

## Capitolo 1

### *Intelligenza artificiale e lavoro*

SOMMARIO: Sez. I – *Questioni definitorie e primi tentativi di inquadramento*. – 1. Individuazione dell’ambito della ricerca: rischi e potenzialità dell’IA. – 2. Un tentativo definitorio: l’IA imita il comportamento umano. – 3. Lo stato dell’arte dal punto di vista regolativo: dalla anomia alle disposizioni generali europee. La carta dei diritti fondamentali. – 4. (*segue*). L’intelligenza artificiale per l’Europa (comunicazione del 25 aprile 2018) e il piano coordinato per l’intelligenza artificiale (comunicazione del 7 dicembre 2018). – 5. (*segue*). Creare fiducia nell’intelligenza artificiale antropocentrica (Comunicazione dell’8 aprile 2019): i sistemi di *machine learning* e l’applicazione acritica della tecnologia. – 6. (*segue*). Il Libro Bianco sulla intelligenza artificiale del 19 febbraio 2020: il tentativo di condizionamento delle iniziative nazionali e il problema del c.d. «ecosistema di fiducia». – 7. (*segue*). La proposta di Regolamento del Parlamento europeo e del Consiglio del 21 aprile 2021. – 8. (*segue*). La proposta di direttiva per il miglioramento delle condizioni di lavoro, nel lavoro mediante piattaforma. – 9. *Chatbots* e mercato dei servizi: riflessioni introduttive e rinvio. – Sez. II – *Rischio tecnologico e tecniche di tutela: da un sistema di liability ad un sistema di accountability?* – 1. Verso la creazione di un modello di sicurezza. Il problema della supervisione umana e la progettazione *ex ante* di sistemi dal comportamento prevedibile: *security by design; ethics by design; logging by design; reverse engineering*. – 2. (*segue*). Fattori giuridici, tecnici, organizzativi e aree di rischio. – 3. *AI in the UK: ready, willing and able?* – 4. (*segue*). A proposito delle cc.dd. aree di impatto ad alto rischio: «*What do we know so far?*». Posti di lavoro ed effetto c.d. sostitutivo. Brevi considerazioni e rinvio.

#### Sez. I – *Questioni definitorie e primi tentativi di inquadramento*

##### 1. *Individuazione dell’ambito della ricerca: rischi e potenzialità dell’IA.*

L’evoluzione tecnologica produce effetti significativi sulle organizzazioni produttive e sui rapporti di lavoro: non si tratta di una novità ed è ormai

esistente una copiosa letteratura che – da diversi punti di osservazione – cerca di analizzare il fenomeno provando a coglierne dimensione, contenuti, criticità<sup>1</sup>.

È (condivisibilmente) avvertito finanche il timore di una vera e propria de-costruzione del quadro giuridico relativo ai rapporti di produzione e del contenuto sociale e identitario delle relazioni di lavoro<sup>2</sup> così come è (sempre condivisibilmente) diffuso il timore di dover riscrivere regole di ingaggio e processi organizzativi per far fronte alle nuove sfide correlate al massiccio impiego di «decisori automatici»<sup>3</sup>.

In letteratura circola una affermazione molto diffusa: la digitalizzazione<sup>4</sup>

---

<sup>1</sup> Sul tema la letteratura è sterminata e non è certamente possibile confrontarsi con tutta la produzione esistente: si rinvia, senza alcuna pretesa di esaustività, a: G. SARTOR, *L'intenzionalità dei sistemi informatici e il diritto*, in *Riv. trim. dir. proc. civ.*, 1, 2003, p. 23 ss.; V. MAIO, *Il diritto del lavoro e le nuove sfide della rivoluzione robotica*, in *Arg. dir. lav.*, 6, 2018, p. 1414 ss.; M. FAIOLI, *Mansioni e macchina intelligente*, Torino, 2018; E. DAGNINO, *Dalla fisica all'algoritmo: una prospettiva di analisi giuslavoristica*, in ADAPT University Press, 2019; A. DONINI, *Il lavoro attraverso le piattaforme digitali*, Bologna, 2019; V. FILÌ, F. COSTANTINI (a cura di), *Legal Issues in the Digital Economy Technologies in the Labour Market*, in ADAPT University Press – Cambridge Scholars Publishing New Castel (UK), 2019; M. PERSIANI, *Osservazioni sulla nuova vicenda dei riders*, in *Giur. it.*, 2020, p. 1802 ss.; S. MAINARDI, *Rivoluzione digitale e diritto del lavoro*, in *Mass. giur. lav.*, 2, 2020, p. 341 ss.; P. PERLINGIERI, S. GIOVA, I. PRISCO (a cura di), *Rapporti civilistici e intelligenze artificiali: attività e responsabilità. Atti del 15° Convegno Nazionale SISDiC*, Napoli, 2020; F. GARIBALDO, M. RINALDINI (a cura di), *Il lavoro operaio digitalizzato. Inchiesta nell'industria metalmeccanica bolognese*, Bologna, 2021; A. PERULLI, *Oltre la subordinazione. La nuova tendenza espansiva del diritto del lavoro*, Torino, 2021; A. PIZZOFERRATO, *Digitalisation of work: new challenges to labour law*, in *Arg. dir. lav.*, 2021, p. 1329 ss.; A. FAILLI, *Dal digitale al mondo fisico. Le condizioni di lavoro nel sito Amazon di Passetto Corese*, Roma, 2022; M. NOVELLA, P. TULLINI (a cura di), *Lavoro digitale*, Torino, 2022; G. GAUDIO, *L'algorithmic management e il problema della opacità algoritmica nel diritto oggi vigente e nella Proposta di Direttiva sul miglioramento delle condizioni dei lavoratori tramite piattaforma*, in *Lavoro Diritti Europa*, 2022, 1; S. BORELLI, V. BRINO, C. FALERI, L. LAZZERONI, L. TEBANO, L. ZAPPALÀ, *Lavoro e tecnologie. Dizionario del diritto del lavoro che cambia*, Torino, 2022; L. VALENTE, *La tutela del lavoro ai tempi dell'intelligenza artificiale*, in *lavoce.info*, 7 febbraio 2024.

<sup>2</sup> Così P. TULLINI, *La digitalizzazione del lavoro, la produzione intelligente e il controllo tecnologico nell'impresa*, in ID. (a cura di), *Web e lavoro. Profili evolutivi e di tutela*, Torino, 2017, p. 6. Per un'analisi di più ampio respiro v. G. SANTORO-PASSARELLI, *Ricerche trasformazioni socio-economiche e nuove frontiere del diritto del lavoro civiltà giuridica e trasformazioni sociali nel diritto del lavoro*, in *Dir. rel. ind.*, 2, 2019, p. 417 ss.

<sup>3</sup> Gli studiosi dei «decisori automatici» affrontano il tema con approccio interrogativo: cfr. F. ROSSI, *Il confine del futuro. Possiamo fidarci dell'intelligenza artificiale?*, Milano, 2019.

<sup>4</sup> Termine qui volutamente utilizzato in modo generico, quale sinonimo di tutto ciò che è tecnologico/informatico, ecc. e che indica un insieme di strumenti diversi dai tradizionali strumenti analogici. Come si cercherà di dire più avanti, l'approccio squisitamente definitorio rischia di essere limitato e limitante.

sta cambiando le strutture della società e dell'economia<sup>5</sup>. Si presuppone, quindi, un'alterità tra la struttura socio-economica e gli strumenti che consentono lo svolgimento dei rapporti umani a qualsiasi livello; si dà per scontato che lo *shock* tecnologico operi come un *outsider*, penetrando all'interno di relazioni e processi, determinandone un adeguamento forzato non solo nel metodo ma anche dal punto di vista contenutistico, in qualche modo inducendo un'interazione bidirezionale tra l'uomo e le macchine<sup>6</sup>.

