

21



QUADERNI UNSIIC

Intelligenza artificiale: profili etici e giuridici

Giuseppe Mosa

QUADERNI



Intelligenza artificiale: profili etici e giuridici

Giuseppe Mosa*

*Avvocato - Consulente di imprese, organizzazioni sindacali ed enti pubblici

Indice

1. Sinossi	4
2. L'intelligenza artificiale (IA): definizione, origine e declinazioni concettuali.....	6
3. Etica e diritto nell'intelligenza artificiale: il quadro europeo	10
4. La responsabilità civile nell'uso dei sistemi di IA	16
5. IA e diritto del lavoro	24
6. Problemi e prospettive	40

1. Sinossi

I sistemi di intelligenza artificiale pongono nuovi e inaspettati problemi nel mondo del diritto e della società in generale.

Ci troviamo davanti a nuove entità, le cui decisioni e risultati non sono pienamente l'effetto di azioni umane, ma derivano da una serie di processi che, dopo l'istruzione della macchina, hanno autonoma capacità decisionale e in alcuni casi non giustificabili a posteriori, in quanto non è possibile comprendere come il sistema sia giunto ad assumere una certa decisione.

Di chi è dunque la responsabilità giuridica delle loro azioni? Quali sono gli ambiti più critici? Come si stanno muovendo, in ambito regolatorio, i paesi attivi nello sviluppo di sistemi di AI?

Il presente contributo nasce dalla determinazione dell'Ufficio Studi di Fondolavoro di approfondire lo studio dei moderni sistemi di intelligenza artificiale e delle relative implicazioni anche da una prospettiva giuridica.

L'impegno che si intende approfondire mira a stimolare la comunità scientifica e, più in generale, gli *stakeholders* e gli attori istituzionali, a una meditata presa di coscienza sulle conseguenze che derivano dalla diffusione di nuove tecnologie su larga scala che offrono inedite opportunità ai cittadini, alle imprese e alle pubbliche amministrazioni ma al contempo, ove non adeguatamente governate, possono dare

origine a rischi per i diritti e le libertà fondamentali.

Si impone, altresì, la necessità di assicurare che il progresso tecnologico si svolga in armonia con le esigenze di tutela individuali e collettive, nel rispetto di una dimensione antropocentrica¹. Occorre così delineare le modalità di intervento che minimizzino le esternalità negative sullo sviluppo delle nuove tecnologie garantendo al contempo la salvaguardia del nucleo duro dei diritti fondamentali. Con il presente lavoro si è inteso effettuare una mappatura dei tentativi sinora condotti a titolo sia di *soft law*, sia di *hard law*, in campo giurisprudenziale e in ambito dottrinale, per offrire un congruo inquadramento giuridico dell'intelligenza artificiale.

Si è cercato, altresì, di illustrare quali principi dovrebbero orientare e guidare i prossimi sviluppi. Svoltata una mappatura preliminare, infatti, il saggio procede secondo una logica de jure condendo, effettuando una rassegna di problematiche attraverso ideali carotaggi entro i vari settori del diritto che appaiono attinti dalle innovazioni tecnologiche di cui si discute.

In quest'ottica particolare attenzione viene dedicata all'ambito giuslavoristico (Cfr. paragrafo 5), che si presenta delicatissimo a causa della sua stessa natura.

Prima di affrontarlo specificatamente è, tuttavia, opportuno effettuare talune notazioni preliminari sull'intelligenza artificiale.

¹ Sui profili etici dell'I.A., in chiave ermeneutica e fenomenologica, si rimanda a F. Abbate, Paul Ricoeur, un'etica per la medicina e l'intelligenza artificiale, in Paul Ricoeur tra moderno e postmoderno, "In Circolo. Rivista di Filosofia e Culture", n.12/2021, Milano, pp.135-152.

2. L'intelligenza artificiale (IA): definizione, origine e declinazioni concettuali

L'intelligenza artificiale (IA) può essere definita come la **disciplina che studia "se e in che modo" si possano riprodurre i processi mentali più complessi mediante l'uso di un computer².**

Tale ricerca si sviluppa secondo due percorsi complementari: da un lato l'intelligenza artificiale cerca di avvicinare il funzionamento dei computer alle capacità dell'intelligenza umana, dall'altro usa le simulazioni informatiche per fare ipotesi sui meccanismi utilizzati dalla mente umana.

Il tentativo di simulare le funzioni umane è stato presente fin dagli inizi

² La formula dell'IA ricorre per la prima volta a metà degli anni Cinquanta del secolo scorso, con l'intento di indicare un «attempt» per «to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves». A questa prima fase, è seguito un periodo in cui l'obiettivo perseguito dalla comunità nella costruzione della macchina non è più rappresentato dalla sua idoneità a riprodurre il cervello dell'uomo (Artificial General Intelligence, AGI), ma piuttosto nella soluzione di specifici problemi. Oggi vi è chi definisce la IA come «the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable». Altri, invece, la definiscono come «la scienza della produzione di macchine e sistemi volti all'esecuzione di compiti che, qualora realizzati da essere umani, richiederebbero l'uso dell'intelligenza per risolvere problemi di apprendimento e conoscenza, di ragionamento e pianificazione»⁴⁸. Ancora, di recente, si è sostenuto che per IA si debba intendere il «field that studies the synthesis and analysis of computational agents that act intelligently». Anche con riferimento al piano normativo e para-normativo, è dato registrare una notevole pluralità di soluzioni. Tra queste, particolare attenzione ha ricevuto quella proposta a livello europeo lo scorso anno e che così recita: «Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting their behavior by analyzing how the environment is affected by their previous actions». Orbene la mancanza di un consenso attorno a una determinata definizione è spiegata con la difficoltà di affermare «a bright-line distinction between what constitutes AI and what does not» ed è impiegata per suggerire ai legislatori di «to find specific definitions which could prove useful to address narrowly identified problems posed by AI applications» the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt.

dell'informatica.

Nel 1956 J. McCarthy organizzò a Dartmouth un convegno, nel corso del quale fu coniato il termine i. artificiale nel senso usato oggi, che segna la nascita di questa disciplina come settore autonomo; durante il convegno vennero presentati alcuni programmi capaci di comportamenti intelligenti, quale per es. il *logic theorist*, in grado di dimostrare teoremi di logica matematica³.

In una **prima fase** di evoluzione l'i. artificiale si è rivolta alla soluzione di problemi di tipo logico, relativamente ben formalizzati, registrando significativi successi; nel 1957 apparve il programma *General Problem Solver*, destinato a emulare il comportamento umano nella soluzione di problemi di tipo generale; nel 1959 H. Gelertner presentò un programma per la dimostrazione di teoremi di geometria e, subito dopo, uno per l'integrazione simbolica.

In una **seconda fase**, sull'onda del successo, si è cercato di espandere la gamma di problemi cui l'i. artificiale potesse dare risposta, muovendosi verso l'emulazione del sistema di percezione umano e verso l'interazione con l'ambiente, registrando viceversa un sostanziale insuccesso. In effetti ci si rese presto conto che per i problemi di

³ Durante il convegno di Dartmouth ebbe un ruolo fondamentale il lavoro di Alan Turing, considerato uno dei padri dell'informatica moderna: già nel 1936 aveva posto le basi per i concetti di calcolabilità, computabilità e la macchina di Turing. Nel 1950, lo stesso Turing, scrisse l'articolo intitolato "Computing machinery and intelligence", in cui proponeva quello che sarebbe divenuto noto come test di Turing. Secondo il test, una macchina poteva essere considerata intelligente se il suo comportamento, osservato da un essere umano, fosse considerato indistinguibile da quello di una persona. Grazie al lavoro di Turing, il tema dell'Intelligenza Artificiale ricevette una forte attenzione da parte della comunità scientifica e nacquero diversi approcci: fra tutti, i principali furono la logica matematica (per la dimostrazione di teoremi e l'inferenza di nuova conoscenza) e le reti neurali (nell'ultimo decennio la loro tecnologia è stata implementata e oggi vengono applicate nell'ambito del Deep Learning, un ramo del Machine Learning).

gestione efficace di informazioni incomplete e incerte mancavano i modelli concettuali che potessero supportare l'attività di sistemi automatizzati.

In una **terza fase** il settore si è articolato in indirizzi specifici che, con forti interazioni con altri settori, hanno cercato di affrontare alcuni problemi delimitati, registrando alcuni interessanti successi e contribuendo comunque all'approfondimento dei complessi problemi che stanno dietro a molte attività umane, solo apparentemente semplici.

Oggi l'Intelligenza Artificiale rappresenta uno dei principali ambiti di interesse della comunità scientifica informatica, con temi di ricerca come il *Machine Learning*, l'elaborazione del linguaggio naturale e la robotica.

Da un lato questo insieme di tecnologie aprirà mercati completamente nuovi e offrirà prospettive professionali anche a chi oggi è escluso da molte possibilità, fornendo soluzioni a problemi oggi insormontabili.

