INTELLIGENZA ARTIFICIALE E DIRITTO: UNA RIVOLUZIONE?

A CURA DI
ALESSANDRO PAJNO, FILIPPO DONATI E ANTONIO PERRUCCI

VOLUME I
DIRITTI FONDAMENTALI, DATI PERSONALI E REGOLAZIONE

SOCIETÀ EDITRICE IL MULINO

Alla pubblicazione di questa ricerca ha contribuito il Gruppo AlmavivA, che Astrid vivamente ringrazia

ISBN 978-88-15-29967-3

Redazione e produzione: Edimill srl - www.edimill.it

CAPITOLO UNDICESIMO

ALCUNE RIFLESSIONI SUL CONCETTO DI AUTONOMIA DECISIONALE DELLA MACCHINA E SULLE SUE IMPLICAZIONI REGOLAMENTARI

1. *L'autonomia decisionale della macchina: il contesto*

Nautilus non è solo il nome del sottomarino ideato e comandato dal capitano Nemo nel romanzo *Ventimila leghe sotto i mari* di Jules Verne, ma anche quello dell'intelligenza artificiale creata dai ricercatori dell'Università dell'Illinois e che, già dieci anni fa, è stata provocatoriamente definita «capace di predire il futuro» perché in grado di prevedere il nascondiglio di Osama Bin Laden e i fatti della Primavera araba[1]. Questa capacità, che a un primo sguardo potrebbe sembrare soprannaturale, non ha in realtà nulla di magico. Nautilus, infatti, ottiene le sue previsioni grazie a una potenza di elaborazione di 8,2 teraflops che gli permette di analizzare ogni anno più di cento milioni di articoli di giornale, esaminando parole e concetti ricorrenti e generando reti di milioni di collegamenti tra testi. Sono dunque la grande disponibilità di dati e la capacità di Nautilus di correlarli tra loro che rendono possibile prevedere con elevata probabilità di successo il verificarsi di determinati avvenimenti.

È proprio in questa tendenza, ormai consolidata da decenni, che è inquadrabile lo sviluppo dell'intelligenza artificiale: i dati in circolazione e le prestazioni dei sistemi che li trattano sono in aumento esponenziale. Allo stesso tempo, si assiste a una costante riduzione dei costi da

Questo capitolo è di Giuseppe D'Acquisto, Carmine Andrea Trovato e Ludovica De Benedetti.

[1] *The computer that predicts the future*, in «The Guardian», disponibile al link: https://www.theguardian.com/commentisfree/2011/sep/11/charliue-brooker-computer-predicts-future.

sostenere per memorizzarli e analizzarli[2]. Per farsi un'idea della portata del fenomeno, ogni due anni il volume dei dati trattati nel mondo raddoppia, mentre il costo per ottenere le stesse prestazioni da una macchina si dimezza[3]. Questo processo è inarrestabile e l'effetto che oggi rileviamo di questa tendenza è la capacità delle macchine di funzionare attraverso algoritmi che non necessitano di una supervisione dell'uomo[4].

Non possiamo prescindere da questo elemento se vogliamo comprendere per quale motivo, rispetto al passato, l'intelligenza artificiale, e in particolare il *machine learning*, rappresenta oggi una rottura di paradigma: qualcosa che cambia le nostre consuetudini e le cambierà per sempre. Grazie alla crescente disponibilità di informazioni e allo sviluppo tecnologico, infatti, la macchina non ci consentirà soltanto di ottenere dei risultati attraverso l'elaborazione dei dati con cui viene alimentata e sulla base di una teoria sviluppata dall'uomo, cosa che facciamo già da decenni, ma renderà anche possibile estrarre il significato di questi dati e, questa la vera novità, lo farà in completa autonomia.

