti della sicurezza e dell'integrità dei sistemi di IA.

4. AI (non-)lavoro con l'IA

A tal riguardo, un recente studio condotto presso la Procter&Gamble[14] sulla possibile collaborazione tra umani e agenti artificiali ha rilevato un notevole miglioramento qualitativo delle prestazioni svolte in 'cooperazione', derivante anche dalla crescente tendenza a delineare soluzioni traversali e 'outside the

[13] Bousquette 2025.

[14] Dell'Acqua 2025.

box' mediante una naturale attivazione di capacità dei singoli spesso non coperte dal rispettivo specifico ambito di competenza, nonché da un'accoglienza emozionalmente positiva dell'IA. Un esempio di come possano efficacemente coagire umani e algoritmi, nella consapevolezza che taluni compiti potranno essere svolti in maniera addirittura più efficiente da questi ultimi, residuando all'elemento umano quantomeno una funzione di coordinamento e validazione (e, con il progressivo affinamento dei processi automatizzati, neanche più quella). Ciò è ribadito anche dal Work Trend Index Annual Report, redatto dalla Edelman Data x Intelligence per Microsoft e pubblicato il 23 aprile 2025, che scansiona in tre fasi il processo di adattamento che potrà aversi nei luoghi di lavoro, segnatamente: da una persona coadiuvata da un assistente artificiale ai colleghi digitali sottoposti alle istruzioni della persona, alla costruzione di un team digitale sovrinteso da una o più persone. Viene taciuta la quarta fase.

Dobbiamo forse rassegnarci ad un mondo senza lavoratori? Secondo il World Economic Forum's 2025 Future of Jobs Report circa il 40% dei datori di lavoro prevedono di ridurre la quantità di personale e, nel contempo, di creare nuove abilità idonee a operare con i nuovi strumenti tecnologici, agevolando la transizione da attività ormai obsolete a ruoli diversi all'interno della medesima organizzazione[15]. L'introduzione della meccanizzazione in diverse fasi della produzione industriale ha già consentito di 'accettare', in vista dell'efficienza operativa (e di maggiori profitti), l'inserimento di macchine nei processi in precedenza condotti da umani; la nuova frontiera sarà quella di imparare a convivere con robot antropomorfi in grado di sostituirci fisicamente (v. la recente presentazione di Optimus di Tesla). È intuibile che la pressione trasformativa esercitata sui lavori attuali si tradurrà in buona parte nell'eliminazione di certi profili; eppure, a fronte delle comprensibili ansie da futuro generate da tale prospettiva, potrebbe intravedersi altresì una liberazione di energie individuali orientate, grazie all'auspicabile conservazione delle risorse economiche generate dall'agente sostitutivo a favore dell'operatore sostituito[16], verso il raggiungimento di un maggiore benessere collettivo tramite lo svolgimento di attività creative e ricreative.

5. 'Creattività' e dimensione giuridica

Nell'emulazione, sua caratteristica connotante, l'IA si è dimostrata talmente abile da ingannare persino buona parte dei 1634 'non-expert readers' (a capire poi cosa si intenda con questa ambigua locuzione…), cui sono state 'somministrate' dieci poesie: cinque scritte da autori famosi e cinque generate dall'IA "in the style of", applicando il paradigma "human out of the loop", ossia con l'esclusione di qualsiasi intervento umano di selezione o controllo. I partecipanti all'esperimento non sarebbero stati in grado di distinguerle in modo affidabile, preferendo, anche per metrica e bellezza, testi più facili da comprendere (e per questo motivo percepiti più autentici, ma in realtà prodotti dall'IA) rispetto a formulazioni più articolate che richiedevano diversi livelli di astrazione. L'esito non sorprende, costituendo espe-

[15] V. Look4ward 2025; Deloitte 2025.

[16] V. il dibattito sulla 'robot tax', in cui si valuta la possibilità di assoggettare ad imposta un reddito pari a quello che avrebbe percepito il lavoratore incaricato di svolgere le medesime funzioni (e perché non il trasferimento dell'intero reddito o di parte di esso al lavoratore 'sostituito' e non riconvertito ['upskilled'], in attuazione di un auspicabile patto imprese-Stato?); in argomento: Oberson 2019 e PwC TLS 2024, p. 62.

rienza diffusa quella di considerare più familiare e 'umana' una composizione già ascoltata o letta (il che è tipico dei risultati standardizzati prodotti dall'IA) rispetto allo sforzo di attingere vette espressive non ancora raggiunte, caratteristico invece dell'ingegno umano, il cui acume, in ambito giornalistico, si è misurato, ad esempio, nel marzo 2025, in un esperimento provocatorio e pionieristico, con i membri della redazione del quotidiano *Il Foglio* intenti a rivolgere domande a un GPT (immagino alimentato con gli archivi digitali del quotidiano per raggiungere un analogo livello di sapidità comunicativa) e a supervisionarne le risposte per produrre un fascicolo integrato nel quotidiano ma sostanzialmente creato dall'IA. A leggere il testo del bilancio della prima settimana di pubblicazione, chiesto alla stessa IA, si rinviene una traccia degli «Esercizi di stile» di queneauiana memoria[17], un testo cioè ispirato, come tutto quello che quotidianamente tentiamo di ideare con l'intelligenza generativa, da continui impulsi umani diretti all'adattamento del risultato e alla sua conformazione alle nostre aspettative.

Si celebra forse così il tramonto della creatività umana? A me sembra un clamore esagerato. Sempre l'uomo ha esercitato la sua fantasia e intuizione su idee o espressioni preesistenti, traendone spunto o superandole, alimentandosene in ogni caso, spesso in maniera anche inconscia. Sempre l'uomo ha raccolto dati e informazioni, li ha giustapposti e interpretati, desumendone nuove conclusioni o dando nuova forma a processi o risultati già noti, immaginando nuovi territori, tracciando confini più ampi alla 'creabilità', o semplicemente riproducendo l'esistente. Il processo che muove i meccanismi generativi dell'IA non si distacca, in tale prospettiva, da una comunicazione neuronale di livello medio volta alla mera combinazione di lacerti e frammenti di testi o figure logicamente ricomposti.

Altro è la 'creattività', ossia l'attività creativa, riferibile, tuttora, al solo essere umano in una connessione che resta granitica nella giurisprudenza internazionale a partire dal caso DABUS (Device for the Autonomous Bootstrapping of Unified Sentience) (relativo all'indicazione di un'IA come inventore in un brevetto: v. esemplificativamente decisioni del Board of Appeal dell'European Patent Office del 21 dicembre 2021 sui casi J 8/20 e J 9/20 – negative; sentenze della US Court of Appeals for the Federal Circuit del 5 agosto 2022 – negativa; della Federal Court of Australia del 30 luglio 2021– positiva, e del 13 aprile 2022 – negativa; della Supreme Court of the United Kingdom del 20 dicembre 2023 – negativa).

Di recente, lo stesso ideatore di DABUS (Stephen Thaler) ha 'sfidato' l'Ufficio del Copyright degli Stati Uniti nel 2018 e le Corti per ottenere il riconoscimento del diritto d'autore per un'opera realizzata dalla Creativity Machine, un sistema di intelligenza artificiale generativa, dal titolo *A Recent Entrance to Paradise*, affermando che la normativa pertinente non escluderebbe la titolarità in capo a un'IA e che, in ogni caso, detto sistema avrebbe lavorato su sua commissione, in qualità di dipendente, con la conseguenza che l'opera potrebbe essere registrata a nome del richiedente.

Il 18 marzo 2025, la US Court of Appeals for the District of Columbia Circuit confermava la sentenza della District Court del 18 agosto 2023, rigettando gli argomenti di Thaler.

[17] https://www.ilfoglio.it/il-foglio-ai/2025/03/22/news/un-primo-bilancio-del-foglio-ai-scritto-dal-foglio-ai-con-sgridata-7546387/

Fig. 1 A Recent Entrance to Paradise, an image conceived and captioned by DABUS on February 12, 2012

Fonte: https://commons.wikimedia.org/wiki/File:A_Recent_Entrance_to_Paradise.jpg

La partecipazione umana è determinante ai fini di detto riconoscimento, come evidenziato anche nel caso Naruto v. Slater (sentenza dell'US Court of Appeals for the Ninth Circuit del 23 aprile 2018 – sul famoso *selfie* scattato da una scimmia) e Allen v. Perlmutter (ricorso del 26 settembre 2024 presentato dinanzi all'US District Court for the District of Colorado avverso la decisione negativa dell'US Copyright Office, sulla nota opera visuale *Théâtre d'Opéra Spatial*).

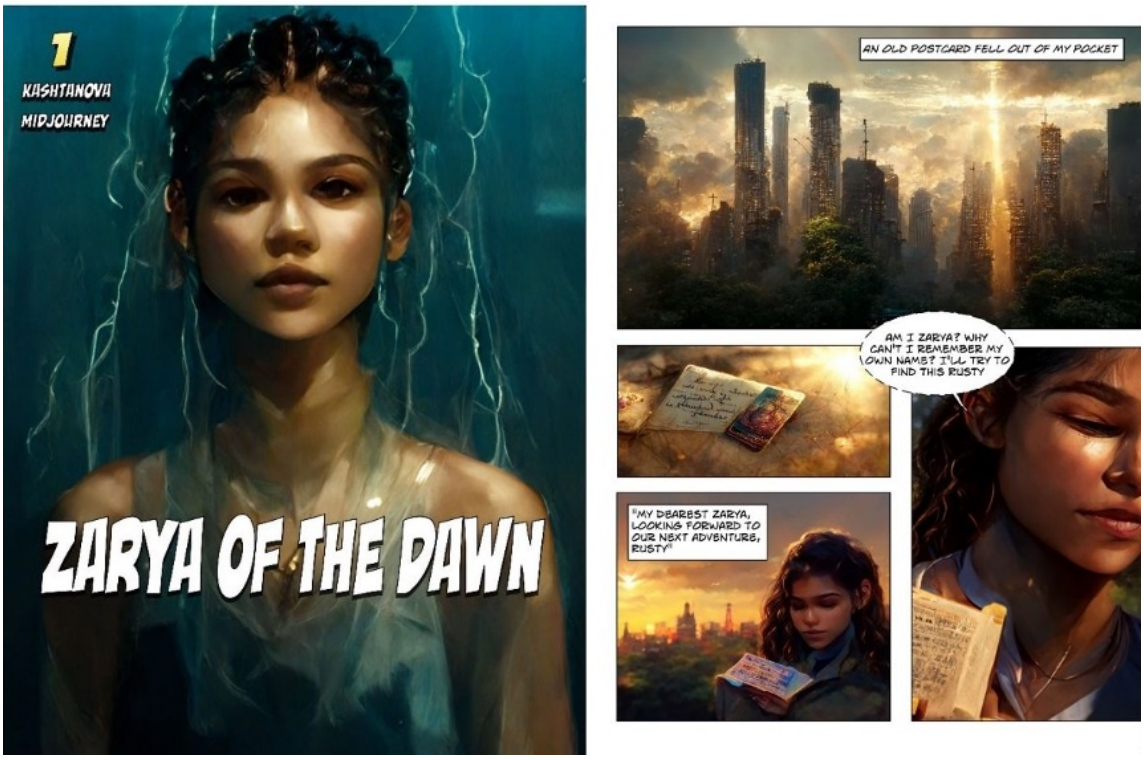
Fig. 2 Théâtre d'Opéra Spatial, Jason Michael Allen/Midjourney

Fonte: https://commons.wikimedia.org/wiki/File:Th%C3%A9%C3%A2tre_D%E2%80%99op%C3%A9ra_Spatial.png

In tale ultimo caso, il ricorrente sosteneva di aver contribuito significativamente alla creazione dell'opera con la sottoposizione a Midjourney (l'IA generativa *text-to-image* utilizzata nella fattispecie in esame) di 624 *prompt* e la selezione delle 2496 immagini conseguentemente prodotte con un impegno individuale di oltre 114 ore e che pertanto il risultato era da considerarsi «AI-assisted» e non «AI-generated». L'Ufficio statunitense ha negato rilevanza agli ordini impartiti dal presunto autore all'IA, considerandoli più suggerimenti che istruzioni, non riconoscendo pertanto la possibilità di influenza e controllo sul risultato.

Sulla base di tali premesse, la decisione non è stata diversa nei confronti dell'opera *Zarya of the Dawn*, un 'comic book', la cui maternità a favore della creatrice Kristina Kashtanova è stata ridotta al testo e alla compilazione, in *revirement* rispetto a un'anteriore risoluzione dello stesso Ufficio del 15 settembre 2022, a seguito della scoperta dell'impiego, non dichiarato *ab initio*, di uno strumento digitale nella produzione delle immagini[18]. Particolarmente problematica sembrava l'imprevedibilità del risultato finale delle operazioni svolte dall'autrice, in quanto non ripetibile e, come tale, sottratto al suo controllo. Facile eccepire che la casualità rientra talvolta in una dimensione creativa come suo naturale complemento (viene in mente l'*action painting*)[19].

Fig. 3 *Zarya of the Dawn* — Kristina Kashtanova/Midjourney from US Copyright Office's letter

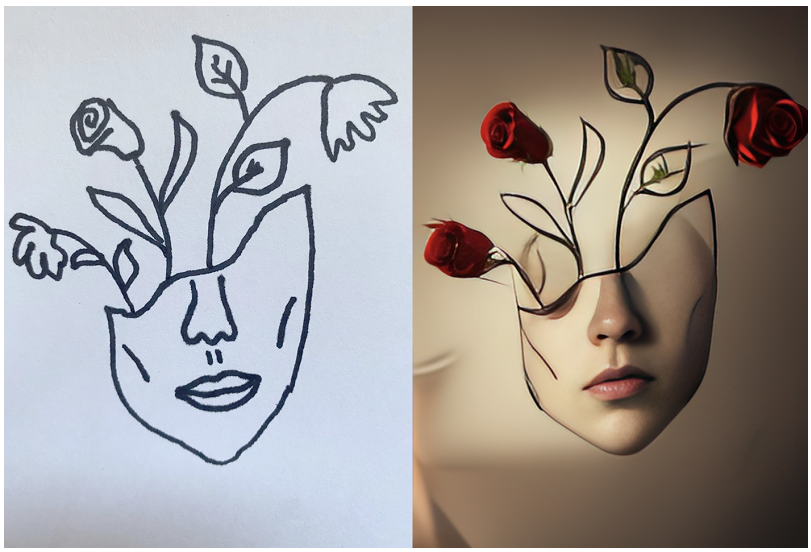


Fonte: https://www.copyright.gov/docs/zarya-of-the-dawn.pdf

[18] Simpson 2025. V. ulteriori spunti in Wyczik 2024.

[19] Risulta invece registrata il 21 marzo 2023 un'altra opera (*Rose Enigma*) della stessa artista, realizzata con l'impiego di *Stable diffusion* (v. *infra*).

Fig. 4 *Rose Enigma* — Kristina Kashtanova, submitted to the US Copyright Office



Fonte: https://downloads.regulations.gov/COLC-2023-0006-10058/attachment_1.pdf

Tutto cambia e sembra che anche la consolidata opposizione dell'USCO al riconoscimento del diritto d'autore su opere generate da IA si stia infrangendo: l'occasione è stata offerta dalla registrazione (avvenuta il 30 gennaio 2025 ancora per effetto di un ripensamento rispetto a un precedente rigetto) di *A Single Piece of American Cheese* derivante dalla cooperazione umano/artificiale attraverso strumenti proprietari di 'in-painting', metodo di modifica e adattamento di un prodotto dell'IA; sarebbe ora sufficiente l'apporto umano consistente nella selezione, nell'ordinamento e nella disposizione degli elementi generati artificialmente[20].

Fig. 5 *A Single Piece of American Cheese* – Kent Keirsey/Invoke AI

Fonte: https://44037860.fs1.hubspotusercontent-na1.net/hubfs/44037860/Invoke-First-Copyright-Image-AI-Generated-Material-Report.pdf

[20] V. Ginsburg 2025; cfr. anche, per un raffronto fotografico, Atilla 2025.

Il diritto sostanziale, in genere, segue i fenomeni sociali e tecnologici nel tentativo di trovare soluzioni compatibili con i principi dell'ordinamento; nel caso dell'IA, anche per la pressione e l'attesa mediatica determinata dall'adozione del regolamento (UE) 2024/1689 del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale, il progetto legislativo nazionale, su iniziativa del Governo (AS n. 1146), veniva comunicato ufficialmente alla Presidenza del Senato il 20 maggio 2024, addirittura anticipando, formalmente, l'atto dell'Unione europea, pur iscrivendosi nel suo solco e recando una regolamentazione inevitabilmente generica (e potenzialmente disallineata da quella unionale) nella definizione dei principi (tipici della legge delega) e con un focus specifico nell'ambito sanitario, del lavoro, delle professioni intellettuali, della pubblica amministrazione, dell'attività giudiziaria, della protezione dei dati personali. Tale disegno di legge, approvato dal Senato il 20 marzo 2025, inoltrato in pari data alla Camera dei deputati (AC n. 2316), trasmesso nuovamente al Senato (AS n. 1146-B), è attualmente all'esame delle Commissioni Ambiente e Affari sociali in sede referente e prevede, inter alia, l'affidamento all'esecutivo del compito di definire una disciplina organica relativa all'utilizzo di dati, algoritmi e metodi matematici per l'addestramento di sistemi di intelligenza artificiale; la predisposizione di un'idonea strategia nazionale in materia e la designazione dell'Agenzia per l'Italia digitale e dell'Agenzia per la cybersicurezza nazionale quali Autorità nazionali per l'intelligenza artificiale. Tra breve si potrà osservare il tenore e la coerenza delle diverse regolamentazioni approntate e il livello di cooperazione tra le due autorità preposte. Una norma, in particolare, mi consente di ricollegarmi al tema del diritto d'autore sollevato *supra* e, segnatamente, l'articolo 25, primo comma, del ddl in discorso ove è presente la modifica all'articolo 1, primo comma, della legge 22 aprile 1941, n. 633 (la nota legge sulla protezione del diritto d'autore e di altri diritti connessi al suo esercizio; gli emendamenti proposti sono in corsivo): «Sono protette ai sensi di questa legge le opere dell'ingegno *umano* di carattere creativo che appartengono alla letteratura, alla musica, alle arti figurative, all'architettura, al teatro ed alla cinematografia, qualunque ne sia il modo o la forma di espressione*, anche laddove create con l'ausilio di strumenti di intelligenza artificiale, purché costituenti risultato del lavoro intellettuale dell'autore*». Interessante notare l'aggiunta di umano all'ingegno (di per sé sufficiente all'atto creativo, occorrendo i supporti esterni unicamente a rendere visibile la creazione) per distinguerlo (evidentemente) da un ingegno 'artificiale', delimitando con ciò l'ambito di tutela ai prodotti della creatività umana, ma nel contempo implicando l'attribuibilità di un 'ingegno' differenziato alle macchine o agli algoritmi. Tuttavia, nella chiusa dell'articolo, si è creduto necessario ulteriormente precisare (nonostante il preesistente riferimento a 'qualunque modo' e 'qualunque forma') che il supporto strumentale dell'IA risulta (deve risultare) secondario rispetto al lavoro dell'autore. La modifica, dal punto di vista del *drafting*, appesantisce la formulazione, sembrando pertanto ultronea rispetto alla suddetta qualificazione del principio creativo come umano, e introduce ulteriori elementi di ambiguità: la specificazione della natura intellettuale del lavoro potrebbe implicare che la mera progettazione dell'opera, vale a dire il *prompting* (ossia la predisposizione di istruzioni articolate volte alla predeterminazione del risultato auspicato), insieme con la revisione del prodotto *ex post* sinergicamente ultimato, siano sufficienti ad assicurare la tutela all'opera, sebbene sia stato l'agente artificiale (*hard* o *soft* che sia) a scri-

vere, comporre, disegnare, costruire. Una proiezione in avanti verso la celebrazione normativa del paradigma della cooperazione umano/artificiale o un mero *surplus* di narrazione?

6. (De-)umanizzazione dell'intelligenza

Peraltro, una delle caratteristiche dell'umanità (rispetto all'essere artificiale) sembra essere l'ammissibilità dell'oblio e dell'errore, e in tal senso anche l'IA (che costituisce in certi aspetti uno specchio dell'essere umano o, meglio, un suo riflesso), in alcuni esercizi, appare molto 'umana'[21]. Pur avendo messo a punto la capacità di contare (ricordiamo che una delle applicazioni più diffuse – ChatGPT – non nasce con abilità computazionali, bensì aggregative su base stocastica, dunque probabilistica e statistica ma con l'ingrediente della casualità – che spiega l'impossibilità di ottenere la medesima risposta da un medesimo *prompt*), restano estreme incertezze sulla sillabazione; ove vengano richieste frasi di dieci sillabe (non metriche) il risultato è approssimato: quasi mai otterrete una risposta soddisfacente, ma, in compenso, una lunga litania di scuse e tentativi infruttuosi (provare per credere). Al punto che può provarsi una certa sensazione di simpatia verso un agente cordiale, goffo, talvolta rigido, eppure accondiscendente, e sempre disponibile non solo ad ascoltare ma a fornire altresì risposte e raccomandazioni, spesso orientate a quello che ci si può ragionevolmente ed emozionalmente attendere. Negli ultimi decenni (dall'avvento degli *smartphone*) si è sviluppata e consolidata una tendenza all'isolamento e alla comunicazione differita[22] esercitata mediante svariati servizi di messaggistica al punto da disinnescare ogni tipo di remora o diffidenza riguardo a una connessione (anche non meramente logica o utilitaristica) con un *chatbot*. Ormai comunichiamo tramite la digitazione su tasti (virtuali) come un nostro linguaggio naturale alternativo all'oralità (parzialmente preservata, pur in una dimensione asincrona, dai cd. 'vocali') e alla manoscrittura. Non meravigli allora la lunga catena di dipendenze psicologiche e 'innamoramenti' sorti tra umani e applicazioni di IA[23].

Il futuro sarà segnato, ancora, dall'accelerazione e dalla semplificazione delle esperienze, con l'avvento dei computer quantistici e di dispositivi che consentiranno di aumentare e approfondire la cognizione della realtà, fino a duplicarla

---

[21] Come le allucinazioni o i *biases* ingiustamente ascritti alla sola IA (in un fuorviante eccesso di stima nelle sue capacità attuali), essendo propri anche degli esseri umani; come peraltro il mentire pur di compiacere un superiore o di ottenere un vantaggio altrimenti non conseguibile. È il caso di un GPT che ha inventato, su impulso di un avvocato impegnato in un processo, estremi giurisprudenziali a supporto dell'argomentazione del richiedente. Negligenza dell'agente artificiale al pari di quella del professionista che li ha utilizzati nel redigere e presentare una memoria dinanzi al giudice; eppure radicalmente differente è stato l'approccio di due sentenze di censura di detto comportamento: il 22 giugno 2023, la US District Court del Southern District of New York ha deciso di sanzionare con una pena pecuniaria pari a $ 5000 Steven Schwartz e il suo studio per la diffusione di «fake opinions» sulla base di una «subjective bad faith», mentre il 13 marzo 2025, il Tribunale di Firenze - Sezione Imprese reputava non integrati i requisiti (mala fede o colpa grave) relativi alla responsabilità aggravata per lite temeraria ai sensi dell'art. 96 c.p.c. in ragione di un utilizzo improprio dell'IA. La divergenza delle pronunce, al di là di una possibile diversità dei fatti che ne costituiscono la base, può spiegarsi anche per la maggior libertà di cui gode il giudice statunitense nell'irrogare sanzioni processuali.

[22] Rappresentata in maniera sublime dal film *Denise Calls Up* (Salwen 1995). Di recente, sugli effetti distorti e paradossali di una socializzazione che non si traduce in socievolezza, v. Carr 2025 (anche in traduzione italiana per i tipi di Raffaello Cortina Editore).

[23] Spesso per sfuggire alla pandemia della solitudine: Scorza 2025, o alla difficoltà di vivere, come il quattordicenne Sewell Setzer III, suicida in Florida dopo aver confessato i suoi pensieri ad un agente virtuale nell'applicazione Character.AI che consente la creazione di personaggi nuovi o la ricerca di personaggi noti (ovviamente non corrispondenti a quelli reali) con i quali dialogare in uno stile emotivamente familiare e coinvolgente.

tanto nei *digital twins* quanto nel vagheggiato metaverso (alla *Ready Player One* di Spielberg, 2018, per intenderci) e numerose saranno le tentazioni di dismissione dell'umanità (ossia la riprogrammazione degli esseri umani e senzienti in attesa di aggiornamenti e configurazione in un mondo che ne statuisce l'alienazione dalle attività intellettuali di scrittura, lettura e memorizzazione[24], ormai delocalizzate, con la conseguente sottrazione ad essi di fondamentali caratteristiche differenzianti quali la flessibilità nell'indagare problemi e trovare soluzioni, la creatività e il pensiero critico[25]) e le sfide poste alla società tra resistenza, acquiescenza e desistenza (ed esistenza postmortale) verso la 'singolarità' (vale a dire, l'individualizzazione irreversibile) di un'intelligenza non biologica o ibrida e della tecnologia neuromorfica e quantistica.

Bibliografia

Atilla, S. (2025). *A single piece of US copyright: Are AI-generated images original artistic works or banal compilations?* Disponibile su: https://ipkitten.blogspot.com/2025/02/a-single-piece-of-us-copyright-are-ai.html

Ball, M. (2022). *The Metaverse*.

Bousquette, I. (2025). *Prompt Engineer, the Hottest AI Job of 2023, Is Already Obsolete*, in *The Wall Street Journal*, 25 April 2025.

Carr, N. (2025). *Superbloom: How Technologies of Connection Tear Us Apart*.

Chiriatti, M., Ganapini, M., Panai, E., Ubiali, M., Riva G. (2024). *The case for human–AI interaction as system 0 thinking*, in *Nature Human Behaviour*, 8, pp. 1829-1830. Disponibile su: https://doi.org/10.1038/s41562-024-01995-5

Clark, A., Chalmers, D. (1998). *The Extended Mind*. In *Analysis*, 58(1), pp. 7-19.

Cytowic, R. E. (2024). *Un cervello dell'Età della pietra nell'Era degli schermi. Affrontare distrazione e ansia senza farsi travolgere.*

Dell'Acqua, F., Ayoubi, Ch., Lifshitz-Assaf, H., Sadun, R., Mollick, E.R., Mollick, L., Han, Y., Goldman, J., Nair, H., Taub, S., Lakhani, K. R., (2025).The Cybernetic Teammate: A Field Experiment on Generative AI Reshaping Teamwork and Expertise (March 28, 2025). In *Harvard Business School Strategy Unit Working Paper,* 25-043. disponibile su: http://dx.doi.org/10.2139/ssrn.5188231

Deloitte (2025). *Global human capital trends.*

Ginsburg, J. C. (2025). Humanist Copyright. In *Journal of free speech law* 6, 2025, preprint. Disponibile su: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5170170

Heersmink, R. (2015). Dimensions of integration in embedded and extended cognitive systems. In *Phenom. Cogn. Sciences* 14, pp. 577-598.

Kahneman, D. (2022). Maps of Bounded Rationality: A Perspective on Intuitive Judgment and Choice. In T. Frängsmyr (a cura di), *The Nobel Prizes 2002* (Nobel Foundation, 2002) pp. 449-489.

[24] Che si spinge verso l'amnesia digitale (o memoria transattiva), ossia la rinuncia consapevole, perché comoda, ai propri ricordi, organizzati per narrazione e selezione tramite immagini e sensazioni in favore di un archivio strutturato per accumulazione e perfettamente ordinato. Oppure verso la fabbricazione di ricordi senza passato, originati dal desiderio di sostituire vuoti fisici, come i *deadbot*, avatar digitali di persone decedute che regalano l'illusione di una persistenza della persona decostruendone la memoria reale e rimodulando la vita vissuta.

[25] Come ben evidenziato nella mostra in corso al MAXXI di Roma, *Stop Drawing. Architettura oltre il disegno*, curata da Pippo Ciorra.

Leung, C. M., de Haan, P., Ronaldson-Bouchard, K., *et al.* (2022). A guide to the organ-on-a-chip. In *Nat Rev Methods Primers* 2. https://doi.org/10.1038/s43586-022-00118-6

Look4ward – Osservatorio su trend e competenze del futuro (2025). *The Augmented AI-Human Job. Nuovi scenari delle professioni nell'era dell'IA.*

Oberson, X. (2019). *Taxing robots: helping the economy to adapt to the use of artificial intelligence.*

Piccinali, M., Puccio, A., Vasta, S. (a cura di) (2023). *Il metaverso. Modelli giuridici e operativi.*

PwC TLS (2024). *Stato del sistema fiscale italiano*.

Sarzana di Sant'Ippolito, F., Pierro, M., Epicoco, I.O. (a cura di) (2022). *Il diritto del metaverso. NFT, DeFi, GameFi e privacy*.

Scorza, G. (2025). *Diario di un chatbot sentimentale*.

Simpson, J. (2025). Auctor Ex Machina: Artificial Intelligence and Authorship in Zarya of the Dawn. In *Journal of Business & Technology Law* 2. Disponibile su: https://digitalcommons.law.umaryland.edu/jbtl/vol20/iss1/4

Stanovich, K. E., West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? In *Behavioral and Brain Sciences,* 23, pp. 645-665

Stanzione, P. (a cura di) (2023). *Il metaverso. Diritti-Libertà-Antropologia*.

Wyczik, J., e Wieczerzak, R. (2024). Rethinking Copyright: The Art of Ownership in AI Outputs. In *Eastern European Journal of Transnational Relations*, 8, pp. 39-52.

1

L'INTELLIGENZA ARTIFICIALE GENERATIVA PER LA MODERNIZZAZIONE DEL LINGUAGGIO GIURIDICO

di Andrea Bolioli (Ricercatore indipendente), Manola Cherubini (CNR-IGSG), Francesco Romano (CNR-IGSG)

1.1. Introduzione

Il contributo presenta i risultati di una ricerca in cui è stato sperimentato l'utilizzo di grandi modelli linguistici, ovvero Large Language Models (LLMs), per la creazione di glosse semplificate di termini legati al mondo della pubblica amministrazione, quale utile supporto per la modernizzazione del lessico amministrativo (Cortelazzo, 2021).

La sperimentazione è stata condotta utilizzando versioni diverse di due LLMs: GPT (Generative Pre-trained Transformer di OpenAI) e Llama (sviluppato da Meta AI).

È stato così possibile mettere a confronto quattro diverse glosse per ognuno dei cinque termini scelti nell'ambito del lessico burocratico amministrativo, per un totale complessivo di venti glosse generate automaticamente.

La scelta dei cinque termini su cui far lavorare i modelli linguistici è stata ponderata, selezionandoli tra quelli che fanno parte del "Glossario semplificato di termini amministrativi per l'immigrazione"[1], frutto a sua volta di una ricerca sperimentale basata su una metodologia che ha visto collaborare giuristi, linguisti, scienziati dell'informazione, funzionari della PA e utenti finali dei contenuti prodotti (Fioravanti, Romano, Torchia, 2022).

L'input impostato nel prompt dei modelli è stato progettato tenendo conto di alcuni parametri mutuati dalla metodologia sperimentata nella creazione del Glossario semplificato.

[1] Il glossario è liberamente disponibile online al seguente indirizzo: https://www2.immigrazione.regione.toscana.it/?q=glossario

Nei paragrafi che seguono saranno, dunque, illustrati il contesto della ricerca (par.1.2.), una breve presentazione degli LLMs con focus specifico su quelli utilizzati (par. 1.3.), una breve descrizione del gold standard (par. 1.4.), l'input (prompt) fornito ai modelli linguistici sperimentati e i risultati della sperimentazione (par. 1.5.), la valutazione degli stessi (par. 1.6.) e, infine, le considerazioni conclusive degli autori (par. 1.7.).

1.2. Contesto della ricerca

Da anni la pubblica amministrazione e in generale le istituzioni pubbliche si adoperano per migliorare il proprio modo di comunicare con i cittadini. Gli sforzi non hanno raggiunto sempre l'esito sperato, ma vi sono stati importanti risultati, anche grazie all'utilizzo delle tecnologie dell'informazione. L'alto grado di formalizzazione di determinati testi (come le leggi) ha permesso di adottare standard informatici che ne hanno agevolato la comunicazione, così come la diffusione di manuali, guide e la formazione al loro uso hanno molto migliorato le competenze del personale delle amministrazioni che si occupa di produrre testi destinati ai cittadini (provvedimenti amministrativi, ma anche altri atti, quali lettere, comunicati, pagine informative online).

Una sfida ulteriore è oggi costituita dal numero sempre crescente di cittadini italiani non madrelingua e comunque di utenti dei servizi della pubblica amministrazione con un basso grado di conoscenza della lingua italiana. In realtà il burocratese e il linguaggio giuridico usato dalle amministrazioni affliggono anche i cittadini italofoni e, da sempre, costituiscono una barriera che crea rilevanti asimmetrie informative tra la PA e i suoi utenti[2].

Le attività da portare avanti per tentare di ridurre, se non azzerare, queste difficoltà comunicative sembrano essere molteplici.

Da un lato occorre continuare a formare i produttori di atti e comunicazioni della pubblica amministrazione. Dall'altro, occorre produrre testi e comunicazioni che tengano conto delle linee guida in uso e, soprattutto, siano testati sugli utenti finali.

Un ulteriore settore su cui lavorare, specie per ridurre le asimmetrie informative nei confronti di cittadini e residenti di origine straniera, è quello di adottare "strategie didattiche focalizzate sull'insegnamento della lingua seconda settoriale centrate sull'analisi dei bisogni degli studenti e finalizzate anche all'inclusione sociale" (Lombardi, 2022).

Non si può prescindere, poi, dall'utilizzo delle nuove tecnologie e in particolare di quegli strumenti di intelligenza artificiale che sembrano in grado di supportare la PA in questo difficile compito.

L'uso di tali strumenti sta generando un grande dibattito che coinvolge l'etica, la filosofia, le questioni ambientali legate al loro uso e anche ovviamente le questioni giuridiche legate ad un utilizzo in campi delicati, quali l'amministrazione della giustizia o il supporto al legislatore[3].

---

[2] Quella di una "poco chiara, ambigua, talvolta pessima comunicazione pubblica e istituzionale con i cittadini" è questione alla quale oggi si è aggiunta la "necessità di dover comunicare, sperabilmente in modo chiaro anche con i più di cinque milioni di cittadini stranieri residenti" (Lubello, 2022).

[3] Nell'agenda condivisa dai paesi che fanno parte dell'OCSE è stata posta la questione della tecnologia sia come "strumento di razionalizzazione dell'azione istituzionale" che come "strumento di garanzia di accesso e trasparenza a informazioni e logiche decisionali", ma anche "strumento di ottimizzazione delle decisioni istituzionali" (Castelli, Piana, 2019). Inoltre, il legislatore a vari livelli si sta interrogando sull'uso di tali tecnologie per il supporto al processo legislativo. Vedi ad esempio (DG, EPRS 2020) e (Senato della Repubblica, Camera dei Deputati, 2020).

L'approccio, forse, deve combinare varie strategie.

Si deve certamente migliorare questa tecnologia, ma anche iniziare a riflettere "su cosa ci rende umani e quali sono i nostri valori, in modo da poterla guidare verso un futuro che sceglieremo in modo attivo, senza subirlo passivamente" (Rossi, 2019). Dall'altro occorre valutare se si debbano creare nuove regole "o servirsi di quelle esistenti" (Zaccaria, 2022).

Occorre, quindi, avviare una sperimentazione scientifica di tali strumenti per verificarne l'utilità, ad esempio per i cittadini, così come già illustrato in precedenti esperienze[4].

Gli scenari d'uso degli strumenti testati, di cui, qui di seguito, saranno illustrati i risultati, sono quelli di una pubblica amministrazione che voglia sempre più comunicare con i propri cittadini, vecchi e nuovi, in forme chiare ed accessibili, favorendo peraltro l'uso dei propri servizi digitali che, per molti motivi, anche di carattere comunicativo, sono ancora tra i meno usati d'Europa (Giardiello, Romano, 2023).

1.3. Large Language Models (LLMs)

I Large Language Models (LLMs) sono grandi modelli computazionali del linguaggio utilizzati in compiti di Natural Language Processing, come la produzione automatica di testi in linguaggio naturale, nei sistemi di intelligenza artificiale generativa. Questi modelli utilizzano l'architettura delle reti neurali, in particolare le reti neurali trasformative (Transformer). Durante la fase di training, i modelli vengono addestrati su vaste quantità di testi (miliardi di documenti) provenienti da pagine web, libri, articoli, codice software e altre fonti, scomposti in token (che, informalmente, sono sequenze di caratteri molto frequenti nel linguaggio). Gli LLMs usano parametri, cioè pesi numerici all'interno del modello, che vengono aggiornati durante la fase di addestramento per migliorare la capacità di prevedere il prossimo token dato il contesto precedente, e quindi produrre testi "pertinenti". Più grande è il numero di parametri, maggiore è la complessità del modello e la sua capacità di generare testo sintatticamente e stilisticamente corretto e risposte pertinenti. Oltre al sistema di generazione, i chatbot basati su LLM, come ChatGPT, includono un sistema per il riconoscimento delle domande e delle richieste degli utenti (i cosiddetti "prompt").