Questa alterità però non sembra così netta e appare, a ben vedere, sempre più sfumata.

Gli strumenti a disposizione sono il frutto di complessi processi evolutivi molto diversi da quelli che hanno portato – per fare alcuni esempi – dal lavaggio dei panni al fiume alla lavatrice o che hanno consentito di passare dalla produzione manuale dei libri al torchio da stampa di Gutenberg<sup>7</sup>: nel giro di pochi anni la tecnologia si è sviluppata in modo da andare ben oltre il condizionamento delle vecchie strutture sociali (inducendo processi di adeguamento), giungendo sostanzialmente a codificare nuovi rapporti, con nuove regole e nuove relazioni che non possono prescindere dalla tecnologia quale elemento fondante<sup>8</sup>.

Non si tratta, in altri termini, di interrogarsi solo sull'impatto della tec-

---

<sup>5</sup>T. TREU, *La digitalizzazione del lavoro: proposte europee e piste di ricerca*, in *Conversazioni sul lavoro a distanza da ottobre 2021 a giugno 2022*, promosse e coordinate da V.A. POSO, 5 ottobre 2021, in <http://www.bollettinoadapt.it/wp-content/uploads/2021/10/Tiziano-Treu-La-digitalizzazione-del-lavoro-Conversazioni-sul-lavoro-a-distanza-05-10-2021.pdf>. Più in generale, v. G. SARTOR, *Artificial intelligence and human rights: between law and ethics*, in *Maastricht Journal of European and Comparative Law*, 27, 2020, issue 6, p. 705 ss.; E. CATERINI, *L'intelligenza artificiale «sostenibile» e il processo di socializzazione del diritto civile*, Napoli, 2020; G. PASCERI, *Intelligenza artificiale, algoritmo e machine learning*, Milano, 2020. Per un approccio filosofico al tema si rinvia (anche per i riferimenti bibliografici) a P.B. HELZEL, A. NAPOLI, voce *Algoritica*, in *Enciclopedia della Bioetica e Scienze Giuridiche, Aggiornamento*, 2022. Per un'analisi dei profili relativi alla sicurezza sociale dei lavoratori v. V. FILÌ, *Platform Workers' Needs and Social Security Challenges*, in V. FILÌ, F. COSTANTINI (a cura di), *Legal Issues in the Digital Economy Technologies in the Labour Market*, cit., p. 105 ss.

<sup>6</sup>Sul rapporto tra il comportamento delle persone e la tecnologia si potrebbero fare tanti esempi. Un caso singolare – e particolarmente eccentrico – è quello citato da M. MITCHELL nel suo libro *Artificial Intelligence. A guide for thinking humans*, London, 2019, p. 37, a proposito della abitudine – sorta poco più di 10 anni fa – che hanno sviluppato alcune persone che amano caricare su *YouTube* video di gatti: il sistema – implementato nel 2012 da un *team* di Google – ha reagito sfruttando le reti neurali e, dopo qualche tempo, ha iniziato automaticamente a riconoscere video che riproducevano azioni di gatti (*self-taught cat-recognition machine*).

<sup>7</sup>R. DONKIN, *Il futuro del lavoro*, Milano, 2010, p. 25.

<sup>8</sup>Sul tema si soffermano ampiamente P.V. MOORE, M. UPCHURCH, X. WHITTAKER (ed. by), *Humans and Machines at work. Monitoring, Surveillance and Automation in Contemporary Capitalism*, London, 2018.

nologia su rapporti tradizionalmente intesi (e, quindi, di interrogarsi sul pur importante percorso di adattamento di tali rapporti a nuovi strumenti); si tratta anche di confrontarsi con la tecnologia intesa quale patrimonio genetico di rapporti e persone (si pensi alle nuove generazioni, che non conoscono molti strumenti utilizzati nel millennio scorso per gestire affari, relazioni, beni, ecc.: esse conoscono solo i nuovi strumenti).

Non appare allora inutile provare a cogliere questo nuovo paradigma analizzando in particolare l'IA per provare a capire se nel passaggio dalla società cambiata da un certo tipo di tecnologia a quella che si sviluppa utilizzando la tecnologia quale elemento fondativo si è perso (o si rischia di perdere) qualche «pezzo» per strada.

Si tratta di capire, sia pure con un ragionevole margine di approssimazione, se e quale sia il punto di equilibrio: da un lato, la scontata efficienza che deriva dall'utilizzo della IA e, più in generale, della tecnologia disponibile; dall'altro, l'inevitabile rischio (per le persone, per le organizzazioni, per le relazioni) che ne deriva<sup>9</sup>.

In disparte, al momento, la definizione di IA (cfr. *infra*) si assiste alla moltiplicazione di sistemi (si diceva, poc'anzi, «decisori automatici») che evolvono molto velocemente e che ricorrono sempre più al *machine learning* o al *deep learning*<sup>10</sup> e che riescono finanche a «imparare» autonomamente, prendere decisioni ed eseguirle facendo a meno dell'uomo.

Si tratta di sistemi che decidono sulla base di processi molto complessi e connotati da un certo grado di imprevedibilità (è sempre più diffusa l'espressione «decisioni algoritmiche»<sup>11</sup>), non scrutinabili (o, in alcuni casi, parzialmente scrutinabili) dall'essere umano.

Da tali preliminari e superficiali considerazioni già emerge, però, un elemento che in qualche misura orienta la riflessione sul tema: non potendo arrestare il progresso tecnologico e lo sviluppo di determinati strumenti, si

---

<sup>9</sup>Per rimanere sul piano generale (e riservando al prosieguo l'analisi dei fattori di rischio in ambito prettamente lavorativo) si pensi ai rischi da manomissione dei sistemi di IA deputati allo svolgimento di attività particolarmente delicate (ad esempio, i c.d. veicoli autonomi); ai rischi relativi alla gestione di infrastrutture critiche (in ambito militare, finanziario, ecc.); ai rischi concernenti la gestione della sicurezza e dell'ordine pubblico (tramite, ad esempio, il c.d. riconoscimento facciale), ecc.

<sup>10</sup>Cfr. *infra*.

<sup>11</sup>Cfr. G. D'ACQUISTO, *Decisioni algoritmiche. Equità, causalità, trasparenza*, Torino, 2022. Parla di «potere algoritmico» S. CIUCCIOVINO, *La disciplina nazionale sulla utilizzazione della intelligenza artificiale nel rapporto di lavoro*, in *Lavoro Diritti Europa*, 1, 2024, p. 3. Sul tema si sofferma anche F. SEGHEZZI, *La nuova grande trasformazione. Lavoro e persona nella quarta rivoluzione industriale*, ADAPT University Press, 2017.

deve necessariamente trovare il modo di governarlo e per farlo, dal punto di vista lavoristico, si deve provare a capire quale possa essere l'impatto sulle tradizionali categorie del diritto del lavoro, cercando di comprendere – per un verso – quali possano essere le opportunità che l'IA offre e – per altro verso – quali possano essere le *chances* di gestione delle inevitabili criticità.

Criticità che, ove fossero confermati i timori di chi prevede una certa area di totale ingovernabilità di alcuni sistemi<sup>12</sup>, si scontrerebbero con il dilemma posto dal *computer Joshua* in *Wargames: the only winning move is not to play*<sup>13</sup>.