Dunque, la persona non si troverà più di fronte a una macchina che si limita a svolgere compiti e analizzare dati sulla base di istruzioni assegnate, ma a una che decide, determina ed estrae senso dai dati autonomamente. È proprio l'autonomia decisionale della macchina che, rispetto al passato, rappresenta l'elemento distintivo dell'intelligenza artificiale e su questo cambio di paradigma si sono concentrate sia la nuova Proposta di Regolamento sull'intelligenza artificiale del 21/4/2021 da parte della Commissione

[2] *Cisco's 2020 Global Trends Report*, https://www.cisco.com/c/dam/m/en_us/solutions/enterprise-networks/networking-report/files/GLBL-ENG_NB-06_0_NA_RPT_PDF_MOFU-no-NetworkingTrends-Report-NB_rpten018612_5.pdf.

[3] The Hamilton Project Press Oil Booking Institution, *The Cost of computing power equal to an iPad2*, disponibile al link: https://www.hamiltonproject.org/charts/cost_of_computing_power_equal_to_an_ipad2.

[4] G. D'Acquisto e M. Naldi, *Big data e Privacy by Design, Anonimizzazione, Pseudonimizzazione, Sicurezza*, Torino, 2017, p. 15.

europea[5] sia i testi che l'hanno preceduta in tema di etica[6], responsabilità[7] e sicurezza[8].

Se da un lato, dunque, esistevano e continueranno a esistere macchine che non decidono autonomamente e che funzionano sulla base di una teoria sviluppata dall'uomo, dall'altro avremo sempre più macchine il cui funzionamento è completamente autonomo e basato su un solo schema: l'elaborazione di dati e la ricerca di correlazioni tra questi. Entrambi i paradigmi comportano sia benefici che rischi ma, mentre nel primo caso esistono già strumenti di tutela che l'uomo ha sviluppato nel tempo per mitigare gli effetti malevoli derivanti dall'uso della tecnologia[9], nel secondo si aprono nuovi scenari che devono essere affrontati dal «regolatore» con un approccio innovativo per valorizzare gli effetti benefici e mitigare i rischi che l'autonomia decisionale della macchina comporterà.

È infatti pacifico che l'intelligenza artificiale si diffonderà sempre di più nei prossimi anni[10] perché essa risponde a

[5] *Proposal for a 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*, Bruxelles, 21/4/2021 COM(2021) 206 final 2021/0106 (COD).

[6] *The Assessment List For Trustworthy Artificial Intelligence (ALTAI)*, High-Level Expert Group on Artificial Intelligence set up by the European Commission, July 2020.

[7] *European Parliament Resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence*, Parlamento europeo, 20/10/2020.

[8] ENISA, l'agenzia europea per la cybersecurity, ha analizzato le possibili implicazioni dell'autonomia decisionale della macchina sugli aspetti di sicurezza, all'interno del documento *AI Threat Landscape*, 15/12/2020.

[9] Si pensi alla possibilità di ricondurre, sulla base di schemi prescrittivi già esistenti, la responsabilità derivante da un uso malevolo della tecnologia all'essere umano. Ciò avviene proprio in virtù del fatto che la macchina agisce sulla base delle istruzioni che le vengono date dall'essere umano stesso.

[10] Grand View Research, *Artificial Intelligence Market Size, Share & Trends Analysis Report By Solution (Hardware, Software, Services), By Technology (Deep Learning, Machine Learning), By End Use, By Region, And Segment Forecasts, 2020-2027*. Il valore del mercato globale dell'intelligenza artificiale è stato quantificato in 39,9 miliardi di dollari nel 2019 e ci si aspetta che cresca a un tasso annuo del 42,2% fino

un bisogno dell'uomo: quello di essere sollevato dall'ultima fatica da cui si può essere sollevati dopo quelle meccaniche, ovvero la fatica rispetto ad azioni ripetitive o a quelle che richiedono un carico computazionale superiore alle nostre capacità. Infatti, l'uomo è un animale computazionalmente limitato, ha una scarsa capacità di memorizzazione e di calcolo e, conseguentemente, decide tenendo in considerazione un numero limitato di fattori. Se è la macchina che svolge il processo di analisi delle informazioni e di determinazione dei risultati, analizzando una mole incommensurabilmente superiore di dati, è ragionevole prevedere che l'uomo sarà ben disposto ad accettarne l'autonomia per superare tali limiti.