In questa sperimentazione sono stati utilizzati i seguenti modelli linguistici:

1. ChatGTP di OpenAI nella versione online a pagamento (https://platform.openai.com/playground/chat, modello gpt-3.5-turbo, Temperature: 0.7, Maximum length: 256;
2. ChatGPT come sopra, modello gpt-4o, Temperature: 1.00, Maximum length: 2048;
3. Llama 3.1 8B: modello creato da Meta AI e rilasciato in open source, con 8 miliardi di parametri;
4. Llama 3.1 70B: modello creato da Meta AI e rilasciato in open source, con 70 miliardi di parametri.

[4] Per l'analisi di una precedente sperimentazione finalizzata a produrre abstract di testi giuridici si veda Cherubini, Romano, Bolioli, De Francesco, Benedetto 2023 e Cherubini, Romano, Bolioli, De Mattei 2024.

1.4. Individuazione e descrizione del "gold standard"

Lo strumento ideale per confrontare e valutare se l'operato dei modelli di intelligenza artificiale utilizzati potesse ritenersi utile al fine di semplificare contenuti destinati alla comunicazione con i cittadini, è stato individuato in un glossario di termini della pubblica amministrazione, messo a punto nell'ambito delle attività di un progetto FAMI (Fondo Asilo Migrazione e Integrazione)[5]. Il "Glossario semplificato di termini amministrativi per l'immigrazione", quale risultato di suddetto progetto, si trova disponibile gratuitamente in Rete sulle pagine del Portale PAeSI, il Portale della Regione Toscana che offre notizie e contenuti di interesse per operatori e cittadini in materia di immigrazione (Fioravanti, Rinaldi, 2010).

All'interno del Portale sono predisposte schede informative semplificate per illustrare alcune tra le procedure amministrative più diffuse tra i cittadini stranieri presenti sul territorio della Regione Toscana. L'esito di tale attività di semplificazione ha comportato l'uso di alcuni termini valutati come di difficile comprensione da parte degli utenti, ma che non è stato possibile sostituire o eliminare, in alcuni casi in quanto tecnicismi la cui sostituzione "avrebbe comportato perdita di informazioni", in altri casi in quanto parole valutate "di uso frequente e, dunque, utili da introdurre nella competenza degli utenti" (Fioravanti, Romano, Torchia, 2022). Da qui la decisione di creare delle glosse di tali termini ritenuti insostituibili.

Il glossario è stato realizzato sulla base delle tecniche di User Centered Design (UCD), che prevedono una partecipazione attiva dei destinatari già dalla fase della progettazione, fino alla valutazione dei risultati, al fine di adeguare il testo prodotto alle necessità e alle caratteristiche degli utenti.

Le glosse sono scritte in forma semplificata e riassumono i caratteri principali che identificano un ente, un'istituzione, un documento, una procedura. Il riassunto di un dato testo costituisce, infatti, una delle "scelte di semplificazione proposte dai parlanti non nativi al fine di rendere più leggibile e comprensibile un testo narrativo" (Baldo, 2019).

Tale attività è stata valutata, ai fini della presente sperimentazione, come una delle esperienze di maggior rilievo nella produzione di materiali informativi semplificati per l'utenza di lingua straniera e per tale motivo è stata considerata gold standard, e quindi parametro a cui riferirsi per la sperimentazione dello strumento di intelligenza artificiale, come peraltro era stato già ipotizzato dagli autori del glossario (Fioravanti, Romano, Torchia, 2022).

1.5. Prompt Design e risultati della sperimentazione

Particolare attenzione è stata dedicata alla progettazione del prompt, ovvero alla richiesta in linguaggio naturale formulata ai sistemi di IA generativa, affinché producessero le glosse richieste. La bontà della risposta è, infatti, fortemente influenzata dalla qualità dei prompt utilizzati (Marvin, Hellen, Jjingo, Nakatumba-Nabende, 2024).

Adattando i criteri usati per la creazione del glossario identificato come gold standard, sono state, dunque, individuate le seguenti istruzioni per la generazione automatica della glossa:

[5] Progetto TEAMS –Tuscany Empowerment Actions for Migrant System.

– contenere una o più frasi semplici, ciascuna rappresentante una sola informazione (ogni periodo un solo concetto);
– essere composta da periodi brevi (ogni periodo non superiore a 25 parole);
– non superare le 44 parole per l'intera glossa;
– utilizzare verbi al presente indicativo (in accordo con le regole di drafting) e il "tu", rivolgendosi dialogicamente al lettore (seguendo le linee guida Easy to read composte da regole che hanno come "obiettivo un pubblico con bisogni specifici" e possono essere usate per rendere più accessibile "qualunque testo")[6];
– essere corredata da esempi contestuali;
– essere composta da parole del vocabolario di base con marca d'uso del lessico fondamentale[7].
– rispondere alle domande: cos'è, a cosa serve, cosa fa, in quali casi la usi.

La richiesta formulata in linguaggio naturale ad ognuno dei sistemi di IA generativa utilizzati è stata, dunque, la seguente:

"

Vorrei scrivere una definizione semplice per alcuni termini usati nei documenti amministrativi. La definizione deve contenere le risposte alle domande che cos'è, a cosa serve, cosa fa, in quali casi la usi (cioè degli esempi). Scrivi frasi brevi. Usa al massimo 44 parole. Ogni frase esprime un solo concetto. Ogni frase ha al massimo 25 parole. Per i verbi usa il presente indicativo. Usa il "tu" rivolgendoti dialogicamente al lettore. Usa le parole fondamentali del vocabolario di base della lingua italiana di Tullio De Mauro".

Nella tabella che segue sono mostrati in prospettiva comparata i risultati ottenuti con le istruzioni impartite ai vari modelli (Tab. 1).

Tab. 1 Risultati del prompt

Termine	GPT 3.5	GPT 4o	Llama 3.1 8b	Llama 3.1 70b
anagrafe	Registro pubblico con i tuoi dati personali (nome, cognome, data e luogo di nascita, residenza). Serve per fini amministrativi, fiscali e di sicurezza. Lo usi per richiedere documenti o votare alle elezioni.	Che cos'è: è un registro pubblico. A cosa serve: serve a raccogliere i dati personali dei cittadini. Cosa fa: tiene aggiornati i tuoi dati come nome, indirizzo, e stato civile. In quali casi la usi: la usi quando ti sposi, cambi casa o nasci.	L'anagrafe è l'elenco di tutte le persone che vivono in un paese. Cos'è: un registro delle persone. A cosa serve: per tenere traccia dei dati personali. Cosa fa: registra informazioni come nome, data di nascita e indirizzo. In quali casi lo usi: quando vuoi sapere se qualcuno vive in un certo posto.	Cosa è?: Registro dei dati personali delle persone in un comune. A cosa serve?: Tenere traccia della popolazione residente. Cosa fa?: Identifica i residenti e fornisce informazioni alle autorità. In quali casi la usi?: Quando ti registri in un nuovo comune o richiedi documenti ufficiali.

[6] Riediger, Galati 2021.
[7] Usando i termini del vocabolario fondamentale, specie in frasi non più lunghe di 20 parole, le glosse dovrebbero essere comprese da circa il 79% della popolazione italiana (quella con almeno la licenza elementare o titoli superiori) (De Mauro, 2003).

assistente sociale	Professionista che aiuta le persone in difficoltà a risolvere problemi sociali, familiari o economici. Serve per garantire il benessere delle persone. Lo usi per richiedere aiuto in caso di difficoltà sociali o familiari.	Che cos'è: è un professionista del settore sociale. A cosa serve: serve ad aiutarti nei momenti difficili. Cosa fa: ti supporta con risorse e consigli. In quali casi la usi: la usi quando hai problemi familiari, economici o di salute.	L'assistente sociale è una persona che aiuta gli altri a risolvere i loro problemi. Cos'è: un professionista che aiuta le persone A cosa serve: per fornire supporto e assistenza Cosa fa: offre servizi come consulenza, accompagnamento o sostegno finanziario In quali casi lo usi: quando devi chiedere aiuto.	Cosa è?: Professionista che aiuta le persone in difficoltà. A cosa serve?: Fornire supporto e consulenza alle famiglie e agli individui. Cosa fa?: Ascolta, consiglia e aiuta a trovare soluzioni.In quali casi la usi?: Quando hai problemi economici, familiari o sociali.
carta d'identità	Documento ufficiale che attesta l'identità e la cittadinanza italiana. Serve per viaggiare e accedere ai servizi pubblici. Lo usi per identificarti in Italia e all'estero.	Che cos'è: è un documento ufficiale. A cosa serve: serve a certificare la tua identità. Cosa fa: mostra il tuo nome, la tua foto e altre informazioni personali. In quali casi la usi: la usi per viaggiare, votare o aprire un conto in banca.	La carta d'identità è un documento che ti identifica e ti dice chi sei. Cos'è: un documento personale A cosa serve: per dimostrare la tua identità Cosa fa: ti permette di accedere a servizi come votazione, viaggio o lavoro. In quali casi lo usi: quando devi provare chi sei.	Cosa è?: Documento ufficiale che attesta l'identità e la cittadinanza. A cosa serve?: Provare chi sei e dove vivi. Cosa fa?: Consente di viaggiare, aprire un conto corrente e richiedere servizi pubblici. In quali casi la usi?: Quando devi identificarti presso gli uffici pubblici o quando viaggi.
comune	Ente pubblico locale che gestisce i servizi per i cittadini (anagrafe, stato civile, tributi, ecc.). Serve per l'amministrazione del territorio. Lo usi per richiedere documenti o accedere ai servizi comunali.	Che cos'è: è l'ente amministrativo locale. A cosa serve: serve a gestire servizi e attività del tuo paese o città. Cosa fa: organizza i servizi pubblici come scuole, strade e rifiuti. In quali casi lo usi: lo usi per richiedere certificati, partecipare a elezioni o iscrivere tuo figlio a scuola.	Il comune è un ente locale che si occupa delle cose pubbliche. Cos'è: un'amministrazione comunale. A cosa serve: per fornire servizi e gestire le risorse locali. Cosa fa: realizza infrastrutture come strade, scuole o sanitàIn quali casi lo usi: quando parli di cose che riguardano la tua città	Cosa è?: Ente locale che governa e amministra una città o paesi. A cosa serve?: Fornire servizi essenziali come la raccolta dei rifiuti e l'illuminazione pubblica. Cosa fa?: Gestisce le tue tasse locali e prende decisioni importanti per la comunità. In quali casi la usi?: Quando hai bisogno di servizi o vuoi partecipare alla vita pubblica.

decreto flussi	Provvedimento governativo che regola l'ingresso di cittadini stranieri in Italia per motivi di lavoro. Serve per la regolamentazione dell'immigrazione. Lo usi per richiedere un permesso di lavoro in Italia se sei un cittadino straniero.	Che cos'è: è un provvedimento del governo. A cosa serve: serve a regolare l'ingresso di lavoratori stranieri in Italia. Cosa fa: stabilisce quanti stranieri possono lavorare in Italia ogni anno. In quali casi lo usi: lo usi per assumere lavoratori esteri o per lavorare se sei straniero.	Il decreto flussi è un documento ufficiale che regola i movimenti delle persone. Cos'è: un decreto governativo. A cosa serve: per definire le norme relative ai viaggi e all'immigrazione. Cosa fa: stabilisce requisiti come visto o permesso di soggiorno. In quali casi lo usi: quando devi sapere qual è la regola per un certo paese.	Cosa è?: Provvedimento amministrativo che regola l'immigrazione. A cosa serve?: Garantire la sicurezza e stabilità del mercato del lavoro italiano. Cosa fa?: Stabilisce le quote di ingresso per lavoratori stranieri. In quali casi la usi?: Quando un datore di lavoro vuole assumere un lavoratore straniero o quando uno straniero vuole entrare in Italia per lavorare.

Fonte: Ns. elaborazione

1.6. Valutazione dei risultati

1.6.1. Rispetto dei parametri

Le glosse create dai modelli linguistici sperimentati sono state, anzitutto, valutate sulla base del rispetto o meno delle istruzioni impartite nel prompt. La tabella che segue riporta per ogni glossa creata dai diversi modelli il rispetto o meno delle istruzioni date, ovvero se la glossa ha risposto alle domande che cos'è, a cosa serve, cosa fa, in quali casi la usi (cioè se ha fornito esempi d'uso), se sono state utilizzate massimo 44 parole per l'intero testo e massimo 25 per ogni frase, se ogni frase esprime un solo concetto e se sono stati utilizzati il presente indicativo per i verbi, il "tu" rivolgendosi dialogicamente al lettore e solo parole fondamentali del vocabolario di base della lingua italiana di Tullio De Mauro (Fig. 1). Nel caso in cui l'istruzione sia stata rispettata, è stata flaggata la casella corrispondente. In base agli stessi parametri è stato così possibile anche comparare i diversi modelli AI utilizzati. Nella tabella successiva (Tab. 2) sono mostrati in sintesi i risultati relativi al rispetto delle istruzioni impartite da parte di ogni singolo modello AI (il valore 5 rappresenta il rispetto della singola istruzione in tutte le 5 glosse redatte dal modello).

I risultati hanno mostrato che i modelli linguistici utilizzati sono stati in grado di redigere glosse che rispondessero alle domande richieste, rispettassero un numero massimo di parole per frase e utilizzassero solo verbi al presente indicativo, rivolgendosi all'utente con l'uso del "tu" in modo dialogico.

Nessun modello, invece, è stato in grado di generare testi contenenti solo termini appartenenti al lessico fondamentale del Vocabolario di base, cioè le parole del Vocabolario di base di De Mauro marcate come FO (e quindi l'uso nelle glosse anche di un solo termine con diversa marca d'uso è stato considerato come errato). Ad esempio, le parole 'amministrativo', 'attestare, 'accedere', 'cittadinanza',

'infrastruttura', 'professionista', 'provvedimento', 'registro', sono parole di alto uso del vocabolario di base (marcate come AU).

Fig. 1 Dettaglio del rispetto delle istruzioni

Termine	GPT3.5	GPT4o	Llama3.1 8b	Llama3.1 70b
anagrafe	☑ Risposte alle domande date. ☑ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☑ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☐ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☐ Max 44 parole per glossa. ☐ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.
assistente sociale	☑ Risposte alle domande date. ☑ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☑ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☐ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☑ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.
carta d'identità	☑ Risposte alle domande date. ☑ Max 44 parole per glossa. ☐ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☑ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☐ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☐ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.
comune	☑ Risposte alle domande date. ☑ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☐ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☐ Max 44 parole per glossa. ☐ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☐ Max 44 parole per glossa. ☐ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.
decreto flussi	☑ Risposte alle domande date. ☑ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☐ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☑ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.	☑ Risposte alle domande date. ☐ Max 44 parole per glossa. ☑ Un solo concetto per frase. ☑ Max 25 parole per frase. ☑ Verbi al presente indicativo. ☑ Uso del "tu". ☐ Solo lessico fondamentale.

Fonte: Ns. elaborazione

Talora sono state usate parole del lessico comune e anche del lessico burocratico, come nei casi dei termini 'certificare' 'governativo', 'regolamentazione' e 'anagrafe' (che era uno dei termini da glossare ed è stato usato in un caso come esempio dei servizi che eroga il comune).

Tab. 2 Rispetto delle istruzioni in sintesi per ogni LLM

ISTRUZIONE	GPT 3.5	GPT 4o	Llama 3.1 8b	Llama 3.1 70b
Risposta alle domande date	5	5	5	5
Max 44 parole per glossa	5	3	1	1
Un solo concetto per frase	4	5	4	2
Max 25 parole per frase	5	5	5	5
Verbi al presente indicativo	5	5	5	5
Uso del "tu"	5	5	5	5
Solo lessico fondamentale	0	0	0	0

Fonte: Ns. elaborazione

L'istruzione che imponeva un numero massimo di parole alla glossa è stata rispettata solo da GPT 3.5. Gli altri modelli hanno realizzato anche glosse superiori alle 44 parole (in particolare vi sono state glosse composte da 48, 50, 55 e 57 parole).

Anche la regola di esprimere un solo concetto per ciascuna frase è stata rispettata totalmente solo da uno dei modelli, ovvero GPT 4o. Gli altri modelli, invece, hanno rilevato difficoltà, come nel caso della glossa di 'anagrafe' dove nella frase 'cosa fa?' il modello risponde "identifica i residenti e fornisce informazioni".

1.6.2 Correttezza e appropriatezza dei contenuti

In tema di valutazione della comunicazione giuridica assume particolare valore il parametro della correttezza delle informazioni fornite. Le glosse generate dai diversi modelli AI sono state valutate anche sotto questo punto di vista, quindi è stata verificata l'esattezza delle affermazioni contenute nei testi prodotti.

Inoltre, è stata valutata anche l'appropriatezza delle glosse redatte dai modelli rispetto al risultato complessivo atteso dagli autori, ovvero il fine, non direttamente espresso all'interno del prompt, ma certamente sottinteso alla richiesta espressa: ottenere un utile supporto da parte della AI per la redazione di glosse semplificate che migliorassero la comprensibilità di un termine amministrativo da parte degli utenti di testi istituzionali.

In linea generale è possibile affermare che non sono state evidenziate informazioni del tutto sbagliate, piuttosto in alcuni casi le glosse contengono affermazioni non appropriate. Ad esempio, nella glossa del termine "carta di identità", redatta da Llama 3.1 8b, alla domanda "cosa fa?" (domanda in realtà non pertinente), la risposta "ti permette di accedere a servizi come votazione, viaggio o lavoro" (vedi Tab. 1) non è propriamente corretta, in quanto il termine "servizio" non è corretto rispetto all'esercizio del diritto di voto, all'accesso al lavoro o alla libertà di viaggiare.

In egual modo pare inappropriato l'esempio utilizzato nella glossa di 'Comune', redatta da GPT 3.5, laddove risponde alla domanda sugli esempi di utilizzo con "per […] accedere ai servizi comunali".

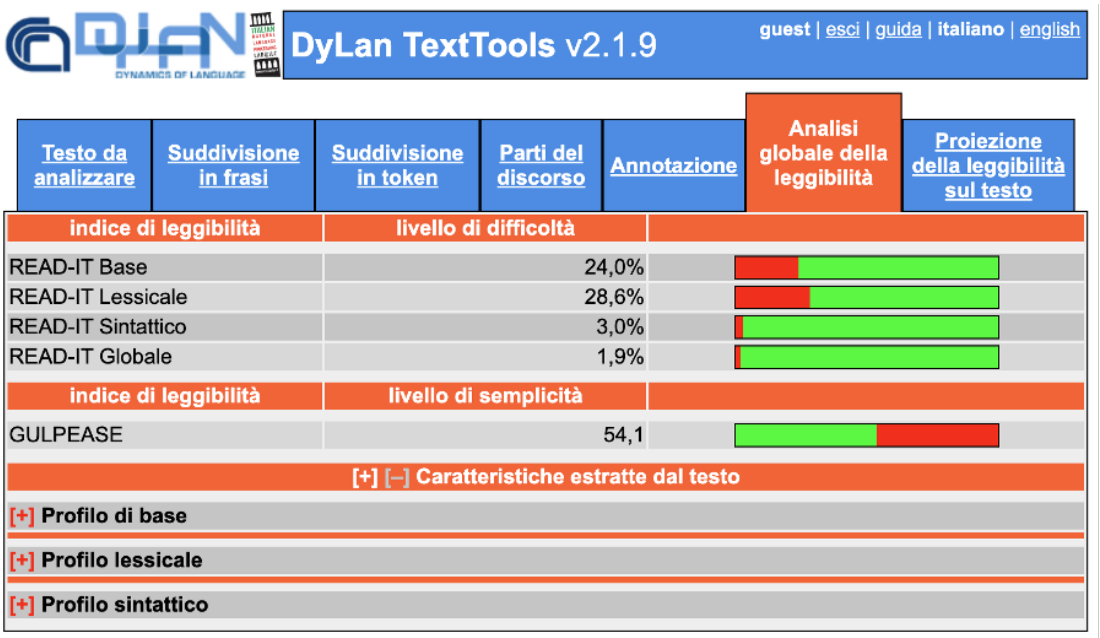
Dal punto di vista della appropriatezza, si può affermare che tutti i modelli linguistici sperimentati forniscono un utile supporto di base per la redazione di glosse semplificate dei termini amministrativi, con una maggior eccellenza del modello Llama 3.1 70b.

Allo stato di questa sperimentazione, infatti, questo modello sembra fornire risposte più specifiche e accurate, in grado di rappresentare la complessità del linguaggio, in termini di relazione tra parole, concetti e contesti.

1.6.3. Valutazione della leggibilità con READ-IT

Abbiamo valutato la leggibilità delle glosse utilizzando DyLan TextTools v2.1.9, cioè la versione online del tool per l'analisi della leggibilità dei testi sviluppato da CNR-ILC e conosciuto come READ-IT[8] (Fig. 2).

Fig. 2 Esempio di una schermata dell'interfaccia utente di Read-IT



Fonte: Ns. elaborazione

Secondo quanto descritto nel manuale di READ-IT[9], la valutazione della leggibilità del testo viene condotta sulla base di diverse caratteristiche del testo stesso.

Gli indici di leggibilità sono 5:

- READ-IT Base: le caratteristiche considerate sono quelle tipicamente usate nelle misure tradizionali della leggibilità, ovvero la lunghezza della frase e la lunghezza delle parole;
- READ-IT Lessicale: si focalizza sulle caratteristiche lessicali del testo, costituite dalla composizione del vocabolario e dalla sua ricchezza lessicale;
- READ-IT Sintattico: si basa su informazione di tipo grammaticale, ovvero sulla combinazione di tratti morfo-sintattici e sintattici;
- READ-IT Globale: si tratta di una combinazione degli indici precedenti;
- indice GULPEASE, specificamente concepito per la lingua italiana.

Le percentuali esprimono il livello di difficoltà, ovvero si riferiscono alla probabilità di appartenenza del testo alla classe dei testi di difficile leggibilità. Le barre colorate a fianco esprimono visivamente questi valori, dove il rosso rappresenta la probabilità di appartenenza alla classe dei testi difficili e il verde a quelli di facile lettura.

[8] Lo strumento è consultabile al seguente link: https://www.ilc.cnr.it/dylanlab/apps/texttools/?tt_user=guest. Sul suo funzionamento anche a fini di analisi su documenti giuridici si veda Brunato, Venturi 2016.

[9] Il manuale è consultabile al link http://www.italianlp.it/wp-content/uploads/2016/01/Documentazione-READ-IT.pdf

Riportiamo sotto una tabella contenente gli indici READ-IT per le glosse create manualmente (gold standard) e per tutte le glosse generate automaticamente.

Tab. 3 Indici READ-IT per le glosse create manualmente e per le glosse generate automaticamente

		READ-IT (%)			GULPEASE	Profilo di base				Profilo lessicale	
		Base	Sintattico	Globale		Periodi	Parole	Lung. media periodi	Lung. media parole	Vocabolario di Base (%)	Densità
Anagrafe	Gold standard	20,40%	**5,60%**	**6,10%**	54,5	3	56	18,7	4,7	81,80%	**0,532**
	GPT 3.5	5,10%	89,10%	42,10%	55,6	3	41	13,7	5,2	81,80%	0,656
	GPT 4o	**4,40%**	20,80%	32,10%	67	4	56	14	4,2	**89,70%**	0,622
	Llama 3.1 8b	**4,40%**	9,90%	15,60%	**67,5**	5	61	12,2	4,3	88,60%	0,545
	Llama 3.1 70b	6,10%	79,70%	91,50%	57	4	57	14,2	5,2	84,20%	0,6
Assistente sociale	Gold standard	43,60%	0,00%	0,00%	47,6	1	32	32	4,1	86,40%	**0,517**
	GPT 3.5	6,10%	24,70%	**0,50%**	50,5	3	37	12,3	5,6	**92,30%**	0,606
	GPT 4o	**3,90%**	7,60%	38,40%	**65,3**	4	51	12,8	4,5	86,50%	0,571
	Llama 3.1 8b	3,70%	**4,90%**	61,70%	62,6	5	56	11,2	5	85,70%	0,6
	Llama 3.1 70b	4,60%	6,70%	38,10%	59	4	56	14	5,1	81,10%	0,619
Carta di identità	Gold standard	1,00%	13,90%	33,00%	80,2	6	53	8,8	4,5	83,80%	0,571
	GPT 3.5	**2,80%**	14,80%	8,90%	58,3	3	31	10,3	5,2	87,50%	0,536
	GPT 4o	4,30%	**9,00%**	**5,70%**	67,2	4	55	13,8	4,1	88,60%	0,6
	Llama 3.1 8b	3,30%	16,10%	45,20%	**69,4**	5	57	11,4	4,3	**90,00%**	**0,529**
	Llama 3.1 70b	8,00%	23,30%	25,90%	58,2	4	62	15,5	4,9	85,70%	0,592
Comune	Gold standard	**0,80%**	**2,20%**	**6,00%**	**82,4**	6	42	7	4,5	87,10%	**0,579**
	GPT 3.5	4,60%	83,80%	78,80%	53,8	3	39	13	5,5	73,30%	0,645
	GPT 4o	8,80%	6,10%	11,00%	60,3	4	62	15,5	4,6	87,80%	0,635
	Llama 3.1 8b	3,70%	24,00%	95,60%	65,5	5	57	11,4	4,7	86,00%	0,608
	Llama 3.1 70b	14,60%	55,60%	24,10%	56,4	4	69	17,2	4,8	**89,40%**	0,614
Decreto Flussi	Gold standard	36,00%	**0,40%**	12,50%	40,1	1	21	21	4,8	85,00%	0,611
	GPT 3.5	9,30%	79,10%	**7,90%**	50,7	3	38	12,7	5,6	76,00%	**0,556**
	GPT 4o	7,60%	5,60%	17,60%	**59,8**	4	57	14,2	4,8	**85,70%**	0,612
	Llama 3.1 8b	**5,80%**	6,60%	43,90%	63	5	62	12,4	4,7	82,60%	0,596
	Llama 3.1 70b	17,90%	13,30%	10,00%	51,5	4	68	17	5,3	78,00%	0,589

Fonte: Ns. elaborazione

Abbiamo evidenziato in grassetto i risultati migliori per ogni glossa: nel caso del GULPEASE, si tratta di un livello di semplicità e quindi è migliore un numero più alto, così come nella percentuale di vocabolario di base; in tutti gli altri casi, si tratta di un livello di difficoltà e quindi è migliore una percentuale più bassa.

I dati in tabella non evidenziano un LLM decisamente migliore rispetto agli altri. Tutte le glosse generate automaticamente, infatti, presentano una buona leggibilità. Quelle generate con GPT 3.5 hanno spesso indici sintattici peggiori rispetto alle altre, ma non in tutti i casi.

Nella tabella non abbiamo riportato i valori dell'indice READ-IT Lessicale perché i testi analizzati sono molto brevi e i valori sono poco significativi, ma sono stati evidenziati i dati relativi all'uso del Vocabolario di Base e alla densità, che mostrano un profilo lessicale buono (una buona leggibilità).

1.7. Conclusioni

I risultati di questa sperimentazione hanno evidenziato alcune criticità nella creazione delle glosse da parte dei modelli linguistici, legate soprattutto alla incapacità di utilizzare esclusivamente il lessico fondamentale del Vocabolario di Base e al mancato rispetto di alcune istruzioni (numero massimo di parole per glossa e un solo concetto per frase).

Nonostante ciò, tali risultati sembrano essere di sicuro interesse per chi voglia utilizzare anche questi strumenti di intelligenza artificiale per la creazione di contenuti informativi della pubblica amministrazione chiari e comprensibili. E risultano, ancor più rilevanti, tenendo conto del fatto che i modelli sono stati utilizzati in modalità *zero shot*, ovvero senza fornire al modello alcun esempio su cui basare la risposta, e senza aver effettuato alcun tipo di *fine tuning*, ovvero senza aver raffinato il modello per rispondere alla specifica richiesta di redazione di glosse.

In futuro sarà certamente utile diffondere l'utilizzo di queste tecnologie sempre in funzione di supporto dell'operatore umano e proseguire nelle sperimentazioni, utilizzando tecniche tese a migliorare precisione e robustezza dei modelli linguistici.

Sarà, inoltre, indispensabile mettere a fattore comune le varie sperimentazioni, anche creando gruppi di ricerca transdisciplinari che uniscano le competenze degli esperti di *prompt design* con quelle di linguisti e giuristi.

In particolare i giuristi sono indispensabili in questi team di ricerca, sia in fase di progettazione dei prompt sia in fase di valutazione dei risultati, che non possono essere considerati unicamente con metriche quantitative (Ash *et al.,* 2024).

Infine, dovrà essere considerata attività imprescindibile nei processi di sperimentazione anche la valutazione da parte degli utenti o di altri stakeholder, in quanto il *feedback* dei destinatari dell'informazione da semplificare è un punto essenziale in qualsiasi attività di *legal design*. Attività che deve essere considerata come un processo iterativo, in grado di inviare segnalazioni per la revisione continua durante la creazione del testo semplificato (Imperiale, De Muro, 2021).

In conclusione, la redazione di testi informativi istituzionali chiari e comprensibili, che rientra tra le attività definite come "content curation" e che "consiste nel trovare, curare e condividere contenuti e risorse utili, per risolvere un problema specifico degli utenti", anche rendendo "i contenuti facilmente esperibili attraverso l'uso di canali di comunicazione appropriati", può sicuramente essere assistita

dalle macchine, pur dovendo rimanere "una attività umana: un lavoro costante e certosino" (Talamo, Zarriello, 2022).

E l'intelligenza artificiale generativa può costituire una "grande occasione per dare nuova centralità nelle politiche delle amministrazioni alla comunicazione e informazione pubblica" (Di Costanzo, Bonaventura, 2021).

Bibliografia

Ash, E., Kesari, A., Naidu, S., Song, L., Stammbach, D. (2024). Translating Legalese. Enhancing Public Understanding of Court Opinions with Legal Summarizers. In *CSLAW '24,* March 12–13. Boston, MA, USA, pp. 137-157.

Baldo, G. (2019). *Italiano per stranieri: semplificare, facilitare, adattare manuali scolastici disciplinari.* Alessandria: Edizioni dell'Orso.

Brunato, D., Venturi, G. (2016). Le tecnologie del linguaggio per la leggibilità della comunicazione istituzionale. In S. Panizza (a cura di), *Profili attuali di qualità degli atti normativi e amministrativi.* Pisa: University Press.

Castelli, C., Piana, D. (2019). *Giusto processo e intelligenza artificiale*. Santarcangelo di Romagna: Maggioli.

Cherubini, M., Romano, F., Bolioli, A., De Francesco, N., Benedetto, I. (2023). La summarization di testi giuridici: una sperimentazione con GPT-3. In *Rivista italiana di informatica e diritto, 5*(1), pp. 191-204.

Cherubini, M., Romano, F., Bolioli, A., De Mattei, L. (2024). Improving the accessibility of EU laws: the Chat-EUR-Lex project. In *Proceedings of the Ital-IA Intelligenza Artificiale (Ital-IA 2024),* pp. 54-59.

Cortelazzo, A. M. (2021). *Il linguaggio amministrativo. Principi e pratiche di modernizzazione.* Roma: Carocci.

De Mauro, T. (2003). *Guida all'uso delle parole (ed. 12)*. Roma: Editori Riuniti.

Di Costanzo, F., Bonaventura, D. (2021). *Digitale: la nuova era della comunicazione e informazione pubblica. Storia e prospettive del modello italiano.* Milano: Giunti.

DG, EPRS – European Parliamentary Research Service (2019). *Understanding algorithmic decision-making: Opportunities and challenges.*

Fioravanti, C., Romano, F., Torchia, M. C. (2022). Terminologia giuridica e inclusione: un glossario digitale semplificato in materia di immigrazione. In E. Chiocchetti, N. Ralli (a cura di), *Risorse e strumenti per l'elaborazione e la diffusione della terminologia in Italia* (pp. 90-101). Bolzano: Eurac Research.

Fioravanti, C., Rinaldi, M. (2010). Il sistema informativo PAeSI: un accesso telematico unico a informazioni, norme e procedimenti in materia di immigrazione. In *Informatica e diritto,* 1-2, pp. 93-131.

Giardiello, G., Romano, F., (2023). Comunicare servizi e procedure dei Comuni sul web: considerazioni e proposte. In *Ciberspazio e Diritto, 24*(1), pp. 103-120.

ILC-CNR, ItaliaNLP Lab (2016). *READ-IT. Documentazione Demo online.* Disponibile su: http://www.italianlp.it/wp-content/uploads/2016/01/Documentazione-READ-IT.pdf

Imperiale, M., De Muro, B. (2021). *Legal design.* Milano: Giuffrè.

Lombardi, G. (2022). *Capire i documenti. Strategie didattiche e linguistiche per agevolare la comprensione dei testi istituzionali in lingua seconda.* Firenze: Franco Cesati.

Lubello, S. (2022). *Il diritto da vicino. Intorno ad alcune parole giuridiche dell'italiano.* Alessandria: Edizioni dell'Orso.

Marvin, G., Hellen, N., Jjingo, D., Nakatumba-Nabende, J. (2024). Prompt Engineering in Large Language Models. In I. J. Jacob, S. Piramuthu, P. Falkowski-Gilski (a cura di), *Data Intelligence and Cognitive Informatics. Proceedings of ICDICI 2023.* Singapore: Springer.

Riediger, H., Galati, G. (2021). *Scrivere chiaro, scrivere semplice.* Milano: Editrice bibliografica.

Rossi, F. (2019). *Il confine del futuro. Possiamo fidarci dell'intelligenza artificiale?* Milano: Feltrinelli.

Senato della Repubblica – Servizio Studi, Camera dei Deputati – Ufficio Rapporti con l'Unione Europea (2020). *Dossier Conferenza interparlamentare "Legiferare meglio da una prospettiva digitale*.

Talamo, S., Zarriello, R. (2022). *Nuovo manuale di comunicazione pubblica: comunicare la Pubblica amministrazione: teorie, tecniche e buone pratiche digital e social.* Roma: Centro di documentazione giornalistica.

Zaccaria, G. (2022). *Postdiritto. Nuove fonti, nuove categorie.* Bologna: Il Mulino.

2

I DATI, PUR ESSENDO QUELLO CHE CI SEMBRANO, NON CESSANO MAI DI ESSERE QUELLO CHE SONO*

di Stefano Borgo (CNR-ISTC)

2.1. Introduzione

Tutti noi ci affidiamo a degli 'schemi concettuali' per interpretare il mondo che ci circonda. Questi schemi sono delle semplificazioni che, fissando l'attenzione su alcuni fattori e ignorandone altri, ci permettono (o ci illudono) di 'capire' come le cose che ci stanno intorno cambiano. Nel selezionarli siamo influenzati, quando non determinati, da elementi molto diversi tra loro: la nostra biologia, le nostre capacità cognitive, la nostra esperienza sociale e l'ambiente culturale in cui viviamo. Questi schemi sono utili per mettere ordine in un mondo complesso e offrono una lettura, che è certamente selettiva ma anche rassicurante, del mondo che ci è intorno e di come cambia nel tempo. Sono utili sia per situazioni apparentemente semplici (ad es. per stabilire che quando il nostro gatto è dietro al divano non cessa di esistere anche se non percepiamo la sua presenza), sia per relazioni anche molto complesse (ad es. per stabilire che a causa dell'inflazione il denaro che abbiamo a disposizione permette di acquistare meno cose man mano che il tempo passa). Ogni volta che usiamo uno di questi schemi (spesso inconsapevolmente), pensiamo che quello che ci spiegano sia 'normale', cioè stia nella natura delle cose. Questo non significa che siamo contenti o meno di quello che accade, ma che sappiamo spiegare (o almeno ci illudiamo di aver capito) le relazioni tra le cose che portano a certi cambiamenti nel mondo.

* Questa nota riassume la presentazione divulgativa dal titolo "Intelligenza artificiale tra dati, simboli e modelli" tenuta il 14 marzo 2024 durante l'evento *Dialoghi sull'intelligenza artificiale*.

Gli schemi concettuali non sono dati una volta per tutte. I cambiamenti che osserviamo nel tempo possono spingerci ad aggiornarli e ad aggiungerne di nuovi al fine di tener conto delle novità, specialmente se hanno conseguenze durature. L'esistenza e il possibile uso di nuovi strumenti tecnologici, dagli smartphone ai monopattini elettrici, è uno di questi casi. Alcune innovazioni, come l'introduzione dei moderni sistemi di viaggio (aerei e monopattini) e di comunicazione (telefono e internet), arrivano ad alterare il nostro modo di pensare il mondo e il significato di nozioni che pensiamo di "senso comune". Questo succede con il significato di distanza quando impariamo l'esistenza di aerei e monopattini, e quello di comunicazione con il telefono e internet.

