

I principi costituzionali nella disciplina dell'Intelligenza Artificiale. Nuove prospettive interpretative

di Marta Fasan

Abstract: *Constitutional Principles in the Regulation of Artificial Intelligence. New Interpretative Perspectives* – Artificial Intelligence is increasingly widespread within the society. This phenomenon is giving rise to relevant benefits for individuals but also to significant risks that could negatively affect fundamental rights. This paper aims at analysing the possible key role of constitutional principles in determining the present and future regulation of Artificial Intelligence in the legal context, focusing more on the new regulatory functions and contents of three specific constitutional principles: the transparency principle, the equality principle, and the self-determination one. After considering how these constitutional principles are interpreted, mainly at the European level, when applied to the Artificial intelligence context, the author analyses the possible emergence of a new constitutional principle, whose role could be fundamental to balance the risks due to the use of this technology: the human oversight principle.

181

Keywords: Artificial Intelligence; Constitutional principles; transparency; equality; self-determination, Human oversight.

1. Introduzione

Negli ultimi anni l'Intelligenza Artificiale (AI) è entrata far parte della vita quotidiana delle persone in modo sempre più incisivo, prospettando funzionalità che mai si sarebbe pensato di poter ricondurre ad un sistema artificiale¹. I progressi realizzati nell'elaborazione di tecniche di apprendimento² quali il *machine learning*³ e il *deep learning*⁴, uniti ad un

¹ Per una rapida ricostruzione dei principali ambiti di applicazione dell'AI nell'epoca contemporanea si veda G. Sartor, F. Lagioia, *Le decisioni algoritmiche tra etica e diritto*, in U. Ruffolo (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milano, 2020, 65; M. U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, in *Harvard Journal of Law & Technology*, 2, 2016, 354 ss.

² In generale sui profili tecnici riguardanti l'AI si veda in questo fascicolo P. Traverso, *Breve introduzione tecnica all'Intelligenza Artificiale*.

³ Nello specifico, il concetto di *machine learning* si riferisce ad un insieme di tecniche che permettono al sistema artificiale di migliorare, attraverso l'esperienza, le prestazioni delle proprie funzioni, secondo la definizione elaborata in T. M. Mitchell, *Machine learning*, Boston, 1997, 2.

⁴ Con il termine *deep learning* si è, invece, soliti identificare una tipologia di *machine learning* basata sul funzionamento di un sistema di reti neurali artificiali, che sono in grado di apprendere in modo non supervisionato da dati non classificati e non

notevole aumento nella disponibilità e nell'accessibilità a dati di varia natura, hanno contribuito allo sviluppo di sistemi di AI in grado di apprendere e di migliorare le proprie capacità tecniche direttamente dai dati esaminati, secondo un meccanismo in cui l'efficacia della prestazione tecnologica tende ad essere direttamente proporzionale all'incremento dell'esperienza del sistema artificiale⁵.

Tale nuovo approccio nella realizzazione dell'AI sta determinando notevoli successi nell'applicazione di questa tecnologia, modificandone fortemente l'impatto prodotto e il ruolo svolto all'interno del contesto sociale. Infatti, le opportunità offerte dai sistemi artificiali in termini di tempestività, efficacia ed economicità nello svolgimento di determinate funzioni stanno dando luogo ad una presenza sempre più significativa di azioni, scelte e decisioni condizionate dal fattore "artificiale", incentivando, di conseguenza, le persone ad affidarsi in modo crescente alle soluzioni proposte dall'AI⁶.

Questo fenomeno tecnologicamente connotato, che sta progressivamente determinando la nascita di un nuovo paradigma di «società algoritmica»⁷ o di «società onlife»⁸, non si presenta privo di

strutturati e, conseguentemente, di analizzare le informazioni e di cercare tra esse schemi di correlazione da utilizzare nel processo decisionale finale. Così in I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, Cambridge (MA), 2016, 8.

⁵ In questo senso A. Simoncini, S. Suweis, *Il cambio di paradigma nell'intelligenza artificiale e il suo impatto sul diritto costituzionale*, in *Rivista di filosofia del diritto*, 1, 2019, 90 ss.

⁶ Tali aspetti emergono in particolar modo in S. Rodotà, *Il diritto di avere diritti*, Roma-Bari, 2012, 401-402, in cui l'autore evidenzia come questo fenomeno riguardi scelte e decisioni di diversa natura, sia relative alla vita quotidiana delle persone, sia relative alla società nel suo complesso.

⁷ Questa nuova categorizzazione del contemporaneo modello di società è così espressa in M. Bassini, L. Liguori, O. Pollicino, *Sistemi di Intelligenza Artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, in F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018, 333 ss.; A. Pajno et al., *AI: profili giuridici. Intelligenza Artificiale: criticità emergenti e sfide per il giurista*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2019, 206-207. In particolare, gli autori sostengono che il successo delle tecnologie di AI e la diffusione della loro applicazione rappresentano un passaggio evolutivo dalla «società dell'informazione» alla «società dell'algoritmo». Essi, infatti, constatano come l'Intelligenza Artificiale dimostri di avere la stessa portata dirompente che aveva caratterizzato l'avvento della rete Internet negli anni '90, fenomeno per cui, allora, numerose categorie giuridiche preesistenti vennero poste in discussione. Tuttavia, nello svolgere queste considerazioni gli autori ricordano come ad oggi l'AI continui ancora ad essere lontana dalla sua fase di piena maturità tecnologica. Per meglio contestualizzare tali riflessioni cfr. Y. Benkler, *The Wealth of Networks. How Social Production Transforms Markets and Freedom*, New Haven-Londra, 2006, 219 ss.

⁸ Il concetto di «società onlife» è frutto dell'elaborazione di Luciano Floridi e di altri ricercatori europei e ha lo scopo di definire e di descrivere la nuova realtà sociale caratterizzata dalla diffusione pervasiva delle tecnologie e dal crescente processo di digitalizzazione in cui ogni giorno si trovano ad essere inseriti gli esseri umani. Sul punto L. Floridi (a cura di), *The Onlife Manifesto. Being Human in a Hyperconnected Era*, Cham, 2015, 7 ss.

conseguenze significative all'interno del contesto giuridico⁹. Il ricorso ad una tecnologia come l'AI, capace di condizionare in modo determinante decisioni produttive di effetti giuridici significativi sulle libertà e sui diritti delle persone¹⁰, costituisce, infatti, un elemento in grado di scardinare l'assetto logico-concettuale che tradizionalmente contraddistingue la dimensione giuridica, dal momento questi sistemi artificiali non si limitano più ad essere meri strumenti per il raggiungimento di uno scopo stabilito, ma diventano fattori determinanti nel definire i risultati del processo decisionale in cui si trovano ad essere utilizzati¹¹. Un simile aspetto deve, quindi, portare la dottrina giuridica contemporanea a valutare l'effettiva tenuta e validità delle categorie giuridiche tradizionali di fronte alle nuove sfide poste dalla diffusione e dall'applicazione dell'AI e, tenuto conto dell'impatto che questa può produrre sui diritti, sulle libertà e sui principi alla base dello Stato costituzionale di diritto, a considerare l'eventualità che siano predisposti nuovi strumenti giuridici per rispondere alle esigenze determinate dall'avvento pervasivo di questa tecnologia.

Il presente contributo si pone l'obiettivo di analizzare le principali problematiche giuridiche poste dall'uso dell'AI, privilegiando le chiavi di lettura fornite dalle categorie del costituzionalismo contemporaneo, e di svolgere alcune prime riflessioni sul ruolo che alcuni principi costituzionali possono avere nell'orientare la disciplina giuridica di questa tecnologia, anche alla luce dei recenti sviluppi normativi all'interno del contesto dell'Unione Europea.

2. L'Intelligenza Artificiale nella dimensione giuridica.

L'uso sempre più diffuso dell'AI in contesti giuridicamente rilevanti, di natura pubblica e privata, non è dovuto semplicemente all'ondata di entusiasmo che circonda questa tecnologia negli ultimi tempi. La ragione principale che sta spingendo verso una maggiore adozione dei sistemi artificiali risiede nella promessa di facilitare e incentivare la realizzazione di azione e l'adozione di decisioni orientate a promuovere il benessere e gli interessi delle persone. Questa tecnologia dimostra, infatti, di poter incidere

⁹ Il collegamento tra le tecnologie rientranti nell'ambito della cibernetica (tra le quali deve essere inclusa anche l'intelligenza artificiale) e il diritto era già stato ipotizzato in V. Frosini, *Cibernetica: diritto e società*, Milano, 1968, 17 ss. In particolare, l'autore sostiene l'esistenza di un'analogia tra il diritto e la scienza cibernetica, in quanto entrambe le discipline sono indirizzate a comprendere e a rendere prevedibili i meccanismi di controllo e di comunicazione che caratterizzano i comportamenti collettivi. Tali profili sono analizzati anche in A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e futuro delle libertà*, in A. D'Aloia (a cura di), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, Milano, 2020, 170 ss.