2. *Vantaggi e rischi dell'autonomia decisionale della macchina*

Uno schema conoscitivo basato sulla decisione autonoma della macchina, che non prevede l'intervento umano, permette di correlare tra loro informazioni in modo molto più efficace e di ottenere risultati inimmaginabili rispetto a quanto sarebbe avvenuto anche solo pochi anni fa.

L'intelligenza artificiale sta conquistando molti ambiti «umani» e quando parliamo di autonomia decisionale della macchina non ci riferiamo a un futuro ipotetico: attualmente, settori come la finanza, la sanità, la difesa, stanno utilizzando il *machine learning* per l'erogazione dei propri servizi e lo svolgimento dei propri compiti. Ancora, la prevenzione delle frodi nei sistemi di pagamento online, così come la medicina digitale e la diagnostica per immagini o il riconoscimento vocale che permette ai sistemi domotici di funzionare sono perlopiù basati su schemi di intelligenza artificiale e ci consentono di beneficiare dei relativi vantaggi.

È proprio in queste situazioni che la macchina può sostituirsi all'azione dell'uomo e individuare autonomamente correlazioni tra i dati che consentano di ottenere nuove informazioni, processo che l'uomo non sarebbe in grado

al 2027. Link: https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market.

di svolgere, o che richiederebbe troppo tempo per essere portato a compimento. Queste nuove informazioni e risultati sono nella maggior parte dei casi esatti e apportano un miglioramento concreto alle nostre vite semplificandole e facendoci fare passi avanti nella conoscenza del mondo; tuttavia, l'autonomia decisionale della macchina comporta anche una serie di rischi per diritti e libertà consolidate che richiedono l'intervento del regolatore.

Gli esempi sono noti. Basti pensare ai casi di utilizzo del riconoscimento facciale tramite videosorveglianza[11], che può sconfinare in una forma di controllo generalizzato. Ancora, si pensi al fenomeno del *deepfake*, una tecnica che si avvale dell'intelligenza artificiale per sovrapporre immagini o video originali che ritraggono più soggetti, o per creare *ex novo* immagini o video completamente falsi eppure talmente realistici da risultare difficilmente distinguibili da fatti realmente accaduti. L'utilizzo con scopi malevoli di tali tecniche può comportare una pluralità di conseguenze negative e perfino vere e proprie alterazioni dei processi democratici.

Tali sistemi, se non progettati, sviluppati e alimentati nel rispetto dei diritti e delle libertà delle persone, possono generare dei veri e propri *bias*, delle discriminazioni, i cui effetti negativi potrebbero talvolta non essere eliminabili *ex post*, causando così dei danni permanenti e irreparabili. Questi rischi hanno un potenziale impatto ancora più alto sulle persone se si pensa che tali fenomeni si verificano in un contesto di asimmetria informativa, ossia senza che la persona sia informata della possibile osservazione o della modificazione della realtà a suo danno, e in fin dei conti che non sia consapevole degli scopi per cui la tecnologia viene utilizzata e delle possibili conseguenze derivanti dal suo utilizzo.

Come per ogni nuova tecnologia, esistono dunque dei rischi derivanti dalle modalità attraverso cui viene utilizzata. Tuttavia, non si deve commettere l'errore di delineare un

[11] Così come definito dall'art. 29 Data Protection Working Party, *Opinion 02/2012 on facial recognition in online and mobile services*, 00727/12/EN, WP 192, p. 2.

contesto catastrofico che tenga conto dei soli rischi e sembri impossibile da regolamentare. In questo caso si finirebbe per desistere dallo sforzo di individuare una regolamentazione che, seppur complesso, merita di essere compiuto proprio per dare certezza ai tanti progetti di pubblica utilità che si basano sull'applicazione di tecniche di intelligenza artificiale. L'assenza di regolamentazione comporterebbe infatti un doppio effetto negativo: da un lato, inibirebbe lo sviluppo di tali progetti a causa dell'incertezza creata dal clima di sospetto e diffidenza nei confronti dell'intelligenza artificiale, dall'altro, creerebbe sacche di illegalità all'interno delle quali le tecnologie continuerebbero a essere sviluppate in un regime deregolamentato e, dunque, senza tutele per i diritti individuali e collettivi.