Dall'altro, tuttavia, esse presuppongono un ripensamento dei tradizionali paradigmi che caratterizzano i diversi ambiti del diritto che vanno rimessi in gioco per poter meglio svolgere una funzione volta a neutralizzare gli eventuali rischi derivanti dalle attività che si compiono in un ambiente interconnesso.

L'utilizzo complesso dell'IA, infatti, è idoneo a determinare dei risultati che pongono in crisi gli ordinari criteri di imputabilità della responsabilità e che comportano inedite forme di lesione di diritti

fondamentali, richiedendo al diritto uno sforzo interpretativo e, in subordine, legislativo.

3. Etica e diritto nell'intelligenza artificiale: il quadro europeo

Come si è visto nel capitolo che precede le odierne tecnologie, volte allo sviluppo dell'intelligenza artificiale, incidono sempre più sulla società e sui costumi, sollevando il problema della elaborazione di strumenti a protezione dei diritti fondamentali, della sicurezza e della protezione delle risorse conoscitive che si estraggono dai dati.

I tradizionali paradigmi che caratterizzano i diversi ambiti del diritto sono rimessi in gioco per poter meglio svolgere una funzione volta a neutralizzare gli eventuali rischi derivanti dalle attività che si compiono in un ambiente interconnesso.

A tal proposito, le istituzioni europee hanno rivolto la propria attenzione precipuamente alle forme di IA in grado di operare in modo autonomo, le quali mimano e riproducono l'intelligenza umana, senza la pretesa tuttavia di superarla. Oggetto di intervento sono stati i *narrow enabled-systems*, sistemi di IA più "limitati" che, grazie alla loro capacità di apprendimento, basata sull'analisi di enormi quantità di dati, pongono delicate questioni in tema di sicurezza, di privacy e, soprattutto, di responsabilità per danni causati a terzi.

Il grado di autonomia del sistema IA è, infatti, uno dei punti chiave dell'analisi sulla responsabilità civile, così come confermato dalla **Risoluzione del Parlamento Europeo** che definisce l'autonomia

dell'Intelligenza artificiale di un robot (dispositivo/macchina) come "la capacità di prendere decisioni e metterle in atto nel mondo esterno, indipendentemente da un controllo o un'influenza esterna"; precisato che "tale autonomia è di natura puramente tecnologica e il suo livello dipende dal grado di complessità con cui è stata progettata l'interazione di un robot con l'ambiente".

I primi spunti regolatori in tema di Intelligenza artificiale non possono essere considerati alla stregua di vere e proprie norme, bensì una semplice, primordiale risposta all'esigenza di **controllare** i dispositivi/macchine basati su sistemi di IA, senza reprimere o limitare le loro azioni, evitando che queste possano prendere il sopravvento sull'essere umano.

Le note "**tre leggi**" **della robotica formulate** dallo scrittore e biochimico russo Isaac Asimov rappresentano un primo tentativo volto a minimizzare il rischio di danni causati dai dispositivi/macchine dotate di IA attraverso una programmazione delle macchine compatibile con il rispetto della legge e con un codice etico⁴.

⁴ Le tre leggi della robotica sono state formulate dal biochimico e scrittore sovietico (naturalizzato statunitense) Isaac Asimov nel 1942, nello stesso racconto in cui appare per la prima volta la parola 'robotica' (*Runaround*, incluso nella raccolta *I Robot*, pubblicata nel 1950). Esse regolano il funzionamento del cervello dei robot:

1. Un robot non può recare danno agli esseri umani, né può permettere che, a causa del suo mancato intervento, gli esseri umani ricevano danno.
2. Un robot deve obbedire agli ordini impartiti dagli esseri umani, tranne nel caso che tali ordini contrastino con la Prima Legge.
3. Un robot deve salvaguardare la propria esistenza, purché ciò non contrasti con la Prima e la Seconda Legge.

In *I Robot e l'Impero* (1985), Asimov aggiunge poi la Legge Zero, che però è accettata solo dai robot più sofisticati. Questa legge è anteposta, in ordine di importanza, alle altre, permettendo una maggiore efficienza ai robot: un robot non può recare danno all'Umanità, né può permettere che, a causa del suo mancato intervento, l'Umanità riceva un danno.

Per quanto futuristiche, vaghe e poco adatte ad essere definite *tout court* delle leggi, le formulazioni di Asimov contengono importanti spunti di riflessione circa la relazione tra etica e diritto, particolarmente avvertita in materia di nuove tecnologie e di sviluppo dei sistemi IA. Numerosi sono, invero, i documenti elaborati nel contesto europeo e mondiale in tale direzione.

Tra questi si annoverano la Risoluzione del Parlamento europeo del 16 febbraio del 2017, che dedica una specifica sezione ai principi etici ed il documento "Orientamenti etici per un'IA affidabile"⁵, che promuove la realizzazione di sistemi IA "antropocentrici", ideati e strutturati quali strumenti nelle mani dell'uomo e al servizio del bene comune, per aiutarlo a migliorare il proprio stato di benessere.

La centralità dell'essere umano assume, in questa prospettiva, una prioritaria rilevanza rispetto al profilo tecnico dell'affidabilità e del buon funzionamento del dispositivo/macchina. Prospettiva, peraltro, condivisa dalla Santa Sede, come emerge dal documento "Call for AI Ethics", varato dalla Pontificia Accademia per la Vita nel 2020, ove si pone l'accento sulla necessità che sia l'uomo a guidare la tecnologia, e non viceversa.

A testimonianza della nevralgica importanza che i profili etici

⁵ L'obiettivo degli orientamenti è promuovere un'IA affidabile. Un'IA affidabile si basa su tre componenti che dovrebbero essere presenti durante l'intero ciclo di vita del sistema: a) legalità, l'IA deve ottemperare a tutte le leggi e ai regolamenti applicabili, b) eticità, l'IA deve assicurare l'adesione a principi e valori etici, e c) robustezza, dal punto di vista tecnico e sociale poiché, anche con le migliori intenzioni, i sistemi di IA possono causare danni non intenzionali. Ciascuna componente in sé è necessaria ma non sufficiente per realizzare un'IA affidabile. Idealmente le tre componenti operano armonicamente e si sovrappongono; qualora, nella pratica, si dovessero verificare tensioni tra di esse la società dovrebbe adoperarsi per risolverle.

dispiegano in materia di nuove tecnologie, non è superfluo precisare che, anche fuori dai confini europei, il dibattito scientifico si è contraddistinto per l'opportunità di stabilire delle linee guida e di indirizzo.

In California, ad esempio, la conferenza *Beneficial AI 2017*, promossa dall'Istituto Future of life, è culminata nell'approvazione di **ventitré principi di Asilomar**, i quali, benché non vincolanti, denotano l'urgenza di definire, in qualche misura, i confini di un tema complesso e sfaccettato⁶. Il profilo che accomuna i documenti e le proposte sopra citati è, all'evidenza, il timore di un uso improprio e distorto dei sistemi di IA, sebbene sia (ancora) radicata la convinzione che i dispositivi non possano essere definiti soggetti autonomi di diritti e doveri. Nella cultura giuridica occidentale non è infatti la tecnologia in sé a sollevare questioni di ordine etico e giuridico, quanto il suo (cattivo) utilizzo. Di qui la urgenza – già sottolineata dalla Risoluzione del Parlamento europeo del 2017 – di addivenire ad una standardizzazione della materia nel contesto europeo, con particolare riguardo ai profili della responsabilità, della protezione dei dati personali e della prevenzione dalla pirateria informatica. Il Parlamento europeo ha esaudito le richieste volte a delineare un quadro uniforme in materia di robotica, approvando, nell'ottobre del 2020, tre importanti Risoluzioni: **la prima**

⁶ All'esito del ciclo di incontri, l'ente ha prodotto una lista di ventitré principi, successivamente indicati come "Principi di Asilomar" dal nome del luogo dove i seminari si erano tenuti, i quali hanno raccolto a oggi più di cinquemila sottoscrizioni. Principi suddivisi in tre categorie, ossia "Questioni oggetto di ricerca", "Questioni a lungo termine" e "Etica e Valori".

(A9- 0186/2020) affronta il tema “costituzionale” dei presìdi etici che le applicazioni di Intelligenza artificiale dovranno garantire per assicurare sicurezza, trasparenza e presa di responsabilità, evitare la creazione di pregiudizi e di discriminazioni, stimolare la responsabilità sociale e ambientale ed assicurare il rispetto dei diritti fondamentali; **la seconda (A9-0178/2020) riguarda il regime della responsabilità civile per danni e pregiudizi arrecati da sistemi di IA; la terza (A9-0176/2020) concerne i diritti di proprietà intellettuale**, con particolare riferimento al rilascio di licenze ed ai processi creativi.

A conferma della nevralgica importanza che le istituzioni europee annettono al tema dell'intelligenza artificiale, considerata uno snodo cruciale nella futura trasformazione digitale della società, nel Green Deal europeo e nel rilancio dell'economia post COVID-19, il Parlamento ha, altresì, emanato una ulteriore risoluzione, avente ad oggetto “questioni relative all'interpretazione e applicazione del diritto internazionale nella misura in cui l'UE è interessata relativamente agli impieghi civili e militari e all'autorità dello Stato al di fuori dell'ambito della giustizia penale” e, insieme al Consiglio, ha avanzato una proposta di Regolamento della materia (21 aprile 2021), la quale aspira alla creazione di un mercato europeo dell'IA rispettoso dei valori dell'Unione⁷.