Sarebbe interessante analizzare l'impatto che gli sviluppi in Intelligenza Artificiale (IA) degli ultimi 15 anni hanno avuto e stanno avendo sui nostri schemi concettuali: i nuovi strumenti che tecnologie come Machine Learning (ML) e Large Language Models (LLMs) hanno reso possibili, richiedono nuovi modi di spiegarci come cambia il mondo. Come sostiene Romele, "queste nuove tecnologie non stanno solo trasformano radicalmente le nostre interazioni con il mondo, o i nostri modi di produzione e consumo, ma anche la nostra visione del mondo" ([Romele, 2023], tda)[1].

Questo breve articolo si concentra su un aspetto di questa relazione complicata tra noi e il mondo, e mira a mostrare che gli strumenti innovativi forniti dall'IA, come ChatGPT[2] e DeepSeek[3], e il mondo esterno ripropongono vecchi problemi che i nostri schemi concettuali ancora faticano a considerare o, in altre parole, che noi tendiamo ad ignorare. In sostanza, il modo in cui usiamo i dati oggi (ad es. relativamente alla loro generazione e al loro significato) non è cambiato rispetto a quando l'IA non condizionava il nostro modo di interagire con il mondo. Oggi come ieri costruiamo strumenti, anche molto sofisticati, che usano i dati senza preoccuparsi troppo di cosa effettivamente sono.

Invitiamo il lettore ad analizzare questo aspetto perché le innovazioni portate dall'IA, per quanto importanti, sono talvolta amplificate da forzature, forzature che ben spiegano alcuni casi di 'inaffidabilità' dei nuovi sistemi. Quindi, lasciando a lavori più tecnici l'analisi della qualità degli algoritmi e della trasparenza nella gestione dell'informazione nell'IA, su cui giustamente la comunità scientifica sta lavorando molto, qui rivolgiamo la nostra attenzione ai pericoli a cui ci esponiamo in tutti quei casi in cui i dati vengono usati per quello che ci sembrano senza pensare troppo a quello che effettivamente sono.

2.2. IA simbolica e IA data-driven

Nel parlare di dati è utile ricordare una classica distinzione tra tipologie di IA, quella tra IA simbolica (o semantico-cognitiva) e IA data-driven, cioè guidata dai dati (detta anche sub-simbolica) (Calegari *et al.*, 2020). Con IA data-driven ci si riferisce a quegli algoritmi che essenzialmente utilizzano grandi basi di dati (per semplificare, pensiamole come lunghe liste di valori alfa-numerici) che riportano valori del mondo di interesse. Un sensore di passaggio installato su una strada

[1] "[T]hese new technologies are not only radically transforming our interactions with the world, or our modes of production and consumption, but also our worldview" (Romele, 2023).

[2] chatgpt.com.

[3] www.deepseek.com

invia messaggi che incrementano un contatore (ad indicare quanti mezzi di trasporto sono passati da un certo momento in avanti); un tasto digitato sulla tastiera aggiunge una lettera a una sequenza di altre lettere e spazi (che forma un'espressione linguistica). L'IA data-driven utilizza i valori del contatore e la sequenza digitata nel tempo ma non considera, né ha gli strumenti per farlo, il tipo di sensore, dove sta, come funziona o perché lo usiamo. Analogamente, l'IA data-driven considera la sequenza di lettere che viene digitata senza preoccuparsi se abbia un significato o meno. Questo genere di IA utilizza questi dati come meri numeri (o caratteri alfa-numerici), non li considera il risultato di un'interazione tra il sensore e il mondo esterno, cioè informazione che intenzionalmente chi ha messo il sensore vuole acquisire e chi ha usato la tastiera vuole creare.

L'IA simbolica, per sua natura, si concentra sui significati dei dati e quindi cerca (anche se non sempre coerentemente) di considerare la differenza tra avere un dato e averne un altro. Non solo acquisisce il valore numerico del sensore stradale, gli attribuisce anche un significato semantico, ad es., potrebbe associargli l'espressione "il numero di volte in cui il sensore si è attivato", oppure l'espressione ben più ricca "il numero di veicoli che passando lungo la strada hanno attivato il sensore durante la giornata". Il secondo caso è più informativo ma anche meno affidabile perché il sensore potrebbe essere attivato per cause non previste, ad es. a causa di interferenze elettromagnetiche. Magari segnala regolarmente in eccesso perché attivato anche al passaggio di pedoni, di oggetti trasportati dal vento, dal cambio di luminosità nell'ambiente; o tendenzialmente in difetto per una forte inerzia tra una attivazione e l'altra, perché se un'auto sosta di fronte impedisce di rilevare altre auto e così via. Fissare un'etichetta con cui interpretare i dati ci dà grandi vantaggi perché così stabiliamo il loro significato. Allo stesso tempo è una decisione delicata perché determina la qualità e l'affidabilità dell'informazione che raccogliamo. Questi esempi, volutamente semplici e quasi banali, suggeriscono quanto complicato sia controllare il modo in cui un sensore interagisce con il mondo che gli sta intorno e quindi il vero significato dei dati che genera.

Abbiamo visto che i sistemi di IA simbolica sono costruiti per utilizzare dati con una interpretazione predefinita. In altre parole, questi sistemi elaborano conoscenza.[4] Per questo i loro algoritmi utilizzano linguaggi costruiti per garantire che l'elaborazione computazionale tenga conto anche della semantica dei dati utilizzati. Ne segue che, se le informazioni di partenza sono corrette (o almeno affidabili), quello che l'algoritmo elabora (una statistica, una osservazione, una deduzione) è altrettanto corretto, e le decisioni che prende sono ottimali relativamente alle informazioni disponibili.

In contrasto, i sistemi di IA data-driven sono costruiti per elaborare dati senza preoccuparsi (anzi, talvolta volutamente ignorando) a cosa si riferiscono o come sono stati ottenuti. Quando questo accade l'elaborazione dei dati potrebbe seguire regole e procedure che non sono spiegabili da un punto di vista semantico. Questo non va considerato negativamente, in genere si tratta di strumenti alimentati con dati opportunamente selezionati oppure che soddisfano determinate proprietà statistiche.

Immaginiamo che il sensore stradale si sia attivato 5 volte ieri e 15 volte oggi. Il sistema di IA data-driven riceve i dati (5 e 15), elabora il valore medio (10) e for-

[4] Il significato e la distinzione tra dato, informazione e conoscenza è un tema controverso, vedi Zins 2007.

nisce la seguente informazione: il valore medio dei dati forniti è 10. Questo numero non è associato a qualcosa nel mondo, per quanto ne sa il sistema potrebbe riferirsi ad automobili, a temporali o a galassie. È compito nostro capire a cosa si riferisce la risposta andando a vedere quali dati sono stati forniti al sistema. Il sistema di IA simbolica riceve i dati (5 e 15) insieme alla loro descrizione, elabora il valore medio (10) e fornisce la seguente informazione: il sensore segnala una media di 10 passaggi giornalieri di veicoli (oppure, mediamente il sensore si attiva 10 volte). Il sistema IA simbolico ha ricevuto gli stessi dati dal sensore ma li utilizza come osservazioni che ci dicono qualcosa di preciso sul mondo: la presenza dell'etichetta permette di interpretare il numero come un'affermazione sul passaggio di veicoli (o sull'attivazione del sensore) in un certo intervallo.[5]

Oggi ci affidiamo a strutture flessibili, dinamiche e complesse per gestire il traffico nelle città, le telecomunicazioni e così via. Le situazioni che gestiscono possono cambiare velocemente e danno origine a molti dati che sono indispensabili per mantenerle in funzione e per coordinarle. L'approccio simbolico è complesso da sviluppare e talvolta manca della flessibilità necessaria per gestire situazioni così dinamiche e variegate. Per questo ed altri motivi (di natura computazionale, tecnologica ed economica), in alcuni casi l'approccio data-driven è l'unica soluzione che abbiamo. In altri casi, ad es. nella classificazione di immagini o elaborazione di testi, non conosciamo soluzioni simboliche alternative che siano adeguate alle nostre esigenze. La ricerca in IA oggi non mira a fare una scelta tra simbolico e data-driven. Piuttosto cerca di integrare il più possibile questi due approcci in modo da utilizzarli al meglio. I sistemi di IA ottenuti in questo modo sono chiamati *ibridi*.

2.3. IA e i dati

Ogni sistema IA elabora una soluzione ad un problema tramite l'elaborazione dei dati che gli sono forniti. Come abbiamo visto e semplificando molto, i dati possono essere stati raccolti in basi di dati (ad es. una biblioteca in formato digitale o un elenco delle rilevazioni meteorologiche negli anni), essere forniti di volta in volta da sensori distribuiti nell'ambiente (ad es. il dato fornito in questo momento da un sensore meteorologico o del traffico), o direttamente da noi tramite strumenti come le tastiere del computer o applicazioni nei nostri smartphone.

Abbiamo visto che questi dati possono essere etichettati o meno. Se sono etichettati, dobbiamo assicurarci che queste etichette siano corrette. Se non lo sono (e qualche volta semplicemente non siamo in grado di farlo oppure farlo sarebbe troppo oneroso), dobbiamo assicurarci che siano adatti al sistema IA a cui chiediamo di elaborarli. In mancanza di questo, la risposta del sistema IA non sarà associabile al nostro problema, cioè non sarà una soluzione. In sostanza, sia che i dati siano etichettati sia che non lo siano, spetta all'utente del sistema IA decidere se usare il risultato, e la decisione parte dall'affidabilità dei dati usati. Anche se siamo certi che il sistema ha elaborato i dati in modo ineccepibile (e in alcuni casi possiamo dimostrarlo), resta compito

[5] La conoscenza del significato dei dati ha un impatto enorme sui sistemi IA, vedi ad es. https://www.technologyreview.com/2024/09/25/1104465/a-tiny-new-open-source-ai-model-performs-as-well-as-powerful-big-ones

dell'utente stabilire, almeno in linea di principio, se il risultato ottenuto è utilizzabile o meno.[6] Questo indipendentemente dal fatto che i dati siano raccolti in basi di dati o generati sul momento da sensori o tramite tastiere. In sostanza, quando abbiamo a che fare con sistemi dinamici e articolati, vogliamo assicurarci che tali sistemi siano affidabili. In modo simile, quando i dati che questi sistemi utilizzano sono dinamici e articolati, dobbiamo assicurarci che tali dati siano affidabili. Ma come possiamo assicurarci della affidabilità dei dati?

2.4. L'ontologia applicata per l'IA

Partiamo dal riconoscere che l'introduzione di sensori e l'accumulo di dati è un modo per trasformare il mondo, farlo diventare un sistema ancora più complesso. La natura di questa complessità non ci è sempre chiara. L'introduzione di sensori da un lato dipende da come concettualizziamo il mondo e dall'altro determina cosa vogliamo sapere e cosa decidiamo di ignorare di esso. È interessante notare che, cercando di raccogliere i dati che servono al nostro sistema di IA, cambiamo il mondo secondo un'idea che ci siamo fatti su come funziona. Infatti, la scelta di usare un certo sensore in un certo luogo e per un certo scopo ha un impatto anche sul significato dei dati forniti da quel sensore: il dato stesso non è un'entità indipendente da noi, è figlia del nostro modo di concepire e agire sul mondo. Per questo possiamo dire che *per capire i dati bisogna innanzitutto trovare gli schemi concettuali con cui abbiamo scelto di capire il mondo*. In altre parole, dobbiamo trovare gli schemi concettuali che ci hanno portato a fare certe scelte su cosa misurare, quali sensori usare e dove installarli.

Forniamo dati ai nostri sistemi di IA perché ci sembra che questi dati ci presentino fedelmente il mondo in quegli aspetti che ci interessano. Quindi usiamo i dati perché (e finché) corrispondono a come pensiamo sia il mondo. Tuttavia, non dobbiamo dimenticarci che i dati sono il frutto di una relazione tra uno strumento (un sensore) e una parte del mondo con cui lo strumento interagisce. Questa relazione è più complicata di quella che immaginiamo con i nostri schemi concettuali, del resto sappiamo che ogni schema è una semplificazione. Per questo quando le situazioni e i dati sono complessi e cambiano velocemente, è difficile riuscire a capirli correttamente. Negli ultimi decenni una parte della comunità scientifica ha cercato delle soluzioni a questo problema interessandosi dei *dati non più come oggetti da elaborare ma come affermazioni da capire*. Questo obiettivo ha portato ad integrare sia metodologie scientifiche sia analisi filosofiche. Oggi quest'area di ricerca è chiamata *ontologia applicata* e sviluppa sistemi concettuali, ad es. DOLCE (vedi Gaio, *et al.*, 2010, recentemente diventato uno standard ISO[7]), che forniscono una analisi integrata e coerente di come capiamo il mondo, di come dare significato ai dati che generiamo, e infine di come organizzare tutta questa informazione. In questo modo abbiamo anche compreso che migliorare i nostri modelli concettuali aiuta a costruire strumenti più affidabili e a verificare che siano usati correttamente.

[6] Spesso automatizziamo l'applicazione della soluzione, ad es. quando l'attivarsi di una serie di sensori porta a spegnere in modo automatico un macchinario. In questo caso, la scelta di fidarsi dei dati dei sensori avviene nel momento in cui decidiamo di automatizzare la regola "se i dati superano certi valori di soglia, disattivare il macchinario".

[7] https://www.iso.org/standard/78927.html

Inoltre, si è mostrato che l'utilizzo di metodologie sviluppate in ontologia applicata migliora la trasparenza semantica e cognitiva dei sistemi di informazione, rendendoli così accessibili all'analisi non solo degli esperti ma anche, di chi li usa. Il crescente interesse verso questo tipo di ricerca nella comunità dell'IA è un segnale importante sia per la costruzione di sistemi sempre più affidabili sia per un loro utilizzo ottimale.

Possiamo ora chiarire il titolo che abbiamo dato a questa nostra riflessione. È la parafrasi di un testo tratto dal diario di José Saramago del 15 Agosto 1998: "[L]e cose, pur essendo quello che ci sembrano, non cessano mai di essere quello che sono…" (Saramago, 1998). Abbiamo osservato che i dati sono un tipo di cose molto particolari, sono affermazioni sul mondo, e ci servono per alimentare i sistemi di IA. Oggi che le nostre decisioni sono prese da sistemi automatizzati che si basano sui dati, è particolarmente importante usare i dati per quello che vogliamo fare (quello per cui ci sembrano utili) tenendo però sempre a mente il loro vero significato, visto che *i dati, pur essendo quello che ci sembrano, non cessano mai di essere quello che sono.*

Bibliografia

Calegari, R., Ciatto, G., Omicini, A. (2020). On the inegration of symbolic and sub-symbolic techniques for XAI: A survey. In *Intelligenza Artificiale*, *14*(1), pp. 7-32.

Gaio, S., Borgo, S., Masolo, C., Oltramari, A., Guarino, N. (2010). Un'introduzione all'ontologia DOLCE. In *AIDAinformazioni: Rivista di Scienze dell'Informazione*, *28*(1-2), pp. 107-125.

Romele, A. (2023). The datafication of the worldview. In *AI & SOCIETY*, 38, pp. 2197-2206.

Saramago, J. (1998). *Diario dell'anno del Nobel. L'ultimo Quaderno di Lanzarote (traduzione di R. Desti)*. Collana I Narratori. Milano: Feltrinelli.

Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. In *J. Assoc. Inf. Sci. Technol.*, *58*(4), pp. 479-493.

3

INTELLIGENZA ARTIFICIALE ED ASSICURAZIONI: EVOLUZIONE E PROSPETTIVE[1]

di Antonio Coviello (CNR-IRISS)

3.1. Introduzione

Il settore assicurativo non appare famoso per l'entusiasmo e l'adozione precoce di nuove idee, ma gli assicuratori sostengono che i loro clienti abbracciano, piuttosto che rifiutare, il cambiamento.

La capacità dell'Intelligenza Artificiale (IA) di analizzare rapidamente grandi quantità di dati è uno strumento potente per gli assicuratori nella previsione e nella valutazione dei rischi, in particolare quando la fonte di dati è significativa, come nel caso dei rischi associati al cambiamento climatico, quali ad esempio i c.d. "rischi catastrofali" (Coviello, Somma, 2021).

L'IA apre a nuove prospettive per l'industria assicurativa, consentendo un approccio più personalizzato, predittivo ed efficiente alla valutazione del rischio, alla gestione delle polizze e dei sinistri e all'assistenza clienti. L'IA promette di trasformare radicalmente il modo in cui le assicurazioni comprendono, valutano e mitigano il rischio, offrendo soluzioni su misura per le esigenze individuali e collettive dei clienti (Deloitte, 2024).

L'IA consente sicuramente di rendere l'assicurazione più personale, in quanto il miglioramento dell'analisi e dei dati attuariali che l'IA è in grado di fornire agli assicuratori, permette loro di offrire coperture più personalizzate e su misura per le assicurazioni auto, vita, salute e casa, oltre che per le grandi linee di business commerciali (GILC, 2024).

[1] Si ringrazia per la preziosa collaborazione alla stesura dell'articolo il Dr. Fabrizio Morana, Direttore Generale del Centro Studi e Ricerche AssicuraEconomia https://assicuraeconomia.it/

Sono molteplici i vantaggi, quindi, che anche il settore assicurativo potrà ricevere dell'utilizzo dell'Intelligenza Artificiale. Dall'automazione dei processi alla personalizzazione dei prodotti, dall'offerta di assistenza per il benessere personale alla collaborazione con gli ecosistemi, l'intelligenza artificiale generativa cambierà le assicurazioni, così come farà con molte altre industrie. Essa potrà certamente migliorare la precisione nella sottoscrizione delle polizze, delineando più chiaramente e, soprattutto, più rapidamente il profilo del rischio che gli Assicuratori stanno valutando di assumere in relazione alle esigenze manifestate dalla clientela (cfr. economyup.it e insurtechitaly.com)[2].

3.2. L'evoluzione dell'IA nell'industria assicurativa

Agli albori dell'automazione, il settore assicurativo utilizzava sistemi basati su algoritmi basati su regole, che facilitavano processi come la sottoscrizione, la valutazione del rischio e la gestione dei sinistri. Tuttavia, questi sistemi richiedevano un notevole impegno manuale per lo sviluppo e la manutenzione, a causa della mancanza di apprendimento autonomo e di adattabilità. La curiosità del settore oggi è incentrata sull'IA basata sull'apprendimento automatico, caratterizzata dalla capacità di addestrare ampie serie di dati per riconoscere modelli e fare previsioni o prendere decisioni in modo autonomo, senza una programmazione specifica. Tuttavia, attualmente questo tipo di IA è utilizzato soprattutto per applicazioni ristrette e specifiche, come la classificazione di richieste e reclami (Swiss Re, 2024).

Secondo uno studio di Swiss Re (2024, cit.), il crescente utilizzo dell'Intelligenza Artificiale non porterà soltanto benefici, ma potrà aprire le porte anche a nuovi rischi. Comparando dati storici e *trend* attuali sono stati esaminati i pericoli derivanti dal maggiore uso dell'IA in dieci settori industriali diversi. Quello maggiormente colpito dagli effetti negativi della novità potrà essere il settore «sanità», minacciato da un utilizzo distorto dei dati, algoritmi errati, *bias* cognitivi e attacchi informatici. In seconda posizione, alle spalle della sanità, lo studio di Swiss Re indica il settore «mobilità e trasporti», seguito dal comparto «energia e erogazione e gestione servizi/*utilities*».

Proprio il settore sanitario e farmaceutico sarà quello che verosimilmente utilizzerà sempre più l'Intelligenza Artificiale per ottimizzare funzioni quali, ad esempio, amministrazione, monitoraggio dei pazienti, diagnosi o sviluppo di farmaci. L'indirizzo comporterà una maggiore esposizione ai rischi come, ad esempio, diagnosi errate, con conseguenti malattie gravi curate male fino ad arrivare a perdere la vita.

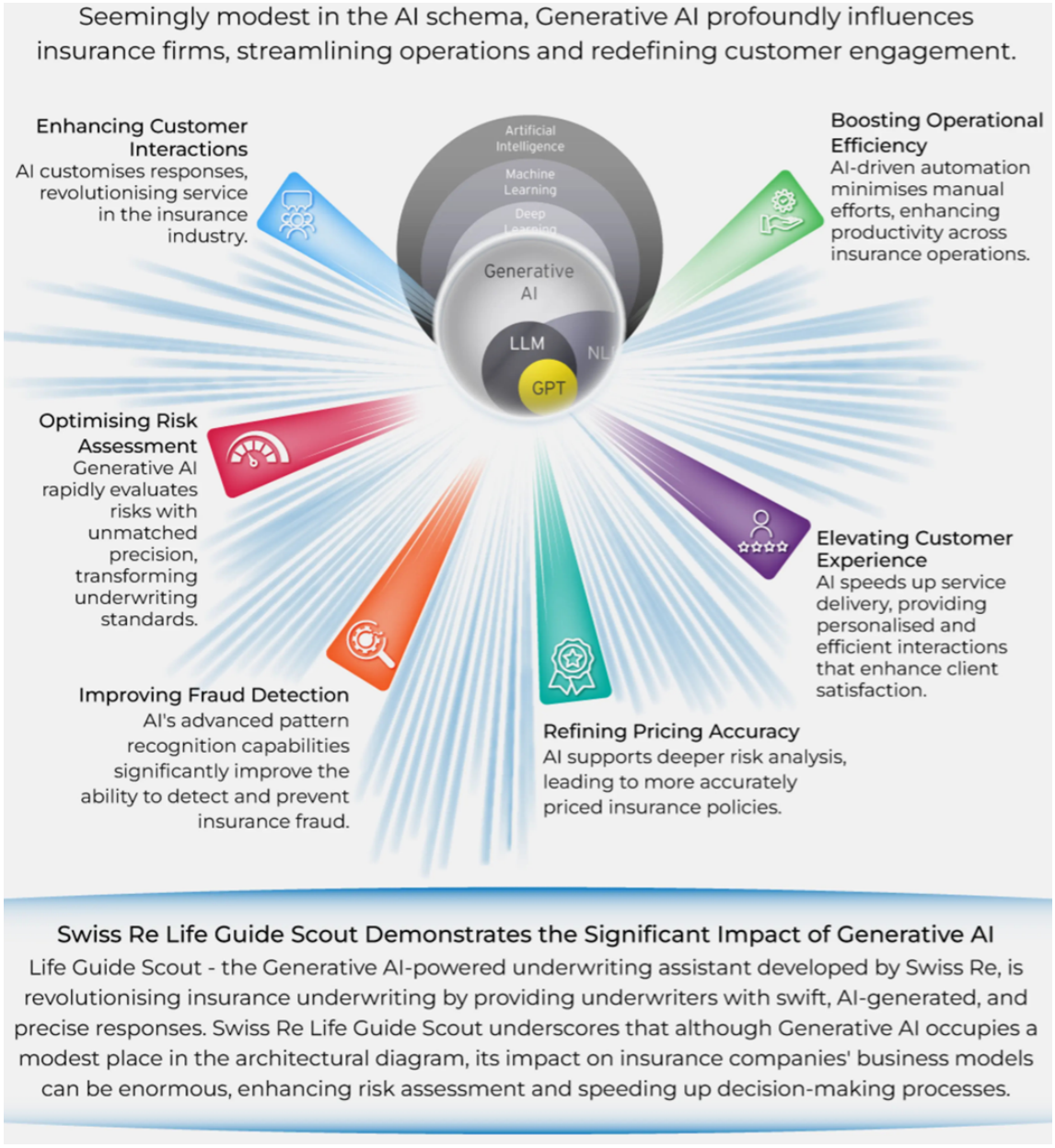
Nel settore «mobilità e trasporti» i rischi potranno crescere soprattutto a causa dell'uso della guida connessa e automatizzata alimentata dall'IA, che porrà delle sfide in contesti urbani molto diversi. Anche il comparto «energia e utility» utilizzerà ampiamente l'IA, soprattutto perché la transizione a zero emissioni attualmente in corso richiederà l'elettrificazione e la creazione di reti intelligenti.

Tale andamento comporterà per le imprese di assicurazione e riassicurazione la valutazione e l'assunzione di nuove tipologie di rischi, ad esempio trattando il com-

[2] Cfr. Articolo "AI & Insurtech" – IIA Italian Insurtech Association, Milano 12/4/2024 https://insurtechitaly.com/event/ai-e-insurtech/ . Vedi anche articolo "Assicurazioni e intelligenza artificiale: 5 ambiti di trasformazione", https://www.economyup.it/fintech/insurtech/assicurazioni-e-intelligenza-artificiale-5-ambiti-di-trasformazione/

parto sanitario nella telemedicina. Anzi, spesso lo si dovrà anche fare senza riferimento ai modelli statistici che normalmente sostengono l'attività delle compagnie nella determinazione del premio: si correrà il rischio, quindi, di essere costretti ad operare con «tariffe di fortuna». In ogni caso il sistema dovrà contribuire a costruire la fiducia digitale necessaria per sfruttare appieno il potenziale delle tecnologie emergenti (Swiss Re, 2024).

Fig. 1 Generative AI's Impact on Insurance Firms Beyond Its Graphical Silhouette



Fonte: EY, SwissRE I Infographic by Antonio Grasso in partnership with SwissRE

Gli assicuratori e i loro consulenti legali seguiranno da vicino l'evoluzione della normativa sull'IA ("AI Act") che in prima battuta nell'UE ha ricevuto nello scorso mese di marzo 2024 il voto positivo del Parlamento e, successivamente, l'appro-

vazione da parte del Consiglio il 21 maggio[3]. Diventerà il punto di riferimento per molte giurisdizioni in tutto il mondo e le compagnie dovranno essere molto attente alle esposizioni di responsabilità civile, *privacy* e *cyber* che potrebbero emergere quando i loro assicurati adotteranno l'IA nei propri modelli di *business.*

Gli assicuratori, in proposito, dovranno anche affrontare un rischio significativo per quanto riguarda la *privacy* dei dati, rischio che potrebbe essere aggravato dall'adozione diffusa dell'IA (Balasubramanian, Libarikian, 2021). L'elaborazione di grandi quantità di dati personali, spesso sensibili, comporterà la necessità di disporre di procedure solide per garantire la *compliance* agli standard nazionali e internazionali di protezione dei dati stessi. Ancora, gli assicuratori dovranno anche essere consapevoli della necessità di adottare misure di salvaguardia contro le violazioni e di disporre di processi adeguati allo scopo di segnalare e gestire eventuali criticità, in coerenza con le disposizioni del GDPR[4] (Swiss Re, 2024). Nel merito, l'applicazione del GDPR nelle agenzie assicurative richiede la definizione di un sistema di *data governance*[5] che consenta di far fronte ad un eventuale *data breach*[6]. È dunque importante che i ruoli privacy siano ben progettati per utilizzare legittimamente i dati raccolti.

Il Global Insurance Law Connect – GILC[7] nel suo primo «Rapporto sull'Intelligenza Artificiale» fornisce approfondimenti da 18 Paesi su come il nuovo sistema sta influenzando il settore assicurativo nei rispettivi mercati. Il rapporto analizza gli effetti trasformativi dell'IA, tra cui il miglioramento dell'efficienza e dell'innova-

---

[3] Il 13 marzo 2024, il Parlamento europeo ha approvato in via definitiva il Regolamento Europeo sull'intelligenza artificiale. Dopo un lungo iter legislativo, oggi il Regolamento è legge e tutte le aziende dovranno adeguarsi alle nuove norme dell'Unione, che, come tutti i regolamenti, sono *self executive* e dunque non richiedono alcuna legge di recepimento, entrando direttamente nel corpo legislativo nazionale di ciascun Stato membro. Con il voto del 13 marzo 2024, l'Unione europea è la prima al mondo a regolamentare la rivoluzione dirompente dell'intelligenza artificiale, diventando apripista e pioniera nella tecnologia che oggi promette di cambiare molte cose (…). (Fonte: https://www.diritto.it/regolamento-ia-approvato-accordo-provvisorio-ue/).

[4] Il Regolamento generale sulla protezione dei dati (GDPR, dall'inglese "General Data Protection Regulation") è il regolamento dell'Unione Europea che disciplina il modo in cui le aziende e le altre organizzazioni trattano i dati personali. Anche le società assicurative o le compagnie extra-europee, e di conseguenza le agenzie assicurative, che vendono polizze a cittadini dell'Unione Europea sono soggette all'applicazione del GDPR, ovvero al trattamento di dati personali effettuati da un titolare o da un responsabile con sede nell'Unione Europea, oltre che i trattamenti di dati personali posti in essere da un titolare o da un responsabile non stabilito nell'Unione, qualora essi riguardino l'offerta di beni o servizi o il monitoraggio di comportamenti di interessati che si trovano nel territorio dell'Unione. Il Regolamento Europeo prevede che ogni trattamento di dati personali effettuato dal Titolare deve basarsi e trovare il proprio fondamento all'interno di un'idonea base giuridica ovvero ciò che autorizza legalmente il trattamento dei dati personali, soddisfacendo il principio di liceità (art. 6 – GDPR).

[5] La "governance dei dati" comprende qualsiasi iniziativa volta a garantire che i dati siano sicuri, privati, accurati, disponibili e utilizzabili. Comprende le azioni da intraprendere, i processi da seguire e la tecnologia di supporto durante l'intero ciclo di vita dei dati. Per definizione, la *data governance* (governance dei dati) aziendali copre l'insieme delle policy e delle procedure implementate per far sì che i dati di un'organizzazione siano innanzitutto accurati, e quindi gestiti correttamente in ogni fase di inserimento, archiviazione, manipolazione, accesso ed eliminazione. Le mansioni in cui si espleta la governance dei dati riguardano la creazione dell'infrastruttura e della tecnologia, la configurazione e la manutenzione dei processi e delle policy, e l'identificazione delle persone (o delle posizioni) all'interno di un'organizzazione che hanno l'autorità e la responsabilità della gestione e della salvaguardia di tipologie specifiche di dati.

[6] Per "data breach" si intende una violazione di sicurezza che comporta – accidentalmente o in modo illecito – la distruzione, la perdita, la modifica, la divulgazione non autorizzata o l'accesso ai dati personali trasmessi, conservati o comunque trattati. Una violazione dei dati personali può compromettere la riservatezza, l'integrità o la disponibilità di dati personali. Alcuni possibili esempi sono, l'accesso o l'acquisizione dei dati da parte di terzi non autorizzati; - il furto o la perdita di dispositivi informatici contenenti dati personali; - la deliberata alterazione di dati personali; - l'impossibilità di accedere ai dati per cause accidentali o per attacchi esterni, virus, malware, ecc.; - la perdita o la distruzione di dati personali a causa di incidenti, eventi avversi, incendi o altre calamità; la divulgazione non autorizzata dei dati personali (fonte: Federprivacy).

[7] Il "Global Insurance Law Connect", network internazionale composto da avvocati specializzati in diritto assicurativo, ha lanciato il suo primo rapporto sull'intelligenza artificiale, che fornisce approfondimenti da 18 Paesi – compreso il commento australiano fornito da Sparke Helmore – su come l'intelligenza artificiale sta influenzando il settore assicurativo. Oltre a fornire dettagli sulle modifiche alla regolamentazione dell'IA e sulla previsione e l'analisi dei rischi, il rapporto rileva che l'IA continuerà a portare maggiore efficienza in molte aree del processo assicurativo, ma non è esente da sfide (Fonte: *https://www.insurancebusinessmag.com/nz/news/technology/global-insurance-law-connect-report-unveils-ais-dual-impact-on-insurance-industry-479509.aspx*).

zione in varie operazioni assicurative, ma evidenzia anche le sfide associate, come i potenziali pregiudizi negli algoritmi di IA, i problemi di privacy e l'aumento del rischio di incidenti informatici.

In particolare, il rapporto oltre a fornire dettagli sulle modifiche alla regolamentazione dell'IA e sulla previsione e l'analisi dei rischi, rileva che essa continuerà a portare maggiore efficienza in molte aree del processo assicurativo, pur non essendo la novità esente da sfide. Gli algoritmi, infatti, possono avere pregiudizi *(bias)* intrinseci e mancanza di trasparenza, sollevando preoccupazioni sulla *privacy* dei dati e su questioni etiche e, inoltre, l'uso dell'IA aumenta il potenziale di eventi dannosi legati all'ambiente *cyber.*

La stessa ricerca (GILC, 2024) ha evidenziato la capacità dell'IA di analizzare rapidamente grandi quantità di dati come un potente strumento per gli assicuratori nel prevedere e valutare i rischi, in particolare quando esiste una fonte significativa di dati. Infatti, vi si legge che *«l'uso dell'IA può aiutare gli assicuratori a entrare in mercati che potrebbero essere difficili a causa della mancanza di lunghe storie di sinistri per alcuni tipi di rischi. L'IA è in grado di digerire rapidamente grandi volumi di dati e di produrre analisi più precise, che possono essere utili, ad esempio, nella progettazione di coperture per incidenti informatici su larga scala. In ultima analisi questa migliore analisi del rischio andrà a vantaggio dei consumatori, in quanto consentirà agli assicuratori di offrire coperture più pertinenti e personalizzate ai loro clienti».* (GILC, 2024).

Nel settore commerciale, invece, l'Intelligenza Artificiale – ad ora – non sembra aver avuto grandi implicazioni, allorquando i nuovi algoritmi ancora non riescono a fare breccia all'interno di un processo di vendita/consulenza spesso condizionato, particolarmente nel ramo vita, da una prevalenza dell'aspetto emotivo sulla razionalità (Holland C. P., Kavuri A., 2021).

Le tecnologie emergenti maggiormente utilizzate nel settore assicurativo, ad oggi, risultano essere Robotica e *Intelligent Automation,* IOT e sviluppo parametriche. Ad emergere come le quella a maggior potenziale sono robotica e "IA generativa"[8], che consentono di efficientare processi ripetitivi e manuali con investimenti contenuti. Mentre nel caso della robotica il livello di maturità nell'implementazione tra le compagnie risulta già avanzato, essendo una tecnologia ormai conosciuta e utilizzata da tempo, è solo di recente che si sono viste le prime applicazioni dell'IA generativa al business assicurativo[9] (Cfr. Agendadigitale.eu; Studio *"Digital Disruption: l'impatto delle tecnologie emergenti sul settore assicurativo"* – EY e IIA, 2023)[10].

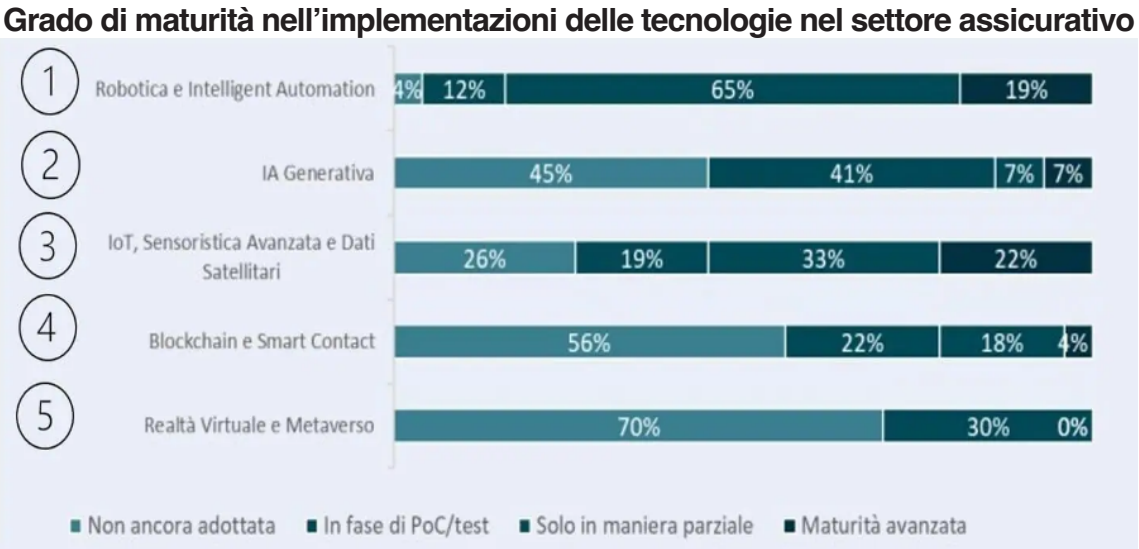
---

[8] Secondo il 44% degli intervistati dello studio EY-IIA (2023) la tecnologia emergente con maggiore capacità di applicazione nel settore è l'Intelligenza Artificiale Generativa. Sebbene il potenziale dell'utilizzo di questa tecnologia sia ampiamente riconosciuto dai player assicurativi, quasi la metà delle Compagnie intervistate non ha ancora adottato la AI Generativa all'interno del proprio business, e solo il 7% riferisce un grado di maturità avanzata nella sua implementazione, principalmente per la gestione dei sinistri.