Il presente lavoro non ha l'ambizione di analizzare analiticamente ogni singolo – e possibile – impatto dell'IA su rapporto e mercato del lavoro.

Ci si concentrerà su aspetti che riguardano più da vicino la diffusione dell'IA e dei sistemi che decidono automaticamente, approfondendo alcuni temi che catalizzano l'attenzione dei legislatori europeo e nazionale: il primo capitolo è dedicato alle questioni definitorie e ai tentativi di inquadramento europei, con particolare attenzione alle proposte di Regolamento che stabilisce regole armonizzate sull'intelligenza artificiale e di direttiva per il miglioramento delle condizioni di lavoro, nel lavoro mediante piattaforma. Nella seconda sezione del primo capitolo si analizzano, inoltre, alcuni problemi concernenti la supervisione umana e le tecniche utilizzabili per imporre ai sistemi che utilizzano l'IA specifiche condotte a tutela delle persone.

Il secondo capitolo muove dai profili di metodo che caratterizzano i tentativi di governo dell'impatto dell'IA sui rapporti di lavoro e si sviluppa, poi, su alcuni aspetti specifici: la selezione del personale; l'organizzazione delle piattaforme e le decisioni algoritmiche; l'*algorithmic management* nella gestione dell'impresa e dei rapporti di lavoro, con particolare riferimento al rapporto tra l'art. 2103 c.c. e l'opacità delle decisioni assunte dall'IA; il fenomeno della disintermediazione, con specifico riferimento alle relazioni sindacali.

Il terzo capitolo è dedicato ai c.d. *Large Language Models (LLM)* e a *ChatGPT*: il tema è l'impatto che questi sistemi possono provocare sui rapporti di lavoro, soprattutto dal punto di vista sostitutivo; le riflessioni si estendono, peraltro, ad alcuni casi che riguardano l'ambito legale.

---

<sup>12</sup>Ne fornisce ampia notizia E.J. LARSON, *The Myth of Artificial Intelligence. Why computers can't think the way we do*, Cambridge (USA), London (UK), 2021.

<sup>13</sup>*Wargames* (Giochi di guerra), 1983, diretto da John Badham e presentato fuori concorso al 36° Festival di Cannes.

## 2. Un tentativo definitorio: l'IA imita il comportamento umano.

Lo studio dell'impatto della IA sul lavoro e sui rapporti di lavoro impone l'individuazione e la definizione degli elementi in gioco, in modo da poter chiaramente attribuire a diversi fattori diversi significati e, quindi, in modo da poter impostare diversi punti di osservazione dei fenomeni avendo una certa idea dei rapporti causa/effetto.

Il tema è però particolarmente sfuggente e mette a dura prova il giurista<sup>14</sup>, costretto a maneggiare termini che non conosce, la cui genesi (spesso) non comprende, il cui impatto non riesce a misurare<sup>15</sup>.

Si tratta, peraltro, di tema al quale si dedicano discipline diverse e che spesso si presenta non privo di ambiguità: basti pensare al disinvolto utilizzo dei termini *algoritmo*, *intelligenza artificiale*, *machine learning*, spesso utilizzati quali sinonimi<sup>16</sup>.

L'ambiguità è tradizionalmente maneggiata dai giuristi facendo ampio ricorso alla lettera della legge e alle regole poste a presidio della sua interpretazione<sup>17</sup>, che è naturalmente attraversata da quella tensione ispirata da valori e principi fondamentali indispensabili per l'interprete, che gli consentono di perseguire il fine della equità, della sostenibilità, della giustizia<sup>18</sup>.

Nel caso della IA il discorso è invece molto più complesso.

<sup>14</sup> E non solo: V. BAVARO, *Questioni in diritto sul lavoro digitale, tempo e libertà*, in *Riv. giur. lav.*, n. 1, 2018, p. 35, segnala come la filosofia arrivi «troppo tardi» nel rincorrere i processi di formazione della realtà. L'A. fa espresso riferimento a G.W.F. HEGEL, *Lineamenti di filosofia del diritto*, Bari, 1965, p. 17.

<sup>15</sup> Per una ampia ricognizione delle diverse criticità v. A. PAJNO, M. BASSINI, G. DE GREGORIO, M. MACCHIA, F.P. PATTI, O. POLLICINO, S. QUATTROCOLO, D. SIMEOLI, P. SIRENA, *Intelligenza artificiale: criticità emergenti e sfide per il giurista*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2019. Cfr. inoltre G. SARTOR, *Cognitive automata and the law: Electronic contracting and the intentionality of software agents*, in *Artificial Intelligence and Law*, 2009, 17, 4, p. 253 ss.

<sup>16</sup> Si soffermano su questo profilo V. FERRARI, L. PALOPOLI, *Rapporti civilistici e intelligenze artificiali: attività assicurativa*, in P. PERLINGIERI, S. GIOVA, I. PRISCO (a cura di), *Rapporti civilistici e intelligenze artificiali: attività e responsabilità. Atti del 15° Convegno Nazionale SISDiC*, Napoli, 2020, p. 192 ss.

<sup>17</sup> Sul tema la letteratura è sterminata: si rinvia a P. GROSSI, *Della interpretazione come invenzione (la riscoperta pos-moderna del ruolo inventivo della interpretazione)*, in *Quaderni fiorentini per la storia del pensiero giuridico moderno*, vol. 47, 2018, p. 9 ss.; E. CATERINI, *L'intelligenza artificiale «sostenibile» e il processo di socializzazione del diritto civile*, Napoli, 2020 (in particolare i paragrafi 1, 2 e 14).

<sup>18</sup> Cfr. N. LIPARI, *Costituzione e diritto civile*, in *Riv. trim. dir. proc. civ.*, 4, 2018, p. 1273.

In linea generale si tende ad utilizzare una particolare definizione spendibile al di fuori di quel contesto *iper-specialistico* riservato agli addetti ai lavori: ai fini della più ampia divulgazione, infatti, si definisce IA quella disciplina scientifica «che mira a definire e sviluppare programmi o macchine (software e/o hardware) che mostrano un comportamento che verrebbe definito intelligente se fosse esibito da un essere umano»<sup>19</sup>. Si tratta di definizione apparentemente accessibile che, tuttavia, poco o nulla dice del contenuto della disciplina, presupponendo che siano noti concetti particolarmente tecnici (macchine; *software*; *hardware*) e che siano noti i processi che consentono lo sviluppo di quel comportamento definibile *intelligente*.

Andando più a fondo e provando a risalire alle fondamenta della disciplina, ci si imbatte nei lavori di Turing<sup>20</sup> e von Neumann<sup>21</sup> e, inevitabilmente, negli studi di McCarthy.

Nel 1956 il matematico John McCarthy propose ai colleghi Minsky, Shannon e Rochester di avviare un *summer workshop* al Dartmouth College: «a 2 month, 10 man study of artificial intelligence to be carried out during the summer of 1956»<sup>22</sup>. Per la prima volta si fa riferimento alla IA allo scopo di individuare l'oggetto del *workshop* e distinguerlo dal tema che all'epoca era già ben più noto, ossia quello della *cibernetica*<sup>23</sup>. McCarthy descrisse la sua proposta in questi termini, ipotizzando una sorta di obiettivo minimo: «To write a calculator program which can solve intellectual problems as well as or better than a human beings».