¹⁰ Sul punto cfr. A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e futuro delle libertà*, cit., 169 ss.

¹¹ Questa particolare funzione che, negli ultimi tempi, viene sempre più spesso attribuita alla tecnologia costituisce un'espressione del c.d. *Law 3.0*, illustrato in R. Brownsword, *Law 3.0*, Abingdon-New York, 2020, 28 ss.

positivamente all'interno di scelte particolarmente complesse¹², migliorandone qualità ed efficienza, e di poter raggiungere, intervenendo sia direttamente che indirettamente, risultati vantaggiosi e importanti benefici in termini di benessere e produttività nella risoluzione delle sfide che le persone e la società devono affrontare quotidianamente¹³.

Tale aspetto risulta particolarmente chiaro alla luce di tre principali vantaggi riconducibili all'uso dell'AI. In primo luogo, i sistemi artificiali possono contribuire a ridurre significativamente i tempi normalmente richiesti per l'adozione di determinate decisioni o per lo svolgimento di determinate funzioni, garantendo, quindi, benefici in termini di maggiore rapidità grazie alla potenza computazionale di analisi delle informazioni e alla capacità di individuare correlazioni rilevanti tra i dati esaminati¹⁴. In secondo luogo, l'AI costituisce uno strumento vantaggioso nel migliorare l'efficacia delle soluzioni proposte, sia nella prospettiva di ridurre possibili margini di errore¹⁵ sia nell'ottica di personalizzarne i contenuti e renderli sempre più adatti agli interessi e ai desideri delle persone¹⁶. Infine, i sistemi di AI possono contribuire in modo significativo dal punto di vista economico, diminuendo il costo dei processi predittivi e decisionali e incentivando,

¹² Così espresso in G. Sartor, F. Lagioia, *Le decisioni algoritmiche tra etica e diritto*, cit., p. 77; G. Resta, *Governare l'innovazione tecnologica: decisioni algoritmiche, diritti digitali e principio di eguaglianza*, in *Politica del diritto*, 2, 2019, 213. In questo ultimo contributo l'autore evidenzia come l'adozione di processi decisionali parzialmente o interamente automatizzati consenta di raggiungere ragguardevoli risultati dal punto di vista dell'uniformità, dell'affidabilità e della controllabilità della decisione stessa, coerentemente ai criteri di efficienza e di economicità che, nell'ordinamento italiano, assicurano il buon andamento della pubblica amministrazione.

¹³ Tali aspetti benefici sono evidenziati in OECD, *Artificial Intelligence in Society*, Parigi, 2019, 15, in www.oecd.org/publications/artificial-intelligence-in-society-eedfee77-en.htm.

¹⁴ In generale, sui vantaggi determinati dall'uso dell'AI in termini di rapidità si veda A. K. Agrawal, J. Gans, A. Goldfarb, *Prediction Machine. The Simple Economics of Artificial Intelligence*, Boston, 2018, 105 ss.; E. Fast, E. Horvitz, *Long-Term Trends in the Public Perception of Artificial Intelligence*, in *AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 2017, 963-969. Per quanto concerne, poi, i benefici temporali che questa tecnologia può avere in settori specifici come la medicina e la pubblica amministrazione cfr. E. J. Topol, *Deep Medicine. How Artificial Intelligence Can Make Healthcare Human Again*, New York, 2019; I. A. Nicotra, V. Varone, *L'algoritmo, intelligente ma non troppo*, in *Rivista AIC*, 4, 2019, 86 e ss.; C. Coglianese, D. Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, in *The Georgetown Law Journal*, 1147, 2017, 1152-1153.

¹⁵ Così sottolineato in C. R. Sunstein, *Algorithms, Correcting Biases*, in *Social Research: An International Quarterly*, 2, 2019, 499-511; J. Kleinberg et al., *Human Decisions and Machine Predictions*, in *The Quarterly Journal of Economics*, 1, 2018, 237 e ss.; D. Kahneman, *Thinking, Fast and Slow*, New York, 2011.

¹⁶ Per un'analisi delle specifiche potenzialità espresse dall'AI in termini di profilazione degli utenti si rimanda a F. Lagioia, G. Sartor, *Profilazione e decisione algoritmica: dal mercato alla sfera pubblica*, in *Federalismi.it*, 11, 2020, 89 ss.; K. Yeung, *Why Worry about Decision-Making by Machine?*, in K. Yeung, M. Lodge (a cura di), *Algorithmic Regulation*, Oxford, 2019, 34-35; M. Ainis, *Il regno dell'Uroboro. Benvenuti nell'era della solitudine di massa*, Milano, 2018, 15.

grazie alle dimostrate capacità di analisi e di correlazione dei dati, una corretta ed equilibrata allocazione delle risorse disponibili¹⁷.

Tuttavia, l'applicazione dell'AI può dare luogo anche a questioni problematiche di un certo rilievo e che, in ragione dei rischi che possono determinare, devono opportunamente essere prese in considerazione, soprattutto se si considera che la portata degli svantaggi riconducibili a questa tecnologia può essere tale non solo da vanificare tutti gli aspetti positivi dati dall'uso dell'AI, ma anche da dare luogo a risultati inferiori rispetto a quanto realizzabile senza il supporto dei sistemi artificiali. E tali aspetti assumono ancora più importanza alla luce delle conseguenze giuridiche che ne possono derivare, dal momento che un'applicazione dell'AI i cui rischi non vengano sufficientemente controllati potrebbe causare un'ingiustificata compressione dei principi e delle garanzie poste a tutela degli interessi, delle libertà e dei diritti delle persone.

2.1 L'Intelligenza Artificiale nella dimensione giuridica. Primi profili critici

Gli svantaggi che emergono dall'uso di questa tecnologia sono, ad oggi, principalmente determinati da fattori che caratterizzano la natura, lo sviluppo, la produzione e il funzionamento dell'AI e possono realizzarsi anche nei casi in cui tali strumenti siano creati e impiegati a presunto beneficio dei singoli soggetti e della collettività¹⁸. Nello specifico, è possibile individuare tre principali tipologie di rischi riconducibili all'applicazione dell'AI.

In primo luogo, le tecniche di apprendimento utilizzate dai sistemi artificiali, che tanto stanno contribuendo al progresso di questa tecnologia e al raggiungimento di risultati promettenti, non permettono una piena capacità di comprensione delle modalità e dei motivi in base a cui l'AI, partendo dagli input ricevuti, arriva a produrre uno specifico risultato¹⁹. Tale fenomeno, comunemente definito *black box problem*²⁰, comporta un elevato

¹⁷ Cfr. A. K. Agrawal, J. Gans, A. Goldfarb, *Prediction Machine. The Simple Economics of Artificial Intelligence*, cit., 7 ss.; J. Furman, R. Seamans, *AI and the economy*, in *NBER working papers series*, giugno 2018, 2 ss., in www.nber.org/system/files/working_papers/w24689/w24689.pdf.

¹⁸ K. Yeung, *Why Worry about Decision-Making by Machine?*, cit., 23-24.

¹⁹ In M. Ebers, *Regulating AI and Robotics: Ethical and Legal Challenges*, in M. Ebers, S. Navas (a cura di), *Algorithms and Law*, Cambridge, 2020, 48, l'autore evidenzia come, oltre ai processi e ai meccanismi che hanno determinato il risultato finale potrebbero essere sconosciuti, o noti solo parzialmente, anche gli input stessi ricevuti dal sistema intelligente. Tali aspetti sono analizzati anche in Y. Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, in *Harvard Journal of Law & Technology*, 2, 2018, 901 ss.