Di fronte a questo tipo di scenario, si comprende bene la necessità di predisporre regole per lo sviluppo delle tecnologie. Peraltro, se si guarda alla storia dello sviluppo tecnologico per trarre indicazioni sul futuro, lo scenario è molto meno pessimistico di quanto si potrebbe immaginare: l'uomo ha da sempre convissuto con lo sviluppo tecnologico e con i suoi potenziali rischi e i benefici che ne ha ricavato sono decisamente maggiori rispetto ai corrispondenti eventi dannosi. Le tecnologie non sono intrinsecamente buone o cattive e non hanno lo scopo di migliorare la vita dell'essere umano o, al contrario, di coartarne la libertà. La tecnologia è neutra ed è l'uso che ne viene fatto che la connota[12].

Se facciamo dunque un bilancio complessivo dell'uso che è stato fatto della tecnologia nella storia umana e dei conseguenti effetti positivi e negativi, è un dato di fatto che i primi prevalgano sui secondi[13]. È l'uso che faremo dell'intelligenza artificiale che determinerà la sua connotazione in senso positivo o negativo ed è per questo motivo che

[12] E. Severino, *La tendenza fondamentale del nostro tempo*, Milano, 2008.

[13] H. Rosling, O. Rosling e A. Rosling Rönnlund, *Factfulness: ten reasons we're wrong about the world – and why things are better than you think*, Paris, 2019.

una sua efficace regolamentazione assume un'importanza fondamentale.

3. *Possibili conflitti tra regole e funzionamento delle macchine*

Una sfida importante per lo sviluppo dell'intelligenza artificiale è quella di riuscire a costruire un sistema regolatorio che permetta di mantenere la centralità dell'uomo e creare un clima di fiducia per quanto riguarda l'utilizzo delle nuove tecnologie.

A questo fine, in assenza di un pieno controllo sul funzionamento della macchina, è bene estendere la portata della regola verso una sempre maggiore *accountability* da parte del progettista e dell'utilizzatore di tecnologie, anche attraverso il perseguimento *by design* di principi etici che permettano di indirizzare l'autonomia della macchina a beneficio della persona. Occorre però una riflessione sullo spazio per questi interventi in considerazione dei vincoli tecnologici. Siamo, infatti, in presenza di sistemi logico formali.

È, dunque, necessario che i principi etico-giuridici alla base delle decisioni autonome assunte dalla macchina siano compatibili con le regole della logica. Una macchina etica richiede il rispetto di alcune regole minime. La prima regola è impedire che le decisioni autonome possano danneggiare l'uomo[14] (principio di non maleficenza); la seconda è indirizzare l'autonomia decisionale della macchina a vantaggio dell'uomo (principio di beneficenza). Tutto in un quadro di dominanza dell'uomo sulla macchina. Sono obiettivi che mitigherebbero i rischi che derivano da macchine completamente autonome. Tuttavia essi sono molto difficili da conseguire in pratica.

Il principio di non maleficenza presuppone l'implementazione di regole che impediscano che un sistema di intelligenza artificiale possa danneggiare l'uomo da un punto

[14] European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103 [INL]).