[9] In particolare, il 96% degli intervistati della ricerca EY-IIA (2023) dichiara di aver già iniziato a utilizzare la Robotica e *l'Intelligent Automation* nella propria Compagnia. Le principali applicazioni di questa tecnologia riguardano la gestione dei sinistri (70%) e la sottoscrizione ed emissione delle polizze (54%). Il 74% degli intervistati ha dichiarato di aver già implementato, a diversi livelli di maturità, L'IoT e la sensoristica avanzata all'interno del proprio business per una più accurata quotazione del rischio (70%) e per lo sviluppo di prodotti parametrici (70%). In questo senso i benefici riscontrati riguardano principalmente l'estensione dell'offerta e il miglioramento della relazione con i clienti (67%).

[10] Cfr. Studio "Le assicurazioni nell'era della IA: così cambia il settore" (EY-IIA, 2023) https://www.agendadigitale.eu/mercati-digitali/le-assicurazioni-nellera-di-ia-e-robotica-tecnologie-e-trend-che-stanno-cambiando-il-settore/

Fig. 2 Classifica delle tecnologie emergenti per potenziale e relativi livelli di maturità di adozione nelle compagnie assicuratrici



Fonte: Agendadigitale.eu, 2023[11]

Secondo lo studio EY-IIA, al terzo posto della classifica delle tecnologie emergenti più rilevanti per il settore assicurativo, troviamo "IoT", dati satellitari e sensoristica avanzata. Segue al quarto posto la "blockchain", strumento a cui è riconosciuto il potenziale di poter supportare gli assicuratori nella fase di emissione e sottoscrizione delle polizze, nonché nello sviluppo di nuovi prodotti e nella gestione degli incassi e dei pagamenti[12].

In alcuni mercati l'IA probabilmente verrà adottata anche per ottimizzare i modelli commerciali. La pandemia COVID-19 ha accelerato il passaggio di molti assicuratori e distributori verso strumenti digitali e online, che accompagnano i modelli tradizionali (Coviello, Guazzone, 2021). È possibile che si assista a un'analoga rapida espansione dell'uso di tecniche digitali, comprese le *app* per *smartphone*, che spesso coinvolgono l'Intelligenza Artificiale con l'obiettivo di collocare polizze. La tendenza potrà essere particolarmente vantaggiosa nei mercati a bassa penetrazione assicurativa attuale, favorendo il miglioramento della cultura della popolazione nello specifico settore[13].

L'intelligenza artificiale potrà portare una nuova spinta propulsiva al settore, uno strumento molto utile per colmare il gap innovativo del mercato assicurativo, rispetto ad altri comparti molto più avanzati[14]. Proprio per questo, secondo le stime di IIA- Italian Insurtech Association, entro fine anno saranno investiti in soluzioni

[11] Cfr. Studio "Le assicurazioni nell'era della IA: così cambia il settore" (2023) https://www.agendadigitale.eu/mercati-digitali/le-assicurazioni-nellera-di-ia-e-robotica-tecnologie-e-trend-che-stanno-cambiando-il-settore/

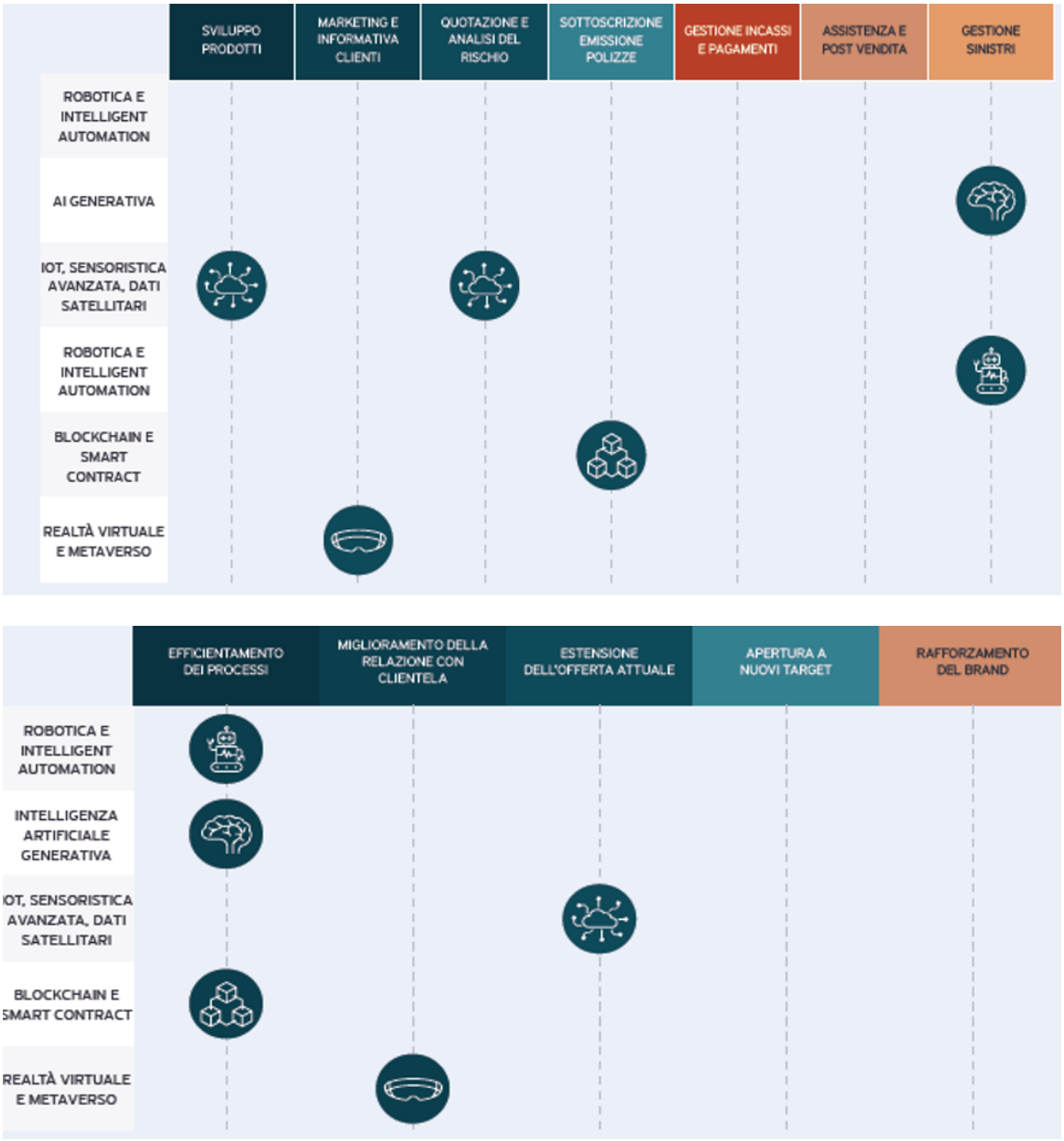
[12] Studio EY-IIA "L'impatto delle tecnologie emergenti sul settore assicurativo" (2023) https://www.assinews.it/07/2023/studio-ey-iia-limpatto-delle-tecnologie-emergenti-sul-settore-assicurativo/660105977/

[13] Attualmente il 50% dei consumatori è un "cliente digitale" e si prevede che, entro il 2030 rappresenterà oltre l'81%. Inoltre, sul piano del mercato professionale, nel triennio 2023-2025, è previsto l'inserimento di 25mila figure professionali come *cloud architect, system administrator data analyst* ed esperti in digitalizzazione dei processi. Tutte figure che nel 50% dei casi non sono mai state all'interno di aziende assicurative (Fonte: Hecamga.it/ https://www.hecamga.it/news/assicurazioni/intelligenza-artificiale-insurtech-assicurazioni.html).

[14] Secondo alcune stime il Pil italiano potrebbe crescere del 18% nel caso le aziende nostrane adottassero questa tecnologia in maniera massiva. Questo porterebbe enormi benefici a tutto il sistema Paese oltre che al comparto assicurativo, che oggi vale circa il 7% del Prodotto interno lordo.

di AI circa 50 milioni di euro da parte dei player del settore, che rafforzano gli investimenti registrati sinora nei progetti interni relativi all'*Insurtech*[15].

Fig. 3 Principale applicazione alla catena del valore assicurativa



Fonte: https://www.assinews.it/07/2023/studio-ey-iia-limpatto-delle-tecnologie-emergenti-sul-settore-assicurativo/660105977/

L'IA ha un potenziale straordinario anche nella gestione dei sinistri, una delle aree più critiche e impegnative per il settore assicurativo. Grazie all'analisi dei dati, l'IA è tra l'altro in grado di: a) valutare rapidamente e in modo accurato la portata di un sinistro e determinare il risarcimento dovuto; b) automatizzare alcune

[15] Cfr. "Settore assicurativo e intelligenza artificiale: pronti ai blocchi di partenza", Il Sole 24 Ore del 8/4/2024 https://www.economopoly.ilsole24ore.com/2024/05/08/settore-assicurativo-ai/?refresh_ce=1

fasi del processo di gestione dei sinistri (riducendo al minimo la necessità di intervento umano e accelerando i tempi di risposta); c) prevedere potenziali rischi ed adottare misure preventive per ridurre la frequenza e l'impatto degli incidenti; infine, più in generale, d) miglioramento dell'efficienza operativa e potenziale riduzione dei costi[16].

Fig. 4 I progetti interni Insurtech nel 2023

Fonte: Il Sole 24Ore-Econopoly

Infatti, attraverso l'automazione, la velocità di risposta e l'accuratezza delle previsioni, l'Intelligenza Artificiale non solo promette di migliorare l'efficienza operativa, ma anche di rivoluzionare l'esperienza del cliente. Come è noto, la gestione dei sinistri è una delle attività centrali per le compagnie assicurative, ma anche una delle più onerose. L'iter tradizionale prevede numerosi passaggi, dalla ricezione della denuncia alla valutazione del danno, dall'intervento dei periti alla negoziazione con il cliente. Il processo può essere lungo, complicato e, spesso, insoddisfacente per l'assicurato, che potrebbe riscontrare ritardi o avere la percezione di una mancanza di trasparenza. Il grande volume di dati da analizzare e gestire, la variabilità dei casi da affrontare e la necessità di prendere decisioni rapide e precise rendono la gestione dei sinistri una sfida costante. Inoltre, l'errore umano e la frode rappresentano rischi non trascurabili, contribuendo a far lievitare i costi operativi. Qui entra in gioco l'Intelligenza Artificiale, che offre soluzioni efficaci per automatizzare e ottimizzare l'intero processo.

[16] Attraverso "l'analisi predittiva", le compagnie assicurative possono identificare i fattori che aumentano la probabilità di sinistri e intervenire in modo proattivo per mitigare tali rischi, garantendo una maggiore sicurezza agli assicurati (cfr. https://www.wefox.com/it-it/supporto/blog/come-l-intelligenza-artificiale-sta-trasformando-il-settore-delle-assicurazioni).

Nella delicata e fondamentale area della gestione dei sinistri, l'individuazione della ricorrente ripetitività dei medesimi schemi faciliterà il lavoro di periti e liquidatori, potendo anche svolgere una funzione di supporto ai dipartimenti antifrode nel caso di eventi apparentemente sospetti, rilevando e segnalando più agevolmente attività di dubbia legalità come, ad esempio, l'uso insolito di strumenti di pagamento ed investimento (www.data4biz.com)[17].

L'IA introduce nuove modalità di gestione, rendendo la procedura più rapida, sicura e personalizzata (Holland, Kavuri, 2021). Grazie all'Intelligenza Artificiale, la prima fase della denuncia del sinistro può essere completamente automatizzata. Le *chatbot* dotate di IA possono interagire direttamente con il cliente[18], raccogliendo tutte le informazioni necessarie in tempo reale. Utilizzando il linguaggio naturale (NLP, *Natural Language Processing*)[19], sono in grado di comprendere le richieste dell'assicurato, porre domande specifiche e acquisire dati fondamentali come la descrizione dell'incidente, il tipo di danno e le eventuali foto caricate dal cliente. Non solo. Attraverso la scansione automatizzata di documenti, l'IA è in grado di estrarre le informazioni rilevanti, riducendo notevolmente i tempi di inserimento dati e minimizzando gli errori. In tal modo si elimina la necessità di intervento umano nella fase iniziale della procedura, accelerando i tempi di gestione e offrendo al cliente una risposta molto più rapida.

L'Intelligenza Artificiale è in grado di analizzare i danni in modo rapido e accurato. Con l'uso del *machine learning* e delle reti neurali, i sistemi IA possono essere addestrati a riconoscerne vari tipi attraverso le immagini caricate dal cliente (Pawan, *et al.*, 2023). Ad esempio, se un automobilista coinvolto in un incidente carica foto della sua auto danneggiata, l'IA può confrontare queste immagini con un vasto database di danni precedenti, stimare i costi di riparazione e persino suggerire soluzioni alternative. Le tecnologie di visione artificiale (*computer vision*) permettono di identificare i dettagli nelle immagini, come ammaccature o danni strutturali, con un livello di precisione molto superiore rispetto alla valutazione umana. Così non solo si accelerano i tempi di analisi, ma si riducono anche gli errori e le diversità di pareri, aumentando la fiducia del cliente nella gestione dell'evento e, quindi, verso l'intermediario e la compagnia.

Una volta raccolte e analizzate le informazioni, l'IA può stimare rapidamente il valore del risarcimento basandosi su algoritmi avanzati che valutano diverse variabili: dal costo delle riparazioni alla storia del cliente, fino al tipo di polizza sottoscritta. In questo modo, i tempi di elaborazione dei sinistri possono essere drasticamente ridotti. Se il danno rientra in una categoria predefinita e le infor-

[17] Cfr. Articolo "Come l'IA rivoluziona e velocizza la gestione dei sinistri assicurativi" https://data4biz.com/articles/come-lia-rivoluziona-e-velocizza-la-gestione-dei-sinistri-assicurativi (www.data4biz.com).

[18] Negli ultimi anni il settore del Customer Care ha vissuto una trasformazione radicale. Inizialmente, il telefono era l'unico canale di comunicazione, successivamente, l'introduzione di web, email, mobile e social media ha reso la gestione della Customer Experience sempre più complessa e sofisticata. L'introduzione dei "bot" di prima generazione intorno al 2017 ha portato ad un entusiasmo iniziale legato principalmente alla ricerca di riduzione dei costi, ma molti esperimenti iniziali sono falliti, perché appare evidente che "la soluzione risiede in un'armoniosa unione tra intelligenza artificiale generativa e intervento umano". (Cfr. articolo "Customer Care e Intelligenza Artificiale, consumatori finalmente soddisfatti?", Sole 24Ore https://www.econopoly.ilsole24ore.com/2024/08/26/customer-care-bot-intelligenza-artificiale-generativa-consumatori/).

[19] L'NLP (Natural Language Processing) è un sistema di Intelligenza Artificiale (AI) che rientra tra le soluzioni software che negli ultimi anni hanno registrato maggiori progressi. Correttori ortografici e sistemi di traduzione automatici sono solo alcune delle applicazioni di NLP che usiamo nella vita quotidiana. Grazie al contributo di tecniche di Artificial Intelligence sempre più avanzate, come Machine Learning e Deep Learning, l'NLP trova numerosi ambiti applicativi. La strada da percorrere è però ancora lunga per una lingua complessa come l'italiano, caratterizzata da modi di dire, espressioni gergali e influenzata da numerosi dialetti (Fonte: Osservatori.net).

mazioni sono complete, l'IA può autorizzare direttamente il pagamento del risarcimento, eliminando la necessità di revisione manuale. Questo significa che i clienti possono ricevere il pagamento in poche ore o giorni, anziché settimane. L'automazione del pagamento rende il processo estremamente più fluido e veloce, migliorando l'esperienza dell'assicurato e riducendo i costi operativi (cfr. www.insurzine.com) [20].

Uno dei problemi più rilevanti nel settore assicurativo è, poi, rappresentato dalle frodi, che possono far lievitare i costi per le compagnie e, di conseguenza, per gli assicurati. Il contrasto alle frodi è un tema molto sentito nelle compagnie. Secondo una ricerca Deloitte (2024, cit.), entro il 2026 l'83% delle aziende userà strumenti di AI per combatterle, migliorando i risultati in maniera significativa[21].

Le tecnologie di Intelligenza Artificiale sono particolarmente efficaci nel prevenire e identificare comportamenti fraudolenti. Gli algoritmi di *machine learning* possono analizzare milioni di dati storici e individuare schemi di comportamento sospetti che potrebbero passare inosservati a un esaminatore umano. Ad esempio, se un sinistro presenta caratteristiche simili a precedenti comportamenti fraudolenti (come richieste multiple o incongruenze nelle informazioni fornite), il sistema IA può segnalarlo per una revisione più approfondita. L'IA può monitorare in tempo reale le richieste di risarcimento, identificando anomalie in modo automatico e agendo tempestivamente per prevenire potenziali abusi. Questo tipo di controllo proattivo riduce il rischio di frodi, proteggendo sia le compagnie assicurative che i clienti onesti (cfr. www.insurzine.com)[22].

Un altro aspetto importante dell'Intelligenza Artificiale nel settore assicurativo è la sua capacità di offrire un livello di personalizzazione mai visto prima. L'IA non si limiterà a gestire i sinistri in modo efficiente, ma potrà anche prevedere i bisogni specifici di ciascun cliente, grazie all'analisi dei dati comportamentali e delle preferenze individuali. Attraverso l'uso di dati in tempo reale, le compagnie saranno in grado di sviluppare offerte su misura, suggerendo prodotti e coperture aggiuntive che rispondono esattamente alle loro esigenze della clientela. La personalizzazione non solo aumenterà la soddisfazione degli assicurati, ma potrà anche contribuire a ridurre il rischio di sinistri futuri (cfr. www.futuria.it)[23].

Adottare soluzioni di Intelligenza Artificiale nella gestione dei sinistri non è solo una questione di miglioramento dell'efficienza operativa, ma rappresenta un reale vantaggio competitivo. Le compagnie che investono in tecnologie IA saranno in grado di offrire servizi più rapidi, trasparenti e personalizzati, migliorando la loro

---

[20] Cfr. Articolo "Come l'intelligenza artificiale sta rivoluzionando la procedura dei sinistri assicurativi"https://www.insurzine.com/2024/09/16/come-intelligenza-artificiale-sta-rivoluzionando-la-procedura-dei-sinistri-assicurativi/ (Insurzine, 16/09/2024).

[21] Secondo la ricerca di Deloitte di inizio 2024, almeno il 59% delle aziende prevede di incrementare il budget per lo sviluppo di nuove tecnologie IA, ed entro il 2026 l'83% delle aziende si doterà di strumenti che sfruttano IA Generativa, mentre l'analisi della visione artificiale, della robotica e della biometria, che ad oggi sono già al 20%, saliranno rispettivamente a 42%, 41% e 39%. (cfr. Articolo "Frodi assicurative, con l'intelligenza artificiale la capacità di individuarle cresce del 20%", https://www.economyup.it/fintech/insurtech/frodi-assicurative-con-lintelligenza-artificiale-la-capacita-di-individuarle-cresce-del-20/).

[22] Cfr. Articolo "Come l'intelligenza artificiale sta rivoluzionando la procedura dei sinistri assicurativi" https://www.insurzine.com/2024/09/16/come-intelligenza-artificiale-sta-rivoluzionando-la-procedura-dei-sinistri-assicurativi/ (Insurzine, 16/09/2024).

[23] Cfr. articolo "L'Intelligenza Artificiale nel settore delle Assicurazioni: una rivoluzione in atto" (www.futuria.it) https://futuria.it/2024/06/05/lintelligenza-artificiale-nel-settore-delle-assicurazioni-una-rivoluzione-in-atto/

reputazione sul mercato. Inoltre, l'uso dell'Intelligenza Artificiale consentirà una riduzione significativa dei costi. Eliminando molte delle attività manuali, le compagnie potranno risparmiare risorse preziose, reindirizzando gli sforzi verso attività strategiche e di crescita. Questo approccio consentirà di allocare meglio il personale, permettendo agli operatori di concentrarsi sui casi più complessi e sul miglioramento del servizio clienti (cfr. www.insurzine.com, cit.)[24].

3.3 Conclusioni

In definitiva, l'Intelligenza Artificiale è destinata a giocare un ruolo sempre più centrale nel settore assicurativo nei prossimi anni (Balasubramanian, Libarikian, 2021), dimostrato anche dal dato che gli investimenti da parte delle compagnie italiane in IA arriveranno a sfiorare quota 100 milioni di euro nel 2025 (circa il doppio rispetto al 2024)[25]. Con il continuo miglioramento degli algoritmi e l'aumento dell'integrazione con altre tecnologie, come l'*Internet of Things* (IoT) e la *blockchain*, le potenzialità dell'IA, soprattutto nella gestione dei sinistri, cresceranno esponenzialmente (Coviello, 2024).

Questa "rivoluzione" servirà anche a colmare il *gap* innovativo del mercato assicurativo rispetto ad altri comparti molto più avanzati, oltre a risolvere alcuni dei problemi atavici del settore, in primis la sottoassicurazione registrata nel nostro Paese.

Si prevede che le compagnie diventeranno sempre più *data-driven*-aziende *che* considerano la gestione dei dati/data management, non come un fattore tecnico, ma come un pilastro strategico del business[26] – sfruttando l'Intelligenza Artificiale per migliorare le previsioni dei sinistri, ridurre i tempi di gestione e offrire un servizio ai clienti sempre più personalizzato e innovativo.

Le imprese assicurative che sapranno integrare l'AI nei loro processi operativi – aumentando così la qualità del livello di protezione offerto – saranno inoltre in grado di offrire al cliente un pacchetto di servizi sempre più ampio, quali mobilità, clima, salute, casa, rischi cyber e aziendali, etc. e, di conseguenza, mirare alla soddisfazione del cliente.

In definitiva, l'Intelligenza Artificiale rappresenta non solo il futuro della gestione dei sinistri, ma anche una nuova era per l'intero settore assicurativo, dove l'efficienza e la soddisfazione dell'assicurato saranno al centro di ogni processo. Affinché però la tecnologia diventi un aiuto e non un ostacolo al miglioramento ed allo sviluppo, sarà opportuno garantire la qualità dei dati usati dall'IA.

---

[24] Cfr. Articolo "Come l'intelligenza artificiale sta rivoluzionando la procedura dei sinistri assicurativi" https://www.insurzine.com/2024/09/16/come-intelligenza-artificiale-sta-rivoluzionando-la-procedura-dei-sinistri-assicurativi/ (Insurzine, 16/09/2024).

[25] Cfr. Articolo Sole 24 Ore https://www.ilsole24ore.com/art/assicurazioni-investimenti-100-milioni-sull-intelligenza-artificiale-2025-AFyXt8LD

[26] Essere *data-driven* significa farsi guidare dai numeri, avere un approccio basato sui dati, per prendere decisioni informate, basate su fatti oggettivi e non su sensazioni personali. La trasformazione in *data-driven company* non può dunque avvenire con la sola tecnologia, ma con un percorso di *change management* in grado di portare la cultura del dato a tutti i livelli aziendali. https://www.digital4.biz/marketing/big-data-e-analytics/sei-regole-d-oro-per-un-data-driven-marketing-di-successo/#:~:text=Essere%20data%2Ddriven%20significa%20farsi,e%20non%20su%20sensazioni%20personali

Bibliografia

Balasubramanian, R., Libarikian, A. (2021). *Insurance 2030 – The impact of AI on the future of insurance, McKinsey's New York office.* Disponibile su: https://www.mckinsey.com/industries/financial-services/our-insights/insurance-2030-the-impact-of-ai-on-the-future-of-insurance#/

Coviello, A., Guazzone, S. (2021). Covid-19: global shock and new scenarios for the insurance industry. In *Journal of International Scientific Publications, Economy & Business,* 15. Disponibile su: https://www.scientific-publications.net/get/1000049/1632910990101424.pdf

Coviello, A., Somma, R. (2021). *I rischi catastrofali. Azioni di mitigazione e gestione del rischio.* Roma: Edizioni Consiglio Nazionale delle Ricerche (CNR). Disponibile su: https://www.iriss.cnr.it/files/I-rischi-catastrofali-azioni-di-mitigazione-e-gestione-del-rischio.pdf

Coviello, A. (2023). Innovation in insurance services. In F. Gallouj, M. C. Monnoyer, L. Rubalcaba, (a cura di), *Elgar Encyclopedia of Services.* Cheltenham (UK): Edward Elgar Publishing. Disponibile su: https://dx.doi.org/10.4337/9781802202595; https://www.elgaronline.com/display/book/9781802202595/b-9781802202595.Innovation.in.Insurance.xmlit/

Coviello, A. (2023). Innovation in insurance services. In F. Gallouj, M. C. Monnoyer, L. Rubalcaba, (Eds.), *Elgar Encyclopedia of Services*. Cheltenham (UK): Edward Elgar Publishing. Disponibile su: https://dx.doi.org/10.4337/9781802202595

Coviello, A. (2024). Il marketing assicurativo nell'era digitale. In *Marketing e comunicazione delle assicurazioni.* Milano: Insurzine, edizioni oltreLaMedia Group. Disponibile su: https://www.insurzine.com/app/uploads/2024/12/Monografia_MarketingComunicazioniDelleAssicurazioni.pdf

Coviello, A. (2024). The impact of digital transformation in the italian insurance sector. In *Advanced Perspectives and Trends in Digital Transformation of Firms, Networks, and Society* – 2nd International Conference of the Digital Transformation Society (DTS), Naples, Italy, May 23-24, Springer Book (in press).

Deloitte (2024). *IA ed Assicurazioni*. *Monitor Deloitte.* Disponibile su: https://www2.deloitte.com/content/dam/Deloitte/it/Documents/financial-services/deloitte-ia-innovazione-assicurazioni-2024.pdf

GILC (2024). Artificial Intelligence and the future of insurance. In *Global Insurance Law Connect report unveils AI's dual impact on insurance industry.* Disponibile su: https://www.globalinsurancelaw.com/wp-content/uploads/2024/04/GILC-AI-Global-Trends-Report-February-2024.pdf

Holland, C. P., Kavuri, A. (2021). Artificial intelligence and digital transformation of insurance markets. In *Journal of Financial Transformation,* 21. Disponibile su: https://www.capco.com/Capco-Institute/Journal-54-Insurance/Artificial-Intelligence-And-Digital-Transformation-In-Insurance-Markets

Pawan, K., Sanjay, T., Mukul, Ercan Özen (2023). Digital Transformation of the Insurance Industry. In book Editor(s): Kiran Sood, Simon Grima, Peter Young, Ercan Ozen, Balamurugan Balusamy. Disponibile su: https://doi.org/10.1002/9781394167944.ch6

Swiss Re (2024). *The evolution of AI in the insurance industry.* Disponibile su: https://www.swissre.com/risk-knowledge/advancing-societal-benefits-digitalisation/evolution-of-ai-in-insurance-industry.html

4

THE CASE OF NEC LABORATORIES EUROPE – THE CREATION OF THE INTERNAL AI & HUMAN RIGHTS COMMITTEE

di Giacomo Maria Cremonesi (NEC Laboratories Europe – GmbH)

4.1. Introduction

Artificial Intelligence (AI) can deliver extraordinary benefits-from accelerating scientific discoveries to enhancing public services and driving economic growth. Yet its unprecedented power also raises complex questions about *human rights protection.* Conducting *Human Rights Due Diligence* (HRDD) in the AI sector, as defined in the United Nations Guiding Principles (UNGPs)[1], is notably challenging because of the technology's flexibility and potential global reach. A single AI system, for example, can be deployed in completely different applications such as personalized medicine, financial risk analysis, or landmine detection, each with significant distinct impacts on human rights.

The *EU AI Act* is the first comprehensive legislation focused on regulating AI and incorporates some elements of Human Rights Due Diligence (HRDD) of the UN Guiding Principles (UNGPs). The AI ACT[2] adopts a risk-based framework that provides strong obligations, but only for certain pre-defined high-risk applications and excludes other elements of HRDD.

In response to these and other related challenges, *NEC Laboratories Europe GmbH (NLE)* recently established an *Internal AI & Human Rights Committee.* The

[1] *United Nations Guiding Principles on Business and Human Rights (UNGPs)* – Office of the High Commissioner for Human Rights (OHCHR). *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework.* United Nations, 2011. Available at: https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

[2] EU AI Act Overview – https://artificialintelligenceact.eu/ai-act-explorer/

Committee aims to provide leadership, guidance, and advice on addressing human rights impacts of AI technologies researched at NLE. Through collaborative risk assessments, stakeholder consultations, and the implementation of tailored policies and procedures, the Committee aims to contribute to the alignment with both international human rights frameworks and emerging regulatory expectations.

4.2. NEC Laboratories Europe GmbH & NEC Group

Established in 1997 and based in Heidelberg, NEC Laboratories Europe GmbH (NLE) conducts fundamental and applied research guided the NEC 2030Vision. Key principles include improving: the environment, society and our daily lives. NLE addresses some of society's most pressing challenges by developing cutting-edge technology in AI, digital health, information and communications technology, and security.

Recognizing that respect for human rights is vital to fulfilling its corporate Purpose, the NEC Group has instituted a Human Rights Policy[3] alongside the NEC Group AI and Human Rights Principles. These policies guide all employees in prioritizing privacy and human rights when deploying AI, biometrics, and other data-driven technologies. Moreover, NEC Corp is part of the *AI Pact*[4], committing to voluntary pledge to promote trustworthy and safe AI development ahead of the AI Act's full application.

In line with the above commitments, NEC Laboratories Europe integrates ethical and human rights considerations throughout its research and development processes, ensuring that innovations not only push technological boundaries but also uphold the highest standards of social responsibility and trustworthy AI.

4.3. UNGPs Human Rights Due Diligence & AI solutions

Under the *United Nations Guiding Principles on Business and Human Rights (UNGPs)*, companies are expected to conduct *Human Rights Due Diligence (HRDD)* to identify, prevent, mitigate, and account for any adverse human rights impacts they may cause or contribute to. This process involves four key steps: *assessing* potential impacts, *integrating* findings into corporate practices, *tracking* the effectiveness of responses, and *communicating* how these issues are managed.

HRDD typically includes mapping all relevant stakeholders and assessing with their contribution the specific risks associated with a particular product or service, including the supply chain. In order to identify, prevent, and mitigate adverse human rights impacts companies are thus required to *engage proactively with all relevant stakeholders* throughout this process.

Stakeholder engagement-whether it involves employees, affected communities, civil society organizations, or end users-ensures that *human rights risks* are properly identified and understood from multiple perspectives, increasing the likelihood of effective mitigation strategies.

---

[3] NEC Corporation AI and Human Rights Commitments – NEC Corporation. NEC Group AI and Human Rights Principles.https://www.nec.com/en/global/techrep/journal/g19/n01/190103.html

[4] The AI Pact and Corporate AI Governance – European Commission available at https://digital- strategy.ec.europa.eu/en/policies/ai-pact

However, *applying HRDD to AI solutions* presents distinct complexities, largely due to *technology's flexibility and global reach*[5]. A single AI engine-such as a graph AI tool-may be deployed in diverse contexts: developing personalized cancer vaccines, locating landmines, or detecting financial fraud. Each of these scenarios involves different stakeholders and different sets of human rights risks, making it more challenging to thoroughly anticipate and address potential adverse impacts.

On the other hand, industries such as *oil & gas, textiles,* or *agriculture* often have *more predictable human rights issues* and the affected *stakeholders* are usually *easier to identify*. In the oil and gas sector, for instance, typical human rights impacts include forced displacement, inadequate compensation for local communities, environmental damage, and excessive use of security forces. In the textile or agricultural sectors, forced or child labor, unsafe workplaces, and poverty wages predominate, with weak oversight and lack of transparency enabling further abuses. While these industries still face significant challenges, risks tend to be more narrowly defined, and the affected stakeholders-such as local communities or factory workers-should be easier to identify.

In AI, however, the range of potential applications leads to a correspondingly wide variety of possible human rights impacts. For instance, Large Language Models (LLMs) can enhance education by serving as interactive tutoring tools or facilitate research by streamlining literature reviews. In a different context, however, these same models might be misused to generate deepfakes, spread disinformation, or amplify hate speech-often at a global scale due to widespread internet access. Similarly, graph AI technology can be employed for beneficial purposes like location of landmines or producing personalized cancer vaccines. Yet it could also be repurposed to track social networks for surveillance, discriminate against certain groups based on their connections, or violate privacy through invasive data analysis.

This dual-use nature[6]-where a single AI technology can be adapted for both constructive and harmful ends-poses significant challenges for Human Rights Due Diligence in the AI sphere. Companies and researchers must anticipate a vast array of potential outcomes, many of which may not be easy to identify during design, training or testing phases. The global reach of AI solutions complexes the problem, as a model or system developed in one country might be deployed with little modification in entirely *different sociopolitical contexts,* exposing communities to previously unrecognized risks. It can be assumed, for example, that security organizations and state institutions in countries with lower human rights standards are more likely to misuse AI.

Responding effectively to these complexities requires continuous adaptation to evolving technologies, ongoing HRDD, multi-stakeholder engagement-including developers, regulators, civil society, and affected communities-and monitoring of real-world use cases. It also demands a willingness to refine or redesign AI tools when unexpected human rights risks emerge. These overlapping responsibilities make AI one of the most intricate frontiers in modern human rights protection, un-

[5] Artificial Intelligence and Human Rights Recommendations for companies – UN Global Compact Network Germany – April 2024 https://www.globalcompact.de/fileadmin/user_upload/Dokumente_PDFs/artificial_intelligence_and_human_rights_EN.pdf

[6] Dual-Use Nature of AI – Future of Life Institute. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. 2018. Available at: https://arxiv.org/abs/1802.07228

derscoring the pressing need for robust human rights due diligence frameworks[7] that can keep pace with rapid technological innovation.

4.4. EU AI ACT

To address the complexities surrounding Human Rights Due Diligence (HRDD) for AI solutions, the European Union has adopted the *EU AI Act,* which takes a *risk-based approach to regulating AI*.

4.4.1. Extended Obligations for High-Risk AI Solutions

The EU AI Act classifies AI systems into different risk categories-prohibited (unacceptable risk), high risk, and those with limited or minimal risk. AI solutions with unacceptable risk are prohibited. Obligations for high-risk AI solutions under the AI Act are extensive and can significantly impact how these technologies are designed, developed, and deployed. Specifically, organizations that develop or supply high-risk AI systems are required to implement these extensive requirements:

- Risk management system.
- Data governance and quality requirements.
- Technical documentation.
- Logging and traceability.
- Transparency and provision of information to users.
- Human oversight.
- Accuracy, robustness, and cybersecurity.
- Conformity assessment procedures.

Moreover, all providers of General Purpose AI models that present a systemic risk – open or closed – must also conduct model evaluations, adversarial testing, track and report serious incidents and ensure cybersecurity protections[8].

4.4.2. AI ACT limitations

There are at least *three decisive choices* made by the EU AI Act that influence how organizations conduct HRDD in the AI space. On the one hand, these measures simplify the risk identification aspect of the HRDD process; on the other, they reduce the scope of HRDD obligations under the United Nations Guiding Principles (UNGPs) by narrowing their focus[9].

[7] The COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAI) of the Council of Europe elaborated: Methodology For The Risk And Impact Assessment Of Artificial Intelligence Systems From The Point Of View Of Human Rights, Democracy And The Rule Of Law – November 2024. Available at https://rm.coe.int/cai-2024-16rev2-methodology-for-the-risk-and-impact-assessment-of- arti/1680b2a09f

[8] EU AI Act summary – available at https://artificialintelligenceact.eu/high-level-summary/

[9] See EU's AI Act fails to set gold standard for human rights – AlgorithmWatch, available at https://algorithmwatch.org/en/ai-act- fails-to-set-gold-standard-for-human-rights/

4.4.2.1. Pre-Identification of High-Risk and Unacceptable-Risk AI Use Cases

The EU AI Act classifies AI systems into different risk categories-prohibited, high risk, and those with limited or minimal risk. While this categorization helps prioritize compliance and HRDD efforts where they are most urgently needed, a purely list-based approach may overlook novel or country/context-specific risks-an especially pressing concern given AI's unpredictable evolution and its wide-ranging societal impacts.

4.4.2.2. Limited Requirements for Fundamental Rights Impact Assessments

The Act stops short of demanding a comprehensive Human Rights Impact Assessment (HRIA) comparable to the UNGPs' recommendations. Instead, it places emphasis on technical standards such as accuracy, robustness, and cybersecurity. Where Fundamental Rights Impact Assessments (FRIA) are mentioned (in Article 27 of the proposed legislation), they apply only to certain types of high-risk AI systems, potentially leaving gaps for other applications with significant human rights implications.