²⁰ Questo concetto è stato usato per la prima volta in F. Pasquale, *The Black Box Society. The Secret Algorithms That Control Money and Information*, Cambridge (MA), 2015. Secondo l'autore la metafora della *black box* rappresenterebbe correttamente la realtà contemporanea, in cui le persone sono sempre più controllate e sorvegliate dalle società private e dai governi ma non sono consapevoli di come informazioni e dati riguardanti la loro vita vengano divulgate e impiegate da questi soggetti. In merito ai fattori che

livello di opacità per quanto riguarda il funzionamento di questa tecnologia, arrivando a rendere le decisioni, le scelte e le azioni realizzate grazie all'AI sostanzialmente "inspiegabili", dal momento che le persone che si avvalgono di questi sistemi non sono in grado di comprendere concretamente le ragioni che hanno dato luogo ad un determinato risultato finale²¹. Questo aspetto costituisce un rischio non trascurabile e apre a problematiche considerevoli, soprattutto dal punto di vista giuridico. L'incapacità di spiegare il processo che ha portato all'adozione di una determinata decisione può dare origine, innanzitutto, ad una diminuzione di fiducia verso le abilità di questi strumenti tecnologici. Secondariamente, l'inspiegabilità rappresenta un limite rilevante per testare il funzionamento del processo decisionale tecnologicamente connotato e controllare se il sistema di AI abbia rispettato gli standard e le norme giuridiche vigenti nel contesto in cui si trova ad operare. Infine, questo dato si traduce in una grave carenza in termini di conoscibilità della motivazione posta alla base della decisione e, conseguentemente, della sua legittimità, dal momento che l'impossibilità di conoscere tale elemento può ridurre significativamente l'accesso agli strumenti giuridici normalmente preposti a porre rimedio ad eventuali illegittimità o irregolarità verificatesi durante il processo decisionale²².

In secondo luogo, lo stretto legame che intercorre tra il funzionamento dell'AI e l'ampia disponibilità di informazioni da analizzare può dare luogo a rischi rilevanti qualora vengano meno i requisiti di quantità e qualità che dovrebbero caratterizzare i dati utilizzati da questi sistemi. Infatti, l'impiego di dati imprecisi, viziati, inaccurati e carenti²³ può condizionare così

alimentano il fenomeno della *black box* cfr. T. Wischmeyer, *Artificial Intelligence and Transparency: Opening the Black Box*, in T. Wischmeyer, T. Rademacher (a cura di), *Regulating Artificial Intelligence*, Cham, 2020, 80-81; V. Dignum, *Responsible Artificial Intelligence. How to Develop and Use AI in a Responsible Way*, Cham, 2019, p. 59 e ss.; D. Pedreschi et al., *Meaningful Explanations of Black Box AI Decision Systems*, in *Proceeding of the AAAI Conference on Artificial intelligence*, n. 1, 2019, pp. 9780-9784; M. Ebers, *Regulating AI and Robotics: Ethical and Legal Challenges*, in M. Ebers, S. Navas (a cura di), *Algorithms and Law*, Cambridge, 2020, 44-5 ss.

²¹ Cfr. K. Yeung, *Why Worry about Decision-Making by Machine?*, cit., 28; G. Noto La Diega, *Against the Dehumanisation of Decision-Making. Algorithmic Decisions at the Crossroads of Intellectual Property, Data Protection, and Freedom of Information*, in *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 1, 2018, 9 ss.

²² Queste tre diverse tipologie di rischi, giuridici e non, riconducibili all'opacità dell'AI sono così individuati in M. Ebers, *Regulating AI and Robotics: Ethical and Legal Challenges*, cit., 48. Oltre a ciò, in G. Orsoni, E. D'Orlando, *Nuove prospettive dell'amministrazione digitale: open data e algoritmi*, in *Istituzioni del Federalismo*, 3, 3019, 607 ss., gli autori sottolineano come in queste specifiche tecnologie intelligenti la necessità di trasparenza rappresenti una precondizione per la garanzia e la protezione del principio di legalità, in quanto permetterebbe di comprendere se il sistema di AI sia stato creato e impiegata rispettando la normativa vigente.

²³ Per quanto riguarda i principali fattori che possono viziare i risultati prodotti dall'AI cfr. A. Tischbirek, *Artificial Intelligence and Discrimination: Discriminating Against Discriminatory Systems*, in T. Wischmeyer, T. Rademacher (a cura di), *Regulating Artificial Intelligence*, Cham, 2020, 105; V. Eubanks, *Automating Inequality. How High-Tech Tools Profile, Police, and Punish the Poor*, New York, 2017; I. N. Cofone, *Algorithmic*

negativamente l'operatività di un sistema di AI da scardinare gli obiettivi perseguiti attraverso l'uso di questa tecnologia. Le potenziali mancanze relative ai dati usati dall'AI non solo possono portare a risultati inferiori dal punto di vista della qualità, dell'efficacia e dell'efficienza che, invece, dovrebbero poter garantire, ma rischiano anche di incrementare il numero di decisioni e azioni discriminatorie nei confronti di categorie e gruppi di persone già ampiamente emarginate all'interno della società²⁴. In tale modo, l'AI non sarebbe più uno strumento funzionale al raggiungimento di una maggiore neutralità, correttezza e oggettività, ma potrebbe, invece, diventare un mezzo per la realizzazione di trattamenti iniqui verso le persone che si relazionino con questo tipo di tecnologia. Questo profilo, che già di per sé presenta rischi inaccettabili, è ulteriormente aggravato dal velo di neutralità che apparentemente caratterizza i sistemi di AI ma che, invece, rischia di celare le pratiche discriminatorie che potrebbero essere causate da un *dataset* viziato e pregiudizievole²⁵, rendendo così ancora più difficoltosa l'individuazione, la contestazione e la sanzione delle discriminazioni perpetrate attraverso l'uso di questa tecnologia.

Infine, l'ultimo tra i rischi principali che interessano l'uso dell'AI riguarda la natura prevalentemente privata dei soggetti che negli ultimi anni hanno finanziato e sostenuto lo sviluppo e la realizzazione di questa tecnologia. Tale aspetto, nonostante non abbia necessariamente una connotazione *tout court* negativa, può portare a svantaggi rilevanti secondo una duplice prospettiva. Innanzitutto, la creazione di questi sistemi intelligenti, oltre a costituire un'occasione di progresso per le conoscenze in questo settore, rappresenta anche la realizzazione di un prodotto commercializzabile e il cui successo all'interno del mercato dipende fortemente dalla competitività offerta rispetto a quanto proposto dalle società concorrenti. In questo contesto, in cui le aziende e le società devono sia coprire le spese sostenute per la creazione dei sistemi di AI sia guadagnare degli utili grazie alla produzione di questa tecnologia, i produttori hanno un interesse economico nell'avere e nel mantenere una posizione di vantaggio e di prevalenza nel mercato globale di questi prodotti²⁶, favorendo, quindi, un

Discrimination Is an Information Problem, in *Hastings Law Journal*, 6, 2019, 1402 ss.

²⁴ In N. Criado, J. M. Such, *Digital Discrimination*, in K. Yeung, M. Lodge (a cura di), *Algorithmic Regulation*, Oxford, 2019, 84-85; E. Stradella, *Stereotipi e discriminazioni: dall'intelligenza umana all'intelligenza artificiale*, in Aa. Vv. (a cura di), *Liber Amicorum per Pasquale Costanzo – Diritto costituzionale in trasformazione. Vol. I – Costituzionalismo, Reti e Intelligenza artificiale*, Genova, 2020, 391-400. Per quanto riguarda gli effetti discriminatori prodotti dall'AI nella prospettiva della parità di genere cfr. M. D'Amico, *Una parità ambigua. Costituzione e diritti delle donne*, Milano, 2020, 313 ss.

²⁵ Per quanto concerne i problemi determinati dall'apparente neutralità degli strumenti di AI si veda L. Giacomelli, *Big Brother is "Gendering" You. Il diritto antidiscriminatorio alla prova dell'intelligenza artificiale: quale tutela per il corpo digitale?*, in S. Dorigo (a cura di), *Il ragionamento giuridico nell'era dell'intelligenza artificiale*, Milano, 2020, 207 ss.; A. Chander, *The Racist Algorithm?*, in *Michigan Law Review*, 6, 2017, 1040.