di vista fisico o limitandone ad esempio privacy, dignità o sicurezza. Risulta, però, impossibile raggiungere l'obiettivo di una non maleficenza assoluta: infatti una macchina può sempre essere utilizzata per scopi ulteriori rispetto a quelli per cui è stata progettata[15]; inoltre, nel caso di sistemi particolarmente complessi, non è sempre possibile prevedere tutte le azioni che la macchina potrà compiere ed è possibile che alcune fra queste azioni non previste risultino dannose per l'uomo. Entrambi questi rischi non potrebbero essere efficacemente esclusi nemmeno nel caso ci si avvalesse di «macchine etiche» in grado di individuare da sole quali azioni o decisioni possano essere dannose, o di una progettazione etica[16] che si avvalga di una classificazione preliminare di tutti gli stati che la macchina può assumere, in quanto resterebbe sempre un margine di rischio legato all'incompletezza delle conoscenze e all'impossibilità di prevedere tutte le modalità di utilizzo della macchina stessa. L'uomo potrà allora porsi un obiettivo di non maleficenza progressiva andando a implementare nel tempo regole che neutralizzino i rischi di cui, man mano, si viene a conoscenza. Questa implementazione progressiva dovrebbe lasciare anche la possibilità di rivedere le regole nel senso contrario, permettendo l'uso di tecnologie che in un primo momento sono state classificate come potenzialmente nocive: infatti può essere controproducente l'imposizione di regole troppo rigide che rischierebbero di impedire l'utilizzo di tecnologie di intelligenza artificiale o di alcune loro potenzialità, che in determinate condizioni sono state considerate potenzialmente nocive o pericolose, anche per nuove finalità e in contesti diversi in cui potrebbero invece andare a beneficio dell'uomo. In fase di progetto, è necessario che si tenga conto di questi rischi e si blocchino quelle funzioni che possono essere dannose, ma è altrettanto

[15] K. Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, in «Monatshefte für Mathematik und Physik», 38, 1931, pp. 173-198, citato in J. van Heijenoort, *From Frege to Gödel*, Cambridge, Mass., 1971.

[16] The European Commission's High-Level Expert Group on Artificial Intelligence, *Draft ethics guidelines for trustworthy AI*, 18 December 2018.

importante, per una positiva evoluzione tecnologica, che si preveda la possibilità di ripristinare tutte le funzionalità della macchina qualora, per le mutate condizioni, queste possano garantire nuovi benefici.

Una volta raggiunto, per quanto possibile, l'obiettivo di non maleficenza, le raccomandazioni sui diritti civili dell'Unione europea (UE) riguardanti la robotica richiedono un ulteriore passo avanti: l'implementazione di regole che possano indirizzare l'autonomia decisionale della macchina a vantaggio dell'uomo (obiettivo di beneficenza della macchina).

La maggiore difficoltà che si riscontra in questo secondo caso è quella di riuscire a stabilire quali siano i principi etici da implementare in quanto universalmente considerati benevoli per l'uomo. Non solo, infatti, il concetto di «bene» varia molto nelle diverse culture ed è quindi estremamente difficile da oggettivizzare, ossia rendere misurabile e idoneo ad essere confrontato secondo le leggi della logica formale, ma anche all'interno di una stessa comunità vi sono spesso conflitti inconciliabili fra bene individuale e collettivo (si pensi anche solo al conflitto fra il diritto del singolo alla proprietà privata e l'interesse pubblico all'esproprio per la costruzione di un ospedale o di un'autostrada) e possono esistere interessi personali divergenti. Anche qualora si trovasse il modo di oggettivizzare il più possibile il concetto di bene o di correttezza[17], la macchina potrebbe interpretare in modo differente dall'uomo tali principi. In questi casi le controversie valoriali tra uomo e macchina possono essere risolte o attraverso criteri di misurabilità tipici della macchina o attraverso l'esercizio della forza tipico dell'uomo. Il Regolamento generale europeo in materia di protezione dei dati personali intuisce questa tensione logica-forza, e la risolve in una chiave prevalentemente difensiva per l'uomo, che ha sempre il diritto di opporsi a decisioni

[17] H. Elzayn, S. Jabbari, C. Jung, M. Kearns, S. Neel, A. Roth e Z. Schutzman, *Fair Algorithms for Learning in Allocation Problems*, ACM Conference on Fairness, Accountability and Transparency, 2019.

automatizzate particolarmente limitanti e di intervenire nel processo di decisione.