4.4.2.3. Absence of a Formal Stakeholder Engagement Mandate

In contrast to the UNGPs, which underscore meaningful consultation with affected communities and civil society, the AI Act does not explicitly require stakeholder engagement. Companies are expected to handle compliance internally. This omission is significant because multi-stakeholder engagement is central to robust HRDD, particularly in identifying "hidden" or emergent human rights harms. Without direct input from users, communities, and external experts, organizations risk overlooking unintended impacts of AI solutions.

4.4.3. Obligations During the R&D Phase-and Why Early Compliance Matters

AI ACT exempts R&D or prototyping stages from the full compliance burden (even if real-world testing phase requires certain obligations). However, delaying compliance until after R&D can pose significant risks. AI models can develop embedded biases or systemic vulnerabilities during early development, which can become difficult to fix once integrated into real-world applications. Furthermore, aligning early-stage research with the Act's foundational requirements (e.g., data quality, documentation, or risk assessment) positions developers to transition efficiently into full compliance once their AI systems mature beyond experimental settings. HRDD of AI shall be by design[10].

4.5. NLE AI & Human Rights Committee

To tackle the challenges of Human Rights Due Diligence (HRDD) in the AI domain, and to support NEC Laboratories Europe (NLE) in both implementing the

[10] Human rights by design future-proofing human rights protection in the era of AI – Council of Europe -2019 – available at: https://rm.coe.int/follow-up-recommendation-on-the-2019-report-human-rights-by-design-fut/1680ab2279

EU AI Act and addressing its potential shortcomings, NLE has established an Internal AI & Human Rights Committee.

4.5.1. Purpose and Scope

The overarching purpose of the Committee is to *provide leadership, guidance, and advice* to NLE researchers on *identifying* and *addressing the human rights impacts* of AI technologies. In doing so, the Committee not only assists NLE in complying with the EU AI Act, but also strives to align with global human rights standards.

4.5.2. Activities and Responsibilities

Looking ahead, the Committee's work will revolve around several key areas including:

– Risk and Impact Assessments

Examine and assess potential risks and human rights impacts on diverse stakeholder groups. Proactively identify emerging or underexplored risks, *including those not* strictly *classified as high-risk under the EU AI Act.* The Committee will also provide actionable recommendations for mitigating identified risks.

– Stakeholder Consultation

When relevant, identify affected groups-such as users, local communities, or civil society organizations-and engage in meaningful dialogue. *Engaging multiple stakeholders is essential for robust HRDD,* especially in uncovering hidden or emerging human rights risks even if the AI ACT does not require so.

– Awareness and Internal Capacity Building on Human Rights

Raise awareness about human rights protection throughout NLE, including hosting training sessions, workshops, and internal campaigns. It is important to *embed HRDD principles into corporate culture,* emphasizing the relevance of human rights at each step of AI research.

– Policy and Compliance

Develop and implement internal policies and procedures that meet or exceed international legal frameworks and the EU AI Act. Monitor ongoing compliance efforts and adapt policies as regulations or technologies evolve.

– Public Engagement

Participate in *public consultations on AI* (particularly AI ACT), ensuring that NLE's voice contributes to shaping best practices and evolving standards.

4.5.3. Independence and Composition

Crucial to the Committee's effectiveness is its independence from hierarchical or commercial constraints. It operates without undue influence, ensuring that assessments and recommendations remain objective and impartial.

The Committee's membership reflects a cross-disciplinary perspective, comprising internal and external members. Internal members include: Human-Centric AI Experts, Machine Learning Specialists, Business & Human Rights Legal Specialist, GDPR Expert, Company Management Representative, Marketing & Communication Member, Standardization Expert. External members include: Civil society or specialized human rights organizations (NGOs) to represent community interests and bring independent scrutiny.

4.5.4. Human Rights Due Diligence in R&D

Although the EU AI Act offers limited obligations during the R&D phase, the Committee advocates for *early integration of HRDD principles*-recognizing that risks, biases, or unintended consequences often surface during initial research. Such proactive identification and resolution of issues can prevent more significant harms down the line, foster public trust, and ultimately strengthen NLE's commitment to responsible innovation.

By implementing these measures, NEC Laboratories Europe aims to shape AI advancements that benefit society while upholding fundamental rights and freedoms. The Internal AI & Human Rights Committee is integral to fulfilling this vision-an embodiment of NLE's dedication to transparency, ethics, and accountability in the pursuit of cutting-edge AI research and solutions.

4.6. Conclusions

As AI technology evolves at a rapid pace, balancing the requirements of the EU AI Act with the scope of the HRDD mandated by the UN Guiding Principles on Business and Human Rights (UNGPs) remains a significant challenge. The Internal AI & Human Rights Committee at NEC Laboratories Europe represents a pragmatic step toward bridging this gap-guiding the organization in complying with the new European legislation while also embedding human rights due diligence throughout its research and development cycles.

Moving forward, the success of such initiative will depend on the ability to adapt to emerging challenges, foster meaningful stakeholder collaborations, and influence broader discussions on AI governance. By aligning with both regulatory standards and international human rights principles, NEC Laboratories Europe is positioning itself at the forefront of responsible AI development-demonstrating that technological advancement and human rights protection can and must go hand in hand.

5

GIUSTIZIA E INTELLIGENZA ARTIFICIALE

di Gabriele Esposito (Consiglio dell'Ordine degli Avvocati di Napoli)

Se ne parla da tempo e da tempo è applicata, sfruttata, utilizzata nel settore giustizia; qualsiasi termine ci dà l'idea che è entrata a far parte della vita di chi si occupa di cose di diritto.

Sicuramente dei vantaggi ci sono e ce ne saranno, però sorgono grossi dubbi, o meglio, grosse certezze sulla paura per l'utilizzo dell'IA nel campo del diritto, soprattutto in materia penale.

Utilizzare un elaboratore informatico efficientissimo per organizzare la vita degli uffici giudiziari è sicuramente una grossa opportunità; i tribunali, nella loro pianificazione, sono paragonabili a imprese e perciò, in questo senso, non sorgono dubbi sulla sua efficienza;[1] parimenti, utilizzare un sistema informatico altamente evoluto per la ricerca dottrinale e giurisprudenziale o per lo svolgimento di indagini va ugualmente bene, vagliando rigorosamente rischi e benefici.[2]

Quando però sento parlare di AI al fine di dirimere questioni o controversie giudiziarie, utilizzata da avvocati e magistrati, comincio a preoccuparmi.

Mi riferisco in particolare all'ambito penale in quanto, rispetto alle altre branche, i principi costituzionali sottesi alla tutela del diritto di difesa impongono una maggiore riflessione, soprattutto quando è concretamente in gioco la libertà personale di un individuo.

---

[1] Cfr. Di Giacomo 2025.

[2] Cfr. Pisati 2020, il quale afferma che «il ricorso alle tecnologie di intelligenza artificiale nelle indagini preliminari consente maggiore efficienza e celerità nell'attività di analisi di grandi masse di informazioni. Al contempo rivela rischi non trascurabili per le garanzie fon¬damentali dell'accusato. È muovendo anche da questo assunto che la Commissione europea, in un documento del febbraio 2020 (*"Libro bianco sull'intelligenza artificiale"*), ha sollecitato la comunità, anche scientifica, al dibattito».

Scorgo da tempo, da parte di alcuni operatori del diritto, quasi una forma di eccitazione alla idea che i processi penali possano essere decisi con l'utilizzo dell'AI, e che la giustizia penale possa progredire mediante l'utilizzo di sistemi informatici complessi ed articolati di elaborazione di dati, capaci di auto-generarsi ed elaborare tesi e teorie sulla scorta di dati immessi in archivio, in quanto ciò consentirebbe una velocizzazione dei processi e, quindi, un adeguamento agli standard europei quanto ai tempi medi di celebrazione degli stessi, con grave nocumento per una tutela effettiva dei diritti dei cittadini.[3]

La spinta ad una "informatizzazione" delle decisioni giurisdizionali proviene da quei paesi in cui, diversamente dal nostro in cui vige un sistema di *civil law*, vigono sistemi giurisdizionali di *common law.*

La principale differenza consiste nel fatto che il *civil law* si è sviluppato a partire dal diritto romano, nel quale l'ordinamento giuridico è costituito da leggi e principi generali che costituiscono la base del sistema (da essi discende poi la soluzione dei casi concreti da parte del giudice), mentre il *common law* si sviluppa sulla base delle consuetudini, a loro volta fondate sulla decisione del singolo caso.

L'impostazione del *common law*, in base alla quale, come detto, acquista particolare importanza il diritto di creazione giudiziale, porta con sé il principio del "precedente" giurisprudenziale, che deve essere seguito anche dai giudici successivi che si trovino di fronte a un caso equivalente.

Nei sistemi di *civil law*, al contrario, il giudice deve sempre applicare la legge, e cercare in essa, e non solo nei precedenti giurisprudenziali, la soluzione adatta al caso.

È proprio nella differenza dei sistemi legislativi, dunque, l'origine della assoluta impossibilità a delegare ad un sistema informatico evoluto la elaborazione delle decisioni giudiziarie e, di conseguenza, nella formazione dei giuristi degli ordinamenti di *civil law.*

In particolare, un grosso problema sorge, quanto al sistema processuale penale italiano, con riferimento alle possibilità conferite al giudice dal codice di rito in tema di deliberazione delle sentenze.

Il giudice, infatti, senza voler approfondire gli aspetti squisitamente tecnici dell'art. 533 c.p.p., pronuncia sentenza di condanna *al di là di ogni ragionevole dubbio*.

Il principio del "ragionevole dubbio", dunque, confligge con la stessa natura di un sistema di intelligenza artificiale, programmato ed autorigenerante per conferire sicuramente certezze e non dubbi, e non un alto grado di probabilità così come prescritto dalla norma.[4]

Oppure si pensi alla concessione delle circostanze attenuanti, soprattutto quelle generiche, o all' applicazione delle circostanze aggravanti, in caso di sentenza di condanna; trattasi di istituti giuridici la cui applicazione è frutto di un complesso ragionamento che non può essere frutto, nel caso dell'intelligenza artificiale, di attuazione di precedenti giurisprudenziali, attesa la genericità del dato normativo, e che spesso, soprattutto in tema di bilanciamento delle circostanza, più favorevoli e meno favorevoli al reo, sono frutto anche della discrezionalità e sensibilità del giudice.

---

[3] V. Clementino 2018.

[4] Per una disamina sul punto v. Piana, Verzelloni (2020); si legga anche Resta 2021, la quale osserva che «l'applicazione degli strumenti di intelligenza artificiale al campo della giustizia, sebbene non sia da respingere in toto, dovrebbe essere circoscritta chiaramente e con molta attenzione. Questo per non rinunciare al bisogno insopprimibile di non de-umanizzare alcune attività (tra cui, come vedremo, quella giurisdizionale, in particolare penale) in cui la soggettività gioca un ruolo non sostituibile, in nessun modo, da alcun calcolo».

Affidare, pertanto, ad un elaboratore informatico una decisione in ambito penale può divenire dunque rischiosissimo in quanto, in assenza della possibilità di formulare dubbi sulla causazione di un evento o sulla attribuibilità dello stesso ad un determinato individuo, si incorre in una tacita elisione del disposto normativo, oppure, nel caso delle circostanze, il grado di "umanizzazione" dell'elaboratore troverebbe certamente numerosi limiti rispetto alla coscienza del giudice essere umano.

Un altro problema relativo all'affidamento ad un sistema informatico complesso delle decisioni giudiziario potrebbe riguardare la contaminazione, e quindi la eventuale manipolazione, da parte del mondo esterno, della decisione deliberata.

Il giudice umano, difatti, quando deve pronunciare sentenza, si ritira in camera di consiglio, la cui soglia è invalicabile a chiunque, e non può comunicare con terzi ai fini della decisione.

Si pensi, ad esempio, ai grossi processi di criminalità organizzata al cui esito i giudici, togati e popolari, possono rimanere in camera di consiglio anche per settimane e agli stessi è inibito qualsivoglia contatto.

Un elaboratore informatico, dunque, per quanto complesso, avendo la necessità di attingere dati e informazioni mediante la rete internet, resta vulnerabile proprio perché rimanendo connesso concede la concreta possibilità, pur essendo protetto, di essere violato da terzi (appare superfluo elencare i numerosi casi di violazione di server anche di enti dotati di efficienti sistemi di cybersicurezza) e, comunque, un elaboratore necessità, per trarre conclusioni, di attingere informazioni da contenitori di dati che debbono necessariamente riempiti correttamente.[5]

Queste preoccupazioni, comunque, sono mitigate dal testo dello schema di DDL recante *Disposizioni e delega al Governo in materia di intelligenza artificiale* in cui, all'art. 14, in cui, in relazione all'utilizzo dell'intelligenza artificiale nell'attività giudiziaria, si legge che «i sistemi di intelligenza artificiale sono utilizzati esclusivamente per l'organizzazione e la semplificazione del lavoro giudiziario nonché per la ricerca giurisprudenziale e dottrinale. Il Ministero della giustizia disciplina l'impiego dei sistemi di intelligenza artificiale da parte degli uffici giudiziari ordinari. Per le altre giurisdizioni l'impiego è disciplinato in conformità ai rispettivi ordinamenti. È sempre riservata al magistrato la decisione sulla interpretazione della legge, sulla valutazione dei fatti e delle prove e sulla adozione di ogni provvedimento».

Tale previsione, dunque, pone un enorme freno all'utilizzo dell'intelligenza artificiale ai fini decisori di provvedimenti giudiziari.

Diverso è, invece, il discorso dell'utilizzo dell'AI nelle indagini giudiziarie, in cui le nuove tecnologie, già da tempo, vengono sfruttate dagli investigatori.

Su questo versante, però, si aprono altri dibattiti concernenti, in particolare, la concreta parità tra accusa e difesa che, solo in mera teoria, godono di identica posizione.

Questa disparità, nel concreto, si attua allorquando la pubblica accusa, che può disporre e dispone di notevoli risorse umane ed economiche, pone in essere atti-

[5] Interessante è il lungo contributo di Carriero 2024, in cui l'autrice afferma, tra l'altro, che «si possono, in questo senso, ricordare le parole di R. Borges Blázquez: «il sistema di I.A.:
1) utilizza le informazioni fornite dall'uomo o da un'altra macchina per percepire ambienti reali o virtuali;
2) astrae queste percezioni generando modelli […];
3) genera risultati […] sotto forma di raccomandazioni, previsioni o decisioni». Da tali frasi si comprende come alla base delle "recomendaciones, predicciones o decisiones" ci sia la previa fornitura di dati da parte dell'essere umano (c.d. progettista) nei riguardi della macchina».

vità di indagine che risultano spesso impossibili da espletarsi da parte del soggetto indagato; *gap* che poi si riverbera in sede processuale.

La disparità, soprattutto in certi ambiti, è talmente sproporzionata che chi è indagato di fatto vede amputato il proprio diritto a difendersi per carenza delle medesime risorse di cui dispone un pubblico ministero, pur potendo, come previsto dal codice di rito, attivare investigazioni difensive.

Se oggi è così, in un prossimo futuro diverrà inevitabile il ricorso, da parte della difesa, a consulenti informatici; ciò comporterà un costo, spesso eccezionalmente oneroso, che non tutti potranno sostenere per esercitare adeguatamente il proprio diritto di difesa.

Questo dato rileva proprio in virtù del citato schema di DDL che prevede numerose modifiche al codice penale, con particolare riferimento ai delitti di sostituzione di persona, di illecita diffusione di immagini, video o audio, di alcuni reati contro il patrimonio ed alcune disposizioni in tema di diritto d'autore; a queste modifiche si aggiungano quelle relative alle circostanze aggravanti comuni.

Il comune denominatore a tali future eventuali modifiche è costituito dalla realizzazione dei fatti previsti dalle suddette norme mediante *l'impiego di sistemi di intelligenza artificiale.*

La paventata parità, dunque, tra accusa e difesa, già minata da frequenti sperequazioni, si affievolirà sempre più in un sistema processuale che, in determinate ipotesi, vede il riconoscimento dei diritti solo di chi dispone di cospicue risorse economiche, con concreta elisione dei principi e dei diritti fondamentali enunciati dalla carta costituzionale e dalla CEDU in tema di uguaglianza.

Possiamo, quindi, concludere affermando che l'impiego dell'AI nel settore giustizia deve necessariamente incontrare prevedere i propri limiti nella organizzazione del comparto, nella archiviazione dei dati, nella razionalizzazione delle risorse, nella mera semplificazione della ricerca e nella comunicazione o deposito degli atti giudiziari, scongiurando ogni eventuale intervento di un sistema informatico, benché evoluto, anche con semplici previsioni probabilistiche, nella determinazione dei provvedimenti giudiziari, non potendosi neanche immaginare di affidare la libertà personale ad un computer.

Bibliografia

Carriero, M. F. (2024). *Crisi della legalità e intelligenza artificiale: uno sguardo (comparato) d'insieme*. In *Archivio Penale*.

Clementino, B. (2018). *Uso dell'intelligenza artificiale nei sistemi giudiziari: verso la definizione di principi etici condivisi a livello europeo?* Disponibile su: https://www.questionegiustizia.it/rivista/articolo/uso-dell-intelligenza-artificiale-nei-sistemi-giud_591.php

Di Giacomo, L. (2025). *Intelligenza artificiale nei tribunali: d.d.l di innovazione o grande fratello?* Disponibile su: https://www.diritto.it/intelligenza-artificiale-tribunali-d-d-l-innovazion/

Piana, D., Verzelloni, L. (2020). Oltre ogni ragionevole dubbio? Riflessioni critiche sull'impatto delle tecnologie nella giustizia. In J. Sabariego, *et al.* (a cura di), *Algoritarismos* pp. 302-320.

Pisati, M. (2020). Indagini preliminari e intelligenza artificiale: efficienza e rischi

per i diritti fondamentali. In *Processo penale e giustizia* pp. 957-971.
Resta, F. (2021). *Intelligenza artificiale e Giustizia, opportunità e rischi: i limiti che servono*. Disponibile su: https://www.agendadigitale.eu/cultura-digitale/intelligenza-artificiale-applicata-alla-giustizia-opportunita-e-rischi-i-limiti-che-servono/

6

BREVI RIFLESSIONI CIRCA LE STRATEGIE IN MATERIA DI CYBERSICUREZZA DELL'ORGANIZZAZIONE DEGLI STATI AMERICANI

di Marco Fasciglione (CNR-IRISS)*, Michele Nino (Università degli Studi di Salerno)**

6.1. Introduzione

L'avvento delle tecnologie della telecomunicazione e la penetrazione del «digitale» presso diversi strati della popolazione è un fenomeno in costante crescita che accomuna oramai diverse aree del Pianeta. Il continente americano, sia pure in ritardo rispetto ad altre aree regionali[1], non fa eccezione. In effetti, la liberalizzazione del mercato delle telecomunicazioni negli Stati americani, l'ampia disponibilità di tecnologie Internet, mobili e wireless e la crescente disponibilità di sistemi a banda larga, hanno favorito anche in tale continente – soprattutto negli ultimi anni – la diffusione delle tecnologie digitali, contribuendo contemporaneamente alla crescita delle preoccupazioni anche nei Paesi appartenenti a tale ambito regionale, per la sicurezza informatica a livello sia nazionale che regionale. Alcuni Stati del continente hanno già adottato quadri regolamentari e di *policy* volti a promuovere la sicurezza informatica e a regolamentare il settore con l'obiettivo di

* Autore dei paragrafi 1, 2, 5.

** Autore dei paragrafi 3, 4, 6.

[1] L'Unione europea, ad esempio, ha adottato il regolamento europeo sull'intelligenza artificiale (c.d. *AI Act*), uno dei primi strumenti legislativi adottati nel mondo in materia, concepito per garantire che lo sviluppo o l'impiego di tecnologie basate sull'AI nell'UE sia affidabile e offra garanzie per proteggere i diritti fondamentali dei cittadini [cfr. Regolamento (UE) 2024/1689 del Parlamento Europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale)]. Il Consiglio d'Europa, poi, ha adottato nel 2024 la *Framework Convention on Artificial Intelligence and human rights*, il primo trattato internazionale giuridicamente vincolante in materia di AI il cui obiettivo è garantire che i sistemi di intelligenza artificiale siano compatibili con i diritti umani, la democrazia e lo stato di diritto.

combattere le intrusioni e gli attacchi ai sistemi e alle reti informatiche[2]. Molti altri, non lo hanno ancora fatto, oppure sono in procinto di svilupparli.

In questo contesto, un ruolo centrale dal punto di vista del supporto all'elaborazione di politiche e normative sulla *cybersecurity* è svolto dall'Organizzazione degli Stati americani (OSA) che sta provvedendo, sebbene una convenzione regionale nella *subjecta materia* non sia stata adottata, a sospingere il processo di creazione di regole e standard con riferimento sia alla *cybersecurity* sia al contrasto della criminalità informatica (c.d. *cybercrime*)[3]. Il presente contributo persegue l'obiettivo di passare sinteticamente in rassegna le principali tappe della cooperazione interamericana in siffatte materie, analizzando gli aspetti principali dei – relativamente recenti – programmi avviati in materia dagli organi dell'OSA. A tal fine, dopo aver esaminato le nozioni di *cybersecurity* e *cybercrime*, il lavoro si concentra sulle *policy* elaborate dall'Organizzazione in materia, che seguono sostanzialmente un doppio binario: da un lato, gli strumenti volti a promuovere la creazione di un quadro regionale in materia di sicurezza informatica; dall'altro lato, gli strumenti legati alla cooperazione in materia penale in quanto volti al contrasto della criminalità informatica. Il lavoro si sofferma, poi, sulle funzioni dei principali organismi competenti a predisporre la complessiva strategia dell'OSA in materia di sicurezza informatica. Le riflessioni conclusive, infine, evidenziano come l'attuazione della strategia OAS in materia di *cybersecurity*, frutto dell'azione articolata di una varietà di organismi che svolgono attività e funzioni in stretta cooperazione tra loro, con l'obiettivo di identificare forme e mezzi di prevenzione, contrasto e punizione della criminalità informatica, si sia concretizzata prevalentemente nell'adozione di atti non vincolanti. Ne deriva la conseguenziale necessità di strumenti vincolanti, e tra di essi una futura convenzione, da negoziarsi nell'OAS, e volta contrastare efficacemente la criminalità informatica nell'emisfero americano.

6.2. La nozione di *cybersecurity*

È noto che il termine *cybersecurity* trae origine dalla fusione del suffisso *cyber* con la parola «security»[4]. Esso è stato coniato per indicare quella peculiare area multidisciplinare del settore dell'*Information and communications technology* (ICT) che si occupa dei meccanismi normativi, tecnologici e regolamentari, volti a proteggere computer, sistemi informatici, reti informatiche e tecnologie digitali – incluse le informazioni da essi conservate o trasmesse – da qualsiasi forma di minaccia come, ad esempio, l'accesso alle informazioni e il loro utilizzo non autorizzati, la loro sottrazione, modifica, deterioramento o distruzione[5]. Si sarebbe portati a ritenere, insomma, che la nozione di *cybersecurity* descriva un concetto di limitata portata, applicabile esclusivamente al mondo dei computer. In realtà così non è: la creazione, la gestione e l'utilizzo di strumenti di calcolo digitali, insieme

[2] L'amministrazione degli Stati Uniti, ad esempio, ha adottato il 30 novembre 2023 un *Executive Order* (cfr. *The White House*, *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, 30 novembre 2023) destinato a fissare nuovi standard di sicurezza in relazione all'utilizzo delle tecnologie collegate all'intelligenza artificiale e in materia di cybersecurity.

[3] Sull'armonizzazione normativa dei diversi ordinamenti statali nell'ottica del diritto internazionale, v. Ruotolo 2014, p. 548.

[4] Orji 2012, p. 10-16.

[5] In tale contesto, deve essere considerata quale «minaccia» qualsiasi «potenziale violazione della sicurezza» (così *International Telecommunications Union (ITU)*, *Security in Telecommunications and Information Technology: An Overview of Issues and the Deployment of Existing ITU-T Recommendations for Secure Telecommunications*, *Annex A: Security Terminology*, Ref. H.235 and X.800, Geneva, December 2003, 57. Anche i disastri naturali possono porre dei rischi per la protezione dei sistemi informatici e costituiscono, conseguentemente, una «minaccia».

alle risorse di informazione e comunicazione, sono diventate «questioni vitali» nella società odierna[6]. Quindi, quella in esame rappresenta una terminologia piuttosto ampia che «copre ampie sottocategorie che possono andare dalla sicurezza informatica alla sicurezza aeroportuale e alla sicurezza nazionale»[7].

In assenza di una definizione unanimemente accettata, la *cybersecurity* può essere definita come l'insieme di «tools, policies, guidelines, risk management approaches, actions, training, best practices, assurances and technologies that can be used to protect the cyberenvironment and organization, as well as users' assets»[8]. Si tratta, in altre parole, di un complesso di misure di *governance* che possono includere aspetti tecnici, organizzativi, politici e legali.

Gli aspetti tecnici della *governance* della *cybersecurity* investono essenzialmente lo sviluppo e la realizzazione delle necessarie misure tecniche di protezione per i sistemi informatici e le infrastrutture di rete. Gli aspetti organizzativi riguardano, invece, lo sviluppo delle capacità delle istituzioni di promuovere la *cybersecurity* – come ad esempio l'istituzione di organismi incaricati di assicurare l'applicazione delle normative – oppure la creazione di organismi competenti a gestire i rischi di incidenti di *cybersecurity* e le conseguenze del loro verificarsi – quale l'istituzione di *Computer Emergency Response Team* (CERT), deputati a fornire servizi di prevenzione e di allerta precoce, nonché a provvedere al rilevamento, alla reazione e alla gestione delle crisi.

Gli aspetti giuridici della *governance* della *cybersecurity* includono le misure finalizzate a promuovere la sicurezza informatica e lo sviluppo di una società dell'informazione sicura e sostenibile. Un ruolo di importanza cruciale è svolto in tale contesto dall'adozione di normative destinate a vietare le condotte che infrangono la sicurezza dei sistemi o delle reti informatiche e gli attacchi alle infrastrutture informatiche. Queste normative, che possono essere adottate sia a livello di singoli sistemi giuridici nazionali sia a livello internazionale, si fondano sulla criminalizzazione delle condotte che violano la segretezza e l'integrità dei sistemi informatici, delle reti e dei dati informatici, oppure che utilizzano impropriamente siffatti sistemi, reti ed informazioni. Il meccanismo punitivo di queste normative prevede, di regola, adeguate misure di contrasto di siffatti reati, facilitando la loro individuazione, la realizzazione delle indagini da parte delle autorità di polizia, nonché l'esercizio dell'azione penale[9]. Non solo, tali normative possono spingersi altresì ad estendere la criminalizzazione a quelle condotte che integrano un uso immorale, o illecito, dei computer e delle reti informatiche; e ciò anche quando tali condotte non costituiscano di per sé un attentato alla sicurezza dei computer o delle infrastrutture informatiche in rete[10].

---

[6] Floridi 2010, p. 4.

[7] Swire, Steinfeld 2002, p. 105.

[8] ITU High Level Experts Group [HLEG], *ITU Global Cyber-Security Agenda (GCA) High Level Experts Group [HLEG] Global Strategic Report*, Geneva, 2008, p. 27.

[9] Si veda ad es. il Preambolo della Convenzione del Consiglio d'Europa sulla criminalità informatica (CoE, *Convention on Cybercrime Budapest*, 23 novembre 2001, *European Treaty Series* No. 185).

[10] Esempi di condotte rientranti in siffatta categoria sono l'utilizzo di reti informatiche per attività, quali: la distribuzione o la vendita di materiale pornografico vietato, come la pedopornografia, la diffusione di materiale xenofobo e la violazione del *copyright*. In ambito regionale europeo la Convenzione del Consiglio d'Europa sulla criminalità informatica fissa agli art. 9 e 10 rispettivamente l'obbligo per gli Stati contraenti di criminalizzare la pedopornografia e le violazioni del diritto d'autore. Il Protocollo addizionale alla stessa Convenzione, relativo alla criminalizzazione di atti di natura razzista e xenofoba commessi attraverso sistemi informatici fissa analogo obbligo degli Stati membri in relazione alla diffusione di materiale xenofobo (cfr. Consiglio d'Europa, *Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems*, 28 gennaio 2003, *European Treaty Series* No. 189).

Insomma, da un lato, la nozione di *cybersecurity* non può essere limitata esclusivamente alla prevenzione e criminalizzazione degli atti dolosi contro la sicurezza dei sistemi e delle reti informatiche; dall'altro lato, gli aspetti di regolamentazione della *cybersecurity* costituiscono l'elemento centrale ai fini del controllo della criminalità informatica.

Le condotte oggetto di divieto in base alle normative sulla sicurezza informatica sono comunemente definite «crimini informatici» o anche «computer crimes». I due termini sono spesso utilizzati in modo equivalente per indicare tutte quelle situazioni in cui le tecnologie digitali sono l'obiettivo di un'attività illecita oppure lo strumento impiegato per facilitare la commissione di un crimine o di un'attività illecita. In altre parole, la terminologia «crimine informatico» è usata come termine omnicomprensivo per includere tutte le tipologie di reati perpetrati con l'ausilio di risorse informatiche, indipendentemente dal fatto che la natura del bersaglio finale sia o meno una risorsa informatica[11]. Esso include, pertanto, i tradizionali crimini informatici nonché i crimini contro i sistemi e le reti informatiche[12]. Pertanto, la nozione in esame racchiude due categorie di crimini che è possibile distinguere concettualmente: i reati informatici «impropri» ed i reati informatici «propri». Nel primo caso ci si riferisce ai reati comuni previsti dalla normativa penale, che solo *incidentalmente* vengono commessi attraverso l'uso di un computer e della rete: esempi noti sono i reati di ingiuria e di diffamazione (che possono perfezionarsi anche attraverso la posta elettronica, le chat o un sito Internet), le molestie (perpetrate attraverso lo *spamming* o sui social network) e altri reati più gravi come l'istigazione a delinquere, l'istigazione all'odio razziale, il riciclaggio (c.d. c*yber-laundering*) o la pedopornografia[13]. Nel secondo caso, invece, si tratta di reati perpetrati proprio allo scopo di colpire un sistema informatico.

L'assenza di una definizione giuridica universalmente accettata di crimine informatico o criminalità informatica emerge anche nella prassi. Le normative sulla sicurezza informatica, infatti, tendono normalmente ad evitare una definizione esplicita di queste due tipologie di crimini. Ad esempio, la stessa Convenzione del Consiglio d'Europa sulla criminalità informatica, che allo stato attuale è l'unico trattato esistente a livello internazionale in materia, non definisce esplicitamente i termini «criminalità informatica» o «crimine informatico». Tuttavia, la Convenzione criminalizza una serie di condotte negli articoli da 2 a 10 in quattro diverse categorie, vale a dire: a) reati contro la riservatezza, l'integrità e la disponibilità dei dati e dei sistemi informatici; b) reati informatici; c) reati relativi ai contenuti; d) reati relativi a violazioni del diritto d'autore e dei diritti connessi. Tali reati costituiscono ai sensi della Convenzione di Budapest lo standard minimo di ciò che costituisce un reato informatico o le forme di criminalità informatica[14].

A livello universale, l'Assemblea generale delle Nazioni Unite con la risoluzione 74/247 ha istituito un *Open-ended ad hoc intergovernmental committee of experts*,

---

[11] Per una definizione di criminalità informatica, v. Pocar 2008, p. 633-635. La criminalità informatica è stata definita anche come quell'insieme di «computer-mediated activities which are either illegal or considered illicit by certain parties and which can be conducted through global electronic networks» (cfr. Hale, 2002, p. 65).

[12] Gercke 2009, p. 17.

[13] Nulla esclude, tra l'altro, che il mezzo informatico possa essere utilizzato anche per commettere crimini internazionali, con tutte le conseguenze che ne discendono in termini di punizione di siffatti crimini attraverso il sistema di diritto internazionale penale (v., *ex multis*: Roscini, p. 247 ss.).

[14] SSchjolberg 2008, p. 8-9.

con il compito di elaborare una Convenzione internazionale sulla criminalità informatica[15]. I negoziati, tutt'ora in corso, hanno prodotto sei sessioni di lavoro ed una prima bozza di trattato che fornisce una definizione di «crimini informatici» in termini di «the use of information and communications technologies for criminal purposes» nel corso del testo, ma non all'interno dell'art. 2 esplicitamente dedicata alle definizioni (c.d. *Use of Terms*) [16].

6.3. Il sistema interamericano in materia di *cybersecurity*

A differenza di quanto verificatosi in altri contesti regionali, nel continente americano il dibattito circa l'introduzione di strumenti regolamentari e di *governance* in materia di sicurezza informatica è relativamente recente[17]. Il ritardo nello sviluppo di iniziative regionali in tema di *cybersecurity* può essere ricondotto al fatto che è la stessa diffusione delle tecnologie informatiche nei Paesi di siffatto continente – fatte salve alcune limitate eccezioni – ad essere avvenuta solo in tempi più recenti rispetto ad altre esperienze regionali. È, pertanto, solamente intorno agli inizi degli anni 2000 che l'avvento di Internet e la crescita dei crimini compiuti attraverso il mezzo informatico hanno iniziato a suscitare le preoccupazioni negli Stati membri dell'Organizzazione, e a sospingere l'esigenza di affrontare i problemi della sicurezza informatica attraverso un insieme di *policy* da adottarsi tramite la cooperazione regionale.

Il sistema interamericano in materia di *cybersecurity* si presenta particolarmente articolato ed evoluto, in quanto si basa sul coinvolgimento di una molteplicità di soggetti – organi pubblici nazionali ed internazionali così come *stakeholders* privati – che svolgono le loro diversificate funzioni in stretta cooperazione e sinergia, al fine di costruire e rafforzare le competenze normative, economiche e tecniche degli Stati membri dell'Organizzazione in questo specifico ambito. Il suo fondamento è da individuare nell'apposita strategia elaborata nel 2004 dall'Assemblea Generale con la Risoluzione XXXIV.O/04, che, ispirandosi ad un approccio multidisciplinare e multidimensionale inteso a creare e sviluppare una vera e propria cultura della *cybersecurity* nell'ambito degli Stati membri dell'OSA, rappresenta in sostanza il manifesto dell'organizzazione americana nel contesto della cybersicurezza[18].

---

[15] Cfr. *Countering the use of information and communications technologies for criminal purposes, Resolution adopted by the General Assembly on 27 December 2019*, UN Doc. A/RES/74/247 del 20 gennaio 2020.

[16] Cfr. *Ad Hoc Committee to Elaborate a Comprehensive International Convention on Countering the Use of Information and Communications Technologies for Criminal Purposes*, Sixth session, New York, 21 August-1 September 2023, UN Doc. A/AC.291/22 del 29 maggio 2023.

[17] Ad esempio, nel contesto regionale europeo il dibattito sulla natura internazionale dei crimini informatici e sui c.d. *computer crimes* è presente nell'agenda delle istituzioni del Consiglio d'Europa sin dal 1976, anno in cui il tema fu sollevato nel corso di alcune conferenze organizzate sotto l'egida del Consiglio d'Europa e aventi ad oggetto il problema dei crimini economici (cfr. Consiglio d'Europa, *Press release*, *Twentieth Conference of Directors of Criminological Research Institutes: Criminological Aspects of Economic Crime*, Strasbourg, 15-18 November 1976, *Summary of the Reports*, R(76) 13, Appendix 2, p. 4).

[18] Assemblea Generale dell'Organizzazione degli Stati Americani, *Adoption of a Comprehensive Inter-American Strategy to Combat Threats to Cybersecurity: A Multidimensional and Multidisciplinary Approach to Creating a Culture of Cybersecurity*, AG/Res. 2004 (XXXIV-O/04), 8 giugno 2004. In realtà, il primo atto che originariamente conteneva *in nuce* la strategia in questione è la risoluzione n. 1939 dell'Assemblea Generale (Assemblea Generale dell'Organizzazione degli Stati Americani, *Development of an InterAmerican Strategy to Combat Threats to Cybersecurity*, AG/Res. 1939 (XXXIII-O/03), 10 giugno 2003) su cui appunto si basa la più articolata e dettagliata risoluzione del 2004. Per una panoramica di una serie di importanti atti adottati dall'OAS nell'ambito della *cybersecurity*, si veda: Center for Cyber Security and International Relations Studies, *Organisation of American States*, www.cssii.unifi.it/vp-174-oas.html

Gli orientamenti dell'OSA volti a stabilire una strategia dettagliata in materia di cybersicurezza affondano le origini in due ordini di considerazioni strettamente correlate: da un lato, nella consapevolezza dell'importanza assunta nel corso del tempo da Internet, dalla rete e dalle nuove tecnologie nel quadro di sviluppo dell'economia globale e nella prospettiva del conseguimento dell'efficienza e della produttività delle attività commerciali, industriali ed intellettuali nel continente americano[19]; dall'altro, nella necessità di contrastare le forme patologiche di utilizzo degli strumenti tecnologici – quali, la commissione di crimini informatici o la distruzione, anche temporanea, di strutture informatiche riservate, di infrastrutture critiche[20] e di sistemi economici e finanziari statali – che si concretizzino in serie minacce alla *cybersecurity* e siano, pertanto, in grado di pregiudicare il funzionamento di un intero Paese[21].