Non si trattò di una iniziativa particolarmente proficua giacché la Rockefeller Foundation (alla quale McCarthy si rivolse per ottenere un finanziamento) erogò poche risorse e il costituendo gruppo di ricerca non riuscì a ricavare risultati apprezzabili<sup>24</sup>.

Tuttavia, spalancò le porte ad iniziative simili e gettò le basi per lo svi-

<sup>19</sup> F. ROSSI, *Il confine del futuro. Possiamo fidarci dell'intelligenza artificiale?*, cit., p. 25.

<sup>20</sup> Cfr. *infra*.

<sup>21</sup> J. VON NEUMANN, *The general and logical theory of automata*, in *The world of mathematics*, New York, 1954, p. 2070 ss.

<sup>22</sup> J. MCCARTHY *et al.*, *A Proposal for the Dartmouth Summer Research Project in Artificial Intelligence*, inviata alla Rockefeller Foundation e della quale rende notizia M. MITCHELL, *Artificial Intelligence. A guide for thinking humans*, London, 2019, p. 4 e in nota 1 a p. 374.

<sup>23</sup> Avente ad oggetto il controllo e la comunicazione delle macchine: cfr. N. WIENER, *Cybernetics*, Cambridge, 1961.

<sup>24</sup> Si veda G. SOLOMONOFF, *Ray Solomonoff and the Dartmouth Summer Research Project in Artificial Intelligence, 1956*, in <http://www.raysolomonoff.com/dartmouth/dartray.pdf>. «Everyone had a different idea, a hearthy ego, and much enthusiasm for their own plan...».

luppo di un approccio scientifico: si unirono a McCarthy e Minsky altri studiosi, veri e propri pionieri della materia, ossia Allen Newell e Herbert Simon; venne fondato a Stanford il c.d. *Artificial Intelligence Project* con l'obiettivo di costruire una macchina completamente *intelligente* entro 10 anni<sup>25</sup>; si gettarono le basi che valsero a Simon – anni dopo – il premio Nobel<sup>26</sup>; venne fondato il MIT AI Lab<sup>27</sup>.

Gli sforzi dell'epoca si infrangevano su una barriera pressoché insormontabile che era banalmente rappresentata dal livello tecnologico delle macchine disponibili, certamente non paragonabile a quello dei *computer* attuali; nondimeno, gli obiettivi rimanevano ambiziosi e spaziavano dallo studio della intelligenza umana (e, quindi, lo studio dei meccanismi biologici utilizzati dall'uomo per risolvere problemi complessi) alla ricerca della matematica necessaria a replicare tali capacità utilizzando una macchina<sup>28</sup>.

Si tratta, in altri termini, di sforzi riferibili ad una sorta di *imitazione* delle capacità umane.

In questa stessa direzione si sviluppano le intuizioni di Turing<sup>29</sup> che negli anni '50 propone – provocatoriamente – la domanda *Can machines think?*

L'accostamento del verbo *pensare* alla parola *macchine* è certamente arduo e, tuttavia, ben rappresenta lo sforzo di sfruttare l'ambiguità di determinati concetti per sfondare il muro della non-replicabilità di alcune capacità umane.

Lo studio di Turing si snoda proprio intorno alla verificabilità di tale ipotesi tramite il noto *imitation game*: nell'esempio che apre il suo *paper* i protagonisti sono solo umani (un uomo, una donna, un verificatore – indifferentemente uomo o donna) e l'obiettivo consiste essenzialmente nel capire se il verificatore – collocato in una stanza separata dagli altri due – possa capire da alcune risposte (scritte) chi sia l'uomo e chi sia la donna; ove a rispondere alle domande sia l'uomo, la donna può interloquire con il verificatore e sviluppare una strategia per provare ad aiutarlo o a confonderlo.

<sup>25</sup> Si può facilmente reperire in rete documentazione concernente il gruppo di Stanford: cfr., ad esempio, [https://stacks.stanford.edu/file/druid:qb147kq8662/SC1041\\_SAIL\\_AIM\\_014.pdf](https://stacks.stanford.edu/file/druid:qb147kq8662/SC1041_SAIL_AIM_014.pdf), riguardante il *Computer Control of a Machine exploring Mars*, 15 giugno 1964.

<sup>26</sup> Per le sue pionieristiche ricerche sul processo decisionale nelle organizzazioni economiche.

<sup>27</sup> Ne raccontano la storia S. SABANOVIC, S. MILOJEVIC, J. KAUR, *John McCarthy [History]*, in *IEEE Robotics & Automation Magazine*, 4, 19, 2012.

<sup>28</sup> Cfr. M. MITCHELL, *Artificial Intelligence*, cit., p. 7.

<sup>29</sup> A.M. TURING, *Computing Machinery and Intelligence*, in *Mind*, vol. 59, 236, 1950, p. 433 ss.



Così costruito il gioco, Turing immagina di sostituire uno dei partecipanti (che fornisce le risposte) con una macchina e prova ad immaginare se e in che misura il verificatore possa distinguere – analizzando solo le risposte – la macchina dall'essere umano. La macchina supera il c.d. *Test di Turing* se il verificatore umano, pur sapendo che uno dei due interlocutori è una macchina, non riesce a distinguerli<sup>30</sup>.

Oggi l'IA disponibile supera abbondantemente le aspettative del gruppo di Stanford o di Turing: i sistemi possono consistere in *software* che operano nel mondo virtuale (un esempio è il *software* che consente il riconoscimento facciale) oppure in sistemi più complessi che consistono in dispositivi *hardware* (*robots*, ad esempio) che interagiscono con gli esseri umani e si adattano al loro comportamento<sup>31</sup>.

Ed è proprio l'ampio raggio di operatività della IA a suggerire diversi tentativi di inquadramento che, muovendo da definizioni più o meno circoscritte, si pongono l'obiettivo di governare l'interazione uomo-macchina.

### 3. Lo stato dell'arte dal punto di vista regolativo: dalla anomia alle disposizioni generali europee. La carta dei diritti fondamentali.

Il tema è nuovo anche per i *policy/law makers*. In linea generale si deve prendere atto di una disarticolazione dei fenomeni: l'evoluzione della tecnologia, nelle sue diverse declinazioni, sembra avere tempi incompatibili con l'evoluzione dei sistemi giuridici, tradizionalmente ancorata a processi lunghi e farraginosi<sup>32</sup>.

Nondimeno, negli ultimi anni l'Unione europea ha avviato iniziative sempre più articolate e finalizzate a costruire le basi della disciplina applicabile all'IA e a ciò che intercetta, in linea generale, l'interazione uomo/macchina.

I principi che animano tali iniziative vanno individuati nella Carta dei Diritti Fondamentali dell'Unione europea (CDFUE, c.d. Carta di Nizza)

---

<sup>30</sup> Si sofferma sul *Test di Turing* G.F. ITALIANO, *Le sfide interdisciplinari dell'intelligenza artificiale*, in *Anal. giur. econ.*, 1, 2019, p. 9 ss.

<sup>31</sup> Per un approccio definitorio e per una analisi casistica cfr. S. BORELLI, V. BRINO, C. FALERI, L. LAZZERONI, L. TEBANO, L. ZAPPALÀ, *Lavoro e tecnologie. Dizionario del diritto del lavoro che cambia*, Torino, 2022, p. 133 ss.