Estremo rimedio ai rischi che l'evoluzione di macchine intelligenti comporta e che, come visto, non possono essere del tutto eliminati, è la possibilità di spegnere la macchina (principio di dominanza dell'uomo sulla macchina). Ma una macchina dotata di autonomia decisionale potrebbe essere molto difficile da spegnere in quanto potrebbe ignorare il comando interpretando il proprio spegnimento come nocivo per l'uomo stesso o per il raggiungimento degli obiettivi per i quali è stata costruita. Sul tema sono molte le ricerche in corso[18], e quello che emerge è che per garantire il principio di dominanza dell'uomo sulla macchina è fondamentale la possibilità di rendere la macchina incerta sul comportamento dell'uomo e sugli obiettivi che vuole raggiungere. Sono temi di ampia portata su cui il dibattito scientifico abbraccia più discipline ed è, in sostanza, ancora aperto.

Una volta compresa la complessità e le criticità nell'implementazione dei principi etico-giuridici, è importante interrogarsi su quali altri strumenti regolatori possano essere messi in campo per garantire uno sviluppo sicuro dell'intelligenza artificiale che permetta di sfruttare i benefici di queste tecnologie, mantenendo la centralità dell'uomo e riducendo i rischi derivanti dall'impossibilità per l'uomo di prevedere o supervisionare il lavoro della macchina.

4. *Quale regolamentazione? Verso delle norme «technology-based»*

Dal punto di vista regolatorio abbiamo già visto che risulterebbe poco utile un approccio basato solo sulla proibizione o su norme eccessivamente rigide o incentrate unicamente sull'azione umana. È, dunque, necessario trovare nuove soluzioni regolatorie che possano efficacemente

[18] D. Hadfield-Menell, A. Dragan, P. Abbeel e S. Russell, *The Off-Switch Game*, International Joint Conference on Artificial Intelligence, 2017.

disciplinare l'uso dell'intelligenza artificiale. Per tali ragioni si sta sviluppando un concetto di *governance* multilaterale, fondato sia su principi giuridici che mettono al centro la persona e la costruzione di un ambiente di fiducia nella tecnologia, sia sull'incentivo ad un impiego della tecnologia come antidoto alla tecnologia stessa ogniqualvolta il contesto sia di tale complessità da rendere insufficiente e inefficace la sola prescrizione giuridica.

Proprio questo è l'approccio che l'Europa vuole mettere in campo con la nuova strategia di *governance* dell'intelligenza artificiale lanciata dalla Commissione europea nel febbraio 2020 con il *White Paper On Artificial Intelligence – A European approach to excellence and trust*[19] e che si sta delineando anche con la nuova Proposta di Regolamento IA, del 21/4/2021, che mira ad armonizzare le norme sull'intelligenza artificiale.

All'interno di questa strategia si inquadra anche il recente report *AI Threat Landscape*[20], pubblicato dall'Agenzia europea per la cybersecurity (ENISA) il 15/12/2020. Si tratta di un documento incentrato sulle minacce di sicurezza per le applicazioni di intelligenza artificiale. Il report parte dall'assunto che l'intelligenza artificiale offre grandi opportunità anche nel campo dell'implementazione della sicurezza informatica, ma allo stesso tempo espone a nuovi rischi, spesso ancora imprevedibili. Solo partendo da questo quadro dei rischi che l'intelligenza artificiale può comportare, si possono analizzare gli strumenti necessari per sviluppare una *governance* efficace di queste tecnologie.

Si può, tentativamente come ipotesi di lavoro, individuare tre fasi di intervento per il «regolatore». La vera e propria fase in cui la macchina elabora la propria decisione autonoma, la fase precedente e la fase successiva alla decisione.

[19] Reperibile al seguente link: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en.

[20] ENISA, *8th annual ENISA Threat Landscape (ETL)*, 2020, reperibile al seguente link https://www.enisa.europa.eu/news/enisa-news/enisa-ai-threat-landscape-report-unveils-major-cybersecurity-challenges.

Sulla fase della decisione la possibilità regolatoria è molto limitata, per le ragioni legate ai volumi dei dati trattati e alla velocità dei tempi della decisione stessa (si tratta a ben vedere di due delle «V» del cosiddetto paradigma dei *big data*). Se così è, bisogna concentrare gli sforzi sulla fase precedente alla decisione e su quella successiva.