²⁶ In questa dimensione assume importanza l'acquisizione del c.d. "surplus comportamentale" che fornisce alle società produttrici dell'AI uno strumento

aumento dei costi e un elevato livello di segretezza sui meccanismi di creazione e di funzionamento dell'AI. Questi aspetti, però, rischiano, da un lato, di escludere tutti coloro che non sono in grado di sostenere economicamente gli elevati costi di questa tecnologia dagli importanti vantaggi che essa promette di realizzare e, dall'altro lato, di vanificare le esigenze di trasparenza e di conoscibilità che, invece, dovrebbero essere garantite nell'uso di questi strumenti intelligenti²⁷. In secondo luogo, le capacità tecniche dimostrate dall'AI in termini di analisi dei dati e di elaborazione di modelli predittivi e decisionali stanno offrendo alle società private, che sviluppano e producono questa tecnologia, la possibilità non solo di conoscere i comportamenti e i desideri delle persone su diversi livelli di azione²⁸, ma anche di influenzare e indirizzare le decisioni e le azioni degli esseri umani²⁹. Questa opportunità, che sta portando alla nascita di uno scenario economico, sociale e giuridico assolutamente rivoluzionario, sta sempre di più trasformando un fenomeno tecnologico ed economico in una nuova forma di potere capace di incidere significativamente nelle scelte delle persone³⁰. In questo contesto, in cui pochi soggetti detengono il potere di sorvegliare e di manipolare le decisioni legate alla sfera privata e alla dimensione pubblica³¹, l'impiego dell'AI rischia di diventare funzionale a

fondamentale per la crescita dal punto di vista del *know-how*. In questo senso A. Vacchi, *Artificial Intelligence e Industria 4.0 tra tecnoetica e tecnodiritto*, in U. Ruffolo (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milano, 2020, 280 ss.

²⁷ Sugli effetti negativi prodotti in termini di *economic divide* e di segretezza si veda Comitato Nazionale per la Bioetica, Comitato Nazionale per la Biosicurezza le Biotecnologie e le Scienze della Vita, *Sviluppi della robotica e della roboetica*, Roma, 2017, p. 21, in bioetica.governo.it/media/1392/p129_sviluppi-della-robotica-e-della-roboetica-gruppo-misto-it-cnb.pdf; F. Pasquale, *New Laws of Robotics. Defending Human Expertise in the Age of AI*, Cambridge (MA)-Londra, 2020.

²⁸ In merito ai pericoli riconducibili a questa acclarata concentrazione del potere di mercato nelle mani di un piccolo gruppo di società si veda M. Martini, *Regulating Algorithms: How to Demystify the Alchemy of Code?*, in M. Ebers, S. Navas (a cura di), *Algorithms and Law*, Cambridge, 2020, 107-108.

²⁹ Sulle conseguenze di questo fenomeno, sia a livello individuale che collettivo, cfr. K. Yeung, *Why Worry about Decision-Making by Machine?*, cit., 35 ss.; M. Martini, *Regulating Algorithms: How to Demystify the Alchemy of Code*, cit., 108. In questo contributo, l'autore, ricordando il noto caso *Cambridge Analytica*, illustra come l'AI non influenzi solo i comportamenti di mercato delle persone, ma possa anche indirizzare anche le scelte politiche degli individui.

³⁰ In riferimento a questo processo di trasformazione si veda O. Pollicino, *L' "autunno caldo" della Corte di giustizia in tema di tutela dei diritti fondamentali in rete e le sfide del costituzionalismo alle prese con i nuovi poteri privati in ambito digitale*, in *federalismi.it*, 19, 2019, 11 ss.; C. Casonato, *Per un'intelligenza artificiale costituzionalmente orientata*, in A. D'Aloia (a cura di), *Intelligenza artificiale e diritto. Come regolare un mondo nuovo*, Milano, 2020, 81 ss.; A. Rouveroy, T. Berns, *Gouvernementalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation?*, in *Réseaux*, 1, 2013 163 ss.

³¹ Per definire questo fenomeno viene comunemente utilizzato il termine «capitalismo della sorveglianza», coniato da Shoshanna Zuboff per identificare questa nuova forma di capitalismo, il cui obiettivo non si limita ad osservare e a controllare i comportamenti delle persone, ma mira, invece, a manipolarne le scelte. Cfr. S. Zuboff, *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Roma, 2019.

favorire il perseguimento e il raggiungimento di interessi e di obiettivi di soggetti terzi, senza che le persone interessate siano effettivamente consapevoli dell'influenza esercitata sulla loro libertà decisionale e acconsentano tale pratica. E tale rischio, oltre a poter incidere sulla struttura giuridica che contraddistingue gli Stati costituzionali di diritto contemporanei³², può impedire alle persone di compiere scelte veramente libere e informate, compromettendo la tutela degli interessi, dei diritti e delle libertà delle persone.

3. L'Intelligenza Artificiale nella prospettiva del diritto costituzionale

I profili appena descritti dimostrano chiaramente quale possa essere l'impatto prodotto da una simile tecnologia all'interno del contesto giuridico e sociale. La crescente tendenza ad affidarsi a sistemi che si caratterizzano per gli svantaggi evidenziati rischia di contribuire all'affermazione di una società algoritmica in cui vengono meno le garanzie, anche giuridiche, che dovrebbero proteggere e tutelare le persone dalle indebite ingerenze di questo nuovo potere tecnologico³³.

In questo scenario, in cui gli effetti dirompenti prodotti dall'applicazione dell'AI si realizzano in un contesto giuridicamente definito che non può non considerare le conseguenze dei rischi prospettati, occorre chiedersi quali categorie e strumenti giuridici possano offrire valide soluzioni alle nuove sfide poste dall'impiego di questa tecnologia e dalle peculiari caratteristiche che la contraddistinguono.

La risposta a tali questioni può arrivare, senza dubbio, dalle categorie, dai principi, dagli strumenti e dalle garanzie appartenenti al diritto costituzionale contemporaneo e questo, in particolare, per tre ordini di ragioni. In primo luogo, le trasformazioni causate dall'avvento dell'AI incidono direttamente su alcuni principi e valori costituzionali che rappresentano le fondamenta e la struttura degli Stati costituzionali di diritto. In secondo luogo, i rischi riconducibili allo sviluppo e alla diffusione dei sistemi di AI riportano l'attenzione all'originaria vocazione e funzione del diritto costituzionale, e cioè la reale e concreta limitazione dei poteri al fine di assicurare un'efficace garanzia e tutela dei diritti³⁴. Infine, le categorie del costituzionalismo contemporaneo possono indicare, ai legislatori nazionali, sovranazionali e ai soggetti regolatori, la direzione da seguire nella realizzazione di una disciplina dell'AI che sappia valorizzare la

³² In B. Caravita di Toritto, *Principi costituzionali e intelligenza artificiale*, in U. Ruffolo (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l'etica*, Milano, 2020, 459 ss., l'autore si interroga sulla capacità degli attuali sistemi giuridici, fondati su un costituzionalismo liberale e sociale, di resistere al potere espresso dai giganti del web e delle tecnologie.

³³ Questi aspetti sono così evidenziati in S. Rodotà, *Il diritto di avere diritti*, cit., 402 ss.

³⁴ Su questi aspetti si veda in questo fascicolo C. Casonato, *L'intelligenza artificiale e il diritto pubblico comparato ed europeo*.

centralità e il rispetto degli esseri umani quali obiettivi e valori centrali nell'uso di questa tecnologia³⁵.

Questi elementi, dunque, attribuiscono al diritto costituzionale un ruolo chiave nell'individuare gli strumenti idonei ad affrontare i nuovi rischi posti dall'uso di una tecnologia come l'AI e nel garantire che gli stessi vengano arginati e mitigati da un impianto normativo e regolatorio volto ad assicurare il pieno rispetto dei diritti e delle libertà delle persone che si trovino ad interagire e ad utilizzare questi sistemi intelligenti. In tale contesto, alcuni tra i principi costituzionali riconosciuti e tutelati dagli ordinamenti giuridici contemporanei assumono una funzione di preminenza rispetto alla prospettiva regolatoria delineata, contribuendo ad individuare i potenziali correttivi da adottare nei confronti delle questioni più problematiche poste dall'AI.

Alla luce di queste considerazioni, appare opportuno approfondire il ruolo particolare che stanno assumendo tre specifici principi costituzionali nell'indirizzare gli interventi e le proposte normative che si stanno adottando, a livello europeo e nazionale, in materia di AI, e ciò proprio in ragione degli effetti che i rischi posti da questa tecnologia possono causare sulla tutela dei diritti e delle libertà delle persone³⁶. Nello specifico si tratta dei principi di trasparenza, di eguaglianza e di autodeterminazione.