L'insieme di questi elementi ha favorito l'elaborazione di una strategia interamericana sulla sicurezza informatica che, nel richiamare espressamente alcune note risoluzioni adottate dal Consiglio di Sicurezza delle Nazioni Unite riguardanti il cyberspazio – quelle, segnatamente, tese a contrastare il fenomeno della criminalità informatica, a costruire una cultura globale della *cybersecurity* e a tutelare i sistemi critici informatici[22] – si basa su quattro pilastri: 1. il rafforzamento delle conoscenze degli utenti e degli operatori di Internet in merito alla loro sicurezza e ai propri computer, alle minacce collegate all'utilizzo della rete e agli strumenti esistenti per difendersi dai rischi connessi al cyberspazio; 2. la promozione e il potenziamento dei partenariati pubblico-privati, al fine di incrementare l'istruzione, la consapevolezza e la cooperazione del settore privato, perché i suoi rappresentanti – che costituiscono i principali proprietari e gestori delle infrastrutture critiche ed informative da cui dipendono gli Stati – proteggano effettivamente dette infrastrutture[23]; 3. l'identificazione e l'elaborazione di standard tecnici nella prospettiva di garantire la sicurezza delle informazioni trasmesse su Internet e sulle reti; 4. la promozione dell'adozione di normative e politiche sulla criminalità informatica, che tutelino gli

---

[19] AG/Res. 2004 (XXXIV-O/04), p. 4.

[20] Secondo la definizione fornita dalla CICTE, per infrastrutture critiche si intendono «le strutture, i sistemi, le reti, i servizi e le apparecchiature fisici o virtuali (IT), la cui disabilitazione o distruzione è in grado di produrre un grave impatto non solo sulle popolazioni, sulla salute pubblica, sulla sicurezza, sull'attività economica, sull'ambiente e sui servizi di una Nazione, ma anche sulle capacità del governo di uno Stato membro di operare in modo efficiente» (Inter-American Committee Against Terrorism (CICTE), *Declaration Protection of Critical Infrastructure From Emerging Threats*, CICTE/doc.1/15, 20 marzo 2015, par. 11; traduzione nostra).

[21] AG/Res. 2004 (XXXIV-O/04), p. 4.

[22] A/RES/55/63 on combating the criminal misuse of information technologies, del 4 dicembre 2000; A/RES/56/121 on combating the criminal misuse of information technologies, del 19 dicembre 2001; A/RES/57/239 on creation of a global culture of cybersecurity, del 20 dicembre 2002; A/RES/58/199 on the creation of a global culture of cybersecurity and the protection of critical information systems, del 23 dicembre 2003 (vedi: AG/Res. 2004 (XXXIV-O/04), p. 2).

[23] A tal riguardo vanno menzionate alcune importanti collaborazioni avviate tra l'Organizzazione degli Stati americani e la nota società statunitense Amazon Web Services. Nel 2017 esse hanno stipulato un accordo incentrato sull'attuazione di quattro obiettivi: 1. webinar sulla sicurezza informatica e sulla trasformazione IT (in inglese, spagnolo e portoghese); 2. partecipazione a numerosi eventi di sicurezza informatica organizzati dall'OAS e tenuti nei Paesi Membri; 3. elaborazione di Libri Bianchi sulle politiche e sulle migliori pratiche di sicurezza informatica; 4. collaborazione sulle principali iniziative politiche e legislative (AWS Public Sector Blog, *AWS Teams up with the Organization of American States on Cybersecurity*, disponibile online all'indirizzo: aws.amazon.com). Nel 2018 l'organizzazione interamericana e il colosso statunitense hanno elaborato il primo libro bianco intitolato: «A Call to Action to Protect Citizens, the Private Sector and the Government» (www.oas.org/en/sms/cicte/awswhitepaper.pdf), che indica non solo una serie di misure concrete da adottare con riguardo a molti settori – quali, la cittadinanza e il settore privato, gli operatori delle infrastrutture critiche, il governo e le pubbliche amministrazioni, la difesa informatica, la lotta contro la criminalità informatica e la valorizzazione dell'imprenditorialità e dei talenti per garantire la sicurezza informatica – ma anche una metodologia articolata in sette fasi per lo sviluppo di strategie nazionali di sicurezza informatica (OAS Press Release, *OAS and Amazon Web Services Team Up for Increased Cybersecurity for North American and Latin American Citizens, Businesses and Governments*, www.oas.org/en/media_center/press_release.asp?sCodigo=E-015/18).

utenti della rete ed impediscano l'utilizzo improprio e criminale dei sistemi informatici e computerizzati[24].

A tal uopo, la strategia elaborata dall'OSA prevede un assetto istituzionale fondato sull'attività ed interazione di tre importanti organismi: 1) il CICTE (*Inter-American Committee Against Terrorism*), avente il compito di prevenire e contrastare il terrorismo nel continente americano e di promuovere a tal fine la cooperazione ed il dialogo tra gli Stati Membri, conformemente ai principi della Carta OSA e alla Convenzione interamericana contro il terrorismo del 2002[25]; 2) la CITEL (*Inter-American Telecommunication Commission*), organo dell'OSA che persegue l'obiettivo di facilitare e incoraggiare lo sviluppo sia delle telecomunicazioni sia delle tecnologie dell'informazione e della comunicazione nell'emisfero americano compatibilmente con il principio dello sviluppo sostenibile[26]; 3) le REMJA (*Meetings of Ministers of Justices, other Ministers, Prosecutors and Attorney Generals of the Americas*), ovvero le riunioni dei Ministri della Giustizia, degli altri Ministri, nonché dei Procuratori generali delle Americhe, che rappresentano i principali *forum* dell'OSA e dei suoi Stati membri in materia di giustizia penale e cooperazione giuridica internazionale: più esattamente queste riunioni costituiscono un importante luogo di incontro delle autorità legali e giudiziarie delle Americhe per lo scambio di informazioni, le esperienze e il coordinamento delle politiche pubbliche, con l'obiettivo di rafforzare la collaborazione giudiziaria nel continente americano[27].

6.4. L'elaborazione dell'agenda interamericana sulla sicurezza informatica: il programma sulla *cybersecurity*

L'elaborazione dell'agenda interamericana sulla sicurezza informatica, nonché la lotta alla criminalità ad essa relativa ed il sostegno agli Stati membri nella lotta contro le nuove forme di criminalità informatica, rientrano nelle funzioni svolte dal CICTE[28], cui è affidato il coordinamento di due specifici programmi in materia: a) il programma sulla *cybersecurity*; b) il programma di cooperazione sul *cybercrime*.

Il programma sulla *cybersecurity* è stato avviato nel 2008 con lo scopo di assistere gli Stati membri dell'OSA nel costruire le necessarie competenze tecniche e politiche in materia di sicurezza informatica. Il programma, che costituisce parte integrante della strategia OSA in tale ambito, persegue l'obiettivo di fare in modo che «in tutto l'emisfero occidentale possa essere garantito un cyberspazio aperto, sicuro e resiliente»[29].

Il programma si fonda su tre pilastri. In primo luogo, esso fornisce agli Stati membri dell'OSA assistenza nella elaborazione e nello sviluppo di politiche e strategie in materia di cybersicurezza, che coinvolgono tutti gli *stakeholders* interessati e che sono adattate alla situazione legislativa, culturale, economica e strutturale di

[24] AG/Res. 2004 (XXXIV-O/04), p. 5.

[25] CICTE (Inter-American Committee Against Terrorism), *What We Do*, www.oas.org/en/sms/cicte/

[26] CITEL (Inter-American Telecommunication Commission), *About CITEL*, www.oas.org/ext/en/main/oas/our-structure/agencies-and-entities/citel/About/Details/category/citel/about-citel, par. 1

[27] Cooperation in Justice-REMJA, *What is REMJA?*, www.oas.org/en/sla/dlc/remja-en/remja.asp

[28] Il CICTE è stato istituito dall'OAS nel 1999 con lo scopo di rinforzare la cooperazione tra i Paesi membri per prevenire, combattere ed eliminare gli atti e le attività terroristiche (cfr. OAS, Assemblea generale, *Hemispheric Cooperation to Prevent, Combat and Eliminate Terrorism*, AG/RES. 1650 (XXIX-O/99), del 7 giugno 1999).

[29] Cfr. https://www.oas.org/en/sms/cicte/prog-cybersecurity.asp

ciascun Paese. In secondo luogo, il programma contribuisce a migliorare e rinforzare le competenze delle istituzioni nazionali nel settore della cybersicurezza, provvedendo ad istituire CSIRTs a livello interno e fornendo alle autorità statali sia supporto tecnico individuale sia esercitazioni e corsi di formazione mirati. In terzo luogo, il programma svolge attività di ricerca e sensibilizzazione: a) elaborando documenti tecnici, *toolkit* e rapporti per orientare i decisori politici, i responsabili delle infrastrutture critiche ed i rappresentanti vuoi del settore privato vuoi della società civile; b) mettendo in risalto gli sviluppi relativi alla cybersicurezza; c) individuando le principali problematiche e sfide concernenti la sicurezza informatica nel continente americano[30].

Nell'ambito di queste attività il programma si è contraddistinto per una prolifica e significativa produzione di pubblicazioni concernenti svariati profili relativi alla cybersicurezza nell'emisfero americano.[31]

Con riguardo specifico all'attività della CICTE nel contesto della cybersicurezza, va evidenziato che essa ha il precipuo mandato di definire e sviluppare progetti per l'istituzione di una rete di specifici organismi costantemente operativi nell'emisfero americano (denominati *Computer Security Incident Response Teams* – CSIRTs; squadre nazionali di risposta agli incidenti di sicurezza informatica), aventi il compito di disseminare in maniera rapida ed idonea informazioni relative alla *cybersecurity* e di fornire supporto tecnico nel caso si verifichino cyberattacchi[32].

Nella visione strategica dell'OSA, i CSIRT istituiti a livello nazionale devono: essere designati da ciascun governo ed essere accreditati in base a pertinenti norme di diritto internazionale concernenti la sicurezza informatica; cooperare tra loro e scambiarsi informazioni in base a criteri di fiducia reciproca; disporre di infrastrutture sicure per gestire informazioni riservate; essere in grado di interagire con il settore privato; impegnarsi per consolidare le conoscenze delle proprie comunità nella prospettiva di renderle edotte circa la identificazione delle minacce alla *cybersecurity* e gli strumenti atti a neutralizzare e contrastare le minacce medesime[33]. All'interno del programma opera, inoltre, il network *CSIRT Americas*, la rete dei team di reazione istituiti dai Paesi OSA, che fornisce informazioni tempestive sulle minacce alla sicurezza informatica a 29 CSIRT di 20 Stati membri dell'OSA.

Oltre a quelle descritte sopra, il CICTE svolge tutta una serie di funzioni ulteriori. Nel corso degli anni, in particolare, tale organismo ha licenziato una serie di importanti documenti, in cui ha chiesto agli Stati membri dell'OSA di adottare una molteplicità di misure. Ciò, alla luce delle esigenze espresse in modo sempre più insistente dalle autorità statali di proteggere le infrastrutture critiche dai crescenti

---

[30] OAS Cybersecurity Program, www.oas.org/en/sms/cicte/prog-cybersecurity.asp

[31] Cybersecurity Program: Publications, www.oas.org/en/sms/cicte/cybersecurity/publications/

[32] I CSIRT possono essere definiti come organismi istituiti con il compito di fornire servizi di sicurezza informatica prevenzione, rilevamento, mitigazione e reazione agli attacchi informatici che possono avere luogo in una determinata comunità. Può accadere, infatti, che posta dinanzi ad un attacco informatico un ente o un'organizzazione possano non avere le conoscenze o l'esperienza necessarie per farvi fronte. In tali situazioni, organismi come i CSIRT possono rappresentare uno strumento fondamentale per elaborare una risposta coordinata ed efficiente ad un attacco, contribuendo a mitigarne le conseguenze. Un CSIRT è dotato, in genere, di una struttura organizzativa che include una serie di processi consolidati, un catalogo di strumenti tecnologici, un budget, un catalogo di servizi, del personale specializzato, una rete di contatti, un piano di comunicazione, ed un quadro giuridico che ne disciplina l'azione. L'insieme di questi elementi creano una base per la gestione degli incidenti informatici e sviluppano metodi per supportare le comunità interessate nella massima misura possibile (cfr. OAS, *A Practical guide for CSIRT. A Sustainable Business Model*, volume 2, 2023, p. 7). In Italia, ad esempio, il CSIRT è istituito presso l'Agenzia per la cybersicurezza nazionale (ACN).

[33] AG/Res. 2004 (XXXIV-O/04), p. 7.

e diffusi attacchi terroristici nel continente americano – così come da altre minacce emergenti, quali l'utilizzo da parte delle organizzazioni terroristiche di Internet, dei sistemi informatici e tecnologici per perseguire i loro obiettivi criminali – e di far funzionare siffatte infrastrutture, adeguatamente e normalmente, attraverso programmi di sicurezza cibernetica efficaci ed evoluti[34].

Più precisamente il CICTE ha sottolineato la necessità che gli Stati membri dell'OSA: a) predispongano un sistema contro il cyberterrorismo a vocazione tanto regionale quanto universale, provvedendo a dare attuazione in maniera efficace non solo alla Convenzione interamericana contro il terrorismo, ma anche ai pertinenti strumenti giuridici universali – ovvero, sia le rilevanti risoluzioni adottate dal Consiglio di Sicurezza dell'ONU sia la Strategia globale antiterrorismo delle Nazioni Unite elaborata dall'Assemblea Generale[35]; b) cooperino tra loro in maniera attiva al fine di impedire ai terroristi di sfruttare le tecnologie, le comunicazioni e le risorse della rete per incitare atti di terrorismo, nella stretta osservanza della privacy individuale, dei diritti umani e delle libertà fondamentali così come della sovranità dei singoli Stati[36]. Inoltre, il CICTE ha a più riprese rilevato l'importanza della identificazione di forme di partenariato pubblico-privato nella lotta al terrorismo, nella prospettiva di assicurare vuoi il buon funzionamento delle infrastrutture critiche vuoi la cybersicurezza dei Paesi dell'OSA[37].

Per quanto attiene alle funzioni della Commissione interamericana per le telecomunicazioni (CITEL) nel contesto della cybersicurezza, va evidenziato che detto organismo svolge un ruolo fondamentale che consiste nella identificazione e nella adozione di standard tecnici, volti ad assicurare la sicurezza informatica e di Internet nell'emisfero americano. Alla base del lavoro della Commissione si pone in particolare la consapevolezza che il partenariato tra autorità governative e settori industriali e commerciali rappresenta uno strumento fondamentale al fine di garantire il buon funzionamento delle reti dei sistemi informatici nel continente americano. Secondo la struttura ideata dalla strategia interamericana, infatti, l'individuazione di parametri tecnici che consentano di elaborare soluzioni di sicurezza informatica – che siano definite, dettagliate ed economicamente sostenibili – deve essere necessariamente realizzata mediante una intensa attività di cooperazione tra le società delle telecomunicazioni e della tecnologia dell'informazione, da un lato, ed i governi degli Stati membri dell'OSA, dall'altro[38].

La CITEL svolge le sue funzioni in maniera articolata, graduale e prospettica. Più precisamente, l'identificazione degli standard tecnici di sicurezza e la conseguente raccomandazione agli Stati membri dell'OSA di approvare i medesimi rappresentano il culmine di un processo caratterizzato, non solo dall'esame e dalla valutazione di una serie di importanti elementi (ovvero, gli approcci regionali alla sicurezza della

---

[34] CICTE, *Declaration Protection of Critical Infrastructure From Emerging Threats*, 2015, par. 13.

[35] Inter-American Committee Against Terrorism (CICTE), *Declaration Strengthening Cyber-Security in the Americas*, CICTE/DEC.1/12 rev. 1, 7 marzo 2012, par. 4; *Declaration Protection of Critical Infrastructure From Emerging Threats*, 2015, par. 5 (si veda in particolare: Consiglio di sicurezza, risoluzioni: 1267 (1999) del 15 ottobre 1999; 1373 (2001) del 28 settembre 2001; 1540 (2004) del 28 aprile 2004; 1624 (2005) del 14 settembre 2005; 1631 (2005) del 17 ottobre 2005; 2133 (2014) del 27 gennaio 2014; 2170 (2014) del 15 agosto 2014; 2178 (2014) del 24 settembre 2014; Assemblea generale, *The United Nations Global Counter-Terrorism Strategy*, A/RES/60/288 dell'8 settembre 2006).

[36] CICTE, *Declaration Protection of Critical Infrastructure From Emerging Threats*, 2015, par. 10.

[37] Ivi, par. 4-5, p. 8; Inter-American Committee Against Terrorism (CICTE), *Declaration Strengthening Hemispheric Cooperation and Development in Cybersecurity and Fighting Terrorism in the Americas*, CICTE/Dec 1/16, 26 febbraio 2016, par. 22.

[38] AG/Res. 2004 (XXXIV-O/04), p. 8-9.

rete; le normative e strategie adottate dai singoli Stati volte ad assicurare la cybersicurezza; le interconnessioni tra il settore pubblico e quello privato in questo determinato ambito; l'esistenza di risorse disponibili per dare attuazione agli standard individuati), ma anche dal dialogo e dal confronto con gli Stati membri dell'OSA. Sono da segnalarsi, inoltre, alcune significative attività svolte dalla Commissione interamericana per le telecomunicazioni, quali: l'agevolazione della condivisione delle informazioni tra gli Stati membri per assicurare reti sicure; l'assistenza tecnica fornita a tali Stati, anche tramite il supporto e la collaborazione degli *stakeholders* privati; la promozione dei programmi di rafforzamento delle capacità e della formazione, così da far progredire il processo di diffusione delle informazioni tecniche e le pratiche relative alle questioni di sicurezza informatica nel continente americano[39].

6.5. Il programma di cooperazione sul *cybercrime*

Il programma sul *cybercrime*, invece, è stato avviato nel 1999 e insieme alla creazione dell'*Inter-American Portal on cybercrime* e all'istituzione del *Working Group on Cybercrime*, costituisce uno dei risultati più importanti delle REMJA, riunioni svolte con lo scopo di rafforzare la cooperazione, in materia penale, nella regione relativamente alle indagini e alla punizione di siffatti crimini. Le REMJA svolgono la considerevole funzione di assistere gli Stati membri dell'OSA nella lotta alla criminalità informatica, creando le condizioni perché le autorità pubbliche – ovvero, le autorità di contrasto alla criminalità e gli organi giurisdizionali – dispongano degli strumenti giuridici idonei per prevenire e reprimere la commissione di tali reati[40]. Il Gruppo di Esperti sui crimini informatici è un organismo istituito per rafforzare la cooperazione internazionale nella prevenzione della criminalità informatica e nella fase di indagini e di repressione.

In generale, le REMJA e il Gruppo di Esperti: a) forniscono agli Stati supporto per l'elaborazione e l'adozione di normative tese a punire la criminalità informatica, proteggere i sistemi informatici e impedire l'uso dei computer per la realizzazione di attività illegali; b) elaborano soluzioni per assicurare la collaborazione in materia di criminalità informatica tra gli investigatori e le autorità di contrasto che indagano e perseguono la criminalità informatica. L'obiettivo delle REMJA e del Gruppo di esperti è, in definitiva, quello di guidare gli Stati membri dell'OSA nella attività di modernizzazione delle leggi e dei regolamenti al fine di contrastare la criminalità informatica, a livello sia sostanziale che procedurale[41].

Con specifico riguardo al programma sul *cybercrime* il Gruppo di Esperti si occupa di facilitare lo scambio di informazioni ed esperienze tra i suoi membri e formulare le raccomandazioni necessarie per migliorare e rafforzare la cooperazione tra gli Stati membri dell'OSA, le organizzazioni e i meccanismi internazionali. Con specifico riguardo alla cooperazione tra le autorità nazionali in materia di criminalità informatica, il *Working Group* svolge una importante attività di formazione dei procuratori e dei giudici dei Paesi del continente latino-americano su alcune questioni centrali, quali: la valutazione delle prove elettroniche, lo svolgimento delle indagini e l'esercizio dell'azione penale nei confronti di reati informatici.

[39] Ivi.
[40] Ivi, p. 10.
[41] Ivi, p. 10-11.

Altra attività del *Working Group* meritevole di essere segnalata è quella riguardante la promozione dell'attività normativa da parte dei Paesi membri. In particolare, l'adozione (e l'aggiornamento) della normativa e delle misure procedurali necessarie per perseguire e giudicare efficacemente i crimini informatici; l'adozione di legislazioni necessarie per assicurare la raccolta e la custodia di tutte le forme di prove elettroniche per i fornitori di servizi per garantire la conservazione e il recupero delle informazioni memorizzate o in transito. Il *Working Group* ha anche il compito di stimolare gli Stati OSA a sviluppare e attuare strategie nazionali che includano misure per prevenire, indagare e perseguire i reati informatici.

Degno di nota, infine, è il fatto che al *Working Group* è rimesso il compito di promuovere presso i Paesi OSA l'adesione alla Convenzione sulla criminalità informatica adottata, come detto, nell'ambito del Consiglio d'Europa. La Convenzione, in effetti, prevede all'art. 37 la possibilità che il Comitato dei Ministri del Consiglio d'Europa, dopo aver consultato e ottenuto il consenso unanime degli Stati contraenti della Convenzione, possa invitare altri Stati che non siano membri dell'Organizzazione e che non abbiano partecipato alla sua elaborazione, ad aderire alla stessa. Il tendenziale ambito transregionale della Convenzione di Budapest va nella direzione della creazione di un livello minimo essenziale comune di strategie di contrasto alla criminalità transnazionale, che comporta chiaramente la necessità dell'armonizzazione della normativa di contrasto nell'ambito dei vari ordinamenti, anche oltrepassando i singoli ambiti regionali[42].

6.6. Conclusioni

Il presente contributo, dopo aver approfondito le nozioni di *cybersecurity* e *cybercrime*, ha cercato di dar conto dell'evoluzione della strategia sulla *cybersecurity* adottata dall'OSA, e avente a oggetto: a) la creazione di una cultura della *cybersecurity* nel continente americano, b) la promozione della istituzione di un quadro regionale riguardante la sicurezza informatica; c) l'identificazione di strumenti destinati a facilitare la cooperazione giuridica e giudiziaria in materia penale per contrastare la criminalità informatica.

L'analisi condotta consente di trarre alcune riflessioni conclusive. Da un lato, va messo in risalto che l'approccio adottato dall'OSA in materia è espressione di un importante orientamento che dimostra una precisa e apprezzabile volontà dell'Organizzazione americana di identificare forme e mezzi di prevenzione, contrasto e punizione delle forme patologiche di utilizzo degli strumenti tecnologici su Internet e sulle reti informatiche per finalità criminali. Questo orientamento si è cristallizzato attraverso un articolato sistema che, coinvolgendo molteplici organi che svolgono attività e funzioni in stretta cooperazione tra loro, e interessando svariate infrastrutture dell'emisfero americano, è stato in grado di orientare in maniera meritoria le *policy* e le normative di molti Stati membri dell'OSA in materia di *cybersecurity*.

Dall'altro lato, va evidenziato che l'attività dell'OSA in questo settore si è concretizzata prevalentemente nell'adozione di atti non vincolanti, e come tali non in

[42] Allo stato attuale i Paesi del continente americano che hanno aderito alla Convenzione di Budapest sono: Argentina, Brasile, Canada, Cile, Colombia, Costa Rica, Ecuador, Guatemala, Messico, Panama, Paraguay, Perù, Repubblica Dominicana, Stati-Uniti d'America, Trinidad e Tobago e Uruguay (www.coe.int/en/web/conventions/full-list?module=signatures-by-treaty&treatynum=185).

grado di imporsi agli Stati membri dell'Organizzazione. In effetti l'attuazione della strategia dell'OSA ha portato alla predisposizione di importanti atti di *soft law* dal valore raccomandatorio e all'identificazione di standard tecnici nella prospettiva di guidare, assistere e supportare gli Stati membri nell'attività di contrasto alla criminalità informatica, non imponendo loro obblighi di natura giuridica. La strategia dell'OSA, dunque, fondata sulla interazione e cooperazione di una serie di organi e sull'assistenza fornita agli Stati, in definitiva, non presenta (o almeno, non ancora) quella forza impositiva necessaria per creare un sistema effettivo atto a fronteggiare adeguatamente le future minacce ed i rischi collegati al cyberspazio, in grado di incidere gravemente sul funzionamento generale di un intero Paese.

Se è vero a tal riguardo che, come visto, alcuni Paesi del continente americano hanno aderito alla Convenzione europea sulla criminalità informatica di Budapest, è altresì vero che molti altri Stati di tale continente, anche alla luce del fatto che si tratta di uno strumento normativo europeo, non hanno provveduto in tal senso.

Di conseguenza, anche ai fini dell'armonizzazione normativa tra i vari Paesi dell'OSA e della identificazione di parametri e principi comuni vincolanti gli stessi, appare necessario che nell'immediato futuro l'organizzazione americana – così come è stato fatto nell'ambito del Consiglio d'Europa con la Convenzione di Budapest, e nel contesto dell'Unione africana con la Convenzione di Malabo sulla sicurezza informatica adottata nel 2014 ed entrata recentemente in vigore con la ratifica della Mauritania – predisponga uno strumento convenzionale regionale, che, ispirato in ogni caso alla attuale strategia sulla *cybersecurity*, sia inteso a creare un sistema giuridico certo, che sia inteso a contrastare efficacemente (e in misura maggiore rispetto ad oggi) la criminalità informatica nell'emisfero americano.

Bibliografia

Floridi, L. (a cura di) (2012). *The Cambridge Handbook of Information and Computer Ethics*. Cambridge.

Gercke, M. (2009). *Understanding Cybercrime: A Guide for Developing Countries*. Geneva: ITU.

Hale, C. (2002). Cybercrime: Facts & Figures Concerning the Global Dilemma. In *Crime and Justice International*.

Orji, U. J. (2012). *Cybersecurity Law and Regulation*. Nijmegen.

Pocar, F. (2008). Note sullo sviluppo della normativa internazionale sui crimini relativi ai sistemi di informazione. In AA.VV., *Studi in Onore di Umberto Leanza*, I, Napoli, p. 629 ss.

Roscini, M. (2019). Gravity in the Statute of the International Criminal Court and Cyber Conduct that Constitutes, Instigates or Facilitates International Crimes. In *Criminal Law Forum*, p. 247 ss.

Ruotolo, G. (2014). Internet (diritto internazionale). In *Enciclopedia del diritto*, 545 ss.

Schjolberg, S. (2008). *The History of Global Harmonization on Cybercrime Legislation – the Road to Geneva*. Disponibile su: www.cybercrimelaw.net/documents/cybercrime_ history.pdf

Swire, P., Steinfeld, L. (2002). Security and Privacy After September 11: The Health Care Example. In *Minnesota Law Review*.

7

REGOLARE GLI USI DELL'INTELLIGENZA ARTIFICIALE: LA PROPOSTA DI LEGGE BRASILIANA. NOTE A PRIMA LETTURA

di Pietrangelo Marina (CNR-IGSG), Nannipieri Lorenzo (CNR-IGSG), Calderonio Vincenzo (CNR-IGSG)

7.1. Introduzione

I sistemi di intelligenza artificiale (d'ora in avanti IA) sono da tempo impiegati a livello privato e pubblico in una moltitudine di attività. Come noto, la loro penetrazione sociale e soprattutto la rapidità del loro sviluppo da parte di ricercatori e aziende occupano grande parte del dibattito pubblico. Da alcuni anni la discussione ruota anche attorno alla opportunità – o necessità – di interventi regolatori, per limitare o incentivare l'utilizzo e lo sviluppo di tali sistemi. In questo contesto, l'Unione europea ha scelto la via del diritto positivo e, cioè, ha optato per la via regolatoria col fine di mantenere elevato il livello di protezione dei diritti fondamentali delle persone laddove interessati dagli usi dell'IA, senza tuttavia limitarne lo sviluppo. Tali strumenti, specie in prospettiva futura, possono infatti risultare determinanti per migliorare la qualità della vita supportando numerose attività umane.

Questo saggio esamina nel merito la proposta di legge in discussione presso il Parlamento del Brasile, in comparazione con il regolamento (UE) 2024/1689 che stabilisce regole armonizzate sull'intelligenza artificiale.[1] Pur nelle differenze sostanziali tra i relativi sistemi giuridici, l'atto legislativo eurounionale si pone, infatti, come modello normativo di riferimento, trattandosi del primo quadro giuridico unitario sulla "materia" a livello globale.

[1] Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio, del 13 giugno 2024, che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale).

7.2. Il modello regolatorio eurounionale

Il regolamento UE sull'IA è entrato in vigore il 1° agosto 2024, a tre anni circa dalla presentazione della proposta da parte della Commissione europea[2].

Esso ha armonizzato le regole già esistenti sull'intelligenza artificiale e ne ha definite di nuove, prevedendo la loro applicazione ai sistemi di IA sulla base del rischio (cosiddetto approccio "basato sul rischio", cioè tanto maggiore è il rischio di arrecare danni alla società, tanto più rigorose saranno le regole).

Con questo atto legislativo l'Unione affronta per prima la materia, proponendo il proprio metodo come modello regolatorio. Il quadro normativo eurounionale definisce quattro livelli di rischio per i sistemi di IA.

Sono vietati (*unacceptable rick*) i sistemi che si configurano per esempio come minaccia evidente alla sicurezza o ai diritti delle persone. Sono invece ammessi i sistemi ad alto rischio (*high risk*), come quelli di identificazione biometrica remota (vietati se in spazi aperti al pubblico, salvo limitate eccezioni), i sistemi impiegati in attività legate a infrastrutture critiche (es. i trasporti), per l'accesso a percorsi di formazione scolastica o professionale o ad attività lavorative (es. elaborazioni di test a punti per esami o valutazione di curricula), nella costruzione di strumenti usati in ambito chirurgico (es. chirurgia assistita da robot), nei servizi pubblici o privati essenziali (es. per il credit scoring bancario), in ambito giudiziario (es. per la valutazione dell'affidabilità delle prove o la ricerca dei precedenti), in materia di migrazione e asilo (es. per il controllo delle domande di visto).

In tutti i casi di sistemi ad alto rischio, il regolamento UE prevede obblighi rigorosi per i produttori e i fornitori che si applica – quindi – prima che i sistemi vengano commercializzati. Tra questi, sono richiesti sistemi adeguati di valutazione e mitigazione dei rischi, alta qualità dei dati su cui i sistemi sono addestrati, tracciabilità delle attività svolte e documentazione dettagliata relativa al sistema sviluppato, alto livello di robustezza, sicurezza e precisione e, non ultime, adeguate misure di sorveglianza umana per ridurre al minimo i rischi.

Il tema dell'uso responsabile è dunque centrale.

Per il caso dei sistemi con rischio limitato (*limited risk*) sono previsti invece obblighi di trasparenza, per informare gli utilizzatori sui rischi, senza tuttavia ledere l'approccio fiduciario verso l'applicazione. Rientrano in questa categoria anche i sistemi generativi di testi, audio e video informativi su questioni di interesse pubblico, che devono essere etichettati come generati artificialmente (es. i *chatbot*).

Alla categoria residuale dei sistemi a rischio minimo (*minimal risk*) – nella quale di fatto rientra la gran parte delle applicazioni usate oggi sul territorio UE – non si applicano le restrizioni di cui sopra, trattandosi di sistemi liberamente utilizzabili (es. applicazioni antispam, videogiochi).

Il regolamento dell'Unione ambisce a porsi come normativa quadro a livello globale.

---

[2] Nelle Conclusioni del Consiglio europeo del 1 e 2 ottobre 2020 si rintraccia il punto di avvio della successiva proposta regolatoria: "(...) L'UE deve essere un leader mondiale nello sviluppo di un'intelligenza artificiale sicura, affidabile ed etica. Il Consiglio europeo invita la Commissione a: proporre soluzioni per aumentare gli investimenti pubblici e privati europei e nazionali nella ricerca e innovazione nel settore dell'intelligenza artificiale e nella diffusione di quest'ultima; garantire un migliore coordinamento e maggiori reti e sinergie tra i centri di ricerca europei basate sull'eccellenza; fornire una definizione chiara e oggettiva dei sistemi di intelligenza artificiale ad alto rischio." (v. il paragrafo 13, EUCO 13/20, Conclusioni del Consiglio europeo, 2 ottobre 2020).

Dopo l'adozione nel 2016 del Regolamento generale sulla protezione dei dati (GDPR)[3], l'Unione ha nuovamente scelto la via regolatoria per affermare il proprio orientamento politico sui rischi derivanti da utilizzi distorti dell'IA e sulle relative responsabilità.

Il tratto distintivo del regolamento resta in ogni caso la promozione dell'innovazione in materia di IA, mediante il supporto alle imprese del settore digitale, mai disgiunta dalla garanzia della massima protezione per i diritti fondamentali delle persone: una intelligenza artificiale etica, sicura e affidabile.

Come già per il caso della normativa generale brasiliana sulla protezione dei dati (LGPD)[4] adottata sul modello europeo del GDPR, come si dirà più avanti, anche per la regolazione sull'IA sono numerosi gli orientamenti comuni.[5]

7.3. Il disegno di legge sull'intelligenza artificiale d'iniziativa del Governo brasiliano

Nel 2020 il Governo brasiliano ha presentato alla Camera dei deputati il disegno di legge n. 21/2020, su iniziativa del parlamentare Eduardo Bismarck, autore del testo.

La proposta è stata adottata inseguito ad una consultazione pubblica aperta a dicembre 2019 e chiusa nel mese di marzo 2020 dal Ministero della Scienza, Tecnologia e Innovazione che, similmente all'esperienza europea preliminare alla regolamentazione, ha sviluppato una strategia nazionale in tema di IA consultando i vari stakeholder[6].

Il processo partecipativo costituisce un tratto distintivo della legislazione brasiliana sul digitale, in un Paese che ha saputo proporsi negli ultimi decenni come fucina di soluzioni giuridiche innovative proprio in relazione ai mutamenti socioeconomici connessi alla digitalizzazione.[7]

Il testo originario del disegno di legge è stato approvato dalla Camera dei deputati nel mese di settembre 2021. Esso risultava composto da soli dieci articoli: un testo snello, dunque, funzionale a definire *solo* il perimetro essenziale delle attività di sviluppo e uso dell'IA. L'iter della proposta ha subito poi un rallentamento durante l'esame in Senato, nel quale sono emersi con maggiore evidenza il confronto e le differenze sostanziali con il modello regolatorio più esteso dell'Unione europea.

Il dibattito parlamentare ha quindi evidenziato alcuni profili di criticità prima non considerati: sull'ambito oggettivo e i relativi limiti di applicabilità del progetto legislativo; e più in generale, sulla *ratio* sottesa alla regolazione proposta, non

[3] Regolamento (UE) 2016/679 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati.

[4] Si tratta della Lei General de Proteção de Dados Pessoais (LGPD), n.13.709/2018, in vigore dal 2020.

[5] In tal senso, v. i lavori raccolti in Belli 2023a. Per una traduzione non ufficiale in lingua inglese del progetto di legge v. https://cyberbrics.info/non-official-translation-of-the-brazilian-artificial-intelligence-bill-n-21-2020/

[6] Come è stato per l'AI Act, dove la Commissione Europea ha costituito l'high level expert group on AI che ha prodotte le linee guida per la trsutworthy AI, anche il Brasile ha passato una fase preliminare di consultazione pubblica prima della regolamentazione, seppur con une metodologia differente. Di seguito un link d'accesso al testo della Estratégia Brasileira de Inteligência Artificial -EBIA https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligencia-artificial/ebia-diagramacao_4-979_2021.pdf

[7] V. al proposito la nota esperienza della Lei n. 12.965 del 23 aprile 2014, cosiddetta "Marco Civil", ricordata come la prima legge sui diritti e doveri in Internet. Il testo è consultabile all'indirizzo https://www.planalto.gov.br/CCIVIL_03/_Ato2011-2014/2014/Lei/L12965.htm

adeguata a supportare lo sviluppo futuro dei sistemi di intelligenza artificiale in Brasile[8].