⁷ Sull'importanza del documento in questione, si veda, in particolare, D.Messina, *La proposta di regolamento europeo in materia di Intelligenza Artificiale: verso una “discutibile” tutela individuale di*

Tale proposta fa parte di un più ampio pacchetto di misure destinate ad affrontare i problemi posti dallo sviluppo e dall'utilizzo dell'IA in Europa. La Commissione ha, di recente, elaborato un "Piano coordinato di revisione dell'intelligenza artificiale 2021", che promuove azioni congiunte degli stati membri volte ad eliminare la frammentazione dei programmi di finanziamento, mentre Parlamento e Consiglio un regolamento in materia di macchine (21 aprile 2001), che dovrebbe sostituire la direttiva 2006/42/CE del 17 maggio 2006, che garantisce la libera circolazione delle macchine all'interno del mercato UE ed assicura un alto livello di protezione per gli utenti e altre persone esposte. Queste iniziative sono il risultato di un vivace dibattito circa la necessità di costruire un mercato europeo dell'IA affidabile, sicuro e rispettoso dei diritti fondamentali.

Uno dei temi di maggior interesse, attiene alla possibilità di prefigurare una **responsabilità civile del robot** (si pensi ai danni prodotti dai veicoli a guida autonoma), tema che pone delicati problemi interpretativi ed applicativi alla luce dei **criteri di imputazione** (colpa e dolo) e delle consolidate **categorie dogmatiche** (responsabilità aquiliana, danno ingiusto) che connotano la nostra tradizione giuridica.

tipoconsumercentric nella società dominata dal "pensiero artificiale", in Media Laws, fasc. 2, 2022, pp. 196 ss., p. 210, a giudizio del quale il c.d. "Rapporto Delvaux", dal nome della parlamentare proponente, rappresenterebbe «una chiara presa di coscienza della forza innovativa del fenomeno dell'IA ed esprime l'urgenza di adottare norme dedicate che tengano conto delle implicazioni e delle conseguenze etiche e legali derivanti da tali tecnologie senza ostacolarne, però, lo sviluppo ormai ritenuto indispensabile».

4. La responsabilità civile nell'uso dei sistemi di IA

Robot, androidi ed altre forme di Intelligenza artificiale convivono, oggi, con l'essere umano, conformando e modificando lo stile di vita di quest'ultimo.

L'inarrestabile processo tecnologico, che non è azzardato definire in termini di autentica "rivoluzione", pone il giurista dinanzi a sfide inedite e lo obbliga a fornire risposte a problemi di ordine teorico, etico e pratico.

Consapevole di questa urgenza, il Parlamento Europeo ha, a più riprese, evidenziato che i moderni robot non solo sono in grado di svolgere attività tipicamente affidate alla manualità dell'essere umano, ma possiedono anche peculiari caratteristiche cognitive come la capacità di apprendere dall'esperienza e di assumere decisioni in totale autonomia; ciò li rende sempre più connessi all'ambiente circostante ed anche, potenzialmente, **in grado di alterarlo** in modo significativo.

Tali capacità inducono, inevitabilmente, ad interrogarsi sulla risarcibilità dei danni causati a terzi per effetti della condotta, commissiva o omissiva, del robot.

A tal riguardo, la già citata Risoluzione del 2017 definisce l'autonomia di una macchina/robot come "la capacità di prendere decisioni e metterle in atto nel mondo esterno, indipendentemente da un controllo o un'influenza esterna". Da tale assunto sembrerebbe discendere, a mo' di logico corollario, che più i robot sono autonomi, meno essi

potranno essere considerati alla stregua di semplici strumenti operativi nelle mani di altri soggetti come il fabbricante, l'operante, il proprietario, l'utilizzatore, etc.

In particolare, il punto 59, lett. F) del documento prevede "l'istituzione di uno status giuridico specifico per i Robot nel lungo termine, di modo che almeno i Robot autonomi più sofisticati possano essere considerati come persone elettroniche responsabili di risarcire qualsiasi danno da loro causato nonché eventualmente il riconoscimento della personalità elettronica dei Robot che prendono decisioni autonome o che interagiscono in modo indipendente con terzi". Previsione che condurrebbe, in prospettiva, ad identificare, accanto alle tradizionali figure della persona fisica e della persona giuridica, una nuova categoria, ovvero quella della "persona elettronica", titolare di diritti e doveri. Tale previsione presta, all'evidenza, il fianco a numerosi rilievi critici.

In primo luogo, considerare i robot quali persone elettroniche postula il riconoscimento in capo ad essi di una autonomia patrimoniale perfetta, tale da consentire loro di assolvere ad eventuali obbligazioni pecuniarie e da essere passibili di procedure, anche di carattere coercitivo, volte ad ottenere il risarcimento dei danni patiti da terzi. Una simile scelta, inoltre, potrebbe incentivare una deresponsabilizzazione degli ideatori e dei produttori di IA, i quali, agendo su scala globale e speculando sulle divergenze ad oggi esistenti tra i diversi ordinamenti giuridici, potrebbero non essere chiamati a rispondere personalmente dei danni

occorsi.

Nel maggio del 2017, pochi mesi dopo la pubblicazione della Risoluzione, il Comitato economico e sociale europeo (CESE), nel parere C-238, ha espresso una serie di perplessità riguardo la proposta avanzata dal Parlamento europeo, in particolare per quanto concerne l'attribuzione della soggettività giuridica a dispositivi/macchine intelligenti. Il Comitato ha, nello specifico, espresso opinione contraria all'introduzione di una forma di personalità giuridica per i robot o i sistemi di I.A., in quanto, a suo avviso, ciò comporterebbe un rischio inaccettabile sul piano morale: "dirottare" la responsabilità civile sul robot significa disattendere la funzione di prevenzione (e di disincentivo) dei comportamenti illeciti o eccessivamente rischiosi, che compete all'illecito aquiliano.

Sulla questione si è espressa anche la Commissione degli episcopati della Comunità europea (COMECE), con un'ampia riflessione intitolata "Robotization of life–Ethics in view of new challenges" (gennaio 2019), frutto del lavoro di gruppo di esperti per i quali la costruzione di uno statuto legale per i robot non è da ritenersi attuabile, così come non è percorribile la via che porta a collocarli allo stesso livello delle persone umane in quanto ciò si porrebbe in insanabile contrasto con l'art. 6 della Dichiarazione universale sui diritti umani, che afferma l'uguaglianza di ogni persona davanti alla legge. La possibilità di estensione della personalità giuridica a macchine autonome, dotate di intelligenza artificiale, andrebbe a stravolgere il concetto di

responsabilità radicata nella personalità giuridica che può essere esercitata solo laddove esista la "capacità di libertà", che è concetto differente da quello di autonomia.

Tali censure sono state solo in parte superate dal Parlamento europeo, che ha, in qualche misura, corretto il tiro, discorrendo di una **sub-specie di personalità giuridica**, la c.d. **personalità elettronica** (o anche artificiale), ma ciò non è valso a fugare i diffusi dubbi, sì che il dibattito è, sul punto, tutt'oggi aperto sia nel contesto europeo, sia negli ordinamenti interni. La stessa Commissione europea si riserva di verificare se i tradizionali strumenti giuridici presenti nelle tradizioni dei singoli paesi membri siano adatti a regolare i rapporti civili nel mondo dell'intelligenza artificiale e si interroga sull'opportunità di introdurre nuove categorie atte a disciplinare il fenomeno.

Al pari del diritto del mare, non disciplinabile con il *nomos* della terra, è diffusa la sensazione che le tecnologie digitali mal si prestino ad essere regolate dagli istituti che presiedono alla responsabilità civile, materia che, peraltro, nel contesto europeo, presenta caratteristiche differenti da sistema a sistema. Gli interventi del legislatore comunitario sono stati, come di consueto, frammentati, avendo egli agito step by step per conferire uniformità a specifici settori: si pensi alla responsabilità dell'internet provider o alla responsabilità da prodotti difettosi (85/374/CEE). Ed è proprio alla direttiva 85/374/CEE che i commentatori hanno rivolto la propria attenzione, interrogandosi sulla possibilità di considerare il robot o il sistema basato su IA quale

“prodotto”, gravando il danneggiato dell’onere di provare il difetto del prodotto, il danno subito ed il nesso di causalità fra difetto e danno (art. 4), lasciando tuttavia impregiudicati “i diritti che il danneggiato può esercitare in base al diritto relativo alla responsabilità contrattuale o extracontrattuale o in base ad un regime speciale di responsabilità esistente al momento della notifica della direttiva” (art. 13).