4. Il principio di trasparenza nella disciplina dell'Intelligenza Artificiale

Il primo principio costituzionale che assume rilevanza nell'orientare la disciplina dell'AI e nel mitigare, soprattutto, i rischi causati dal fenomeno della *black box* è il principio di trasparenza.

Questo principio, sia nella dimensione delle decisioni pubbliche e dell'amministrazione sia nel contesto riguardante il trattamento dei dati personali³⁷, assume un ruolo fondamentale rispetto all'impiego dell'AI, ma la

³⁵ L'importanza di un'analisi basata sulle categorie del diritto costituzionale viene enfatizzata e promossa anche in O. Pollicino, G. De Gregorio, *Constitutional Law in the Algorithmic Society*, in H-W- Micklitz et al. (a cura di), *Constitutional Challenges in the Algorithmic Society*, Cambridge, 2021, 5 ss.; A. Simoncini, *Sovranità e potere nell'era digitale*, in T. E. Frosini et al. (a cura di), *Diritti e libertà in Internet*, Milano-Firenze, 2017, 19 ss.; O. Pollicino, *Forum: Law and Artificial Intelligence. L'impatto dell'AI sul diritto e sui diritti*, in *BioLaw Journal – Rivista di BioDiritto*, 1, 2020, 491-492; C. Casonato, *Per un'intelligenza artificiale costituzionalmente orientata*, cit., 82; O. Pollicino, *L' "autunno caldo" della Corte di giustizia in tema di tutela dei diritti fondamentali in rete e le sfide del costituzionalismo alle prese con i nuovi poteri privati in ambito digitale*, cit., 12.

³⁶ Per un'analisi, invece, dei diritti, anche nuovi, principalmente interessati dalla diffusione e dall'applicazione dell'AI si veda in questo fascicolo L. Rinaldi, *Intelligenza artificiale, diritto e doveri nella Costituzione italiana*.

³⁷ Sull'importanza del principio di trasparenza in questi due contesti si vedano, *ex multis*, E. D'Orlando, *I principi in materia di pubblica amministrazione*, in L. Mezzetti (a cura di), *Principi costituzionali*, Torino, 2011, 428 ss.; S. Wachter, *The GDPR and the Internet of Things: a three-step transparency model*, in *Law, Innovation and Technology*, 2, 2018, 272 ss.

sua concreta effettività risulta messa duramente alla prova alcune questioni tecniche e giuridiche riguardanti l'applicazione di questa tecnologia. L'inspiegabilità delle decisioni adottate con il supporto dell'AI e i meccanismi di tutela dei diritti della proprietà intellettuale rischiano, infatti, di incidere negativamente sulla reale attuazione del principio di trasparenza³⁸, vanificando le garanzie tradizionalmente riconducibili all'applicazione di questo principio.

La constatazione dei limiti che possono riscontrarsi nell'applicazione di questo principio all'AI nella sua accezione tradizionale, insieme alla consapevolezza dell'importanza rivestita dal principio di trasparenza nell'assicurare la tutela dei diritti delle persone che si trovino ad utilizzare questi nuovi sistemi intelligenti, sta spingendo, *in primis* il legislatore europeo³⁹ ma anche gli altri soggetti attivi sul piano nazionale ed internazionale, ad ipotizzare nuove declinazioni normative per questo principio. Nella prospettiva di garantire che esso possa effettivamente costituire una delle basi giuridiche per la realizzazione di una disciplina dell'AI volta ad incentivarne i benefici verso la società⁴⁰, il principio di trasparenza trova concreta attuazione secondo tre specifiche declinazioni.

In primo luogo, il rispetto del principio di trasparenza nel contesto dell'AI si traduce nella garanzia di tracciabilità dei dati e dei procedimenti che portano i sistemi intelligenti all'adozione della decisione finale. Questo significa che tutte le informazioni relative al metodo di raccolta dei dati, alla loro classificazione, al tipo di algoritmo impiegato e i risultati ottenuti devono essere completamente documentate, cosicché tutti coloro che vi abbiano diritto o interesse possano comprendere le modalità di impiego dei dati esaminati, seguirne il flusso e verificare che la decisione finale sia stata elaborata e adottata correttamente⁴¹.

³⁸ Questi profili sono dovutamente analizzati in T. Scantamburlo, A. Charlesworth, N. Cristianini, *Machine Decisions and Human Consequences*, in K. Yeung, M. Lodge (a cura di), *Algorithmic Regulation*, Oxford, 2019, 71 ss., in cui gli autori evidenziano come l'accesso al codice sorgente dell'AI potrebbe non essere sufficiente per la comprensione dei meccanismi adottati da questa tecnologia. Anzi, una simile azione comporterebbe un'indubbia violazione della tutela della segretezza industriale riconosciuta ai produttori di questi sistemi.

³⁹ Il riferimento è alla recente proposta di Regolamento dell'Unione europea in materia di Intelligenza Artificiale, noto anche come *Artificial Intelligence Act*, pubblicata il 21 aprile 2021. Attualmente, questa proposta di Regolamento rappresenta l'intervento normativo più completo e organico che sia stato elaborato a livello europeo in materia di intelligenza artificiale. Per un primo commento ai contenuti di questa importante proposta normativa cfr. C. Casonato, B. Marchetti, *Prime osservazioni sulla proposta di regolamento della Commissione Ue in materia di intelligenza artificiale*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2021, 415 ss.

⁴⁰ Questo obiettivo viene analizzato con maggiore attenzione in L. Floridi et al., *How to Design AI for Social Good: Seven Essential Factors*, in *Science and Engineering Ethics*, 3, 2020, 1771 ss.

⁴¹ Questi aspetti sono, ad esempio, disciplinati all'art. 13 della proposta di Regolamento dell'Unione europea in materia di Intelligenza Artificiale. L'importanza della tracciabilità dei dati e delle informazioni usate dall'AI è, inoltre, affermata in High-Level

In secondo luogo, il principio di trasparenza si concretizza nella previsione, dal punto di vista normativo, del requisito dell'*explainability* della decisione, da intendersi come la capacità di spiegare e motivare i procedimenti tecnici che contraddistinguono il funzionamento dell'AI, le decisioni automatizzate e quelle adottate con il supporto di questa tecnologia⁴². Questa seconda possibile declinazione del principio di trasparenza, però, non si esaurisce nella capacità di accedere alle informazioni riguardanti i meccanismi logici interni al processo decisionale, in quanto questa facoltà potrebbe non essere sufficiente per spiegare i motivi che hanno determinato l'adozione di una determinata decisione da parte del sistema artificiale⁴³. In tale dimensione, la declinazione del principio di trasparenza in termini di *explainability* non si concretizza solo nel ricorso a strumenti giuridici che permettano di spiegare gli schemi di elaborazione della scelta finale o il livello di influenza esercitato dall'AI sulla decisione umana, ma rende indispensabile la predisposizione di soluzioni tecniche che possano arginare le conseguenze di un fenomeno come quello della c.d. *black box*⁴⁴. Da questa prospettiva, il principio di trasparenza trova la sua piena attuazione anche nella realizzazione di modelli di AI interpretabili o nell'impiego di strumenti tecnici che consentano di avere spiegazioni comprensibili sul comportamento di questi sistemi intelligenti⁴⁵.

Infine, il principio di trasparenza, nella dimensione dell'AI, trova concreta attuazione nella garanzia che si svolga un adeguato livello di informazione sull'uso di questa tecnologia nell'elaborazione di una decisione.

Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, Bruxelles, 2019, 18; European Commission, *White Paper on Artificial Intelligence – A European approach to excellence and trust*, Bruxelles, 2020, 19; OECD, *Recommendation of the Council on Artificial Intelligence*, 2019, in legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

⁴² L'importanza del criterio dell'*explainability* viene sottolineato anche in U. Pagallo, *Algoritmi e conoscibilità*, in *Rivista di filosofia del diritto*, 1, 2020, 96 ss.; Council of Europe (CoE), Recommendation CM/REC (2020)1 of the Committee of Ministers to member State on the human rights impact of algorithmic systems, 8 aprile 2020, in search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154; M. Palmirani, *Big Data e conoscenza*, in *Rivista di filosofia del diritto*, 1, 2020, 74 ss. Per una critica all'effettiva esistenza di un diritto alla spiegazione all'interno del dettato normativo del GDPR cfr. S. Wachter, B. Mittelstadt, L. Floridi, *Why a Right to Explanation of AUTomated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *International Data Privacy Law*, 2, 2017, 76 ss.