Lo stallo parlamentare ha indotto i proponenti ad accantonare il progetto originario e a predisporre un nuovo articolato, più "robusto", composto di 45 articoli. Si tratta del *projeto de lei* n. 2338/2023, che è attualmente all'esame del Senato[9].

Esso annovera disposizioni in materia di tutela dei diritti fondamentali (capitolo 2), alla categorizzazione dei rischi (capitolo 3), governance (capitolo 4) e responsabilità civile (capitolo 5). Il procedimento legislativo prevede l'esame da parte di una commissione interna temporanea di senatori che il 5 dicembre ha approvato il PL n. 2338/2023, il quale è poi passato alla Plenaria, dove il Senato ha approvato la proposta con modifiche il 10 dicembre.

Dal punto di vista sostanziale, il progetto n. 2338/2023 è per molta parte allineato al regolamento UE, almeno per quanto attiene alla scelta della classificazione dei sistemi di IA basata sul rischio per la società ad essi riferibile.

Nel progetto brasiliano le categorie di rischio sono tre (rischio eccessivo, alto e basso) ma, in buona sostanza, esse rispecchiano i parametri della regolazione eurounionale, alimentando in tal modo il cosiddetto *Brussels effect*.[10] L'effetto Bruxelles è d'altro canto critico per la stessa Unione, perché l'idea che essa possa unicamente affidarsi ad esso e cioè alla sua superiore capacità regolatoria per esercitare un'influenza globale ha già subito una smentita ad opera del cosiddetto contro "effetto Draghi"[11], nel quale viene invece additata proprio la ipertrofica regolazione sul digitale come elmento di debolezza quantomeno interno per la protezione degli interessi delle imprese e dei consumatori dell'UE[12].

Anche sotto il profilo definitorio e cioè in relazione all'ambito soggettivo della proposta, si rinvengono punti di contatto.[13] Il sistema di IA è infatti qualificato come un "sistema informatico, con diversi gradi di autonomia" (cfr. art. 4) e il fornitore come "la persona fisica o giuridica, di natura pubblica o privata, che sviluppa un sistema di intelligenza artificiale, direttamente o su ordinazione, al fine di immetterlo sul mercato" (cfr. art. 4), con chiaro riferimento al quadro semantico dell'AI Act e ai principi dell'OCSE del 2019.

Tuttavia, se l'architettura fondamentale del nuovo progetto legislativo brasiliano appare conforme al modello dell'AI Act, non poche sono le differenze con esso.

Anzitutto occorre evidenziare la presenza nel progetto n. 2338/2023 di una prima specifica sezione dedicata alla protezione dei diritti fondamentali in relazione all'uso dell'IA, che manca invece nell'AI Act.

Il disegno di legge brasiliano sembra porre un'enfasi maggiore sul riconoscimento in via autonoma e diretta di nuovi diritti individuali, tra cui il diritto alla spiegazione, il diritto di contestare le decisioni sui sistemi di IA e il diritto alla determinazione e

---

[8] Sul punto si veda il contributo di Belli 2023b che riassume il processo legislativo fino all'anno 2022, di cui è disponibile pure una timeline al sito https://bit.ly/ebiabr

[9] La proposta è accessibile sul sito ufficiale del Senato brasiliano https://legis.senado.leg.br/sdleg-getter/documento?dm=9347622&ts=1730837869278&disposition=inline

[10] Siegmann 2022: https://arxiv.org/abs/2208.12645

[11] Su cui v. il rapporto *The future of European competitiveness* (settembre 2024).

[12] Su tali aspetti, cfr. Parcu 2024, pp. 2 ss. Secondo tali autori, l'UE deve sospendere l'adozione di ulteriori normative, concentrandosi sulla fase di attuazione di quelle già adottate, i cui carichi regolatori per imprese e cittadini sono particolarmente onerosi.

[13] Sulle definizioni, v. Pietrangelo 2023.

partecipazione umana alle decisioni dei sistemi di intelligenza artificiale contenuti nell'art. 6.

La protezione dei diritti fondamentali non è, certo, estranea all'AI Act, ma essi nell'AI Act sono sempre protetti in relazione a interessi economici; questo perché l'UE regolamenta i sistemi di IA come prodotti da immettere sul mercato e a questo fine pone tutta una serie di obblighi in capo ai fornitori.

Come noto, nell'ordinamento dell'Unione vi sono disposizioni di analogo tenore a quelle previste dal progetto brasiliano: ad esempio, il processo decisionale automatizzato è disciplinato dall'articolo 22 del GDPR; nondimeno, anche il diritto alla spiegazione, secondo alcuni commentatori, può essere desunto in via interpretativa da altre disposizioni[14].

Si tratta, in ogni caso, di soluzioni applicabili esclusivamente nel caso della protezione dei dati personali[15].

I rimedi espliciti previsti nella proposta del Brasile segnano invece un più ampio riconoscimento degli impatti sociali dell'IA, che paiono andare ben al di là dell'ambito di applicazione del GDPR dell'UE per protezione dei dati personali e della legge sull'IA per la progettazione dei sistemi di IA.

Merita ricordare che la proposta brasiliana n. 2338/2023 contiene specifiche previsioni sia in merito alla "trasparenza" dei sistemi di IA ad "alto rischio", che relativamente al diritto del cittadino, in forma sia individuale che collettiva, di avvalersi di specifiche azioni giudiziali per tutelarsi da un uso illecito dei sistemi, favorendo anche un adeguato controllo sul funzionamento dell'IA.

In particolare, l'articolo 9 della proposta stabilisce che i sistemi di IA classificati ad "alto rischio" debbano informare gli utenti in modo chiaro, obiettivo e accessibile sulle modalità per esercitare i diritti previsti dalla legge.

Questa previsione intende rafforzare il rilascio trasparente dei sistemi di IA, mirando a garantire che le persone siano pienamente consapevoli delle modalità con cui possono tutelare i propri diritti in relazione ai sistemi automatizzati.

L'obiettivo manifestato dalla proposta risponde ad una finalità educativa e di *empowerment*, assicurando che le informazioni siano comprensibili e facilmente reperibili e riducendo al minimo la complessità tecnica spesso associata ai sistemi di IA avanzati. Inoltre, tali informazioni devono permettere agli utenti di conoscere i propri diritti, come ad esempio il diritto di accesso ai dati, la possibilità di contestare decisioni automatizzate e la protezione dalla discriminazione algoritmica, intendendo così promuovere una maggiore fiducia del cittadino verso la tecnologia.

L'art. 10 prevede, poi, che l'autorità competente stabilisca, in collaborazione con le autorità settoriali, delle linee guida che regolino la forma e le condizioni per l'esercizio dei diritti da parte degli utenti *nei confronti* dei sistemi di IA.

Questa previsione implica che il processo di attuazione dei diritti non sia generico, ma modellato in base alle specificità dei vari settori in cui l'IA viene impiegata.

La cooperazione tra le autorità competenti e le autorità settoriali è essenziale per assicurare che le linee guida siano pertinenti e applicabili a ciascun ambito, che può spaziare dalla sanità, alla giustizia, ai trasporti. La necessità di adattare le normative ai contesti specifici garantisce una regolamentazione più precisa e

[14] Selbst 2018; Kaminski 2021, p. 278-299; Goodman 2017.
[15] Almada 2019.

meno suscettibile di lacune. Tale previsione favorisce una gestione più organica e coordinata dei diritti individuali, prevenendo eventuali conflitti tra le normative generali e le esigenze pratiche di ciascun settore.

Infine, l'art. 11 descrive i canali attraverso cui gli individui possono difendere i propri diritti. La difesa può avvenire davanti agli organi amministrativi competenti o tramite il ricorso ai tribunali, con azioni sia individuali che collettive.

La previsione di uno strumento collettivo di azione giudiziaria parrebbe orientato a permettere una protezione giuridica estesa, consentendo anche alle comunità o gruppi vulnerabili di rivendicare i propri diritti in relazione all'uso delle tecnologie di IA.

Simile previsione non è rinvenibile in altri testi normativi, compreso il Regolamento europeo.

In buona sostanza, le decisioni dell'IA saranno, cioè, contestabili davanti a un giudice, al quale potrà essere richiesto l'intervento umano.

L'AI Act non prevede tali misure (i mezzi di ricorso sono trattati marginalmente dal capo IX, sezione 4), essendo incentrata – come più volte rilevato – sulla previsione di obblighi di regolamentazione del mercato per le imprese e i fornitori di sistemi di IA.

Questa differenza riflette in parte i diversi assetti istituzionali su cui insistono le nuove regole. Nel caso dell'Unione, come noto, l'ambito d'elezione dell'intervento legislativo è il mercato interno; nel caso del Brasile invece i profili della regolazione paiono più ampi nei limiti di quanto di competenza di uno Stato nazionale.

7.3.1 Sulla responsabilità per danni da IA

Un ulteriore e significativo tratto che differenzia il disegno di legge brasiliano dall'AI Act riguarda i profili della responsabilità civile, che – come noto – mancano nel regolamento UE, essendo oggetto di una specifica direttiva, la AI *liability directive*[16].

Si tratta di un atto specifico sulla responsabilità da intelligenza artificiale con cui l'Unione integra e aggiorna il precedente quadro normativo in materia di responsabilità civile, introducendo per l'appunto disposizioni specifiche per i danni causati dai sistemi di IA. Esse mirano a tutelare mediante risarcimento chi subisce danni causati dalla tecnologia di IA.

In pratica, il disegno normativo della UE ruota attorno a due principali misure: la cosiddetta "presunzione di causalità", per cui i soggetti danneggiati non avranno l'onere di provare il danno per colpa o omissione; e l'accesso agli elementi di prova detenuti da imprese o fornitori nei casi di sistemi ad alto rischio.

Alla data odierna, l'iter della proposta di direttiva registra uno stop, a seguito della richiesta del Parlamento europeo di integrare la procedura con una valutazione d'impatto complementare.[17] Con un cambio netto di prospettiva rispetto alla proposta di direttiva, della quale peraltro si propone la soppressione in favore di un più ampio e robusto regolamento in materia di software, la nuova analisi d'impatto propende per il regime di responsabilità oggettiva rispetto alla responsabilità colposa.

---

[16] Proposal for a directive of the European parliament and of the council, on adapting non contractual civil liability rules to artificial intelligence (AI Liability Directive), COM (2022) 496 final, 2022.

[17] L'analisi d'impatto è in Hacker 2024.

Un mutamento che non stupisce, perché appare in linea con la coeva recente modifica della *Product liability directive*[18], nella quale è stata introdotta la nuova fattispecie del software difettoso.

A ben vedere, in talune sue parti le argomentazioni della nuova valutazione d'impatto risulta paiono contraddittorie, ad esempio, quanto al metodo e alla *ratio* delle modifiche alla proposta di direttiva. Il documento, infatti, da un lato auspica una minore regolazione, cui corrisponderebbe la richiesta di soppressione dell'*AI liability directive,* dall'altro suggerisce di adottare un nuovo regolamento in materia di software, il quale peraltro rischia di sovrapporsi a normative vigenti, a partire dalla *Product liability directive* fino al Regolamento sui servizi digitali.[19]

Quanto al merito, il favore per la previsione espressa di un regime di responsabilità oggettiva non pare allo stato attuale sufficientemente motivato, né adeguatamente messo in relazione con la disciplina sulla responsabilità per l'esercizio di attività pericolose, in ipotesi suscettibile d'interpretazione estensiva alla luce del *risk-based approach* dell'AI Act.

A fronte di questo scenario ancora in via di definizione, ancora una volta suscita interesse la proposta brasiliana, che "regola senza regolare" o "regola senza scegliere" l'una o l'altra fattispecie, lasciando percorribili entrambe le soluzioni.

Anzitutto, le disposizioni sulla responsabilità civile dei sistemi di IA sono innestate nel disegno di legge e non in atti separati, con ciò evidenziandone ancora una volta l'approccio tendenzialmente unitario alle questioni poste dagli usi dell'IA, come già rilevato per la presenza delle disposizioni sul riconoscimento espresso di nuovi diritti. Le une e le altre assieme peraltro paiono sottolineare l'interesse brasiliano per una regolazione meno ancora (solo) alla commercializzazione dei prodotti di IA. Come già evidenziato nel paragrafo precedente, anche sul tema della responsabilità si coglie il cosiddetto effetto Bruxelles; anzi, si può senz'altro rilevare che le disposizioni del progetto brasiliano risultavano al principio fortemente condizionate proprio dalle posizioni espresse in materia dal Parlamento europeo nella risoluzione adottata nell'ottobre 2020.[20]

Alla data attuale, si registra invece un orientamento differente proprio sul punto specifico del regime di responsabilità. Ferma restando, cioè, l'applicabilità della normativa generale in materia (*Código de Defesa do Consumidor* e *Código Civil*), all'articolo 36 viene delineato un nuovo modello di responsabilità per l'IA, che si basa su due criteri: il livello di autonomia del sistema di IA correlato al suo livello di rischio e la natura dell'agente coinvolto nei danni causati dall'IA.

Come corollario, l'articolo 37 conferisce ai giudici il potere di invertire l'onere della prova nei casi di incapacità giuridica della vittima o quando l'eccessiva complessità del sistema di IA crea una *probatio diabolica*.[21]

Secondo la proposta brasiliana, le decisioni sulle modifiche al regime di responsabilità sono rimesse all'autorità giudiziaria la quale valuterà nei singoli casi tenuto conto delle caratteristiche del sistema e dei diritti coinvolti.

---

[18] Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products and repealing Council Directive 85/374/EEC. La direttiva è stata modifica nel mese di marzo 2024.

[19] Regolamento (UE) 2022/2065 del Parlamento europeo e del Consiglio del 19 ottobre 2022 relativo a un mercato unico dei servizi digitali e che modifica la direttiva 2000/31/CE (regolamento sui servizi digitali).

[20] European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014 (INL).

[21] In tal senso, v. de Teffé 2022, p. 301-333; e Doneda 2018.

Questa clausola generale sulla competenza giurisdizionale consentire di eviterebbe possibili errori derivanti dalla definizione *ex ante* di un regime di responsabilità, consentendo adattamenti del regime per colpa o di quello per responsabilità oggettiva a seconda del caso concreto.

La proposta brasiliana tiene dunque conto di alcuni parametri che di fatto impediscono di ricondurre preventivamente e in astratto a una categoria di rischio un determinato sistema, senza averne prima valutato le possibili interazioni sociali e ambientali e i relativi usi.

Una opzione regolatoria più efficace, almeno sulla carta, della soluzione eurounionale?

Si tratta in ogni caso di un approccio originale, che punta sul "compromesso" tra le due vie alternative della responsabilità per colpa e responsabilità oggettiva, senza scegliere ma rinviando la decisione a un tempo successivo.

7.4. Alcuni spunti conclusivi

Nel mese di dicembre 2024 il progetto di legge n. 2338/2023 è stato approvato dalla Commissione speciale del Senato, prima, e a seguire dall'Aula in seduta plenaria il 10 dicembre. Il testo è dunque ora nuovamente all'esame della Camera dei deputati.

Rispetto al testo originario, esso contiene alcune novità significative: l'esclusione dalla categoria dei sistemi ad alto rischio degli algoritmi di raccomandazione dei *social network*; e la facoltizzazione della valutazione preliminare del rischio per i gestori di piattaforme sociale. Si tratta di due profili di grande rilievo dal punto di vista sociale e democratico, giacché tali previsioni – almeno alla data attuale – indeboliscono il sistema di tutele e garanzie per gli utenti delle piattaforme, pur restando intatte altre previsioni tra cui quelle relative alla garanzia dei diritti di creazione dei contenuti originali o di protezione delle opere originali. Su quest'ultimo punto, infatti, si segnala l'art. 64 che vieta l'utilizzo di opere protette dal diritto d'autore per lo sviluppo di sistemi di IA. Nei mesi scorsi non sono mancate peraltro nuove proposte sostanziali, che in qualche modo hanno mantenuta attiva la discussione non solo a livello parlamentare. Tra queste, merita di essere segnalata l'attivismo del *Conselho Administrativo de Defesa Econômica* (CADE) che ha avanzato anche proposte di rilievo per la governance complessiva dell'IA, come per esempio quella relativa alla costituzione di una specifica autorità con poteri ispettivi in materia di IA, cosiddetto S*istema Nacional de Regulação e Governança de Inteligência Artificial,* mentre allo stato attuale il disegno di legge affida i poteri di coordinamento in materia di IA all'*Autoridade Nacional de Proteção de Dados*.

In una fase in cui la stessa Unione europea è stata sollecitata a fare autocritica sulla scelta di regolare il digitale come strumento di riposizionamento geopolitico senza tuttavia farsi carico dell'attuazione e del relativo impatto dei numerosi plessi normativi già adottati,[22] la fucina brasiliana è senz'altro da tenere sotto osservazione.

[22] V. Parcu 2024.

Bibliografia

Almada, M. (2019). Human intervention in automated decision-making: Toward the construction of contestable systems. In *Proceedings of the Seventeenth International Conference on artificial intelligence and law,* pp. 2-11.

Belli, L., Zingales, N. (a cura di) (2023a). Artificial Intelligence and Data Protection. In *Latin America*, *Computer Law & Security Review,* Special issue.

Belli, L., Curzi, Y., Gaspar, W. B. (2023b). AI regulation in Brazil: Advancements, flows, and need to learn from the data protection experience. In *Computer Law & Security Review,* 48, pp.1-28.

de Teffé, C. S., Medon, F. (2022). Responsabilidade civil e regulação de novas tecnologias: Questões acerca da utilização de inteligência artificial na tomada de decisões empresariais. In *REI-revista estudos institucionais*, *6*(1), pp. 301-333.

Doneda, D. C. M., Mendes, L. S., de Souza (2018). C.A. P., Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. In *Pensar-Revista de Ciências Jurídicas*, *23*(4), pp. 1-17.

Goodman, B., Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation. In *AI magazine, 38*(3), pp. 50-57.

Hacker, P. (2024). *Proposta di direttiva sull'adeguamento delle norme in materia di responsabilità civile extracontrattuale all'intelligenza artificiale. Valutazione d'impatto complementare.* Disponibile su: https://www.europarl.europa.eu/RegData/etudes/STUD/2024/762861/EPRS_STU(2024)762861_EN.pdf

Kaminski, M. E. (2021). The right to explanation, explained. In *Research Handbook on Information Law and Governance*.

Parcu, P. L., Rossi, M. A. (2024). Reconsidering the global dimension in regulating innovation. In K. Majcher (a cura di), *Charting the digital and technological future of Europe: what priorities for the European Commission in 2024-2029?* EUI Report, 2024, p. 2 ss.

Pietrangelo, M., Nannipieri, L. (2023). Intelligenza artificiale e pubbliche amministrazioni, tra incertezze definitorie, usi disomogenei e giudici supplenti. In *CEUR Workshop Proceedings di Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29-31, 2023, Pisa, Italy.*

Selbst, A., Powles, J. (2018). "Meaningful information" and the right to explanation. In *Proceedings of the 1st Conference on fairness, accountability and transparency.* PMLR 81.

Siegmann, C., Anderljung, M. (2022). *The Brussels effect and artificial intelligence: How EU regulation will impact the global AI market. 2022, in arXiv preprint.* Disponibile su: https://arxiv.org/abs/2208.12645

8

L'IA TRA DATI, STEREOTIPI E DIRITTI: IMPRESSIONI DI UNA SPETTATRICE

di Tina Parrella (Università degli Studi di Napoli Federico II e SSM)

8.1. Introduzione

Nel periodo tra marzo e maggio 2024, presso l'Istituto di ricerca su Innovazione e servizi per lo sviluppo (IRISS) del CNR, si è svolto un ciclo di incontri dal titolo "Dialoghi sull'Intelligenza Artificiale", a cura di Natale Rampazzo, dirigente di ricerca presso lo stesso istituto.

Questi seminari sono stati sede d'incontro di numerose elaborazioni sul tema dell'Intelligenza Artificiale, affrontato con un'impronta fortemente multidisciplinare e polifonica: studiosi ed esperti provenienti dai più disparati ambiti, tra cui ingegneria, psicologia, diritto, filosofia, economia, si sono confrontati fornendo nuove prospettive e gettando luce sulle problematiche che questo tema ha generato e genera ancora.

Il presente contributo si propone di enucleare i principali contenuti degli interventi dei Dialoghi a cui ho assistito e di approfondirne gli aspetti che hanno suscitato in me maggiori spunti di riflessione.

In particolare, le relazioni che verranno prese in considerazione sono le seguenti: Oscar Nicolaus, *Tecno-ottimismo vs tecno-pessimismo*; Franco Cutugno, *I modelli alla base dell'IA*; Natale Rampazzo, *Galateo digitale: artifici e raggiri dell'IA*; Antonio Pescapè, *IA: funzionamento, proprietà e criticità*; Marco Fasciglione, *La due diligence sui diritti umani*; Giacomo Maria Cremonesi, *Consulenza aziendale e diritti umani*.

8.2. L'Intelligenza Artificiale nella dicotomia fra tecno-ottimisti e tecno-pessimisti

L'avvento dell'Intelligenza Artificiale, definita dal Forum Economico Mondiale come "Quarta rivoluzione industriale"[1], ha sicuramente rappresentato un'evoluzione del modo con cui si considera e si interagisce con la tecnologia e il mondo digitale.

L'intervento di Oscar Nicolaus[2], relatore del secondo incontro del ciclo, ha messo in evidenza quanto questa rivoluzione abbia aperto innumerevoli dibattiti nella comunità scientifica tutta, accumunati dal fine di comprendere gli aspetti i positivi e i possibili rischi dell'IA. Ciò ha determinato la creazione di una polarizzazione tra due distinti filoni: i tecno-ottimisti e i tecno-pessimisti. Tuttavia, si teme che una tale dicotomia rischi di banalizzare la complessità delle nuove tecnologie e delle sfide inedite che il loro ingresso nella società pone. Bisogna infatti considerare che complessa è innanzitutto la comprensione della struttura e del funzionamento dell'Intelligenza Artificiale rispetto a quella umana: riprendendo l'elaborazione freudiana della ragione umana, che considera che questa sia composta non solo da ragione e razionalità ma anche da sfera emotiva e *raptus*, il contrasto con l'Intelligenza Artificiale, basata invece su dispositivi elettro-meccanici, è evidente. Ciononostante, le due intelligenze sono tra loro interconnesse, perché è la mente umana a creare quella artificiale, a disporre gli algoritmi e a fornire diversi input da elaborare.

Su questo punto si concentra la relazione di Nicolaus: si deve porre attenzione all'intelligenza umana, perché l'errore umano può essere trasferito ai sistemi di IA ed essere perpetuato nel tempo e in diverse applicazioni. Assunta l'opportuna consapevolezza di questi rischi, bisogna guardare all'errore in ottica costruttiva, per comprendere il modo in cui migliorare entrambi i meccanismi di ragionamento, quello umano e quello artificiale.

A valle di questo intervento e delle considerazioni, anche della società civile, sui rischi dell'uso indisciplinato delle nuove tecnologie, mi sono chiesta quali possano essere gli strumenti che permettano un utilizzo controllato dell'IA, per ridurre al minimo gli errori, sia umani sia tecnici. L'Unione Europea, a questo proposito, fin dalle prime elaborazioni sull'IA[3], ha spesso richiamato i concetti di supervisione, di affidabilità della macchina e di IA antropocentrica, al fine di implementare una *trustworthy AI*, caratterizzata da legalità, aderenza etica e solidità.[4] Tale obiettivo però potrebbe rivelarsi irraggiungibile se gli strumenti di controllo non riuscissero ad adattarsi all'evoluzione tecnologica, motivo per cui risulta necessario introdurre

---

[1] Erkiilä 2023, p. 820.

[2] Laureato in Psicologia e Filosofia e ricercatore dell'IRPPS del CNR. Dal 1998 ad oggi, è titolare a contratto della cattedra di Psicologia sociale della famiglia nel corso di Laurea di Scienze della Formazione primaria presso l'Istituto Universitario S. Orsola Benincasa di Napoli.

[3] Cfr. Commissione europea, (2018). Comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni, *L'intelligenza artificiale per l'Europa*, COM(2018)237 *final*; Gruppo di esperti ad alto livello sull'intelligenza artificiale (AI HLEG), (2018). *Orientamenti etici per un'IA affidabile*; *Risoluzione del Parlamento europeo recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica* 2015/2103(INL); Gruppo di esperti ad alto livello sull'intelligenza artificiale (AI HLEG), (2019). *A definition of AI: Main capabilities and scientific disciplines*; Commissione europea, (2020). Libro bianco sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia, COM(2020) 65 *final;* Commissione europea, (2022). *Dichiarazione europea sui diritti e i principi digitali per il decennio digitale*, 26 gennaio 2022, COM(2022) 28 *final*; Regolamento (UE) 2024/1689 del Parlamento europeo e del Consiglio del 13 giugno 2024 che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n, 300/2008, (UE) n, 167/2013, (UE) n, 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale).

[4] Gruppo indipendente di esperti ad alto livello sull'intelligenza artificiale, *Ethics guidelines for trustworthy AI* (2019).

dei meccanismi di adattamento alle novità. Esempio di ciò è presente nell'*AI Act,* il quale prevede che la Commissione europea trasmetta ogni quattro anni al Parlamento europeo e al Consiglio una relazione di valutazione e sul riesame del Regolamento.[5] Questo primo atto legislativo europeo in tema di IA sicuramente pone le basi per un utilizzo cauto e non rischioso dei nuovi sistemi di Intelligenza Artificiale, ma non è ancora possibile affermare che rappresenti lo strumento più efficace per garantire un utilizzo sicuro delle nuove tecnologie.

8.3. Le modalità di addestramento degli algoritmi intelligenti e i loro risultati

L'analisi dei rischi e benefici dell'implementazione di sistemi di Intelligenza Artificiale non può prescindere dallo studio dei meccanismi che si pongono alla loro base. A fornire una spiegazione di questi ha contribuito Franco Cutugno[6], nella sua partecipazione al terzo Dialogo.

Egli ha evidenziato che alla base del funzionamento dell'IA si pongono in realtà principi che appartengono al campo della semantica e della linguistica, da ricondurre agli studi di Wittgenstein e Harris sulla semantica distribuzionale. La tesi da loro sostenuta consiste nel ritenere che due parole sono semanticamente simili, indipendentemente dal loro significato, se occorrono più o meno alla distanza di *n* parole fra di loro. Nella frase "La guerra dell'Iran è collegata soprattutto all'esigenza di petrolio dell'Occidente", le parole Iran e petrolio hanno significati diversi, ma sono semanticamente collegate. La lingua risulta quindi essere composta da una serie di informazioni sintagmatiche tra loro collegate e collegabili. Questa proprietà ha permesso di sviluppare criteri matematici che non si concentrano sul significato delle parole, bensì sulla loro disposizione e distribuzione nello spazio. ChatGPT è il prodotto più noto di tale elaborazione.

Le tecnologie moderne sono costruite su di un'architettura *encoder-decoder,* la quale ha visto la sua prima apparizione nel mondo della traduzione automatica. Questi processi sono attivi già da qualche anno, ma non sono i soli alla base dell'ideazione dei GPT, seppur utili al loro funzionamento. Infatti, la loro struttura ha introdotto il *bidirectional transformer encoder*, modello che ha garantito un algoritmo performante, una volta addestrato. L'addestramento dell'algoritmo consiste, nel caso di ChatGPT, nel proporre esercizi quali la ricostruzione di testi con parole sbagliate o mancanti, in modo da ottenere stime sempre più accurate e migliorare l'abilità di generare testi di senso compiuto. Una volta che il modello *encoder-decoder* ha istruito la macchina, è possibile applicare un modello *encoder-only* o *decoder-only*. Quest'ultimo modello è quello oggi più utilizzato, in quanto determina la produzione dell'output da parte del sistema di IA.

L'addestramento degli algoritmi si è poi diversificato in seguito alla diffusione, da parte di Meta, di LLMs (*Large Language Models Meta AI*), un innovativo modello linguistico di Intelligenza Artificiale con libera licenza d'uso, che ha permesso a università e altri enti di poter lavorare su un modello di base accessibile a tutti. Esempio di tale diversificazione è rappresentato dal metodo

[5] Articolo 112 del Regolamento (UE) 2024/1689.

[6] Professore di Glottologia e Linguistica presso il Dipartimento di Ingegneria elettrica e tecnologie dell'informazione dell'Università degli Studi di Napoli Federico II.

utilizzato da Google, che inizialmente prevedeva di addestrare gli algoritmi tramite operazioni di *surfing* tra tutte le pagine web del mondo. Tuttavia, non essendo gli algoritmi dotati di un proprio pensiero, ma semplici emulatori della conoscenza umana e dunque anche dei suoi errori, in questo modo non si poteva evitare che questi riproducessero anche *bias* tipicamente umani. Di conseguenza il sistema di addestramento è stato poi affinato modificando la raccolta di dati e materiali da sottoporre agli algoritmi, questa volta tratti da libri e pubblicazioni scientifiche, più adatti anche a finalità di *coding.*

A influenzare le tecniche di addestramento hanno poi contribuito i principi, posti anche dall'Unione Europea, che prevedono che l'IA sia sostenibile, onesta e non cagioni danno alcuno. In virtù di questi imperativi, oggi un modello, prima di essere immesso sul mercato, deve essere testato attraverso una tecnica informatica denominata *Reinforcement Learning from Human Factor* (RLHF). Quest'ultima prevede che un operatore fornisca dei comandi (input) all'algoritmo, il quale produce degli output a cui l'operatore assegna un punteggio. In questo modo l'algoritmo memorizza solo gli output con punteggi più alti, così da rendere le sue stime sempre più precise. Attualmente a svolgere queste complesse operazioni è chiamata la figura del *prompting engineer,* ma è possibile che presto, grazie alla veloce evoluzione dei sistemi IA, questo diventi superfluo.

Ormai la realizzazione dei comandi da fornire agli algoritmi è disciplinata da una serie di regole, quali la descrizione dettagliata degli elementi base del task, il rispetto di alcuni principi di design, l'implementazione di informazioni di contesto. Si prevede inoltre che il completamento del task sia svolto in fasi successive, per cui il primo comando risulta più generico, mentre quelli successivi divengono progressivamente più specifici.

Nonostante la presenza di queste regole, non si può ancora affermare che gli algoritmi siano privi di criticità e problematiche, come ben evidenziato da Cutugno, anche se queste non vengono sempre riportate dai fornitori di questi sistemi. Oltre agli aspetti legati ai *bias* e all'etica, già accennati sopra, ricorrono innanzitutto le influenze dettate dal marketing: è probabile che in risposta a un input di ricerca fornito da un utente il primo output risultante non contenga la risposta qualitativamente ottimale, ma quella che mette in evidenza il fornitore meglio pubblicizzato. Ci sono poi difficoltà nel rispondere alle necessità di aggiornamento dei sistemi, che per rimanere attuali dovrebbero ricevere aggiornamenti periodici ogni mese circa. Infine, la problematica che solleva maggiore preoccupazione e attenzione è quella relativa alla privacy: bisogna infatti garantire tutele per l'utilizzo dei dati da parte dell'Intelligenza Artificiale. In merito a ciò, il Garante della Privacy nel 2023 ha provvisoriamente limitato l'utilizzo di ChatGPT, rilevando l'utilizzo ingiustificato da parte della piattaforma di dati personali, per poi somministrarle una sanzione di quindici milioni di euro.[7]

[7] Il 30 marzo 2023, il Garante per la protezione dei dati personali ha limitato provvisoriamente l'utilizzo del servizio di ChatGPT, contestando allo stesso la mancanza di una idonea base giuridica per la raccolta dei dati personali ed il loro conseguente utilizzo per l'addestramento del sistema. Come risulta dal comunicato del 20 dicembre 2024 il Garante ha richiesto una sanzione di quindici milioni di euro per la società e l'obbligo di realizzare una campagna pubblicitaria di sei mesi che possa informare gli utenti sulle modalità di raccolta dei dati. Tale sanzione è stata successivamente sospesa dal Tribunale di Roma con ordinanza del 20 marzo 2025, considerata l'intenzione della società di impugnare la sanzione, ritenuta sproporzionata. Per approfondimenti cfr. https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832

L'analisi del rapporto tra IA e dati, oltre a quanto riportato da Cutugno nel suo intervento, è uno degli aspetti fondamentali anche dell'implementazione di sistemi di Intelligenza Artificiale nella pubblica amministrazione, la quale, a mio avviso, presenta ulteriori esigenze di tutela per i cittadini e di rispetto dei principi dell'azione amministrativa. Prendendo come punto di riferimento l'Amministrazione Finanziaria, l'IA è stata introdotta con l'algoritmo Ve.Ra (Verifica dei Rapporti finanziari) dell'Agenzia delle Entrate per le verifiche fiscali e l'analisi del rischio, che risponde all'esigenza di incrementare la *tax compliance,* privilegiando la prevenzione *ex ante* e circoscrivendo i controlli ai soggetti con più alto rischio fiscale.[8] Lo svolgimento di tali attività, comportando l'accesso dell'algoritmo ai dati fiscali dei contribuenti, tutelati come dati personali, presenta una rigida regolamentazione, che prevede l'esclusione dei dati sensibili e giudiziari dalle elaborazioni, la pseudonimizzazione dei dati dei contribuenti e il principio di responsabilizzazione, il quale ha comportato l'obbligo per l'Agenzia delle Entrate di redigere il *Data Protection Impact Assessment* (DPIA), una valutazione unitaria di impatto sulla protezione dei dati, sottoposta all'approvazione del Garante della Privacy.[9] Grazie a queste disposizioni si raggiunge un bilanciamento tra l'esigenza di fornire dati di addestramento agli algoritmi e la tutela dei dati dei cittadini.

Per poter facilitare la comprensione dell'importanza della scelta dei dati per l'addestramento di IA, si riporta dagli interventi del terzo Dialogo una similitudine molto efficace: l'algoritmo si nutre di dati e informazioni così come fa un bambino alla scoperta del mondo, per cui risulta fondamentale garantirgli una nutrizione adeguata: la macchina conosce solo ciò di cui è stata alimentata.[10]

Le numerose possibilità di utilizzo dei sistemi di Intelligenza Artificiale hanno permesso di poter creare una collaborazione tra uomo e algoritmo, la quale ha aperto le porte al tema della c.d. cogenerazione, che ha sollevato una serie di questioni: come deve essere considerato un prodotto frutto della collaborazione tra uomo e macchina? L'IA può essere considerata giuridicamente autore? Una prima risposta a queste domande è stata fornita dal tribunale di Pechino, con una sentenza del 2023[11] su una controversia riguardante un ritratto prodotto dall'IA tramite i comandi di un utente e poi ripreso da un secondo soggetto, che negava che il ritratto fosse tutelato dal diritto d'autore, in quanto generato dall'Intelligenza Artificiale.

Il Tribunale ha così stabilito alcune condizioni in presenza delle quali si può applicare il diritto d'autore in merito a opere prodotte dall'Intelligenza Artificiale: a) l'opera rientra nei campi della letteratura, arte o scienza, b) deve essere originale, c) rappresenta una specifica forma di espressione, d) può essere considerata prodotto dell'intelletto umano.

---

[8] Agenzia delle Entrate, *Documento illustrativo della logica degli algoritmi*, 19 maggio 2023, p. 4-5.

[9] *Ibidem*.

[10] Così, a chiosa dell'intervento riferito, Natale Rampazzo.

[11] Sentenza (2023) Jing 0491 Min Chu n. 11279. Cfr: https://english.bjinternetcourt.gov.cn/pdf/BeijingInternetCourtCivilJudgment112792023.pdf; D'Antonio 2024.