È appena il caso di precisare, come si evince del resto dai “considerando” che precedono il testo della direttiva, che il regime di responsabilità del produttore prescinde dalla sua colpa. **Il criterio di imputazione del fatto illecito avviene sulla scorta di parametri oggettivi.**

La *no fault liability* è la soluzione – condivisa peraltro dalla Commissione – caldeggiata da un Gruppo di esperti sulla responsabilità delle nuove tecnologie, attraverso la quale si suggerisce un adeguamento delle disposizioni nazionali mediante l’introduzione di regole meno gravose per la vittima circa l’onere della prova⁸. In questa prospettiva, l’operatore è tenuto a dimostrare l’osservanza di cogenti regole sulla sicurezza nel processo di creazione di un sistema IA, la cui inosservanza esonera la vittima dall’onus probandi circa l’esistenza di un nesso causale tra prodotto e danno.

⁸ Nel maggio 2019 è stato pubblicato un Report di un gruppo di esperti, i cui risultati sono stati ripresi nel febbraio 2020 nella Relazione della stessa Commissione “sulle implicazioni dell’intelligenza artificiale, dell’Internet delle cose e della robotica in materia di sicurezza e responsabilità” (Commissione Europea, COM (2020) 64 final, 16 febbraio 2020, Relazione sulle implicazioni dell’intelligenza artificiale, dell’Internet delle cose e della robotica in materia di sicurezza e responsabilità), allegato al Libro Bianco “sull’intelligenza artificiale – Un approccio europeo all’eccellenza e alla fiducia” (Commissione Europea, COM (2020) 65 final, 16 febbraio 2020, Libro bianco sull’intelligenza artificiale – Un approccio europeo all’eccellenza e alla fiducia.

Dal quadro sin qui sinteticamente illustrato sembra emergere il favore per una **valutazione della responsabilità imperniata sulla capacità dell'operatore di gestire e di minimizzare il rischio, e non sul criterio psicologico della colpa.**

La questione che residua è quella della scelta dello strumento più idoneo a disciplinare la materia dello sviluppo e dell'utilizzo dei robot e dell'Intelligenza artificiale. La Commissione ha, inizialmente, optato per strumenti di soft law, quali linee guida e codici di condotta. Le sollecitazioni rivolte dal Parlamento si sono, infatti, concretizzate nella Comunicazione "Intelligenza artificiale per l'Europa", nella quale si rimarca la necessità di contemperare il mondo dell'IA con il rispetto dei diritti fondamentali dell'uomo, soprattutto alla luce dell'entrata in vigore del GDPR sulla protezione dei dati, nella Comunicazione "Creare fiducia nell'Intelligenza artificiale antropocentrica" in cui trovano accoglimento i sette requisiti fondamentali individuati dal Gruppo Indipendente di esperti: intervento e sorveglianza a cura dell'uomo; robustezza tecnica e sicurezza; riservatezza e governance dei dati; trasparenza; diversità, non discriminazione ed equità; benessere sociale ambientale nonché nel Libro bianco sull'Intelligenza artificiale.

Nell'aprile 2021, La Commissione ha, poi, presentato una proposta di regolamento volta a stabilire un quadro giuridico uniforme in materia di intelligenza artificiale ed a creare un mercato europeo per lo sviluppo, la commercializzazione e l'uso dell'intelligenza artificiale, conformemente ai valori dell'Unione. A corredo della proposta, la

Commissione ha varato un “Piano coordinato di revisione dell’intelligenza artificiale” (2021) che stimola azioni congiunte volte ad eliminare la frammentazione dei programmi di finanziamento, delle iniziative e delle azioni intraprese a livello dell’UE e dei singoli Stati membri⁹.

Ad oggi, la regolamentazione esistente, sia a livello nazionale che comunitario, in tema di responsabilità appare non in linea con il progressivo sviluppo delle ultime tecnologie in tema di IA. Per questo motivo, a parere di chi scrive è auspicabile un intervento normativo, pensato ed elaborato per regolamentare l’utilizzo dei sistemi implementati da IA. Lo sviluppo giuridico non riesce a seguire il passo di quello tecnologico come ci rammenta la prima legge di Moore “la complessità di un microcircuito [...] raddoppia ogni 18 mesi e quadruplica ogni 3 anni”¹⁰, oggi si è in grado di costruire un processore con una potenza raddoppiata rispetto ad uno costruito solamente un anno e mezzo prima; pertanto, il legislatore non dovrà commettere di basarsi esclusivamente sul presente.

⁹ Al momento non esiste una versione ufficiale italiana della proposta per cui il titolo è stato liberamente tradotto dall’inglese. Il titolo originario è “the Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts”.

¹⁰ In elettronica e informatica, è indicato come prima Legge di Moore il seguente enunciato: “La complessità di un microcircuito, misurata ad esempio tramite il numero di transistor per chip, raddoppia ogni 18 mesi (e quadruplica quindi ogni 3 anni).” La legge è tratta da un’osservazione empirica di David House, direttore esecutivo di Intel, commentando la precedente osservazione di Gordon Moore, cofondatore di Intel con Robert Noyce: nel 1965, Gordon Moore, che all’epoca era a capo del settore R&D della Fairchild Semiconductor e tre anni dopo fondò la Intel, scrisse infatti un articolo su una rivista specializzata nel quale illustrava come nel periodo 1959-1965 il numero di componenti elettronici (ad esempio i transistor) che formano un chip fosse raddoppiato ogni anno[2]. Moore, grazie alle sue supposizioni poi diventate leggi e conosciute come prima e seconda legge di Moore, è stato dunque tra coloro che hanno dato il via alla corsa all’evoluzione dei processori.

Indubbiamente non è possibile prevedere quanto impiegherà l'IA a superare il test di Turing¹¹, né tra quanto tempo essa farà completamente parte della nostra vita, ma è lecito pensare che ciò accadrà e il diritto avrà l'onere di farsi trovare pronto.

¹¹ Il test di Turing è, da sempre, il riferimento utilizzato per valutare se un'Intelligenza Artificiale esponga comportamenti analoghi a quelli di una intelligenza umana o meno. Tuttavia, il test, sviluppato dal matematico Alan Turing, nasceva negli anni '50 per interrogare sistemi con capacità nemmeno lontanamente comparabili a quelli disponibili attualmente. Mustafa Suleyman, co-fondatore della startup specializzata in intelligenza artificiale DeepMind, ha espresso forti dubbi in merito all'attualità del test di Turing per valutare i sistemi basati su AI.

Suleyman ha motivato le sue perplessità ritenendo il test non più efficace per differenziare i vari sistemi che, infatti, ultimamente superano tutti agilmente il test. Suleyman propone un approccio differente da quello previsto nel test di Turing nel quale si valutava quanto la capacità discorsiva del sistema fosse difforme da quella umana. L'approccio di Mustafa è basato sul porre al sistema un problema talmente complesso da costringerlo a mettere in atto tutta una serie di ragionamenti, legati al mondo reale, che sono tipici delle valutazioni umane.

5. IA e diritto del lavoro

Come si è visto nei paragrafi che precedono, i sistemi intelligenti coadiuvano l'essere umano nel prendere decisioni, tanto da poterlo addirittura sostituire in casi sempre più numerosi: è, questo, un tratto che caratterizza la Società algoritmica.

Ciò non significa, però, che vi siano solo aspetti positivi: tutt'altro.

Pur non essendo questa la sede per effettuare considerazioni generali sulle problematiche che scaturiscono dalla pervasiva diffusione dell'intelligenza artificiale nella società contemporanea, può osservarsi che un aspetto negativo particolarmente rilevante è costituito dalla sua opacità che può giungere alla moltiplicazione di sistemi che possono essere considerati delle vere e proprie "scatole nere", in cui i flussi informativi viaggiano sovente all'insaputa degli interessati grazie ad algoritmi segreti. Ciò rileva, ai fini del presente scritto, in relazione a due aspetti principali.

In primo luogo, chi crea gli algoritmi (e li implementa nei sistemi intelligenti che poi fornisce ai propri clienti e utenti) acquisisce un notevole potere economico che può prosperare grazie alla confidenzialità degli algoritmi e dei codici informatici, che a loro volta sono protetti dalle normative in materia di proprietà intellettuale e industriale (che impedisce accessi e utilizzi non autorizzati), da un lato, e dalla stessa complessità dei sistemi informatici (che rende arduo comprenderne in modo preciso e accurato i "comportamenti"),

dall'altro. In secondo luogo, chi adopera detti sistemi - ancorché forniti da terzi - per delegare parzialmente o totalmente l'esercizio del potere datoriale, gestisce i relativi flussi informativi potendo godere non solo di un controllo più o meno intenso sul sistema di cui trattasi, ma altresì della confidenzialità intrinseca proprio a tali sistemi: spesso è di fatto impossibile, o è molto complesso e costoso, andare a ritroso e ricostruire effettivamente (ossia in quel sistema informatico in un determinato arco temporale) le direttive fornite dal datore al sistema intelligente in un determinato momento, così da comprendere la logica sottostante alle decisioni prese: e, in caso di controversia giudiziale, le relative tempistiche non aiutano di certo.