<sup>32</sup> Cfr., anche per i riferimenti alla qualità dei processi normativi, M. AINIS, *La legge oscura*, Bari, 2010.

del 7 dicembre 2000 (poi approvata il 12 dicembre 2007) che ha oggi, come si sa, valore giuridico di Trattato, in base a quanto previsto dall'art. 6 del Trattato sull'Unione europea (TUE). La Carta si articola in 7 capi e ribadisce, nel preambolo, gli elementi sui quali si fonda l'Unione<sup>33</sup>: «l'Unione si fonda sui valori indivisibili e universali di dignità umana, di libertà, di uguaglianza e di solidarietà; l'Unione si basa sui principi di democrazia e dello stato di diritto. Essa pone la persona al centro della sua azione istituendo la cittadinanza dell'Unione e creando uno spazio di libertà, sicurezza e giustizia».

Ed è proprio a questo assetto valoriale e alle sue specifiche articolazioni nei diversi ambiti (libertà fondamentali comuni; diritti dei cittadini della UE; diritti economici e sociali; tutela dei dati personali e tutela contro le discriminazioni, ecc.) che si riferisce il lavoro elaborato nel 2021 dall'Agenzia dell'Unione europea (FRA) per i diritti fondamentali<sup>34</sup>.

L'Agenzia ha prodotto una relazione su IA e diritti fondamentali nell'ambito di un progetto riguardante anche i cc.dd. *Big Data* (sui quali si tornerà nel prosieguo) finalizzato a valutare le implicazioni positive e negative delle nuove tecnologie sui diritti fondamentali.

La relazione prende spunto da riflessioni già svolte dall'Agenzia su tre grandi temi<sup>35</sup>: *facial recognition technology: fundamental rights considerations in the context of law enforcement; data quality and artificial intelligence – mitigating bias and error to protect fundamental rights; BigData: Discrimination in data-supported decision making*.

Si tratta, complessivamente, di un'analisi che conduce ad una conclusione difficilmente contrastabile: le nuove tecnologie, l'IA e la condivisione dei dati creano importanti e nuove opportunità ma anche sfide e minacce per i diritti umani e i diritti fondamentali.

Gli stati membri devono farsi carico di tali sfide e minacce, anche in forza di quanto previsto dall'art. 51, paragrafo 1, della Carta<sup>36</sup>.

<sup>33</sup> Sulla natura giuridica della Carta e sulla sua interpretazione cfr. S. RODOTÀ, *La Carte europea dei diritti – una vittoria dimezzata*, in "La Repubblica", Roma, 4 ottobre 2000, 17, al quale fa riferimento G. PASCERI, *Intelligenza artificiale, algoritmo e machine learning*, cit., p. 70 e, in particolare, nota 4.

<sup>34</sup> Reperibile in [https://fra.europa.eu/sites/default/files/fra\\_uploads/fra-2021-artificial-intelligence-summary\\_it.pdf](https://fra.europa.eu/sites/default/files/fra_uploads/fra-2021-artificial-intelligence-summary_it.pdf).

<sup>35</sup> Tutti reperibili in <https://fra.europa.eu/>.

<sup>36</sup> «Le disposizioni della presente Carta si applicano alle istituzioni e agli organi dell'Unione nel rispetto del principio di sussidiarietà come pure agli Stati membri esclusivamente nell'attuazione del diritto dell'Unione. Pertanto, i suddetti soggetti rispettano i diritti, osservano i principi e ne promuovono l'applicazione secondo le rispettive competenze».

Si tratta di un principio generale che vale evidentemente anche per l'IA e per la grande varietà di sistemi che vi rientrano. L'Agenzia afferma, a tal proposito, che l'utilizzo di sistemi di IA coinvolge una serie di diritti fondamentali indipendentemente dal campo di applicazione, tra cui figurano – a titolo di esempio – la dignità umana, la vita privata, la protezione dei dati, la non discriminazione e l'accesso alla giustizia, il diritto al lavoro, il diritto alla sicurezza sociale e all'assistenza sociale, ecc.

È, quindi, necessario che il Legislatore europeo e i singoli stati membri condizionino il loro agire garantendo il rispetto della Carta e la salvaguardia dei diritti fondamentali, limitandone la restrizione e/o la compressione solo a casi specifici e nel pieno rispetto del principio di proporzionalità.

4. (segue). *L'intelligenza artificiale per l'Europa (comunicazione del 25 aprile 2018) e il piano coordinato per l'intelligenza artificiale (comunicazione del 7 dicembre 2018).*

Il percorso di condizionamento dell'azione dei *policy/law makers* è sostenuto, innanzitutto, da un importante impegno della Commissione europea che, nel corso degli ultimi anni, ha elaborato documenti ufficiali organizzati proprio sul tema dell'IA e del rispetto dei diritti fondamentali.

La comunicazione della Commissione al Parlamento europeo, al Consiglio, al Comitato Economico e Sociale europeo e al Comitato delle regioni del 25 aprile 2018 ha ad oggetto «l'intelligenza artificiale per l'Europa»<sup>37</sup>.

Nella prima parte la Commissione chiarisce la difficoltà di individuare una definizione di IA e, tuttavia, cerca di definire i confini entro i quali collocare i sistemi di intelligenza artificiale, soprattutto al fine di definire – in corrispondenza – un ambito di operatività, di limiti, garanzie, restrizioni.

Così procedendo, la Commissione definisce sostanzialmente tre obiettivi ai quali dovrebbero tendere le future iniziative europee: impulso alla capacità tecnologica ed industriale dell'UE all'adozione dell'IA in tutti i settori economici; modernizzazione dell'istruzione e dei sistemi di informazione per affrontare adeguatamente i cambiamenti socio-economici apportati dall'IA; preparazione di un quadro etico e giuridico adeguato, basato sui valori dell'Unione e coerente con la Carta dei diritti fondamentali dell'UE.

---

<sup>37</sup> Reperibile in <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=COM%3A2018%3A237%3AFIN>.

Con riferimento al terzo obiettivo, la comunicazione si concentra particolarmente sul Regolamento generale sulla protezione dei dati (GDPR) e sulla necessità di elaborare orientamenti etici<sup>38</sup> per l'IA.

A questa Comunicazione segue poi il Piano coordinato del dicembre 2018<sup>39</sup>.

In disparte i riferimenti agli sforzi economici che gli stati membri dovranno realizzare per implementare i principi sopra richiamati e mettere in atto le strategie europee sul tema, appare particolarmente interessante il riferimento ad un approccio «che pon[ga] le persone al centro dello sviluppo dell'IA (IA antropocentrica)»<sup>40</sup>.

Così procedendo, il Piano individua sostanzialmente un percorso comune e coordinato sullo sviluppo dell'IA, al fine di massimizzare gli investimenti UE e degli stati membri: le macro-aree individuate riguardano essenzialmente il settore pubblico (sanità, trasporti e mobilità, sicurezza, energia) e il settore dell'economia (attività produttive e servizi finanziari).

5. (segue). *Creare fiducia nell'intelligenza artificiale antropocentrica (Comunicazione dell'8 aprile 2019): i sistemi di machine learning e l'applicazione acritica della tecnologia.*

Il riferimento all'IA antropocentrica ritorna nella Comunicazione della Commissione datata 8 aprile 2019<sup>41</sup>.

Il ragionamento si svolge coltivando il Piano del 2018 e sottolineando la necessità di sviluppare una IA che non sia fine a sé stessa ma che sia a servizio delle persone, perseguendo quale fine ultimo il miglioramento del benessere degli esseri umani.

E proprio a tale proposito la Commissione individua un aspetto particolarmente critico: il livello di affidabilità dei sistemi di IA con riferimento alla tutela dei diritti fondamentali. D'altro canto, la Commissione ritiene che l'apparato normativo in possesso dell'Unione sia già abbastanza attrezzato

---

<sup>38</sup> Testualmente, p. 16 della Comunicazione.

