UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN INGEGNERIA INFORMATICA

"INTELLIGENZA ARTIFICIALE E DATI PERSONALI: QUESTIONI APERTE E PROSPETTIVE"

Relatore: Prof. Andrea Loreggia

Laureando: Giovanni Friso

ANNO ACCADEMICO 2021 – 2022

Data di laurea 23 Settembre 2022

"L’Intelligenza Artificiale è forse il sogno fondamentale dell’uomo, che aspira da sempre a trasformarsi da essere creato ad essere creatore. Ma proprio questa speranza disvela fino in fondo la natura irrisolta dell’uomo."(F. Pizzetti, Intelligenza artificiale, protezione dei dati personali e regolazione)

Sommario

Nell'aprile 2021 la Commissione Europea ha pubblicato una Proposta per la regolamentazione dell'Intelligenza Artificiale. Dopo aver brevemente analizzato questa tecnologia, il presente studio prende in esame questa proposta dal punto di vista della protezione dei dati personali, poiché le recenti innovazioni hanno sollevato diversi dilemmi e timori irrazionali legati alla possibilità di ingerenze nella vita privata. Questo argomento è strettamente connesso al regolamento generale sulla protezione dei dati, a cui verrà dato ampio spazio nella tesi. Seguendo l'approccio europeo, anche il governo italiano ha avviato una discussione sull'intelligenza artificiale e la sua regolamentazione. Cambridge Analytica e Strava sono discussi nell'opera come due casi noti relativi a questa tematica.

Indice

Capitolo 1

Introduzione

Negli ultimi tempi i sistemi basati sull'intelligenza artificiale (IA) hanno conosciuto una diffusione sempre più intensa grazie allo sviluppo di nuovi algoritmi basati sul *machine learning* e il *deep learning* e alla possibilità di accedere a una notevole quantità di dati forniti dalla rete [63]. In molti casi questa tecnologia è stata accostata ad altre invenzioni che hanno segnato in maniera indelebile la storia dell'umanità, quali il motore a vapore e l'elettricità [13], che hanno permesso progressi repentini e notevoli, rivoluzionando la società e l'industria. Tuttavia, questi due esempi differiscono in modo sostanziale dall'IA, poiché non sollevavano problemi legati all'etica e alla normativacosì complessi come quelli che quest'ultimo ritrovato tecnologico ha sollevato [73, 79] e continua a sollevare [101, 29].

In questo studio si intende intelligenza artificiale un insieme di *"sistemi che mostrano comportamenti intelligenti analizzando il loro ambiente e intraprendendo azioni con un certo grado di autonomia per raggiungere obiettivi specifici"*, come è stata definita dal Parlamento Europeo [52].

Questa sorprendente spinta dell'innovazione ha in alcuni casi migliorato il benessere di parte della società e continua a stimolare la ricerca, dall'ambito medico [46] a quello farmaceutico [99], dall'industria [53] alla salvaguardia dell'ambiente [102] (tenendo anche in considerazione l'impatto che l'IA può avere su di esso [88]).

Tuttavia, questo incredibile balzo tecnologico nasconde anche molte insidie. Senza arrivare a predire futuri catastrofici e distopici che molta letteratura fantascientifica [3], tanto cinema hollywoodiano e parte dell'informazione mainstream cercano di proporre, è necessario porre attenzione alla regolamentazione da seguire in modo che l'IA non sia uno strumento che possa recare danno all'essere umano o, più in generale, all'umanità [65].

Come ben sintetizzato nel documento dell'*European Parliamentary Research Service "The impact of the General Data Protection Regulation (GDPR) on artificial intelligence"* [83] del giugno 2020 l'IA fornisce immense opportunità, ma anche rischi concreti quali *"la disoccupazione, l'ineguaglianza, la discriminazione, l'esclusione sociale ingiustificata, la sorveglianza e la manipolazione"*.

Lungi dal voler effettuare uno studio completo ed esaustivo della materia che risul-

ta articolata e ramificata in molti settori, da quello legale [34] a quello prettamente filosofico [92], da quello sociologico [44] a quello economico [47], l'argomento che verrà trattato concerne la relazione tra i sistemi basati sull'intelligenza artificiale, la protezione dei dati e l'approccio che si è voluto seguire in Europa [25]. Inoltre, si vuole dareun ulteriore sguardo alla proposta avanzata in Italia [28], come diretta declinazione di quella comunitaria.

La protezione dei dati è un tema strettamente legato alla tecnologia e all'uso chese ne fa, dal momento che mondo reale e virtuale si confondono in quell'*onlife* in cuile informazioni sono alla base dell'economia, e dove *online* e *offline* si distinguono difficilmente [43].