Su tale questione si tornerà più compiutamente nel prosieguo, ma appare opportuno ora evidenziare l'attualità di quanto acutamente osservato da **Stefano Rodotà** circa la nuova forma di arcana imperii che ha preso forma in modo tristemente paradossale: da un lato, le tecnologie dell'informazione e della comunicazione rendono la società più trasparente poiché permettono controlli diffusi su qualsiasi potere, ma gli algoritmi che fondano a loro volta il potere dei soggetti che prestano i relativi servizi tecnologici rimangono segreti, dall'altro¹².

¹² Il saggio di Stefano Rodotà intitolato Il mondo nella rete. Quali i diritti, quali i vincoli, pubblicato di recente da Laterza in collaborazione con "La Repubblica", è composto da una prima parte che riprende il XIV capitolo di un altro libro di Rodotà, Il diritto di avere diritti, e da una seconda parte che contiene interviste già pubblicate su quotidiani e settimanali fatte a personaggi cardine del mondo digitale. Il prologo cita John Perry Barlow e la sua "Dichiarazione di Indipendenza del Cyberspazio", risalente al 1996 ma ancora molto attuale: è vero che Internet non ha sovrano? Che è "invincibile" per natura? Che è libertaria sino all'anarchia? Che ormai costituisce una rete di comunicazioni che nessuno potrebbe bloccare o controllare? Il problema, nota Rodotà, è che, nel corso di questi ultimi vent'anni, la storia spesso ha dimostrato il contrario, soprattutto con riferimento all'ingerenza normativa degli Stati sul

Questo fenomeno assume, oggi, nuove valenze: alla segretezza degli algoritmi del fornitore dei sistemi si aggiunge, potenzialmente, quella delle "indicazioni" fornite dal loro utilizzatore-datore di lavoro e, ciò che

tessuto digitale. Da un lato, scrive l'Autore, si parla di una "sovranità improponibile", dall'altro si critica un "potere invadente". Può esistere, allora, una via di mezzo? Una regolamentazione, sì, ma che sia rispettosa dei diritti anche in un ambiente così particolare, e per molti versi fragile, come il cyberspazio? La prima parte dello scritto si apre con una stretta, e molto opportuna, connessione al mondo della politica: democrazia e cittadinanza digitale sono i due argomenti affrontati da Rodotà, collegati al riconoscimento (anche costituzionale) di un diritto all'accesso a Internet. Si richiama, allora, il mito dell'agorà di Atene, che rinasce nel mondo moderno e in un villaggio globale e piazza virtuale che possono ricostruire il mito della democrazia diretta e, addirittura, potrebbero salvare la morente democrazia rappresentativa per traghettarla verso una democrazia immediata capace di consultare in maniera costante e permanente i cittadini. Ma non tutto è così semplice: un simile quadro, e la considerazione dei diritti fondamentali dell'individuo e del loro valore, portano l'Autore a una riflessione sul concetto di cittadinanza digitale, ossia su quelle situazioni che descrivono la condizione di una persona nel mondo digitale. Lo studio prosegue con riflessioni su neutralità e anonimato, e sul patrimonio di conoscenza in rete da intendersi quale bene pubblico globale. La neutralità della rete si fonda sull'eguaglianza, ed è interpretata come il divieto di ogni discriminazione riguardante i dati e il traffico su Internet che sia basata sul mezzo utilizzato, sulla tecnologia, sui contenuti, sui servizi e sulle caratteristiche delle persone. Il timore che aleggia è, appare chiaro, quello di una "censura di mercato" che arrivi anche a impedire l'accesso alle risorse digitali. Interessante è il passaggio che descrive i concetti di "habeas corpus" e di "habeas data" e che contiene, sul punto del diritto alla privacy, considerazioni importanti. Qui la riflessione si relaziona al delicatissimo tema della trasparenza degli algoritmi e della società e del diritto all'oblio, oblio che contrasta visibilmente con l'estrema tracciabilità presente nella società odierna e con una sorta di "obbligo del ricordo". La parte finale dello scritto dello studioso affronta il problema delle nuove forme di redistribuzione del potere (con l'avvento dirompente della trasparenza radicale e di un processo di controllo diffuso sull'esercizio del potere, sino allo smantellamento degli *arcana imperii*), della "qualità" della democrazia elettronica (siamo certi che la democrazia elettronica porti automaticamente a un'espansione del potere del cittadino o, al contrario, può nascondere rischi di populismo e di confusione sino a portare a rigurgiti di totalitarismo?) e della necessità di una proposta di un *Internet Bill of Rights*, una carta dei diritti fondamentali come rimodulati (o riconosciuti) nell'ambiente digitale. Terminato il saggio di Rodotà, la seconda parte del libro si apre con la traduzione di un'intervista a Edward Snowden, l'ex consulente CIA che ha svelato al mondo, nel giugno del 2003, migliaia di documenti sulle attività di controllo della NSA e ha dato origine al fenomeno del cosiddetto Datagate. Una seconda intervista, sempre sul tema, coinvolge due senior executive di NSA che accennano, in maniera molto cauta, alle attività di controllo dell'agenzia. Seguono le opinioni di Vinton Cerf, uno dei padri di Internet e ora executive di Google, sullo stato attuale del controllo: Cerf è molto ottimista, e sostiene che non esistano "porte segrete" attraverso le quali siamo controllati, anche se riconosce il delicato rapporto tra privacy e sicurezza. Profilo interessante, che è presentato subito dopo, è quello di John Young di CryptoMe, sempre sul tema del Datagate. Young afferma di non credere alla versione ufficiale per cui Obama non sarebbe stato a conoscenza delle attività di spionaggio della NSA, ma che anche lui sarebbe da tempo entrato a pieno titolo nel "mondo della segretezza" che è essenza stessa del potere. Chiudono il breve saggio un appello firmato da alcuni intellettuali contro la sorveglianza, nelle righe del quale si domanda un ritorno alla inviolabile integrità dell'individuo e delle sue comunicazioni e al diritto di non essere osservati e disturbati nei pensieri, nell'ambiente personale e nelle comunicazioni. Una persona sotto sorveglianza, si sostiene, non è più libera, così come una società sotto sorveglianza non è più una democrazia. Le ultime pagine riportano un elenco di dati e alcune mappe: particolarmente utili quella sul livello di libertà in Internet (che suddivide i Paesi per intensità di regolamentazione e "inimicizia" nei confronti del mezzo elettronico), quella sugli attacchi di Anonymous, quella che raffigura i Paesi spiati dalla NSA (sono 37 gli Stati sui quali si sono avute notizie documentate rispetto ad attività di intercettazione delle comunicazioni da parte degli USA) e una cronologia finale della rete.

più conta, dei dati effettivamente trattati dal sistema al fine di prendere le proprie decisioni. In argomento, nonostante alle specifiche disposizioni lavoristiche si affianchino quelle in materia di protezione dei dati personali (sostanzialmente il Regolamento UE n. 679/2016, Regolamento Generale sulla Protezione dei Dati – di seguito GDPR, e il D.lgs. 196/2003, Codice in materia di protezione dei dati personali), l'opacità è pressoché assoluta: non è normalmente possibile sapere quali e quante informazioni siano effettivamente acquisite ed elaborate, né come lo siano o controllarne realmente la circolazione (che hanno però un impatto settoriale e limitato).

Ciò comporta la creazione di un circolo vizioso in assenza, a tutt'oggi, di una reale regolamentazione dell'intelligenza artificiale e delle sue applicazioni, cui consegue un potenziamento dell'opacità. Solo negli ultimi anni i legislatori si stanno risvegliando dal loro torpore: troppo poco, troppo tardi? In ogni caso, mentre le proposte vengono discusse, la tecnologia avanza incessantemente.

Si ripropone, ancora una volta e in altre modalità, il problema delle **norme opache**, che vengono applicate dal giudice senza poterle realmente comprendere ma facendo invece riferimento a ciò che viene detto dagli esperti. In simili ipotesi, l'opacità non è dovuta al tecnicismo in sé, bensì all'**asimmetria cognitiva** fra i legislatori e i giudici, da un lato, e gli esperti, dall'altro, poiché quando il linguaggio tecnico non giuridico è comprensibile dai primi, il diritto non è opaco ed

essi operano nell'ambito dei propri ruoli. Qualora ciò non avvenga, essi non possono compiere una serie di attività intellettuali, di scelte pratiche e di ragionamenti.

La summenzionata asimmetria cognitiva è massima in riferimento ai sistemi intelligenti, per la loro stessa natura. Essa riguarda tutti i loro utilizzatori, siano essi diretti (chi si fa coadiuvare da un sistema, come il datore di lavoro, salvo che non abbia sviluppato internamente il sistema stesso o comunque abbia commissionato *ad hoc* lo sviluppo acquisendo altresì i diritti sul codice sorgente) o indiretti (chi subisce gli effetti dell'utilizzo del sistema, come lavoratori e lavoratrici nel caso che ci occupa): essi sono opachi perché non comprensibili al di fuori della cerchia di chi li ha creati e li gestisce.