Per quanto riguarda la fase precedente, è fondamentale rafforzare le tutele intervenendo direttamente sui dati che la macchina utilizza per formulare le proprie decisioni, in modo da guidarne il risultato: la qualità deve essere adeguata, le fonti attendibili, le variabili che la macchina utilizza non devono portare a risultati discriminatori. Fondamentale, a questo fine, sono i principi di accuratezza, correttezza, minimizzazione e necessità che rappresentano il cardine della protezione dei dati personali. Tali principi devono essere integrati nella tecnologia grazie all'uso di un linguaggio che la macchina possa decodificare e comprendere.

Il Regolamento generale per la protezione dei dati personali[21] (GDPR) è stato un precursore di questo approccio e ha, infatti, introdotto il concetto di *privacy by design* allo scopo di integrare le tutele nei trattamenti. Si tratta, di un processo che coinvolge diverse componenti tecnologiche e organizzative, volto all'implementazione dei principi della privacy e della protezione dei dati fin dalla progettazione di tecnologie e sistemi informatici[22]. Questo fine è perseguito, innanzitutto, attraverso l'obbligo previsto sia dal GDPR che dal nuovo Regolamento della Commissione di effettuare una valutazione di impatto nel momento della progettazione di

[21] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (General Data Protection Regulation). Reperibile al seguente link: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679.

[22] Nel 2014 l'ENISA ha pubblicato il primo report *Privacy and Data Protection by Design*, fornendo un inventario degli approcci di *privacy by design* esistenti. Nel 2015 ha fornito il report *Privacy by Design in Big Data*, volto ad analizzare strategie e strumenti di *privacy by design* nell'era dei *big data*.

ogni nuova tecnologia o servizio che possa rappresentare un rischio elevato per le persone a cui si rivolge la decisione[23].

Vi è quindi la questione della mitigazione dei rischi *ex post*. Da questo punto di vista si devono tenere presenti due fattori: il primo è che, nonostante l'implementazione di regole di sicurezza e principi etici che dovrebbero garantire che la macchina operi a beneficio dell'uomo è impossibile azzerare il rischio di incidenti o situazioni impreviste. Il secondo è che l'aumento dell'autonomia decisionale permette che alcune decisioni vengano prese direttamente in autonomia dalla macchina e sfuggano alla possibilità di comprensione o intervento dell'uomo.

Questo ci porta ad alcune considerazioni: innanzitutto è evidente che non è sufficiente un'azione regolatoria che si svolga unicamente o prevalentemente *ex ante*, ma è necessario prevedere come affrontare gli incidenti. In caso contrario si potrebbe solo vietare qualsiasi sviluppo tecnologico i cui risultati non siano del tutto prevedibili, rischiando di limitare grandemente l'uso dell'intelligenza artificiale anche per tutti gli scopi benefici[24]. Una seconda osservazione riguarda la necessità di mitigare o cancellare gli effetti di una decisione che abbia portato ad un risultato scorretto o nocivo per l'uomo. Questo potrebbe essere fatto attraverso lo strumento della notifica dell'incidente o

[23] Nella sua proposta la Commissione identifica diversi livelli di rischio per i diritti fondamentali e sulla base di tale rischio vieta l'uso di alcuni strumenti di intelligenza artificiale, prevede la possibilità per altri di essere commercializzati e utilizzati solo in via del tutto eccezionale, prevede per altri sistemi solo specifici obblighi di trasparenza. Infine, la proposta consente il libero utilizzo di applicazioni considerate a basso rischio quali videogiochi.