⁴³ Tali profili sono evidenziati in A. Oddenino, *Decisioni algoritmiche e prospettive internazionali di valorizzazione dell'intervento umano*, in *DPCE online*, 1, 2020, 204 ss.; E. Stradella, *AI, tecnologie innovative e produzione normativa: potenzialità e rischi*, in *DPCE online*, 3, 2020, 3361.

⁴⁴ In questo senso J. A. Kroll et al., *Accountable Algorithms*, in *University of Pennsylvania Law Review*, 3, 2017, 657 ss.

⁴⁵ Una soluzione di questo tipo viene proposta in A. Monreale, *Rischi etico-legali dell'Intelligenza Artificiale*, in *DPCE online*, 3, 2020, 3396 ss.; C. Rudin, *Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead*, in *Nature Machine Intelligence*, 1, 2019, 206 ss.; R. Guidotti et al., *Factual and counterfactual explanations for black box decision making*, in *IEEE Intelligent Systems*, 6, 2019, 14 ss.

Questo specifico profilo si traduce, quindi, nella necessità che le azioni e le decisioni prese dall'AI siano identificabili come tale, così da garantire il diritto delle persone ad essere informate e consapevoli della loro interazione con un sistema artificiale⁴⁶. Ciò significa anche che, in questa declinazione del principio di trasparenza, le persone devono essere consapevoli delle particolari capacità e dei limiti che caratterizzano il funzionamento di queste tecnologie intelligenti⁴⁷. E questo aspetto assume maggiore importanza nei casi in cui questo tipo di operazioni debbano realizzarsi nel pieno rispetto delle libertà e dei diritti fondamentali delle persone⁴⁸.

5. Il principio di eguaglianza nella disciplina dell'Intelligenza Artificiale

Il secondo principio costituzionale che viene in rilievo nell'elaborazione di una disciplina giuridica dell'AI è il principio di eguaglianza.

Questo principio, che costituisce uno dei capisaldi dello Stato costituzionale di diritto⁴⁹, assume un ruolo fondamentale nel tentativo di mitigare gli elementi di parzialità e le pratiche discriminatorie che possono emergere attraverso l'uso dell'AI e che possono rendere questa tecnologia uno strumento funzionale alla realizzazione di azioni lesive dell'eguaglianza e dei diritti delle persone⁵⁰.

Similmente a quanto affermato per il principio di trasparenza, anche l'effettiva attuazione del principio di eguaglianza, nella sua accezione tradizionale, subisce delle limitazioni significative alla luce delle peculiarità tecniche e applicative che contraddistinguono l'AI. Anche in questo caso, quindi, si sta riconoscendo la necessità, dal punto di vista normativo, di elaborare nuove declinazioni e interpretazioni di questo principio, così da renderlo concretamente uno strumento per la realizzazione e lo sviluppo di un'AI affidabile. Secondo questa prospettiva, nei principali atti e documenti

⁴⁶ A questo proposito si vedano European Commission, *White Paper on Artificial Intelligence – A European approach to excellence and trust*, cit., 20; Commissioner for Human Rights (CoE), *Unboxing Artificial Intelligence: 10 steps to protect Human Rights*, maggio 2019, 9, rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco/1680946e64

⁴⁷ Questo aspetto trova riconoscimento all'art. 13, par. 3, lett. b), della proposta di Regolamento dell'Unione europea in materia di Intelligenza Artificiale. Sul punto anche B. Mittelstadt et al., *The ethics of algorithms: Mapping the debate*, in *Biga Data & Society*, 2, 2016, 6-7.

⁴⁸ Cfr. E. Stradella, *AI, tecnologie innovative e produzione normativa: potenzialità e rischi*, cit., 3361.

⁴⁹ In generale, sui contenuti e le declinazioni tradizionali del principio di eguaglianza all'interno dell'ordinamento giuridico italiano si vedano, *ex multis*, A. Celotto, *Art. 3, 1° co., Cost.*, in R. Bifulco, A. Celotto, M. Olivieri (a cura di), *Commentario alla Costituzione*, Milano, 2006, 65 ss.; A. Giorgis, *Art. 3, 2° co., Cost.*, in *ivi*, 88 ss.; P. Carretti, *I diritti fondamentali. Libertà e diritti sociali*, Torino, 2002, 89-90 e 151 ss.

⁵⁰ In merito a queste possibili conseguenze cfr. F. Zuiderveen Borgesius, *Strengthening legal protection against discrimination by algorithms and artificial intelligence*, in *The International Journal of Human Rights*, 25 marzo 2020, 13.

normativi riguardanti questa tecnologia vengono, dunque, proposte tre linee direttrici che possano garantire un'applicazione effettiva ed efficace del principio di eguaglianza tra le persone.

Innanzitutto, questo principio costituzionale si concretizza nella necessità di evitare la presenza di pregiudizi all'interno delle tecnologie di AI, operando affinché non si verifichino *bias* ed errori che potrebbero avere effetti discriminatori sugli utenti dei sistemi artificiali. A tal fine, il principio di eguaglianza si concretizza nell'identificazione e nella rimozione di qualsiasi elemento pregiudizievole o errato che possa essere presente nei dati raccolti, senza che tale intervento di pulizia delle informazioni comporti una riduzione dell'efficacia e dell'accuratezza del sistema di AI⁵¹. Oltre a ciò, la promozione di un maggior livello di eguaglianza in relazione all'uso di questa tecnologia può essere garantita dalla predisposizione di procedure di controllo e di validazione orientate a verificare l'assenza di pratiche discriminatorie nei sistemi intelligenti⁵².

La seconda linea d'azione nell'implementazione concreta del principio di eguaglianza riguarda la promozione dell'accessibilità rispetto all'AI. Questo significa che, anche dal punto di vista normativo, deve essere incentivata la creazione di sistemi di AI che possano essere facilmente utilizzati e compresi da un numero quanto più ampio di persone, ponendo maggiore attenzione proprio a quelle categorie di soggetti più vulnerabili che potrebbero beneficiare dall'uso di queste tecnologie⁵³. Inoltre, l'attuazione del principio di eguaglianza in termini di accessibilità deve anche garantire che non si realizzino disuguaglianze economiche nell'accesso all'AI, evitando l'applicazione di questa tecnologia possa perpetrare nuove e più profonde forme di *economic divide*⁵⁴.

⁵¹ Così in Agenzia dell'Unione europea per i diritti fondamentali (FRA), *#BigData: Discrimination in data-supported decision making*, 30 maggio 2018, 8-10, in fra.europa.eu/en/publication/2018/bigdata-discrimination-data-supported-decision-making; Advisory Committee on Equal Opportunities for Woman and Men, *Opinion on Artificial Intelligence – opportunities and challenges for gender equality*, 18 marzo 2020, 11, in ec.europa.eu/info/sites/info/files/aid_development_cooperation_fundamental_rights_opinion_artificial_intelligence_gender_equality_2020_en.pdf. Questo aspetto trova riconoscimento anche all'interno degli artt. 10 e 15 della proposta di Regolamento dell'Unione europea in materia di Intelligenza Artificiale. Tuttavia, appare opportuno sottolineare come tra le disposizioni che la proposta di Regolamento si pone l'obiettivo di introdurre in materia di AI ne manchi una esplicitamente riferita al rispetto del principio di eguaglianza e alle modalità concrete con cui si possa dare effettiva applicazione allo stesso in questo contesto. Questi aspetti, infatti, vengono affrontati in termini molto generali solo all'interno dei considerando della proposta di Regolamento.

⁵² Sul punto European Commission, *White Paper on Artificial Intelligence – A European approach to excellence and trust*, cit., 18-19; Agenzia dell'Unione europea per i diritti fondamentali (FRA), *#BigData: Discrimination in data-supported decision making*, cit., 6 ss.

⁵³ In questo senso anche High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI*, cit., 18 ss.; N. Criado, J. M. Such, *Digital Discrimination*, in K. Yeung, M. Lodge (a cura di), *Algorithmic Regulation*, Oxford, 2019, 94 ss.