Solo il loro output e/o le conseguenze delle loro azioni (materiali o immateriali) vengono percepite, ma operando in ambienti complessi in cui avvengono infinite interazioni diviene estremamente difficile, e talvolta impossibile, comprendere se in essi vi siano anomalie o se essi compiano azioni riferibili ai propri produttori od utilizzatori; in ipotesi, tali informazioni potrebbero essere ricavate soprattutto dallo studio dei rispettivi codici e dei file di log.

Una precisazione appare doverosa: anche chi detiene il potere sui sistemi medesimi può faticare nel comprenderli realmente e renderli trasparenti. Ciò è dovuto al fatto, da un lato, che sistemi tanto evoluti sono oltremodo complessi (del resto svolgono compiti che richiedono intelligenza) e, dall'altro, che operano in modo autonomo, applicando

le conoscenze apprese anche automaticamente sulla base degli algoritmi di apprendimento automatico: anche grazie alla loro crescente "intelligenza", le loro azioni sono imprevedibili dal punto di vista sia teorico sia pratico. I rilievi sin qui svolti fanno intuire l'importanza del superamento, anche entro limiti comunque discutibili (ad esempio, solo al verificarsi di determinate condizioni e cercando un punto di equilibrio rispetto alle legittime pretese dei fornitori e degli utilizzatori dei sistemi intelligenti), dell'opacità: ciò, però, può avvenire esclusivamente mediante una regolamentazione che sia non solo formalmente valida, ma anche efficace – il che richiede, però, di volgere in modo deciso lo sguardo anche alle problematiche tecniche che scaturiscono dalla sua eventuale applicazione pratica, fra cui la complessità dei sistemi informatici che potrebbero dover essere sottoposti a verifica.

Sino ad oggi è indubbio che, nella prospettiva del legislatore, un ambito della società tanto trasversale e importante sia stato lasciato a una tecno-regolazione autoreferenziale, in cui la legge è il codice informatico (per usare la celebre espressione di Lawrence Lessig e andando verso la c.d. *lex informatica*) e il legislatore chi lo crea e lo gestisce (con prerogative più ampie di un legislatore nazionale, avendo un dominio sostanzialmente assoluto sullo spazio cibernetico che va a creare e gestire). Il grado di opacità non pare ridursi neanche nei casi in cui il potere tecno-economico debba cedere e adeguarsi alle decisioni di quello giudiziario, come nei celebri casi Google Spagna e

Schrems I e II ; non abbiamo ancora, infatti, strumenti tecnici che consentano di accertare che determinate operazioni vengano effettivamente svolte su sistemi informatici di elevatissima complessità e magari dislocati in stati diversi. Si consideri che mentre in alcuni casi le azioni dell'agente di un sistema sono percepibili esternamente solo dopo essere state compiute, in altri possono non esserlo: basti pensare alla sorveglianza che viene effettuata non solo e non tanto con "vistose" videocamere quanto, piuttosto, sfruttando funzionalità proprie di dispositivi e servizi di uso comune, dagli smartphone agli autoveicoli, dai computer alle carte di pagamento, dalla videoconferenza alla posta elettronica, e così via.

Fra i vari ambiti da prendere in considerazione, dunque, quello lavoristico appare – intuitivamente – assai delicato.

In ambito lavoristico (e non solo), una simile sorveglianza, poi, può essere particolarmente delicata: ad esempio, come è possibile accertarsi che non venga monitorata la condotta di un dipendente durante le videoconferenze grazie al ricorso a tecniche di intelligenza artificiale, magari al fine di calcolarne la produttività in base a un trattamento automatizzato e imperscrutabile ? In tal senso, la sempre maggiore spinta verso il *cloud computing* rende sempre più difficile giungere a rendere concretamente esercitabile un eventuale diritto alla spiegabilità delle "azioni" dei sistemi intelligenti, oltre che effettivamente accertabile il rispetto della vigente normativa in materia di protezione dei dati personali (v. altresì infra, par. 5): bisognerebbe

forse effettuare una verifica sulla versione specifica del relativo software adoperato in un determinato momento? E come contrastare eventuali fenomeni di elusione o violazione anche di una simile normativa? Come dimostra la prassi, oltretutto, ben difficilmente un giudice imporrà di verificare il codice utilizzato nella prestazione di un servizio reso su scala globale, in quanto basato su segreti industriali e utilizzato per fornire servizi a milioni o addirittura a miliardi di utenti; ciò potrebbe in ipotesi essere una conseguenza della sproporzione fra il valore della causa (e del diritto tutelato), da un lato, e della proprietà industriale e intellettuale, dall'altro, cui si aggiunge la (im)possibilità di raggiungere un risultato accettabile (dal punto di vista probatorio) con minor sforzo e rischi per tali diritti. In tal modo, la tecnica prevale sul diritto e ambito cruciale della società contemporanea viene celato allo sguardo di terzi, inclusi legislatori e giudici. Così, l'opacità potrebbe essere uno strumento per nascondere eventuali trattamenti illeciti di dati personali, atti di discriminazione, anomalie del software, falle di sicurezza, e così via, dal momento che è praticamente possibile, o comunque assai difficile, ricostruire ciò che sia realmente accaduto in un determinato sistema in un momento specifico: il che può assumere rilevanza per valutare la legittimità dell'esercizio del potere direttivo, di controllo e disciplinare del datore di lavoro. Invero, già nel considerando n. 71 del GDPR si legge che un trattamento automatizzato che produce effetti giuridici (o analoghi) sull'interessato (nei casi in cui viene consentito), "dovrebbe essere subordinato a

garanzie adeguate, che dovrebbero comprendere la specifica informazione all'interessato e il diritto [...] di ottenere una spiegazione della decisione conseguita dopo tale valutazione e di contestare la decisione". Sul punto, il Gruppo di lavoro articolo 29 per la protezione dei dati ha osservato che il titolare dovrebbe comunicare in modo semplice all'interessato la logica o i criteri sui quali si basa l'adozione della decisione, "ma non necessariamente una spiegazione complessa degli algoritmi utilizzati o la divulgazione dell'algoritmo completo" .

Del resto, la trasparenza potrebbe essere illusoria anche qualora si potesse visionare tutto il codice informatico adoperato durante i processi decisionali, poiché sarebbe difficilmente comprensibile, anche ove dovesse esserne consentito l'accesso a chi ha il controllo sul medesimo (cui si aggiunge la difficoltà di cristallizzarne una versione a causa di costanti modifiche al codice medesimo).

Anche spiegazioni fornite sulla scorta di quanto osservato dal Gruppo di lavoro articolo 29, ad ogni buon conto, potrebbero essere insufficienti ove non dovesse essere possibile verificarne l'implementazione in un caso concreto. Ragionando in senso contrario, il creatore (o gestore) degli algoritmi godrebbe di una presunzione assoluta di buona fede previo adempimento di un'astratta obbligazione di *disclosure*, cui conseguirebbe una sostanziale irresponsabilità (fatti salvi casi estremi in cui sia possibile dedurre l'inadempimento dall'output del sistema o della piattaforma nel caso in cui dovesse essere palesemente in contrasto con quanto dichiarato).

Negli ultimi anni, però, il quadro sta mutando; la letteratura in materia è oramai sterminata così come i dibattiti e le discussioni a livello della società intera .

Ciò nonostante, i legislatori sono in notevole ritardo: i proclami, i lavori delle commissioni, le raccomandazioni, e così via, non si sono ancora colpevolmente tradotti in testi normativi che, ad oggi, effettivamente e concretamente impongano una evoluzione che si potrebbe definire costituzionalmente orientata della Società algoritmica, che viene ora guidata in modo quasi totalmente esclusivo dai poteri privati a scapito di quelli pubblici. Già le procedure (e i pesi e i contrappesi) che regolano l'azione di quest'ultimi, del resto, appaiono basate su concezioni calate in realtà in cui l'evoluzione avanzava in modo rapido ma non incessantemente vorticoso come accaduto dopo la diffusione del web e il passaggio dalla Società dell'informazione a quella algoritmica.

Già nella **fase prodromica del rapporto lavorativo**, l'I.A. sta acquisendo una sempre maggiore importanza: grande sviluppo sta, infatti, avendo l'*algorithmic hiring*, inteso quale procedura di selezione del personale integralmente o parzialmente affidata ad algoritmi.

La percezione diffusa è che tali procedure automatizzate siano più rapide, affidabili ed economiche rispetto alle selezioni "canoniche", consentendo di individuare efficacemente le caratteristiche e le attitudini personali dei candidati tramite l'analisi di una grande mole di dati raccolti durante le interviste virtuali.

Se da un lato l'I.A. rappresenta una grande opportunità, dall'altro, quando non è adeguatamente controllata, può essere influenzata da una problematica insidiosa, ovvero il pregiudizio umano che si riflette inevitabilmente sugli algoritmi. Richiamando l'A.I. Act sopra citato, sono infatti considerati ad "Alto Rischio":

- i sistemi di AI per lo screening dei candidati;
- la formulazione di classifiche e graduatorie;
- i sistemi di matching;
- i sistemi che supportano la valutazione del candidato nel corso di colloqui o test.