<sup>39</sup> Reperibile in [https://eur-lex.europa.eu/resource.html?uri=cellar:22ee84bb-fa04-11e8-a96d-01aa75ed71a1.0004.02/DOC\\_1&format=PDF](https://eur-lex.europa.eu/resource.html?uri=cellar:22ee84bb-fa04-11e8-a96d-01aa75ed71a1.0004.02/DOC_1&format=PDF).

<sup>40</sup> Testualmente, p. 1 del Piano.

<sup>41</sup> Reperibile in <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:52019DC0168&from=IT>.

zato per cogliere le nuove sfide: ancora una volta compare il riferimento al Regolamento generale sulla protezione dei dati; si fa poi espressamente riferimento al Regolamento sulla *cyber-security*<sup>42</sup>.

Quanto ai sistemi di *machine learning*, la Commissione manifesta una certa preoccupazione e cita, in particolare, alcuni settori e categorie di beni e servizi (*smartphone*, automobili automatizzate, *robot*, applicazioni *online*): il rischio è individuato principalmente nell'inaffidabilità dei dati sui quali si costruiscono i modelli che determinano il *modus operandi* di tali sistemi e sull'applicazione acritica della tecnologia, il che potrebbe esporre le persone a particolari rischi (ma sul punto si tornerà nel prosieguo). Ed è proprio su questo tema che la Commissione ribadisce il riferimento all'etica (sia pure senza riempirlo di particolare contenuto): «Di conseguenza, le applicazioni di IA dovrebbero non solo rispettare la legge, ma anche osservare i principi etici e garantire che le loro attuazioni pratiche non comportino danni indesiderati. La diversità in termini di sesso, razza o origine etnica, religione o convinzioni personali, disabilità ed età dovrebbe essere garantita in ogni fase dello sviluppo dell'IA. Le applicazioni di IA dovrebbero dare potere alle persone e rispettarne i diritti fondamentali; dovrebbero puntare a rafforzare le capacità dei cittadini, non a sostituirsi a loro, e consentire l'accesso anche alle persone con disabilità. Vi è quindi la necessità di elaborare orientamenti etici<sup>43</sup> basati sul quadro normativo esistente e che dovrebbero essere applicati da sviluppatori, fornitori e utenti dell'IA nel mercato interno, stabilendo condizioni di parità sul piano etico in tutti gli Stati membri».

Gli esperti e gli osservatori qualificati hanno elaborato alcuni orientamenti che hanno poi portato all'individuazione di parametri più specifici. Si ritiene che, per ottenere un'intelligenza artificiale affidabile, siano neces-

---

<sup>42</sup> Regolamento (UE) 2019/881 del 17 aprile 2019, relativo all'ENISA, l'Agenzia dell'Unione europea per la *cybersicurezza*, e alla certificazione della *cybersicurezza* per le tecnologie dell'informazione e della comunicazione, e che abroga il Regolamento (UE) n. 526/2013 («regolamento sulla *cybersicurezza*»), reperibile in <https://eur-lex.europa.eu/legal-content/IT/TXT/PDF/?uri=CELEX:32019R0881&from=PT>.

<sup>43</sup> La cui individuazione è rimessa dalla Commissione a esperti, rappresentanti di una *vasta gamma di portatori di interessi*. Si tratta del gruppo di esperti sull'intelligenza artificiale (AI HLEG) che ha prodotto alcuni documenti che sono stati poi utilizzati per l'elaborazione del Regolamento del 2021 (sul quale cfr. *infra*): cfr. *Ethics guidelines for trustworthy AI*, reperibile in <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. È stata inoltre attivata una Alleanza per l'IA, costituita da una piattaforma destinata a consentire a circa 4.000 portatori di interessi di discutere le implicazioni tecnologiche e sociali dell'IA (<https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance>).

sari tre elementi: l'IA dovrebbe rispettare la legge; dovrebbe osservare i principi etici; dovrebbe dimostrare robustezza.

Sulla base di questi tre elementi e dei valori europei sopra citati, sono poi individuati sette requisiti fondamentali che le applicazioni di IA dovrebbero soddisfare per essere considerate affidabili: intervento e sorveglianza umani; robustezza tecnica e sicurezza; riservatezza e *governance* dei dati; trasparenza; diversità, non discriminazione ed equità; benessere sociale e ambientale; *accountability*.

Si tratta ovviamente di orientamenti e parametri non vincolanti; d'altro canto, l'ottica prediletta dalla Commissione è quella di preparare una solida base culturale nella quale far sviluppare future iniziative legislative (UE e degli stati membri) che possano costituire l'architettura – questa volta vincolante – che sorregge un apparato regolativo in grado di garantire la tutela dei diritti fondamentali.

6. (segue). *Il Libro Bianco sulla intelligenza artificiale del 19 febbraio 2020: il tentativo di condizionamento delle iniziative nazionali e il problema del c.d. «ecosistema di fiducia».*

Sulla base di tali premesse si arriva, nel febbraio 2020, al Libro Bianco<sup>44</sup>.

Il documento è diviso in 6 sezioni e sostanzialmente mira a compulsare un approccio comune all'IA tramite il condizionamento delle iniziative nazionali e la paralisi di interventi che potrebbero – tra le altre cose – compromettere la certezza del diritto.

Il documento punta alla creazione: di un quadro strategico per la individuazione di misure utili ad allineare gli sforzi a livello europeo, nazionale e regionale; di un quadro normativo per la creazione di un «ecosistema di fiducia» unico.

Il Libro Bianco, al pari degli altri documenti programmatici sopra citati, insiste sulla necessità di fiducia dell'uomo nell'IA e, parallelamente, sul bisogno di tutela di tale fiducia mediante la creazione di sistemi altamente affidabili, nell'ottica della miglior tutela dei diritti fondamentali.

Si tratta di fiducia che non può essere data per scontata ove solo si considerino i rischi di privazione dei mezzi per la tutela dei diritti e i rischi de-

---

<sup>44</sup> Reperibile in [https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020\\_it.pdf](https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_it.pdf).

rivanti dalle asimmetrie informative che accompagnano i processi decisionali algoritmici (p. 10 del Libro).

A questo proposito si legge nel documento un esplicito riferimento alla ricerca del Consiglio d'Europa del marzo 2018, *Study on the human rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications*<sup>45</sup>.

L'approccio in questo caso è più tecnico rispetto a quanto osservato in precedenza dalla Commissione ed è ancorato alle definizioni elaborate da T. Gillespie, N. Diakopoulos e F. Pasquale<sup>46</sup>: in particolare, il Consiglio prende le mosse da Gillespie e dalla sua affermazione secondo la quale «*algorithms need not be software: in the broadest sense, they are encoded procedures for transforming input data into a desired output, based on specified calculations. The procedures name both a problem and the steps by which it should be solved*»<sup>47</sup> e dalle riflessioni di Diakopoulos, secondo il quale «*algorithms are thus perceived as a series of steps undertaken in order to solve a particular problem or accomplish a defined outcome*»<sup>48</sup>.

Su tali basi il Libro delimita il campo d'indagine agli algoritmi che producono un impatto, con il loro funzionamento, su una generalità di individui elaborando processi decisionali (automatici o semiautomatici) che hanno (potenziali) implicazioni sui diritti umani.

Il *panel* elaborato affonda saldamente le radici in due questioni ben note: l'opacità dei sistemi e la loro imprevedibilità.