⁵⁴ Cfr. S. Quintarelli et al., *AI: profili etici. Una prospettiva etica sull'Intelligenza Artificiale: principi, diritti e raccomandazioni*, in *BioLaw Journal – Rivista di BioDiritto*, n. 3, 2019,

Infine, il principio di eguaglianza trova effettiva applicazione se declinato secondo due ulteriori prospettive. In primo luogo, l'eguaglianza in relazione ai sistemi di AI si concretizza nella previsione di un adeguato livello di rappresentatività nei dati e nelle informazioni analizzati da questi strumenti tecnologici, contribuendo così ad evitare che le decisioni adottate da o con il supporto dell'AI possano dare luogo a risultati parziali e non sufficientemente rappresentativi della società. In secondo luogo, la promozione di un approccio quanto più pluralista nelle fasi di progettazione e di sviluppo di questa tecnologia, assicurando la presenza di programmatori appartenenti a contesti sociali, culturali, nazionali e disciplinari diversi, può rappresentare un'ulteriore garanzia per l'implementazione effettiva del principio di eguaglianza nel contesto dei sistemi di AI⁵⁵.

6. Il principio di autodeterminazione nella disciplina della Intelligenza Artificiale

L'ultimo principio che assume importanza nello sviluppo della disciplina di una *trustworthy AI*⁵⁶, in grado di porre la dimensione umana al centro degli strumenti normativi predisposti, è rappresentato dal principio di autodeterminazione.

Questo principio, che negli ordinamenti contemporanei riconosce alle persone il diritto di compiere scelte autonome, libere e indipendenti e che in molte circostanze svolge una funzione prodromica all'esercizio di altri diritti e libertà⁵⁷, risulta particolarmente rilevante alla luce della deriva deterministica che possono avere le tecnologie di AI. In particolare, le potenzialità manipolative di questi sistemi, insieme all'ingresso di interessi esterni e terzi in dimensioni decisionali private⁵⁸, impongono limiti

189.

⁵⁵ Questi ultimi due profili riguardanti l'applicazione del principio di eguaglianza nella dimensione dell'AI sono così delineati in F. Zuiderveen Borgesius, *Strengthening legal protection against discrimination by algorithms and artificial intelligence*, cit., 11 ss.; V. Dignum, *Responsible Artificial Intelligence. How to Develop and Use AI in a Responsible Way*, cit., 100 ss.

⁵⁶ Per quanto concerne l'obiettivo di realizzare un'AI degna di fiducia cfr. L. Floridi, *How to Design AI for Social Good: Seven Essential Factors*, cit., 1771 ss.

⁵⁷ In generale, sulle caratteristiche e i contenuti del principio di autodeterminazione all'interno dell'ordinamento italiano si vedano, *ex multis*, P. Veronesi, *Fisionomia e limiti del diritto fondamentale all'autodeterminazione*, in *BioLaw Journal – Rivista di BioDiritto*, Special Issue 2, 2019, 27; S. Mangiameli, *Autodeterminazione: diritto di spessore costituzionale?*, 2009, 1, in www.forumcostituzionale.it/wordpress/images/stories/pdf/documenti_forum/paper/0148_mangiameli.pdf; V. Marzocco, *Il diritto ad autodeterminarsi e il "governo di sé". La "vita materiale" tra proprietà e personalità*, in A. D'Aloia (a cura di), *Il diritto alla fine della vita. Principi, decisioni, casi*, Napoli, 2012, 33 ss.

⁵⁸ Questo particolare potere esercitato dai colossi del settore digitale è evidenziato in O. Bracha, F. Pasquale, *Federal Search Commission? Access, Fairness, and Accountability in the Law of Search*, in *Cornell Law Review*, 93, 2008, 1176 ss.; S. Zuboff, *Big other: surveillance capitalism and the prospects of an information civilization*, in *Journal of Information Technology*, 30, 2015, 75 ss.

significativi alla libertà delle persone di autodeterminarsi nelle proprie decisioni, senza che il contrattare di queste restrizioni si esprima nel bisogno di tutelare altri principi, diritti e libertà rilevanti dal punto di vista costituzionale. Tali profili problematici, che rischiano di diventare vettori di nuove forme di paternalismo⁵⁹, evidenziano la necessità di chiedersi come garantire un'effettiva attuazione del principio di autodeterminazione di fronte ai rischi che, in questo senso, sono prospettati dalla diffusione e dall'applicazione dell'AI.

Anche in riferimento a questo principio, gli interventi normativi fino ad ora realizzati nel settore dell'AI, insieme agli spunti dottrinali elaborati, propongono una nuova declinazione applicativa dell'autodeterminazione in un contesto fortemente eterodeterminato.

In primo luogo, una concreta implementazione del principio di autodeterminazione in questo contesto tecnologico richiede che ogni persona coinvolta e interessata dall'applicazione dell'AI abbia l'opportunità di liberarsi dalla determinazione algoritmica. Questa concezione dell'autodeterminazione di traduce, dunque, nel riconoscimento in capo alle persone della possibilità di rifiutare l'uso dell'AI nei processi decisionali o di essere destinatari di decisioni e scelte elaborate da questi sistemi artificiali⁶⁰.

In secondo luogo, l'effettiva attuazione di questo principio richiede che siano intraprese azioni indirizzate ad incrementare la qualità e la quantità di informazioni sul funzionamento di questa tecnologia e sugli effetti che la stessa può arrivare a produrre all'interno dei processi decisionali. Una simile operazione fornirebbe una maggiore consapevolezza nelle persone per quanto concerne le applicazioni dei sistemi intelligenti⁶¹, contribuendo a garantire che l'impiego dell'AI sia concretamente indirizzato a migliorare l'*empowerment* degli individui nelle scelte effettuate, agevolandone la navigazione tra le diverse opzioni decisionali disponibili⁶².

Infine, un ulteriore intervento che potrebbe essere funzionale a liberare le decisioni umane, tecnologicamente connotate, dall'ingerenza delle predeterminazioni e degli interessi di soggetti terzi riguarda i finanziamenti pubblici nel settore dell'AI. L'opportunità di incrementare e di incentivare la spesa pubblica nello sviluppo di questa tecnologia, anche valorizzando e

⁵⁹ In particolare, per alcune critiche alla teoria del paternalismo libertario in relazione all'applicazione dei sistemi di AI si veda K. Yeung, "Hypernudge": *Big Data as a mode of regulation by design*, in *Information, Communication & Society*, 1, 2017, 123 ss.

⁶⁰ In questo senso L. Floridi et al., *AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, in *Minds and Machines*, 4, 2018, 697 ss.; European Group on Ethics in Science and New Technologies, *Statement on Artificial Intelligence, Robotics and "Autonomous" Systems*, Bruxelles, 2018, 16, in ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf.

⁶¹ Cfr. J. Danher, *The Ethics of Algorithmic Outsourcing in Everyday Life*, in K. Yeung, M. Lodge (a cura di), *Algorithmic Regulation*, Oxford, 2019, 112 ss.

⁶² Da questo punto di vista, sarebbe, quindi, più facile consentire il raggiungimento degli obiettivi di navigabilità della decisione promossi e sostenuti in C. R. Sunstein, *Sulla libertà*, Torino, 2020.

promuovendo il più possibile le collaborazioni pubblico-privato in questo settore, faciliterebbe la realizzazione di un'AI che sia coerente con gli interessi tutelati dall'ordinamento costituzionali⁶³. E questa prospettiva garantirebbe alle persone una maggiore possibilità di compiere scelte libere e consapevoli, senza che il paternalismo tecnologico di questi sistemi e dei loro produttori comprometta l'attuazione del principio di autodeterminazione.

7. Il ruolo della sorveglianza umana nella disciplina dell'Intelligenza Artificiale. Verso la definizione di un nuovo principio costituzionale?

Il quadro normativo che si sta lentamente creando in relazione all'uso dell'AI evidenzia un dato fondamentale: i principi costituzionali assumono un ruolo fondamentale nell'indicare i caratteri principali di una disciplina dell'AI che sia in grado di bilanciare efficacemente e concretamente i benefici e i rischi che possono derivare dall'applicazione di questa tecnologia. L'effettiva attuazione dei principi di trasparenza, eguaglianza e di autodeterminazione, secondo le nuove declinazioni descritte e in una dinamica di forte interrelazione tra gli stessi, può costituire uno dei tasselli più importanti e più promettenti nella realizzazione di un approccio giuridico generale all'AI che sappia dirimerne le questioni più problematiche dal punto di vista tecnologico, sociale, economico e giuridico.