Con riferimento ai rischi connessi con l'utilizzo dell'intelligenza artificiale nel contesto lavorativo, è stato infatti rilevato che "durante tutto il processo di assunzione, nonché ai fini della valutazione e della promozione delle persone o del proseguimento dei rapporti contrattuali legati al lavoro, tali sistemi possono perpetuare modelli storici di discriminazione, ad esempio nei confronti delle donne, di talune fasce di età, delle persone con disabilità o delle persone aventi determinate origini razziali o etniche o un determinato orientamento sessuale. I sistemi di IA utilizzati per monitorare le prestazioni e il comportamento di tali persone possono inoltre incidere sui loro diritti in materia di protezione dei dati e vita privata"¹³.

¹³ Cfr. R. Santucci, La quarta rivoluzione industriale e il controllo a distanza dei lavoratori, in *Lav. giur.*, 1, 2021, p. 19.

In base alle modalità di costruzione del *software*, anche l'azienda che non abbia finalità discriminatorie, potrebbe inconsapevolmente introdurre c.d. *bias* nel processo di trattamento, che, con un effetto a catena, condizionerebbero gli esiti del processo, con effetti discriminatori.

Ciò in quanto i *software*, per quanto possano essere artificialmente intelligenti, vengono comunque programmati da esseri umani e risentono quindi delle dinamiche giudicanti dei loro stessi programmatori.

A ciò aggiungasi che i dati inseriti nei *software* rimangono memorizzati all'interno del programma condizionando le analisi predittive future che risulteranno influenzate da dati non aggiornati.

Interessante ricordare, a tal proposito, il noto caso di Amazon. Il famoso colosso statunitense aveva sviluppato un programma sperimentale di *talent finding* automatizzato con lo scopo di valutare i candidati secondo una scala di punteggio graduale. Tuttavia, con specifico riferimento a ruoli IT, il sistema non selezionava le candidature in modo neutrale rispetto al genere: le figure femminili venivano escluse automaticamente. La ragione era dovuta al fatto che il *software* si basava su dati raccolti negli ultimi 10 anni e la maggior parte delle risorse assunte in tale arco temporale in ambito informatico erano, appunto, di genere maschile.

Gli algoritmi hanno quindi individuato e messo in luce i pregiudizi dei loro stessi creatori, dimostrando così che l'addestramento dei sistemi automatizzati su dati imparziali porta a future decisioni non neutrali.

Il caso di Amazon offre un interessante spunto di riflessione sui limiti dell'apprendimento dell'Intelligenza Artificiale e su quanto i c.d. *bias* umani possano riflettersi sui sistemi automatici, condizionandone gli algoritmi.

Oltre alla fase pre-assuntiva, i sistemi di I.A rappresentano un fattore importante altresì nell'**organizzazione del lavoro**: si pensi, ad esempio, ai sistemi per la gestione della logistica nei magazzini nonché alle piattaforme utilizzate per la gestione dei *riders*.

In questi settori, le decisioni in merito alla migliore gestione delle attività e delle risorse umane è sempre più spesso demandata ad algoritmi, in grado di analizzare un'infinita quantità di dati e di individuare la soluzione gestionale ed organizzativa più efficace: algoritmi che determinano l'assegnazione di mansioni in base a determinati parametri, sistemi automatizzati di monitoraggio, sistemi di geolocalizzazione che prevedano segnalazioni o interventi automatici in caso di pericolo.

In tale contesto lavorativo in rapida evoluzione, l'Unione Europea ha sottolineato l'esigenza che i lavoratori siano pienamente e tempestivamente informati in merito alle condizioni essenziali del loro lavoro.

Al fine di garantire al lavoratore e alle organizzazioni sindacali una conoscenza dei sistemi digitali nelle singole organizzazioni imprenditoriali, il legislatore, recependo nell'ordinamento interno la Direttiva (UE) 2019/1152 relativa a condizioni di lavoro trasparenti e prevedibili, ha introdotto a carico del datore di lavoro un obbligo di informativa relativo al caso di utilizzo di sistemi decisionali o di monitoraggio automatizzati (art. 1-bis del D.lgs. n. 152/1997 introdotto dal c.d. Decreto Trasparenza, D.Lgs. 104/2022).

Lo scopo della novella legislativa è stato quello, come si evince dalla lettura delle premesse e dell'art. 1 della Direttiva UE, di «migliorare le condizioni di lavoro promuovendo un'occupazione più trasparente e prevedibile, pur garantendo nel contempo l'adattabilità del mercato del lavoro».

Una traduzione in termini pratici di un linguaggio a tratti ostico è che il lavoratore deve poter conoscere se si usano le tecniche automatizzate, se il datore di lavoro si avvale di decisioni algoritmiche e simili; inoltre, il lavoratore ha diritto di sapere come tali tecniche funzionano, quale ne sia la logica e quale gli impatti, anche in termini di rischi per la sicurezza dei dati personali.

Da una lettura combinata dell'art. 1, co. 1, lett. s) e dell'art. 1-bis, co. 1 del D.lgs. 152/1997, si evince che la predisposizione di tale specifica informativa è richiesta nel caso in cui le modalità di esecuzione della prestazione dei lavoratori siano organizzate tramite l'utilizzo di sistemi

decisionali e/o di monitoraggio automatizzati, destinati a «fornire indicazioni rilevanti ai fini dell'assunzione o del conferimento dell'incarico della gestione o della cessazione del rapporto di lavoro, dell'assegnazione di compiti o mansioni nonché indicazioni incidenti sulla sorveglianza, la valutazione, le prestazioni e l'adempimento delle obbligazioni contrattuali dei lavoratori».

La portata della norma contenuta nell'art. 1-bis del Decreto Trasparenza ha creato dubbi interpretativi e difficoltà applicative relativi all'individuazione di quali sistemi fossero da includere tra quelli soggetti a tale ulteriore informativa da distinguersi dagli strumenti di controllo a distanza, rispetto ai quali gli obblighi informativi sono viceversa regolati, come ampiamente noto, dall'art. 4 della L. n. 300/1970, ossia da una disposizione fatta espressamente salva dalla novella e che sembra mantenere un suo grado di autonomia.

Con riferimento alle tipologie di strumenti da intendersi quali sistemi automatizzati, la Circolare del Ministero del Lavoro e delle Politiche Sociali n. 19/2022 ha tentato di fornire alcune precisazioni sulle novità introdotte dal D.lgs. 104/2022¹⁴.

In particolare, la Circolare ha escluso l'obbligo informativo nel caso di utilizzo di badge, ovvero di strumenti automatizzati per la rilevazione delle presenze dei dipendenti in entrata o in uscita, sempre che tale registrazione non generi automaticamente una decisione datoriale, mentre, a titolo puramente esemplificativo ma non esaustivo, ha previsto tale obbligo nel caso di utilizzo di sistemi automatizzati di

gestione dei turni, di determinazione della retribuzione, di tablet, GPS, *wearables* e altro.

6. Problemi e prospettive

Volgendo alla parte finale del presente lavoro, s'intendono prospettare alcune ipotesi di ricerca circa la relazione tra tecnologia e fattore umano, in particolare per quanto concerne la tensione tra la facoltà decisionale umana ed il processo decisionale automatizzato fornito dall'Intelligenza Artificiale.

L'articolo 14 del **c.d. Artificial Intelligence Act europeo** (rubricato "Sorveglianza umana") fornisce importanti indicazioni con specifico segno al ruolo del fattore umano nell'ambito dei processi decisionali dell'IA.

In particolare, posto in rilievo che i sistemi di IA "ad alto rischio" dovrebbero essere "progettati e sviluppati, anche con strumenti di interfaccia uomo-macchina adeguati, in modo tale da poter essere efficacemente supervisionati da persone fisiche durante il periodo in cui il sistema di IA è in uso", si specifica come la sorveglianza umana sia diretta a "prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali che possono emergere quando un sistema di IA ad alto rischio è utilizzato conformemente alla sua finalità prevista o in condizioni di uso improprio ragionevolmente prevedibile".¹⁴

In tal guisa, seppur supportati da strumenti di Intelligenza Artificiale, i

¹⁴ Sul punto ci si permette di rimandare a G. Mosa, Guida alle regole sulla trasparenza nei contratti individuali di lavoro, in Quaderni Unsic, n. 13/2022, Roma, 2022, p. 8 ss.

processi decisionali dovrebbero essere caratterizzati dalla **presenza costante della supervisione umana**.

Tale supervisione, a ben vedere, dovrebbe essere condotta da un gruppo di specialisti di settori disciplinari diversi e variegati (compresi, evidentemente, anche esperti di area giuridica), al fine di compiere una valutazione complessiva e multidisciplinare, non solo con riferimento alla decisione finale, bensì con riguardo all'intero processo di elaborazione.

Parimenti, l'implementazione degli strumenti di IA nella vita pubblica e sociale richiede una regolamentazione rigorosa, soprattutto a livello europeo, per garantire un uso equo, democratico, trasparente e legale di tali strumenti, nonché un adeguato controllo sulle attività implementate e sui risultati ottenuti per mezzo dell'Intelligenza Artificiale.