Un esempio di tecnologia che nei prossimi anni potrà cambiare le nostre vite, rendendo ancora più radicale questo intreccio, è il Metaverso promosso da Zuckerberg. Di questo progetto, ancora in fase embrionale, si conosce ancora poco. *"Il metaverso dovrà essere "persistente", cioè dovrà rimanere online anche quando l'utente si disconnette e torna nel mondo reale; dovrà essere "in sincrono e dal vivo", che significa che tutte le persone che vi partecipano dovranno fare esperienza delle stesse cose allo stesso tempo, e non dovrà mettere limiti al numero di persone che vi possono partecipare"*[18]. Questo estratto, mutuato dalla spiegazione sul medesimo argomento data dall'imprenditore Matthew Ball[5], ci proietta verso un mondo che sarà, ancor più di adesso, connesso in modo pervasivo, non percependo la differenza tra realtà e realtà aumentata, in cuila nostra esistenza sarà sempre più pervasa da una tecnologia sempre più sofisticata, che, proprio per via della sua complessità, potrebbe creare zone d'ombra normative, dove si possono annidare possibili rischi. Infatti, ciò a cui in questi ultimi anni si è cercato di ovviare, specialmente in Europa, è all'opacità di questi processi tecnologiciè uno dei principali ostacoli che una volta superato renderebbe possibile l'attribuzione delle responsabilità in caso di lesioni, discriminazioni, manipolazioni o, per l'appunto, violazioni della privacy [73].

La governance di questi processi e sistemi è da qualche tempo un obiettivo di ricercatori, policy-makers e altri stakeholders. Proprio per andare in questa direzione, nel 2016 la Commissione Europea ha emanato un regolamento in materia, il *General Data Protection Regulation* (GDPR) [71], con cui si è gettata una base normativa avanzata per la protezione dei dati. Questo sarà il punto di partenza per una riflessione più specifica sui rischi per i dati dell'utente in un sistema di IA.

1.1 Approccio europeo e regolamentazione

Il 21 aprile 2021 la Commissione europea ha emanato una proposta di regolamentazione dell'IA, in modo da *"guidare lo sviluppo di nuove norme globali per garantireche l'IA possa essere considerata affidabile"*, come è stato dichiarato da Margrethe Vestager, Vicepresidente esecutiva per Un'Europa pronta per l'era digitale, ponendo l'accento sulla necessità di una tecnologia etica per una leadership europea a livello mondiale. Come sottolineato da Thierry Breton, Commissario per il Mercato interno,

l'IA è vista come un mezzo, non un fine, quindi, si vuole tutelare con questo insieme di norme il cittadino che ne fa uso, incentivando una IA antropocentrica, sostenibile, sicura, inclusiva e affidabile [26]. La regolamentazione ha seguito un approccio basato sul rischio, definendone quattro gradi, da inaccettabile a minimo.

Si ha un rischio inaccettabile qualora tali sistemi fossero qualificati come *"chiara minaccia per la sicurezza, i mezzi di sussistenza e i diritti delle persone"*, attraverso la manipolazione per aggirare il libero arbitrio dell'essere umano o l'applicazione di un punteggio sociale. Soffermando l'analisi su quest'ultimo sistema di controllo, è noto come in Cina sia stato teorizzato e poi messo in pratica un *Social Credit System* (SoCS), in modo da monitorare il comportamento dei cittadini e delle imprese nel territorio. Questa metodologia è stata proposta già negli anni Novanta, per decretare se una determinata società rispetti o meno i pagamenti e agisca in modo affidabile. Nata quindi in ambito finanziario, il SoCS, messo in pratica a partire dal 2014 in alcune città designate che stilano una lista nera di persone e imprese inaffidabili, si è evoluto comprendendo anche i privati cittadini e spaziando in quasi ogni ambito della loro vita come si può vedere in figura 1.1. Specialmente con l'avvento della pandemia di Covid-19 questo controllo si è reso ancora più capillare: per esempio, nella città-pilota di Anqing si è provveduto ad aggiungere alla lista nera, tutti coloro che hanno postato sui social video di ambulanze che trasportavano casi sospetti di Covid. In altre città si è invece provveduto a inserire nei criteri di determinazione del punteggio sociale il rispetto delle norme di prevenzione della pandemia, come l'utilizzo della mascherina [38]. Essendo questo sistema fortemente limitante i diritti della persona, potrebbe comportare rischi di non poco conto per la libertà e la privacy. Proprio per questo motivo l'Unione Europea ne ha vetato l'utilizzo.