[24] In questo senso sarà molto interessante seguire le modalità di recepimento del *Digital Copyright Act*, in cui sono previste forme proattive di indagine dei contenuti da parte delle piattaforme web per tutelare il diritto degli autori sulle opere del loro ingegno, così come, soprattutto, il dibattito intorno al *Digital Services Act* e al *Digital Markets Act*, che disegneranno le future responsabilità delle stesse piattaforme nel bilanciamento tra libertà di espressione, manipolazione e non-neutralità. Anche qui è lecito aspettarsi che molte tensioni possano ricomporsi su un piano in cui le tecnologie offriranno vie d'uscita e forme di garanzie nuove altrimenti non individuabili.

errore. Questa possibilità è già codificata nel GDPR con una certa dinamica (notifica alle autorità, comunicazione agli interessati). Vista la disuguaglianza di forze in gioco e i tempi rapidi dei processi, questo promettente strumento potrebbe estendersi alla possibilità di notificare direttamente alla macchina in modo che sia sempre possibile per l'uomo contestare le decisioni prese in autonomia dall'intelligenza artificiale senza intermediazioni o chiedendo il supporto di un soggetto *super partes* per sovvertire la decisione.

La tecnologia può, dunque, aumentare l'efficacia del diritto e diventare il primo strumento di salvaguardia dai rischi che il suo stesso uso determina, e ciò è tanto più vero quanto più gli scenari tecnologici diventano complessi e la quantità dei dati cresce. Ciò è, però, possibile solo se si parte da uno studio approfondito della tecnologia che permetta di non banalizzarne il ruolo e la complessità per evitare tutele solo formali o regole eccessivamente rigide.

5. *Conclusioni*

Siamo dunque in presenza di un settore strategico in grandissimo fermento sul piano della ricerca scientifica e che nei prossimi anni, anche in vista di un rilancio post-COVID, vivrà una grande espansione. Infatti, l'impatto che l'intelligenza artificiale avrà sulla produttività globale entro il 2025 è stato stimato dal Parlamento UE in 10.000 miliardi di Euro[25].

Individuare regole che favoriscano un pieno sviluppo di questo settore e che siano in grado di ampliare i benefici dell'intelligenza artificiale mitigandone, allo stesso tempo,

[25] *AI rules: What the European Parliament wants*, disponibile al link: https://www.europarl.europa.eu/news/en/headlines/society/20201015STO89417/ai-rules-what-the-european-parliament-wants. La cifra è particolarmente significativa se si pensa che l'intero PIL mondiale ammonta a circa 100.000 miliardi di euro e che quindi l'impatto dell'intelligenza artificiale sulla produttività inciderà per quasi un decimo del PIL mondiale.

i potenziali rischi è un passaggio fondamentale nella regolamentazione delle future tecnologie.

È stato questo l'approccio adottato dalla Proposta di Regolamento sull'intelligenza artificiale da parte della Commissione UE, che rappresenta un primo tentativo a livello mondiale di regolamentare in modo organico il fenomeno dell'autonomia decisionale della macchina.

Questo processo di regolamentazione, come abbiamo visto, non può prescindere da norme che sappiano comprendere e spiegare l'intelligenza artificiale e da soluzioni tecnologiche che non si sostituiscano al diritto, bensì ne amplino la portata. Tali soluzioni consentirebbero ai princìpi normativi di essere applicati anche in contesti molto ostili che si basano sulla apparente negazione di alcuni paradigmi che diamo per assodati, come il controllo umano sulla decisione. La tecnologia può dunque opporsi (più efficacemente di come è in grado di fare l'uomo) ai rischi che l'uso delle tecnologie determina[26]. Si delinea in questo modo un'intelligenza artificiale che, come affermato dalla presidente della Commissione UE Ursula von der Leyen nel suo discorso sullo stato dell'Unione del 2020[27], pone la persona al centro fin dalla sua fase di progettazione per costruire uno spirito di fiducia e di ottimismo nei confronti di questi strumenti che costituiscono indubbiamente un reale fattore di progresso e di crescita per l'intera umanità.

[26] Basti pensare ai software di *bias detection* in grado di individuare gli eventuali pregiudizi sistemici sviluppati dagli algoritmi e di cogliere l'origine dell'errore. Per l'essere umano, d'altronde, è più difficile individuare il *bias* a posteriori, dopo che la decisione viziata è stata presa.

[27] *Discorso sullo stato dell'Unione* 2020, disponibile al link: https://ec.europa.eu/info/sites/info/files/state-of-the-union-speech_it_0.pdf.