Nondimeno, l'uso dei sistemi di IA, seppur ampiamente diffuso nel settore privato (si pensi, a mo' di esempio, agli algoritmi che "governano" il funzionamento dei Social Networks, sovente suggerendo "contenuti personalizzati" ai singoli utenti, sulla base delle "ricerche" effettuate o dei contenuti maggiormente "visualizzati"), dovrebbe essere adeguatamente **regolato da parte dei pubblici poteri**, al fine di evitare che detta tecnologia possa trasformarsi in un mero volano di potenziali speculazioni finanziarie o, per converso, in uno strumento di monetizzazione dei sistemi predittivi, ai fini di un arricchimento privato o di un uso improprio dei dati connessi all'utilizzo di IA.

Si pensi, in tal guisa, a svariati ambiti di interesse e criticità, tanto economico-finanziarie di marketing pubblicitario, quanto afferenti al piano democratico e sociale, quali, ad esempio, il *nudging*, il Social Credit System o, sotto altro versante, il c.d. micro-targeting elettorale, la profilazione degli individui in funzione della selezione delle informazioni da porre "in evidenza" all'utente e, in estrema ipotesi, taluni elementi di "democrazia predittiva".

Tuttavia, sul punto, altra parte della dottrina ha inteso evidenziare i potenziali effettivi positivi dell'implementazione dei sistemi di Intelligenza Artificiale nel quadro ordinamentale di matrice democratica, specie laddove detti strumenti possano contribuire al miglioramento del processo decisionale pubblico, contrastando così, per conseguenza, anche la potenziale "minaccia" rappresentata dalle formazioni populiste e, più in generale, da tendenze distorsive di una corretta dinamica democratica.

In sintesi, l'utilizzo dei sistemi di Intelligenza Artificiale nella vita privata e pubblica comporta evidenti vantaggi, tra cui l'abbattimento delle barriere fisiche, la reattività delle risposte a sollecitazioni fisiche o telematiche provenienti dall'esterno, nonché una potenziale implementazione di principi quali l'efficienza e la trasparenza nei processi decisionali pubblici.

Al contempo, l'implementazione dell'uso dell'IA nella vita privata e pubblica potrebbe essere foriera di potenziali rischi, quali, ad esempio, l'accrescimento delle diseguaglianze basate sul livello di sviluppo

scientifico e culturale in determinate aree geografiche, l'assenza del fattore umano nell'elaborazione delle decisioni pubbliche ed una potenziale deriva tecnocratica delle istituzioni rappresentative.

In un contesto irto di insidie sotto molteplici angoli visuali, il ruolo regolatorio delle istituzioni pubbliche, involgente altresì la protezione dei diritti e delle libertà dei cittadini (specie nei riguardi delle categorie "maggiormente esposte" ad un potenziale abuso attraverso le tecnologie di IA più avanzate) appare irrinunciabile e, al contempo, doveroso.

In sostanza, ai fini dell'interesse pubblico ed in considerazione della necessità di affrontare il cambiamento climatico e numerose altre emergenze pubbliche (sovente di portata globale), l'Intelligenza Artificiale non dovrebbe essere considerata esclusivamente o primariamente come un "bene di mercato", bensì dovrebbe, per converso, essere qualificata come **un concreto e reale "bene pubblico"**, inteso quale strumento attraverso il quale possano essere attuate le misure necessarie ed efficaci che siano richieste dalla gravità delle condizioni di emergenza. Il riconoscimento dell'IA quale bene pubblico non dovrebbe considerarsi un profilo di secondario momento, attesa l'ampia implementazione di questa tecnologia nella società globalizzata.

Pertanto, le istituzioni pubbliche potrebbero preservare, ottenere e condividere i risultati raggiunti e le misure necessarie per affrontare le questioni più rilevanti (come il cambiamento climatico) solo attraverso

un processo di valutazione dell'Intelligenza Artificiale che sia connotato, quanto a meccanismi di funzionamento e risultati attesi, da principi quali trasparenza, responsabilità e democraticità.

In sostanza, volendosi seguire le prospettive evidenziate in dottrina, con riferimento alla suddetta prospettiva "antropocentrica", il tema in discussione non dovrebbe riguardare esclusivamente l'inserimento della "tecnologia all'interno delle Carte costituzionali", bensì la trasposizione dei "valori costituzionali e la centralità della persona umana all'interno dei sistemi algoritmici".

D'altro canto, appare necessario sviluppare un'ulteriore riflessione, legata al tema della "sorveglianza umana" dei processi decisionali artificiali e al connesso "potere di deroga" umano rispetto a decisioni o suggerimenti o previsioni forniti dagli strumenti di IA.

In linea generale, è possibile accettare un risultato fornito dall'Intelligenza Artificiale in assenza di un successivo controllo umano su tale decisione?

Sarebbe possibile derogare, in caso di procedimento interamente "sorvegliato" da parte umana, ai risultati conseguiti al termine del processo elaborativo dell'IA? E, in caso di deroga, su quali basi ciò potrebbe avvenire? Il potere discrezionale di un decisore pubblico può essere considerato proporzionato ed adeguato al fine di "aggirare" una decisione dell'IA, sulla base di alcuni criteri quali l'approccio scientifico, l'efficienza, la trasparenza e l'imparzialità?

È necessario, in tal guisa, sviluppare un'analisi strategica sulle potenzialità dell'IA e sul ruolo delle istituzioni pubbliche, sulla base del postulato necessario della sorveglianza e della supervisione umana.

In tale prospettiva, il ruolo dell'Intelligenza Artificiale appare chiaro e, nondimeno, potenzialmente positivo ma, allo stesso tempo, la capacità della pubblica amministrazione e della politica di **bilanciare diritti ed esigenze** promananti dalla collettività si pone alla base degli ordinamenti giuridici contemporanei, quantomeno nel campo occidentale. In tal senso, la necessità di individuare un giusto equilibrio tra contrapposti interessi si atteggia quale presupposto essenziale nell'ambito delle relazioni umane, mentre, allo stato attuale, l'IA parrebbe essere idonea a proporre soluzioni efficaci, le quali, tuttavia, richiederebbero, ex post, una valutazione "politica" da parte delle istituzioni, specie in funzione tutoria di principi cardine quali trasparenza, imputabilità, responsabilità.

In conclusione, la regolamentazione dell'uso dei sistemi di Intelligenza Artificiale comporta svariate questioni etiche, legali e teoriche, che dovranno essere risolte dalle istituzioni dell'UE e dagli Stati membri.

L'auspicio che può sorgere è che un uso corretto degli strumenti di IA, sotto il controllo e la supervisione umana, possa tradursi in una condizione migliorativa dell'esistente, ovvero sia di maggiore benessere

collettivo, nonché nel raggiungimento degli obiettivi strategici legati alla protezione ambientale del pianeta¹⁵.

¹⁵ Sul punto si rinvia per un maggior approfondimento a R. Trezza, *Diritto e intelligenza artificiale. Etica, Privacy, Responsabilità, Decisione*, Pisa, 2020. Si veda anche G. Taddei Elmi, A. Contaldo (a cura di), *Intelligenza artificiale. Algoritmi giuridici. Ius condendum o “fantadiritto”?*, Pisa, 2020; U. Ruffolo (a cura di), *Intelligenza artificiale Il diritto, i diritti, l’etica*, Milano, 2020; G. Alpa (a cura di), *Diritto e intelligenza artificiale. Profili generali, soggetti, contratti, responsabilità civile, diritto bancario e finanziario, processo civile*, Pisa, 2020; A. Santosuosso, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto*, Milano, 2020; M. Cupersito, *Intelligenza artificiale e diritto: profili normativi, etici e politici*, in *Opinio Iuris*, 1° giugno 2020, consultabile online; A. Longo, G. Scorza, *Intelligenza artificiale. L’impatto sulle nostre vite, diritti e libertà*, Milano, 2020; A. D’Aloia, *Intelligenza artificiale e diritto: Come regolare un mondo nuovo*, Milano, 2021; U. Ruffolo (a cura di), *XXVI lezioni di Diritto dell’Intelligenza Artificiale*, Torino, 2021; G. Schneider, N. Abriani, *Diritto delle imprese e intelligenza artificiale. Dalla Fintech alla Corpotech*, Bologna, 2021; G. Taddei Elmi, *Il Quid, il Quomodo e il Quid iuris dell’IA. Una riflessione a partire dal volume “Diritto e tecnologie informatiche”*, in *Riv. it. inf. dir.*, n. 2/2021, p. 131 ss.; E. Bassoli, *Algoritmica giuridica. Intelligenza artificiale e diritto*, Modena, 2022; P. Severino (a cura di), *Intelligenza artificiale. Politica, economia, diritto, tecnologia*, Roma, 2022; R. M. Agostino, G. Dalia, M. Imbrenda, S. Pietropaoli (a cura di), *Frontiere digitali del diritto. Esperienze giuridiche a confronto su libertà e solidarietà*, Torino, 2022.