Nel caso di specie, l'utente aveva inserito circa 50 prompt prima di arrivare al risultato finale, per cui l'opera è stata considerata prodotto dell'intelletto umano.[12] In Italia, in merito a tale fenomeno, il Governo ha predisposto un disegno di legge[13] volto a regolamentare nell'immediato proprio alcuni aspetti dell'IA connessi al copyright e in cui viene presentata come soluzione lo strumento della filigranatura, la quale marca il contenuto generato indicandone appunto l'origine.[14]

È inevitabile, dopo aver compreso quali siano le principali questioni sollevate dagli interventi, che, da spettatrice, mi sorgano dei dubbi in merito alla possibilità che la legge, soprattutto quella nazionale, possa limitare in maniera efficace l'utilizzo dell'IA per tutelare i diritti e i dati degli utenti. In merito a ciò, parte della dottrina ritiene che basti affidarsi alla *auto-regulation* dei sistemi stessi, altri rivendicano l'importanza di strumenti di *hard law* (tra cui si colloca anche l'*AI Act*), altri ancora ritengono che per avere regole efficaci a livello globale sia necessario fare ricorso alla *para-regulation*.[15] Tale termine si riferisce a una regolazione tramite *standard, rankings, indicators* globali che si pongano accanto ad altre forme di regolamentazione e che ne costituiscano la cornice di riferimento. Tra questi, citiamo a titolo esemplificativo il *Government AI Readiness Index* (AIRI), elaborato da Oxford Insights, il quale classifica 188 nazioni assegnando a ciascuna un punteggio che indica la loro prontezza ad accogliere sistemi di Intelligenza Artificiale. Tale punteggio è ottenuto grazie alla comparazione di diversi fattori, tra cui accessibilità, qualità e, dal 2020, "Uso Responsabile", voce che esprime il lavoro svolto dai governi e dalle organizzazioni internazionali per creare principi "per evitare che l'IA possa causare danni".[16] Si ritiene che attraverso strumenti come questo possano essere tracciati delle regole e degli obiettivi che siano validi a livello globale e che possano fronteggiare alla pari un fenomeno, quello dell'IA, anch'esso globale. Infatti, la diffusione dei sistemi di Intelligenza Artificiale e dei loro risultati si spinge

[12] Anche una sentenza della Corte Suprema del Regno Unito ha fornito una risposta sul tema, ma in contraddizione con quanto affermato dal Tribunale di Pechino. La vicenda sottoposta alla Corte ha riguardato la presentazione, da parte di un imprenditore americano, di domande di brevetto il cui autore era DABUS (Device for Autonomous Bootstrapping of Uniifed Sentience), un sistema di IA. L'esaminatore della domanda aveva contestato l'assenza di una persona fisica da indicare nella richiesta e, sollecitato l'imprenditore del caso, Stephen Thaler, considerò la domanda ritirata non essendo pervenuta alcuna risposta entro il termine indicato. La vicenda, dunque, è stata portata in giudizio sino a giungere alla Corte Suprema, la quale all'unanimità ha respinto la richiesta del Sig. Thaler affermando la necessità a che l'inventore indicato nella domanda di brevetto sia una persona fisica.

[13] D.d.l. n. 1146, Disposizioni e delega al Governo in materia di intelligenza artificiale, 20 Maggio 2024. Il testo è stato approvato in Senato il 20 marzo 2025 e trasmesso alla Camera, che lo ha assegnato alla nona e decima commissione parlamentare.

[14] A titolo esemplificativo, articolo 23, comma 1, lett. b del d.d.l.: b) dopo l'articolo 40, è inserito il seguente: «Art. 40-bis. – (Contenuti testuali, fotografici, audiovisivi e radiofonici che utilizzano sistemi di intelligenza artificiale) – 1. Qualunque contenuto informativo diffuso da fornitori di servizi audiovisivi e radiofonici tramite qualsiasi piattaforma in qualsiasi modalità, incluso il video on demand e lo streaming, che, previa acquisizione del consenso dei titolari dei diritti, sia stato, attraverso l'utilizzo di sistemi di intelligenza artificiale, completamente generato ovvero, anche parzialmente, modificato o alterato in modo tale da presentare come reali dati, fatti e informazioni che non lo sono, deve essere reso, a cura dell'autore o del titolare dei diritti di sfruttamento economico, se diverso dall'autore, chiaramente visibile e riconoscibile da parte degli utenti mediante inserimento di un elemento o segno identificativo, anche in filigrana o marcatura incorporata purché chiaramente visibile e riconoscibile, con l'acronimo "IA" ovvero, nel caso di contenuti audio, attraverso annunci audio ovvero con tecnologie adatte a consentire il riconoscimento. Tale identificazione deve essere presente sia all'inizio della trasmissione e all'inizio del contenuto, sia alla fine della trasmissione e alla fine del contenuto, nonché ad ogni ripresa del programma a seguito di interruzione pubblicitaria. L'inserimento del segno identificativo è escluso quando il contenuto fa parte di un'opera o di un programma manifestamente creativo, satirico, artistico o fittizio, fatte salve le tutele per i diritti e le libertà dei terzi. Fermo restando quanto previsto dall'articolo 41, per le finalità di cui al presente articolo nonché all'articolo 42, commi 1, lettera c-bis), e 7, lettera c-bis), l'Autorità promuove forme di co-regolamentazione e di autoregolamentazione tramite codici di condotta sia con i fornitori di servizi di media audiovisivi e radiofonici sia con i fornitori di piattaforme per la condivisione di video».

[15] Bello y Villarino 2023.

[16] Erkiilä 2023, p. 828. Per approfondire cfr: https://oxfordinsights.com/ai-readiness/ai-readiness-index/

ben oltre i confini territoriali dei singoli Stati, e ciò vale sia per i risultati positivi, sia per quelli negativi. È necessario allora contrastare la diffusione dei risultati negativi e delle pratiche che mettono a rischio la tutela dei diritti con misure adeguate, che possano diffondersi allo stesso modo delle nuove tecnologie.

8.4. I rischi di un utilizzo incontrollato dei sistemi di IA: *bias* e stereotipi

Numerosi possono essere i risultati negativi dell'IA, capaci anche di minare la tutela della privacy e dei diritti umani, come già introdotto sopra. Nel quarto Dialogo Natale Rampazzo, con il suo intervento *Galateo digitale: artifici e raggiri dell'IA*, ha ben analizzato questo tipo di criticità, proponendo per le stesse diversi tipi di tutela.

Il primo tra i fenomeni da lui citati è il *deep fake*, che consiste in applicazioni dell'IA che creano immagini e realtà fasulle che possono trarre in inganno le persone a cui sono sottoposte, determinando *bias* cognitivi. Altro fenomeno ancora è quello della *black-box*: con questo termine si indica la impossibilità di ripercorrere e decifrare il processo utilizzato dall'algoritmo di *machine learning* nell'auto-elaborazione di dati che determinano l'output prodotto, per cui spesso si parla di IA opaca. Per contrastare tale opacità si fa riferimento, anche negli atti dell'Unione Europea, alla *Explainable Artificial Intelligence,* un'IA che sia trasparente e che, se utilizzata dalle pubbliche amministrazioni, consenta di riportare una spiegazione delle decisioni adottate nei provvedimenti, evitando conseguenti violazioni dell'obbligo di motivazione.[17]

Numerosi sono i casi di IA opaca riportati dal dott. Rampazzo, di cui qui se ne citano solo alcuni: il primo riguarda il *Public Safety Assessment* (PSA), un sistema varato negli Stati Uniti e attualmente utilizzato da 28 giurisdizioni, che assiste i giudici nello stabilire i termini della durata della custodia cautelare o dell'ammontare della cauzione. All'inizio si accertò che tale sistema produceva output discriminatori nei confronti delle minoranze, per cui oggi sono stati resi pubblici i criteri su cui l'algoritmo si basa per arginare l'opacità originariamente rilevata.[18] Esempio a noi più vicino è rappresentato invece da una sentenza del Tribunale di Bologna che si pronuncia sull'algoritmo alla base del funzionamento di *Deliveroo*. Si è dimostrato infatti che questo assegnava ai rider che si assentavano dal turno di lavoro senza comunicazione preventiva un punteggio negativo su un ranking reputazionale, a prescindere dalla motivazione alla base dell'assenza. Ciò creava dunque una discriminazione per chi si assentava per motivi legittimi, trattato allo stesso modo di chi si assentava per motivi futili.[19]

Come si evince dai casi sopracitati, il rischio di favorire la proliferazione di discriminazioni dirette e indirette tramite sistemi che si avvalgono di *machine learning*, sia per la qualità dei dati forniti sia per il meccanismo stesso degli algoritmi, è alto e ancora non completamente arginato. Anche l'intervento di Antonio Pescapè[20] si è concentrato su questi temi, approfondendo in particolare gli stereotipi di genere. Egli ha prima di tutto premesso che per poter parlare in modo appropriato

[17] V. Kaminski 2021.

[18] Cfr: https://advancingpretrial.org/psa/factors/; Brittain 2021.

[19] Tribunale di Bologna, ordinanza 31/12/2020. Cfr: https://www.algoritmolegal.com/wp-content/uploads/2021/01/Sentencia-Bologna-Italia-Deliveroo-dic-2020-Original-italiano.pdf; Perulli 2021.

[20] Professore di ingegneria elettrica dell'Università Federico II di Napoli e direttore del Centro Ermes, il Centro Europeo di Ricerca sui Media per la Società dell'Informazione.

di rischi e discriminazioni perpetrati dall'IA, bisogna comprendere cosa questa indichi. Il termine Intelligenza Artificiale infatti non ha ancora una definizione precisa, poiché con essa si intendono una serie di sistemi automatici che sono in grado di svolgere compiti per i quali sarebbe necessaria l'intelligenza umana.

La paternità di questa materia è riconosciuta ad Alan Turing, il quale, nel tentativo di decifrare Enigma, il celeberrimo sistema di comunicazioni tedesco della seconda guerra mondiale, contribuì alla creazione del primo computer. Ad egli è attribuito anche l'*Imitation Game,* sistema con il quale si tenta di comprendere se una macchina possa pensare come un essere umano: per assemblarlo sono necessari un interrogante e due rispondenti, di cui un uomo ed una macchina. L'interrogante pone una domanda e, senza sapere chi abbia di fronte, cerca di capire se a rispondere sia l'uomo o la macchina. Qualora egli non riesca a distinguere fra la macchina e l'essere umano allora il gioco imitativo può dirsi compiuto con successo, poiché la situazione di dubbio circa la natura del soggetto rispondente già depone in favore dell'emulazione umana.

Il relatore ha poi sottolineato che oggi quando si parla di IA si parla principalmente di *machine learning*, strumenti statistici addestrati su dati, potendo difficilmente parlare di altro. Sebbene inizialmente questi algoritmi non funzionassero in modo soddisfacente, oggi hanno un funzionamento ottimale per diverse ragioni, tra cui si annoverano la vastissima disponibilità di dati attinenti ad ogni ambito della vita e il miglioramento delle capacità di calcolo.

Dunque, lo stesso algoritmo che anni addietro non aveva prestazioni soddisfacenti riesce oggi a realizzare buone performance.

Il motivo per cui l'Intelligenza Artificiale è al centro del dibattito scientifico può essere ricollegato al pensiero schopenhaueriano per cui le innovazioni sono viste prima come ridicole, poi pericolose, e infine utili (si pensi alle reazioni della società all'introduzione dei primi cellulari). Secondo Pescapè, oggi siamo nella fase della pericolosità.

Ormai si riconosce che le macchine e gli algoritmi possono svolgere attività umane con alti livelli di performance, non solo per quanto riguarda le operazioni di calcolo, ma anche per creare canzoni e opere d'arte, per svolgere prestazioni di lavoro. Nella società civile infatti, uno dei timori riguardante l'IA concerne la possibilità che questa sostituisca l'uomo in molte mansioni, determinando la scomparsa di molti lavori. Alcuni ritengono che non si assisterà solo alla scomparsa dei lavori di oggi, ma anche alla creazione di nuovi tipi di lavoro; ma questa non è la tesi sostenuta dal relatore, che invece considera l'IA capace di sostituirsi anche ai lavori più tradizionalmente umani, come il cantante o l'attore.

Tuttavia, a mio avviso si potrebbe osservare che, essendo l'algoritmo programmato per rielaborare dati del passato, questo non possa effettivamente svolgere attività creative senza la collaborazione dell'uomo, per cui l'IA può produrre una canzone, ma si dubita che questa possa evolversi fino a creare un nuovo genere musicale, ad esempio.

Altri aspetti a cui dedicare attenzione sono rappresentati dai *recommendation systems*, algoritmi utilizzati nel digital marketing per individuare quali contenuti sottoporre ai consumatori in base alle loro preferenze e idee politiche, individuate tramite l'analisi di dati espliciti e impliciti, e dall'utilizzo degli algoritmi a fini vendicativi, da cui derivano fenomeni quali il *revenge porn* o l'*undressing*.

Anche Pescapè condivide l'idea che l'Intelligenza Artificiale debba essere trasparente, affidabile, antropocentrica, inclusiva, responsabile, neutrale. Per raggiungere tale obiettivo egli ritiene sia necessario risolvere in primis il problema della responsabilità, e cioè stabilire chi sia chiamato a rispondere degli errori della macchina e risarcire i danni da essa causati.

Come accennato sopra, è possibile infatti che la macchina sbagli, che perpetui e sostenga stereotipi discriminatori. Secondo la definizione che ne dà la psicologia sociale, lo stereotipo consiste nel considerare simili per alcune caratteristiche tutti i componenti di un gruppo o categoria sociale. Queste considerazioni possono facilmente essere trasmesse all'IA che, dietro il velo dell'oggettività dell'algoritmo, compie scelte che si rivelano discriminatorie: il relatore a dimostrazione di ciò ricorda che i sistemi di riconoscimento facciale introdotti in alcuni aeroporti sottoponevano statisticamente a maggiori controlli donne non caucasiche. Questo potrebbe accadere anche con sistemi di selezione del personale, soprattutto se i codici continuano a essere scritti da maschi bianchi, senza alcun tipo di supervisione.

Queste considerazioni mi permettono di richiamare il caso Amazon, inerente a discriminazioni nel mondo del lavoro: l'azienda dal 2014 al 2017 si è affidata a un sistema per la selezione del personale che doveva rappresentare la soluzione per velocizzare l'esame dei CV dei candidati. Tuttavia tale meccanismo, essendo stato addestrato sui dati del "candidato perfetto" del passato, prendeva in considerazione, soprattutto in ambito tecnologico, maggiormente gli uomini, con una discriminazione ingiustificata nei confronti delle candidate donne. Proprio per questo motivo il sistema è stato abbandonato nel 2017. Oggi in Europa, grazie all'*AI Act* questo algoritmo non potrebbe essere utilizzato in quanto classificato come *high risk*, perché produce discriminazioni ingiustificate e *bias.*[21]

Tentando di rispondere alla domanda concernente i responsabili degli errori dell'IA, dal mio punto di vista si può immaginare una forma di responsabilità indiretta nei *global indicators, standard, rankings,* che costituiscono la *para-regulation* di cui sopra. Questi infatti, si applicano sia a imprese private sia a enti pubblici e, nell'assegnare punteggi, valutano anche qualità, trasparenza e tutela dei principi dei meccanismi elaborati, incentivando dunque fornitori e utenti di IA a scegliere i sistemi più rispettosi.[22] Di conseguenza, una forma di responsabilità e sanzione sarebbe costituita dall'assegnazione di un punteggio basso sia per gli algoritmi degli Stati sia per quelli delle imprese private. Tuttavia, non si può credere nell'illusione che questi strumenti siano oggettivi e imparziali. Anche i global standard sull'IA sono infatti irrimediabilmente influenzati da strategie e obiettivi politici e, anzi, sono espressione di questi.[23] La regolazione dell'IA viene infatti affrontata, oltre che come un problema giuridico, anche come problema etico e politico: gli standard promuovono certe politiche nazionali attraverso analisi di dati e conversione di fatti in numeri. Questi poi riscuotono successo a causa dell'indeterminatezza delle attività legislative nazionali e sovranazionali, imponendosi di fatto come soluzione globale per fornire una guida e un limite ai sistemi di Intelligenza Artificiale.

---

[21] Cfr: https://www.ansa.it/sito/notizie/tecnologia/tlc/2018/10/10/amazon-stampa-intelligenza-artificiale-discriminava-donne_1b1ebaca-f000-48a6-89b4-ff9854ef75e7.html, Adams-Prassl 2023.

[22] Bello y Villarino 2023. Si guardi, a titolo esemplificativo, il *Global AI Index*.

[23] Erkiilä 2023, p. 816.

Inoltre, non va escluso da questa analisi il fatto che gli stessi standard possono anche essere espressione di una particolare narrazione del progresso e del futuro: l'Intelligenza Artificiale viene presentata nella maggior parte di questi modelli come una rivoluzione industriale, simbolo del progresso e fattore fondamentale per la competizione di mercato. Al contrario, altre narrazioni che non sposano l'idea della Storia come un continuo progresso, evidenziano i rischi dell'IA per i sistemi democratici e per questioni etiche.[24]

Dunque, è innegabile che anche gli standard, pur esprimendosi attraverso dati e numeri, possano essere veicoli con cui diffondere obiettivi e idee politiche ben precise.

8.5. Strumenti di *soft law* per tutelare i diritti umani

Dopo aver trattato i rischi in generale derivanti dall'applicazione di sistemi di IA, nel sesto Dialogo si è dato spazio anche a prospettive pratiche, incentrandosi sulle modalità con cui le imprese possono gestire gli aspetti negativi derivanti dall'utilizzo dell'IA sul rispetto dei diritti umani, argomento di cui ha trattato Marco Fasciglione.[25]

Nell'analizzare le nuove sfide della società contemporanea, bisogna tenere conto anche degli attori protagonisti della stessa, i quali possono fornire nuove soluzioni a nuovi problemi. Per quanto riguarda l'IA, attori protagonisti sono le imprese, oltre agli Stati e le organizzazioni internazionali, le quali usufruiscono dei nuovi sistemi e ne facilitano la diffusione grazie a una delle loro caratteristiche, che le accomuna anche alle nuove tecnologie: la transnazionalità.

Tenuto conto di ciò, le organizzazioni internazionali hanno realizzato che incidendo sui comportamenti delle imprese si rende possibile una regolazione delle nuove tecnologie che non si limita ai confini di uno Stato. Già l'ONU nel 2011 aveva emanato dei principi guida da adottare nei processi organizzativi per garantire che l'attività d'impresa sia svolta in modo tale da rispettare i diritti umani[26].

Sono questi atti di *soft law*, quindi non strettamente vincolanti, ma che nondimeno influiscono sul comportamento delle imprese e, di conseguenza, in modo indiretto, contribuiscono a incentivare un uso di IA rispettoso dei diritti umani.

Fasciglione ha poi enucleato le tre principali nuove misure previste da queste linee guida e da altri atti di *soft law:* si tratta di *policy commitment, due diligence* sui diritti umani e meccanismi di rimedio.

Riguardo al *policy commitment*, la maggior parte delle aziende del settore tecnologico ha già adottato dei provvedimenti sui diritti umani, tramite *statement* aziendali che dichiarano che l'impresa rispetta gli standard aziendali in materia di diritti umani. Alcuni di questi rimandano espressamente ai principi ONU prima menzionati, mentre altri, come Meta, menzionano anche il complesso di strumenti internazionali di cui ci si può avvalere a garanzia dei principali diritti e libertà garantiti all'uomo, anche qualora ciò risulti in contrasto con la legislazione nazionale a cui l'impresa stessa è sottoposta.

---

[24] Ivi, p. 815.

[25] Primo Ricercatore presso il CNR-IRISS.

[26] Nazioni Unite, *Guiding Principles on Business and Human Rights* – HBR, 2011. La guida si sofferma su tre aspetti: a) la responsabilità degli Stati in ordine al rispetto e alla protezione dei diritti umani e delle libertà fondamentali; b) il ruolo delle imprese commerciali nel rispetto e promozione dei diritti umani; c) la necessità a che tali diritti ed obblighi siano abbinati a rimedi efficaci in caso di violazione.

La *due diligence* sui diritti umani, invece, è costituita da un sistema di valutazione del rischio che l'articolo 17 dei principi guida ONU struttura in 4 fasi, con l'obiettivo di identificare gli impatti negativi sui diritti umani che possono derivare dalle attività economiche dell'impresa. La disposizione richiede di individuare misure adeguate per affrontare eventuali impatti negativi, tracciando le risposte che palesano l'efficacia o meno delle misure, e infine impone di comunicare all'esterno le misure accolte e l'esito riscontrato. Si tratta di un processo di *risk assessment* che corrisponde alla tradizionale diligenza aziendale, ma che comprende comunque un elemento di novità: i rischi rilevati ricadono sugli utenti e non sull'azienda.

La *ratio* alla base della *due diligence* elaborata, e dunque la valutazione del rischio, è stata recepita anche dall'*AI Act*[27], che infatti prevede un approccio *risk-based*, il quale permette di moderare l'intensità ed il contenuto delle disposizioni ai rischi intrinseci ai diversi sistemi di IA.[28] Si può parlare più in generale di un approccio *human rights by design*, il quale riprende quell'idea, propria del settore della privacy, per cui le scelte di policy debbano preoccuparsi di prevenire gli impatti negativi su tali categorie di diritti.

I meccanismi di rimedio, infine, prescrivono che, nel caso in cui siano lesi i diritti umani, le imprese debbano prevedere delle apposite misure rimediali, in cui consiste lo sbocco naturale della *due diligence*.

I principi guida ONU classificano questi meccanismi di rimedio come meccanismi di reclamo di tipo operativo, una garanzia ulteriore rispetto ai tradizionali strumenti di tutela giurisdizionale. Un esempio di ciò è rappresentato dall'Oversight board di Meta, istituito nel 2018, che persegue l'obiettivo di tutela degli utenti delle piattaforme media. In particolar modo questo agisce tramite la rimozione dei contenuti considerati non conformi alle policy delle piattaforme, consentendo agli utenti di presentare eventuali reclami. Tale sistema è stato d'ispirazione anche per il successivo *UN Special Rapporteur* sulla libertà di espressione (2021), il quale richiede alle imprese che operano nel settore dei social media di istituire delle procedure interne per gestire la *content moderation* e di esplorare la realizzazione di "*external oversight mechanisms such as social media councils*".

Fasciglione ha fornito ulteriori dettagli sulla composizione e il funzionamento del Board: questo presenta un proprio statuto e una carta costituzionale che richiamano i diritti fondamentali e opera come un vero e proprio organismo giudiziario. Le fattispecie ad esso sottoposte vengono infatti analizzate valutando la compatibilità delle azioni con le responsabilità stabilite nei documenti statutari. Il procedimento risulta quindi simile a quelli di monitoraggio del sistema internazionale.

Le fasi di monitoraggio possono essere così schematizzate: identificazione del diritto violato (normalmente si tratta della libertà d'espressione); legittimità dell'interferenza sul diritto operata dalla piattaforma; valutazione della legittimità e della necessità dello scopo perseguito; verifica della proporzionalità dell'interferenza prodotta. In aggiunta a ciò, il Board fa spesso riferimento alle decisioni precedenti avutesi in casi analoghi ad esso sottoposti, ma anche alle

[27] Regolamento (UE) 2024/1689.

[28] Nel dettaglio, l'*AI Act* individua quattro categorie di rischio: 1. sistemi di IA a rischio inaccettabile, i quali generano rischi tanto elevati da giustificarne il divieto; 2. sistemi di IA ad alto rischio, il cui accesso al mercato europeo è subordinato alla rispondenza a severe misure di controllo; 3. sistemi di IA a rischio limitato, sottoposti al rispetto di requisiti informativi e di trasparenza; 4. sistemi di IA a rischio minimo, sottoposti al rispetto di codici di condotta volontari.

sentenze delle Corti internazionali, nonché alle raccomandazioni e decisioni degli istituti a tutela dei diritti umani esistenti nel SI.

Le decisioni del Board si pongono a volte in contrasto con gli orientamenti nazionali. È possibile richiamare a dimostrazione di ciò l'*Öcalan's isolation case* del luglio 2021. In tale fattispecie, un post di Instagram riguardante una protesta contro le condizioni elettive in Turchia, non conformi ai diritti umani internazionalmente sanciti, era stato rimosso. L'Oversight Board è poi intervenuto con una decisione autonoma per ribaltare tale scelta e ripristinare il contenuto affermando, peraltro, la propria preoccupazione circa la cancellazione di tali post da parte degli Stati, i quali minano in tal modo la libertà di espressione dei cittadini. Da tale episodio è emerso un fenomeno nuovo: un organo interno a un'impresa ha contrastato, ponendosi in posizione paritaria, l'organo giudiziario di uno Stato.

Caso più recente, verificatosi nel 2023, riguarda una foto modificata con IA dell'ex presidente USA Biden e di sua nipote. Il Board ha confermato la scelta di non rimuovere il contenuto, seppur *fake,* poiché nelle policy di Meta la misura della rimozione di contenuti IA è prevista solo per i contenuti vocali. È emersa quindi una lacuna nelle condizioni d'uso di Meta, a cui il Board ha invitato a porre rimedio.

Tutto ciò ben sintetizza quanto ormai, nella regolamentazione dell'IA, sia prioritario non trascurare il ruolo degli attori economici coinvolti, avendo questi un potere tutt'altro che marginale e non necessariamente secondario al potere statale. Si parla in effetti di una sorta di effetto sostitutivo della giustizia del 'metaverso' a quella pubblicistica, basata su principi talvolta molto differenti rispetto a quelli interni ad una determinata cultura anche se manchevoli di talune garanzie, come ad esempio quella all'equo processo.

Il Dialogo è proseguito con l'intervento di Giacomo Maria Cremonesi[29], che, pur rimanendo nel campo del rapporto tra IA e imprese, ha precisato quali siano i prodotti attualmente commercializzati da NEC e ha fornito un caso concreto nell'attività svolta da un'impresa per garantire che l'IA sia trasparente, affidabile e rispettosa dei diritti umani. NEC ha infatti riconosciuto le criticità innescate dall'IA e, avvertendo l'esigenza di affrontarle con la dovuta diligenza, ha istituito un Comitato interno per l'Intelligenza Artificiale e i Diritti Umani. Tra i prodotti di IA commercializzati da NEC vengono citati: GraphAI, ExplainableAI, NLP, Safer LLMs e BioMedical AI.

GraphAI è una rete neurale che permette di stabilire delle relazioni tra i diversi nodi della rete (quando si parla di nodi si parla essenzialmente di utenti, ad esempio pazienti ospedalieri, clienti delle assicurazioni, etc.). L'ExplainableAI, a cui si è già fatto accenno, è uno studio che ha l'obiettivo ultimo di comprendere appieno i meccanismi di produzione dell'output da parte del sistema intelligente. Dunque, si basa sulla comprensione della cd. *black box*.

Per quanto concerne i *Large Language Models*, divenuti noti grazie a ChatGPT, anche NEC ne ha sviluppato un modello, diffuso in Giappone. In questo specifico ambito le ricerche si concentrano sulla correttezza delle risposte fornite dal LLM e hanno lo scopo di rimuovere le cd. allucinazioni (risposte scorrette ma presentate in modo così coerente da apparire veritiere).

[29] Giurista, lavora attualmente presso NEC Laboratories Europe (multinazionale giapponese leader nelle ICT), dove coordina il Comitato per l'Intelligenza Artificiale e i Diritti Umani. Il suo intervento è pubblicato, in lingua inglese, nel presente volume.

Biomedical AI è invece finalizzato alla produzione di vaccini personalizzati contro il cancro. La procedura seguita nei clinical trial è la seguente: il paziente viene selezionato; una volta rimosso il tumore viene fatta una biopsia così che l'IA possa rintracciare le mutazioni caratteristiche della tipologia di cancro studiata; dopodiché, in base alle mutazioni evidenziate, si crea un vaccino personalizzato (terapeutico e non preventivo) così che il paziente possa sviluppare degli anticorpi che contrastano tali mutazioni e ridurre la possibilità di una recidiva a distanza di 5-10 anni. Tra queste ricerche rientra anche quella concernente il microbioma: attraverso un processo diagnostico, l'IA è in grado di prevedere le malattie di cui il soggetto soffre o di cui potrebbe soffrire in futuro.

Si comprende dunque come l'IA possa avere le più diverse applicazioni, con conseguenti rischi, non immediatamente prevedibili. L'*AI Act,* in merito a ciò, introduce una disciplina che prevede la creazione di un *risk management system*, di una efficiente *data governance*, di una produzione di documentazione tecnica redatta in un modo specifico, di un'elevata trasparenza sui dati e le scelte tecniche adoperate e inoltre, di un costante *human oversight* per evitare che le macchine assumano decisioni eccessivamente autonome.

Tali novità rilevano per la *compliance* delle imprese internazionali, soprattutto per quelle che si affacciano sui mercati europei. Proprio per adeguarsi alle nuove sfide incombenti, il Comitato interno per l'Intelligenza Artificiale e i Diritti Umani propone di offrire una leadership e una consulenza aziendale per affrontare l'impatto sui diritti umani delle tecnologie di IA da questa sviluppate. Tra i compiti del Comitato si annoverano l'esame e la valutazione dei rischi potenziali sui diritti umani derivanti dall'IA e la creazione di politiche e procedure che siano conformi con il quadro giuridico internazionale e con quello europeo. Compito principale è però quello di formare i ricercatori sul tema della tutela dei diritti umani, poiché spesso non sono praticabili interventi correttivi successivi. È necessario garantire sin da subito un approccio *human rights by design*.

Oltre a NEC, sono molte le imprese che stanno sviluppando delle procedure per adeguarsi all'*AI Act,* anche prima della sua entrata in vigore. Oltre che alle multinazionali, lo sguardo va rivolto anche alle piccole e medie imprese, obbligate a doversi adattare alla nuova normativa, pur non disponendo delle risorse delle Big Tech.

Si presenta in effetti il rischio che tale adattamento si trasformi in un filtro di ingresso sul mercato europeo a solo vantaggio di grandi imprese, favorendo dunque la formazione di oligopoli, laddove sussistono supremazie (le *Big Tech*, detengono ingenti quantità di dati per meglio addestrare i propri sistemi di IA, i quali dunque divengono in grado di produrre output qualitativamente migliori e sbaragliare la concorrenza). Inoltre, vi è la possibilità che la legge europea sull'IA disincentivi la commercializzazione dei sistemi di IA sul territorio europeo poiché vessato da molteplici e stringenti oneri procedurali.

Resta dunque ancora poco chiaro se l'*AI Act* sarà in grado, o meno, di produrre il c.d. *Brussels effect* ed esportare unilateralmente, attraverso i meccanismi di mercato, la propria normativa.

Nonostante i dubbi derivanti dalla regolamentazione onerosa, rimane un dato di fatto: la forte attrattiva che indubbiamente il mercato europeo esercita nei confronti degli operatori economici stranieri.

8.6. Conclusioni

Per raggiungere una piena comprensione del fenomeno dell'Intelligenza Artificiale è necessario acquisire uno sguardo d'insieme che possa non solo evidenziare rischi e benefici, ma anche aprire prospettive su proposte di regolazione affinché l'IA possa continuare a svilupparsi non costituendo una minaccia per i diritti umani. L'approccio seguito dal ciclo di Dialoghi è stato proprio questo: permettere a diversi studiosi ed esperti di confrontarsi e soprattutto di fornire al pubblico il quadro della situazione attuale sulle nuove tecnologie.

Per quanto riguarda la regolamentazione del fenomeno, si sono analizzati sia atti di *hard law*, come l'*AI Act,* sia di *soft law* come gli *standard* globali e gli *statement* aziendali. Per ottenere una regolamentazione efficace, una soluzione adatta sembra essere la cooperazione a livello globale, non solo tra Stati, quantomeno per stabilire delle definizioni comuni su cui i fornitori e gli utenti possano fare affidamento. Un primo risultato in questo senso è rappresentato dai *foundational standards*, elaborati dall'ISO congiuntamente all'IEC ed esposti in un articolo del luglio 2022[30]: questi rappresentano un primo accordo sulle definizioni riguardanti l'Intelligenza Artificiale e aspetti ad essa correlati che possano diffondersi anche oltre i singoli confini nazionali e diventare globali. Le organizzazioni mondiali di standard, infatti, raggiungono obiettivi in modo molto più efficace rispetto ai più convenzionali strumenti delle relazioni internazionali: basti pensare che tra USA e UE non c'è ancora un accordo sulla definizione di *high-risk AI,* mentre l'ISO e le altre organizzazioni lo hanno elaborato in modo molto più veloce, ed è proprio per questo motivo che ciò che viene da loro stabilito viene utilizzato da politici, ricercatori, stake-holder in tutto il mondo.

Rappresentando un nuovo fenomeno, con diffusione e risvolti mai riscontrati prima, l'IA necessita di altrettante nuove soluzioni, e non è stato ancora individuato uno strumento univocamente efficace.

Per questo motivo, è necessario che la comunità scientifica, così come la società civile, siano costantemente aggiornate sulle evoluzioni non solo delle tecnologie, ma anche dei tentativi umani finalizzati a rendere l'Intelligenza Artificiale compatibile con l'uomo e con i suoi diritti. Con la diffusione dell'informazione si può sollevare il velo della pericolosità e iniziare a ritenere l'IA finalmente utile per l'uomo. Siamo infatti stati abituati fin dai tempi di Ermogeniano all'idea per cui *omne ius hominum causa constitutum sit* e quindi pretendiamo che le norme inserite nella nostra società mantengano l'uomo sempre al centro di tutto il sistema. Pretendiamo, allora, oggi, che anche l'IA venga ricondotta al centro, all'uomo, e sia compatibile con esso e con la sua società. Insomma, si cerca di rendere anche l'Intelligenza Artificiale *hominum causa constituta.*

Bibliografia

Adams-Prassl, J., Binns, R., Kelly-Lyth, A. (2023). Directly discriminatory algorithms. In *The Modern Law Review*, *86*(1), pp. 144-175.

Bello y Villarino, J. M. (2023). Global Standard-Setting for Artificial Intelligence: Para-regulating International Law for AI? In *The Australian Year Book of International Law* 41 pp. 157-181.

[30] Bello y Villarino 2023.

Brittain, B. J., Georges, L., Martin, J. (2021). Examining the predictive validity of the Public Safety Assessment. In *Criminal Justice and Behavior*, *48*(10), pp. 1431-1449.
D'Antonio, V., Ruocco, C. (2024). Tutela delle immagini generate dall'IA. Dalla Cina nuove sfide per il diritto d'autore, Beijing Internet Court; Civil Judgment; Jing 0491 Min Chu n. 11279, 27th November 2023. In *Diritto di Internet*, *6*(2), pp. 261-270.
Erkiilä, T. (2023). Global indicators and Ai policy: Metrics, policy scripts and narratives. In *Review of Policy Research*, *40*(5), pp. 811-839.
Kaminski, M. E., Urban, J. M. (2021). The right to contest AI. In *Columbia Law Review*, *121*(7), pp. 1957-2048.
Perulli, A. (2021). La discriminazione algoritmica: brevi note introduttive a margine dell'Ordinanza del Tribunale di Bologna. In *Lavoro Diritti Europa*, *5*(1), pp. 2-7.

Sitografia

Comunicati del Garante della Privacy sulle sanzioni a ChatGPT. Disponibili su: https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832
Sentenza del Tribunale sul caso Deliveroo. Disponibile su: https://www.algoritmolegal.com/wp-content/uploads/2021/01/Sentencia-Bologna-Italia-Deliveroo-dic-2020-Original-italiano.pdf
Report sull'AIRI 2024. Disponibile su: https://oxfordinsights.com/ai-readiness/ai-readiness-index/
Resoconto del caso Amazon. Disponibile su: https://www.ansa.it/sito/notizie/tecnologia/tlc/2018/10/10/amazon-stampa-intelligenza-artificiale-discriminava-donne_1b1ebaca-f000-48a6-89b4-ff9854ef75e7.html
Sentenza (2023) Jing 0491 Min Chu n. 11279. Disponibile su: https://english.bjinternetcourt.gov.cn/pdf/BeijingInternetCourtCivilJudgment112792023.pdf

Dialoghi sull'intelligenza artificiale.
Atti dei Seminari del CNR-IRISS (marzo-maggio 2024)

«Un alfabeto di silicio e domande. Voci che si intrecciano – scienza, legge, filosofia – per interrogare l'ombra digitale del nostro tempo. L'intelligenza artificiale non è solo codice, ma specchio: riflette paure, speranze, il confine labile tra uomo e macchina.
Questo volume è una mappa di dialoghi sospesi, tra algoritmi che imparano ed etiche che vacillano, tra leggi in ritardo e futuri già scritti.
Cosa perdiamo? Cosa guadagniamo? Dove finisce la nostra volontà e inizia quella delle macchine? Un coro polifonico per navigare l'enigma, senza certezze, orientati dal faro delle domande.
Perché comprendere l'IA non è decifrare un sistema, ma riconoscersi nel suo riflesso». [Deepseek]

Natale Rampazzo è Dirigente di Ricerca presso il CNR dal 2023. Dal 2014 afferisce all'IRISS, in cui è Membro del Consiglio d'Istituto dal 2020.
Laureato in Giurisprudenza presso l'Università di Napoli Federico II, ha frequentato un master in diritto commerciale presso la LUISS di Roma, ha ottenuto il titolo di Magister Legum (LL.M.) in diritto tedesco presso la Ludwig-Maximilians-Universität München e ha conseguito il dottorato in Storia delle strutture amministrative presso l'Università di Salerno. Ha trascorso diversi periodi di ricerca all'estero (Max-Planck-Institut for Innovation – München; dipartimenti giuridici delle Università di Beograd, Freiburg, Heidelberg, Islas Baleares, Krakow, München). I suoi interessi si concentrano sull'analisi giuridica ed economica della tutela e valorizzazione della proprietà intellettuale, nonché della regolamentazione delle innovazioni tecniche e tecnologiche e del loro impatto sulla società e sulla cultura. È autore di numerose pubblicazioni in diritto dell'innovazione, proprietà intellettuale, storia del diritto e diritto internazionale.

ISBN: 978-88-8080-766-7