Gli esempi che si leggono nel Libro Bianco individuano alcune aree particolarmente esposte a tali questioni: il giusto processo; la protezione dei dati personali; la libertà di espressione; la libertà di associazione; il diritto ad un ricorso effettivo (secondo quanto stabilito dall'art. 13 della Convenzione per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali); il divieto di discriminazione; i diritti sociali; il diritto a elezioni libere e non condizionate, ecc.

Le conclusioni abbozzate in calce al Libro Bianco sono prudenti: gli autori riconoscono che lo stato dell'arte non è ancora ben definito e auspica-

---

<sup>45</sup> Reperibile in <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>.

<sup>46</sup> T. GILLESPIE, *The revealance of algorithms*, in *Media technologies: Essays on communication, materiality and society*, Cambridge, 2014; N. DIAKOPOPOULOS, *Algorithmic accountability*, in *Digital Journalism*, 3, 2015, p. 398 ss.; F. PASQUALE, *The black box society: The secret algorithms that control money and information*, Harvard, 2015.

<sup>47</sup> T. GILLESPIE, *The revealance of algorithms*, cit., p. 167.

<sup>48</sup> N. DIAKOPOPOULOS, *Algorithmic accountability*, cit., p. 400.

no maggiori ricerche e ulteriori sforzi interpretativi (non solo sul piano prettamente teorico). Tuttavia, suggeriscono alcune linee interpretative basate su una *human rights perspective* (p. 43 del Libro): i sistemi che adottano decisioni automatiche o semi-automatiche ad alto impatto dovrebbero essere pienamente scrutinabili dalle autorità, in modo da garantire trasparenza e responsabilizzazione.

Ciò richiede non solo uno sforzo di volontà da parte di chi si avvale dell'IA, ma richiede soprattutto una modifica dei processi di ideazione e produzione dei sistemi di IA che dovrebbero essere modellati ponendo *man beings and human rights at the centre of the technology design process*.

Ancora, e questo rappresenta un elemento di novità rispetto alla narrazione contenuta nei documenti programmatici ai quali si è fatto cenno in precedenza, il Libro Bianco suggerisce di individuare aree da sottrarre all'impatto di decisioni assunte da sistemi di IA: *«Importantly, there may be areas of societal and human interaction where algorithmic decision-making systems are not appropriate. Automated data processing and decision-making systems should not be relied upon heavily to promote societal development or resolve complex new challenges for future generations, as this is likely to do more harm than good. Therefore it is critical to ensure that in key areas where automation is not appropriate from a human rights perspective, it does not take place»* (p. 44 del Libro Bianco).

Si tratta di un'affermazione importante giacché sembra fondarsi sulla impossibilità – allo stato – di considerare affidabile l'IA in determinati settori e, quindi, sull'impossibilità di garantire tutela e garanzia di taluni diritti fondamentali, stante la necessità di garantire quella che viene definita dal Consiglio *primacy of the human being*<sup>49</sup>.

In conclusione, gli autori del Libro Bianco elaborano un'articolata riflessione finale: gli enti pubblici e le organizzazioni indipendenti dovrebbero supportare e finanziare la ricerca per approfondire le complesse interazioni tra sistemi di IA e esseri umani; gli enti pubblici dovrebbero essere resi responsabili per le decisioni assunte sulla base di procedure algoritmiche; lo sviluppo tecnologico dovrebbe essere strettamente monitorato al fine di prevenire impatti negativi con particolare riferimento alle elezioni o alle procedure elettorali; sono di cruciale importanza la consapevolezza

---

<sup>49</sup> A questo proposito il Libro Bianco richiama espressamente *Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality*, Report by the Rathenau Institut commissioned and funded by the Parliamentary Assembly of the Council of Europe, adopted by PACE on 28 April 2017.



pubblica e la divulgazione (dovrebbero essere utilizzati tutti gli strumenti disponibili per informare le persone e le comunità sul funzionamento dei sistemi di IA, in modo da consentire la maturazione di un pensiero critico); si dovrebbero sviluppare specifici meccanismi di certificazione e di controllo per i sistemi di IA in modo da consentire la verifica dei livelli di tutela dei diritti umani (il Consiglio suggerisce lo sviluppo di metodi di progettazione dei sistemi di IA orientati proprio alla tutela dei diritti umani); gli stati dovrebbero astenersi dall'imporre agli intermediari di servizi l'utilizzo di tecniche automatiche per il monitoraggio di informazioni (o per l'archiviazione o per l'accesso), in modo da tutelare integralmente la libertà di espressione; gli enti pubblici dovrebbero interloquire con i soggetti che operano nei diversi settori (assicurazioni, banche, agenzie, ecc.) per sviluppare buone pratiche e linee-guida utili ad affrontare le sfide che pongono i sistemi che assumono decisioni automatiche attraverso algoritmi; stante la complessità della materia, la consapevolezza delle persone non è di per sé sufficiente: è necessario creare nuove istituzioni, networks e spazi di confronto per sorvegliare i sistemi di IA, coinvolgendo chiunque possa rappresentare *stakeholders* qualificati; il Consiglio d'Europa è istituzionalmente orientato alla tutela dei diritti umani ed è il soggetto posto nella posizione migliore per valutare l'impatto dell'IA sui diritti fondamentali (sia nell'ambito pubblico sia nell'ambito privato), anche alla stregua delle sue prerogative e della capacità di interagire con gli stati membri.

#### 7. (segue). *La proposta di Regolamento del Parlamento europeo e del Consiglio del 21 aprile 2021.*

Nel mese di aprile del 2021 vede la luce la proposta di Regolamento che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modica alcuni atti legislativi dell'Unione<sup>50</sup>.

Il testo è molto complesso e articolato ed è dichiaratamente orientato a migliorare il funzionamento del mercato interno definendo un quadro giuridico uniforme, in particolare per lo sviluppo, la commercializzazione e l'uso dell'intelligenza artificiale, conformemente ai valori dell'Unione<sup>51</sup>, sfruttan-

---

<sup>50</sup> Reperibile in <https://eur-lex.europa.eu/legal-content/IT/TXT/?uri=COM:2021:0202:FIN>.

<sup>51</sup> Per un primo commento cfr.: C. CASONATO, B. MARCHETTI, *Prime osservazioni sulla pro-*

do quale base giuridica la disciplina dell'art. 114 del Trattato sul funzionamento dell'Unione europea TFUE (livello di protezione elevato in materia di sanità, sicurezza, protezione dell'ambiente, protezione dei consumatori, ecc.).

Si tratta di un'iniziativa che assume un significato simbolico, prima ancora che tecnico-giuridico: nella materia, infatti, in disparte gli interventi di *soft law* ai quali si è pure fatto cenno in precedenza, si registra un atteggiamento generalmente passivo che ondeggia tra il timore di mettere mano a materia delicatissima e in continua evoluzione e un più pratico *lassaiz faire* (si tornerà sull'argomento nel capitolo 3) che, probabilmente, è utile alle grandi aziende private che investono in ricerca e sviluppo di sistemi di IA sempre più sofisticati<sup>52</sup>.

Si tratta, inoltre, di un'iniziativa specificamente collegata a sistemi di controllo e verifica della sua efficacia nel tempo: per un verso è previsto l'intervento della Commissione (soprattutto sul contenuto degli allegati alla proposta) ai sensi dell'art. 290 TFUE e, per altro verso, è previsto un meccanismo periodico di verifica quinquennale (considerando la disciplina dell'IA *stable but not still*<sup>53</sup>).