Tuttavia, in questo scenario che promette di orientare la regolamentazione di questi sistemi alla luce dei valori costituzionali, la concreta applicazione dei principi indicati, anche secondo le nuove declinazioni e interpretazioni proposte, potrebbe non essere sufficiente a soddisfare le esigenze di garanzia prospettate dall'impiego dell'AI. Nello specifico, appare opportuno chiedersi se i principi di trasparenza, eguaglianza e autodeterminazioni, alla luce delle caratteristiche di questa tecnologia, rappresentino strumenti sufficienti a garantire una forma di controllo adeguata sull'operato di questa tecnologia, soprattutto qualora essa venga utilizzata per l'adozione di decisioni completamente automatizzate.

Queste considerazioni stanno portando, dunque, ad ipotizzare che debba essere istituito un ulteriore elemento di garanzia che, in questo contesto, possa contribuire ad assicurare il rispetto dei diritti e delle libertà fondamentali delle persone. Nello specifico, le elaborazioni dottrinali, legislative e giurisprudenziali⁶⁴ degli ultimi tempi hanno individuato tale

⁶³ L'importanza di tale profilo nella prospettiva delineata viene evidenziato in European Commission, *White Paper on Artificial Intelligence – A European approach to excellence and trust*, cit., 3 ss.; Committee of Experts on Internet Intermediaries (CoE), *Algorithms and Human Rights. Study on the human rights dimension of automated data processing techniques and possible regulatory implications*, marzo 2018, 20, in edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html

⁶⁴ Per quanto riguarda le elaborazioni legislative, il riferimento è all'art. 14 della proposta di Regolamento dell'Unione europea in materia di Intelligenza Artificiale, il

garanzia nel criterio della sorveglianza umana, meglio noto con il termine *human in the loop*⁶⁵.

Da questo punto di vista, l'intervento dell'essere umano e l'esercizio della sua discrezionalità in relazione alle operazioni svolte dal sistema di AI vengono qualificate come le uniche garanzie concretamente azionabili rispetto alle decisioni basate esclusivamente sull'AI e che possono incidere sulle posizioni giuridiche rilevanti riconosciute in capo alle persone. A fronte dei limiti tecnici che contraddistinguono questi sistemi, l'intervento e la sorveglianza della persona umana diventa, quindi, una garanzia costituzionale, affinché l'applicazione di questa tecnologia non si traduca in una violazione dei diritti e delle libertà che lo Stato costituzionale di diritto attribuisce agli esseri umani nella loro dimensione individuale e collettiva.

In questo senso, il criterio dello *human in the loop* diventa espressione di un nuovo principio di rilevanza costituzionale, il principio di non esclusività, in base al quale i sistemi di AI non possono essere l'unico elemento decisivo nei processi decisionali riguardanti i diritti e le libertà fondamentali della persona⁶⁶.

Il riconoscimento di questo principio non risulta però esente dalle conseguenze determinate da un ulteriore aspetto problematico che si pone in relazione al ruolo di garanzia attribuito all'intervento e alla sorveglianza umana. Se è innegabile il valore aggiunto di questo strumento laddove una decisione sia affidata esclusivamente all'AI, permangono, però, alcune perplessità in merito all'effettività e all'efficacia dell'intervento umano in questa dimensione. In considerazione dei limiti che sussistono in relazione alla comprensibilità e alla spiegabilità dei meccanismi logici seguiti dall'AI e alla sua potenziale funzione manipolativa, o meglio di *hypernudging*⁶⁷, risulta

quale disciplina esplicitamente il requisito della sorveglianza umana. Per quanto concerne, invece, le elaborazioni giurisprudenziali, si fa riferimento alla sentenza del Consiglio di Stato, sezione VI, n. 881 del 2020 e in generale al filone giurisprudenziale riguardante il c.d. "caso buona scuola", in cui si è riconosciuta l'esistenza di questo tipo di garanzia nei confronti delle decisioni adottate in modo automatizzato. Infine, per quanto riguarda le elaborazioni dottrinali relative al ruolo dell'intervento e della sorveglianza umana cfr. A. Simoncini, *L'algoritmo incostituzionale: intelligenza artificiale e futuro delle libertà*, cit., 123 ss.; C. Casonato, *Per un'intelligenza artificiale costituzionalmente orientata*, cit., 103 ss.; F. Laviola, *Algoritmico, troppo algoritmico: decisioni amministrative automatizzate, protezione dei dati personali e tutela delle libertà dei cittadini alla luce della più recente giurisprudenza amministrativa*, in *BioLaw Journal – Rivista di BioDiritto*, 3, 2020, 389 ss.; C. Casonato, B. Marchetti, *Prime osservazioni sulla proposta di regolamento della Commissione Ue in materia di intelligenza artificiale*, cit., 415 ss.

⁶⁵ Il termine *human in the loop*, o *human on the loop*, identifica lo standard in base al quale l'applicazione dei sistemi autonomi in contesti critici deve avvenire sempre sotto la supervisione di un operatore umano. Il termine è così definito in S. Quintarelli et al., *AI: profili etici. Una prospettiva etica sull'Intelligenza Artificiale: principi, diritti e raccomandazioni*, cit., 191.

⁶⁶ Per alcune riflessioni sull'esistenza di questo principio di non esclusività si veda A. Simoncini, *Diritto costituzionale e decisioni algoritmiche*, in S. Dorigo (a cura di), *Il ragionamento giuridico nell'era dell'intelligenza artificiale*, Milano, 2020, 54 ss.

⁶⁷ Con il termine si fa riferimento alla tecnica in base a cui le previsioni e le decisioni

opportuno domandarsi se l'intervento di un decisore umano sia sufficiente ad assicurare che non si verifichino i rischi ricollegabili all'applicazione e alla diffusione dei sistemi di AI⁶⁸.

Alla luce di queste considerazioni, emerge chiaramente la necessità che i principi costituzionali abbiano e continuino ad avere un ruolo primario nella determinazione dei contenuti della disciplina dell'AI, in modo tale da assicurare un approccio normativo, in questo settore, orientato al perseguimento dei valori costituzionali che caratterizzano gli ordinamenti giuridici contemporanei⁶⁹. E tale obiettivo si può realizzare non solo dando nuova sostanza ai principi cardine del costituzionalismo contemporaneo, ma anche arrivando ad ipotizzare e ad elaborare nuovi principi, più strettamente legati alle peculiari esigenze determinate dalla diffusione di questa tecnologia.

In ragione di ciò e alla luce delle specifiche caratteristiche tecniche che contraddistinguono l'AI, appare fondamentale la realizzazione di un continuo e attivo dialogo tra diritto, scienza e tecnologia, nel tentativo di arrivare all'elaborazione di una disciplina giuridica dell'AI che sia concretamente attenta alla tutela di tutte le istanze che sono destinate ad emergere davanti all'avvento dei sistemi di Intelligenza Artificiale.

Marta Fasan
Scuola di Studi Internazionali
Università degli Studi di Trento
marta.fasan@unitn.it

ottenute attraverso i sistemi di AI vengono inserite all'interno del modello di architettura della scelta adottato, al fine di influenzare le decisioni dell'utente finale. così definito in K. Yeung, "Hypernudge": *Big Data as a mode of regulation by design*, cit., 118 ss.

⁶⁸ La criticità di questo aspetto, che la proposta di Regolamento Ue definisce "distorsione algoritmica", viene analizzata in A. Simoncini, *Diritto costituzionale e decisioni algoritmiche*, cit., 55 ss.; C. Casonato, B. Marchetti, *Prime osservazioni sulla proposta di regolamento della Commissione Ue in materia di intelligenza artificiale*, cit., 429 ss. In questo ultimo contributo, gli autori ricordano i pericoli legati allo sviluppo di un c.d. *effèt moutonnier* (accogliendo la suggestione elaborata in A. Garapon, J. Lassègue, *Justice digitale. Révolution graphique et rupture anthropologique*, Parigi, 2018, 239), in base al quale il decisore umano tenderebbe comunque a seguire quanto deciso e determinato dal sistema di AI, svuotando così il contenuto garantista dello *human in the loop*

⁶⁹ Sul ruolo del diritto costituzionale nell'orientare la futura disciplina dell'AI si veda C. Casonato, *Per un'intelligenza artificiale costituzionalmente orientata*, cit.