Controlli e verifiche che sono pure ancorati al noto meccanismo delle cc.dd. *sandboxes*, secondo quanto previsto nel considerando n. 72 della proposta: «Gli obiettivi degli spazi di sperimentazione normativa dovrebbero essere la promozione dell'innovazione in materia di IA, mediante la creazione di un ambiente controllato di sperimentazione e prova nella fase di sviluppo e pre-commercializzazione al fine di garantire la conformità dei sistemi di IA innovativi al presente regolamento e ad altre normative pertinenti dell'Unione e degli Stati membri, e il rafforzamento della certezza del diritto per gli innovatori e della sorveglianza e della comprensione da parte delle autorità competenti delle opportunità, dei rischi emergenti e degli impatti dell'uso dell'IA, nonché l'accelerazione dell'accesso ai mercati, an-

---

posta di regolamento dell'Unione europea in materia di intelligenza artificiale, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2021, p. 415 ss.; F. LAMBERTI, *La proposta di regolamento UE sull'Intelligenza Artificiale alla prova della privacy*, in *federalismi.it*, focus LPT, 29 giugno 2022.

<sup>52</sup> Sebbene non manchino, come puntualmente osservano C. CASONATO, B. MARCHETTI, *Prime osservazioni*, cit., p. 417 ss. (cfr. in particolare nota 5), interventi *ad hoc*: negli Stati Uniti alcuni stati e alcune città hanno disciplinato e vietato il riconoscimento facciale. Cfr. in argomento A. CHEN, *Why San Francisco's ban on face recognition is only the start of a long fight*, in *MIT Review*, 2019, reperibile in <https://www.technologyreview.com/2019/05/16/135339/facial-recognition-ban-san-francisco-surveillance-privacy-private-corporate-interests/>.

<sup>53</sup> Cfr. ancora C. CASONATO, B. MARCHETTI, *Prime osservazioni*, cit., p. 419.

che mediante l'eliminazione degli ostacoli per le piccole e medie imprese (PMI) e le start-up».

L'art. 53 della proposta conferisce maggiore concretezza a questo *input* prevedendo, appunto, che: «Gli spazi di sperimentazione normativa per l'IA istituiti da una o più autori competenti degli Stati membri o dal Garante europeo della protezione dei dati forniscono un ambiente controllato che facilita lo sviluppo, le prove e la convalida di sistemi di IA innovativi per un periodo di tempo limitato prima della loro immissione sul mercato o della loro messa in servizio conformemente a un piano specifico. Ciò avviene sotto la guida e il controllo diretti delle autorità competenti al fine di garantire la conformità ai requisiti del presente regolamento e, se del caso, di altre normative dell'Unione e degli Stati membri controllate all'interno dello spazio di sperimentazione».

Ai fini che qui rilevano (si omette volutamente in questa parte il confronto con gli aspetti definatori<sup>54</sup>) va evidenziata l'idea di fondo: individuare regole più stringenti a fronte dell'aumento del rischio di compressione/limitazione di diritti fondamentali.

Ove il rischio sia insostenibile, si tratta di «Pratiche di intelligenza artificiale vietate» (art. 5): sistemi che utilizzano tecniche subliminali e/o che impattano sul grado di consapevolezza delle persone; sistemi che sfruttano le vulnerabilità di determinate categorie di persone; sistemi che sfruttano il c.d. «punteggio sociale» e/o classificano le persone sulla base di determinate caratteristiche; sistemi che utilizzano l'identificazione biometrica remota in tempo reale per l'accesso a determinati spazi, ecc.

Vi sono poi rischi gestibili, seppure con particolari limitazioni: la proposta, infatti, fa espressamente riferimento a «requisiti specifici per i sistemi di IA ad alto rischio e obblighi per gli operatori di tali sistemi».

Si tratta di prodotti «già soggetti alla legislazione europea di cui all'Allegato II del Regolamento ed alla valutazione di conformità da parte di terzi in vista dell'immissione sul mercato o della messa in servizio di tale prodotto ai sensi della stessa legislazione elencata nell'Allegato II e di prodotti elencati nell'Allegato III del regolamento nei seguenti settori: identificazione e categorizzazione biometrica delle persone fisiche; gestione e funzionamento delle infrastrutture critiche; istruzione e formazione professio-

---

<sup>54</sup>In linea generale si deve rilevare che la proposta abbraccia un significato molto ampio di «sistema di intelligenza artificiale, facendo riferimento a *software* sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare *output* quali contenuto, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono» (art. 3).

nale; occupazione, gestione dei lavoratori e accesso al lavoro autonomo; accesso e fruizione di servizi privati essenziali e di servizi e benefici pubblici; applicazione della legge; gestione della migrazione, dell'asilo e dei controlli alle frontiere; amministrazione della giustizia e processi democratici».

In questi casi il Regolamento prevede: l'adozione di un sistema di gestione per valutare e contrastare il rischio; la gestione dei dati affinché si garantisca una loro alta qualità (più alta è la qualità dei dati, più bassi sono i rischi per i diritti fondamentali e il rischio di esiti discriminatori); l'adozione di documentazione tecnica che fornisca tutte le informazioni necessarie per valutare la conformità del sistema ai requisiti e per consentire alle autorità di valutare tale conformità; la conservazione della documentazione tramite registri (*file di log*) per garantire la tracciabilità dei risultati; trasparenza nei rapporti con gli utenti e doveri di informarli; la supervisione umana per controllare e minimizzare il rischio; il garantire robustezza, sicurezza e accuratezza nella gestione dell'IA affinché i risultati non siano corrotti o soggetti ad errori, difetti o incongruenze.

Negli altri casi (non rientranti nelle classificazioni di rischio di cui sopra e che comunque prevedono interazione con persone fisiche, riconoscimento di emozioni, classificazioni biometriche, generazione/manipolazione di immagini, audio, video, ecc.) la proposta impone particolari obblighi di informazione e di trasparenza (che dovrebbero garantire la consapevolezza della interazione con un sistema di IA).

Sul fronte della trasparenza bisogna abbozzare alcune considerazioni (sul tema si tornerà comunque nel prosieguo).

Com'è noto il tema è condizionato dalle cc.dd. *black boxes*, ossia da sistemi di *machine learning* e di *deep learning* sostanzialmente impossibili da tracciare nei passaggi che eseguono per generare *outputs*<sup>55</sup>.

In questi casi si conosce il *data set* iniziale e si conosce lo scopo dichiarato dal creatore del sistema; fin troppo poco o nulla si conosce, però, del percorso che il sistema segue per svolgere il suo lavoro<sup>56</sup>: il termine *black box* si rivela, dunque, particolarmente efficace perché evoca l'inaccessibilità di una scatola chiusa: il sistema costruisce un suo modello affidandosi ad

---

<sup>55</sup> La diffusione di questi sistemi è sempre più pervasiva: cfr. V. FILI, F. COSTANTINI (a cura di), *Legal Issues in the Digital Economy Technologies in the Labour Market*, cit., p. viii ss. (con particolari riferimenti alla «dipendenza» dalle *black boxes*).

<sup>56</sup> F. PASQUALE, *The black box society*, cit.; cfr. anche B. DONALD, E. BAKIES, *A Glimpse Inside the Brain's Black Box: Understanding the Role of Neuroscience in Criminal Sentencing*, in *Fordham Law Review*, 85, 2, 2016, p. 481 ss. nonché G.F. ITALIANO, *Le sfide interdisciplinari dell'intelligenza artificiale*, in *Anal. giur. econ.*, 1, 2019, p. 9 ss.