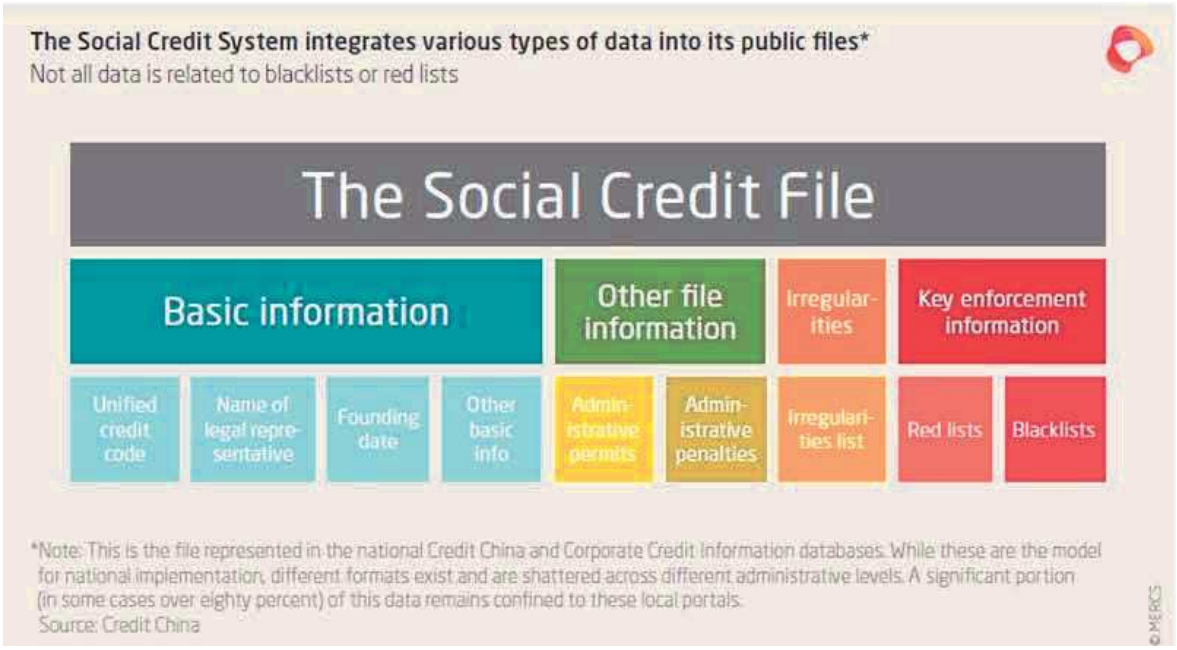

Figura 1.1: Tabella delle informazioni che concorrono a determinare il SoCS in Cina © MERICS

Il rischio diviene alto per tutti quei sistemi concernenti infrastrutture critiche, istru-

zione, componenti di sicurezza dei prodotti, occupazione lavorativa, accesso al lavoro autonomo, servizi essenziali, attività di contrasto, gestione della migrazione e amministrazione della giustizia. In tutti questi ambiti i sistemi di IA, prima di essere immessi nel mercato, dovranno rispettare rigorosi obblighi, cui possono essere ammesse deroghe solo in casi eccezionali.

Si abbassa la potenziale rischiosità dei sistemi di IA a limitata, qualora si esigano determinati obblighi di trasparenza, per esempio deve essere richiesta un'autorizzazione a procedere ad un'operazione, rendendo consapevole l'utente del servizio di interagire con una chatbot.

Infine, il rischio è classificato come minimo per i sistemi che rappresentano un rischio nullo o limitato per la sicurezza dei cittadini, come per i filtri antispam basati sull'IA.

1.2 Piano dell'opera

Lo studio è suddiviso in quattro parti:

Capitolo 2:
In questo primo capitolo si offre una panoramica dell'intelligenza artificiale, dalla sua nascita nel secolo scorso fino alla sua espansione e al suo odierno ampio utilizzo.

Capitolo 3:
Questo capitolo propone l'annosa *quaestio* della tecnologia e della sua regolazione, che, con il suo rapido sviluppo, non ha sempre permesso una altrettanto rapida risposta dell'ente legislativo. Inoltre tale sezione permette di determinare come vada regolatain maniera positiva l'IA, senza ostacolarne lo sviluppo, ma, al contempo, prevenendo possibili criticità che potrebbero emergere dall'uso della stessa. La sezione introduceil focus della trattazione, vale a dire la tutela dei dati personali, che è stata assicurata dall'applicazione del Regolamento Generale sulla Protezione dei Dati (GDPR). Nel merito verranno discussi alcuni articoli che assumono un significato particolare in riferimento alle applicazioni dell'intelligenza artificiale, così pervasivamente presente nella vita odierna.

Capitolo 4:
Il terzo capitolo concerne le più recenti proposte in ambito europeo e italiano, analizzandole in confronto con la protezione dei dati sensibili e la tutela della privacy dei cittadini. In questi ultimi mesi, infatti, sono stati emessi due documenti fondamentaliper il prosieguo nel nostro paese e nella Unione Europea dello sviluppo di sistemi basati sull'intelligenza artificiale: la *"Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'Unione"* datata 21 Aprile 2021, sulla cui scia si è inserito il *"Programma Strategico per l'Intelligenza Artificiale (IA)*

2022-2024", riguardante l'Italia, uscito il 24 novembre scorso.

Capitolo 5:

In questa ultima sezione si analizzano alcuni casi più o meno famosi che, nel recente passato, hanno portato alla luce il problema dell'utilizzo improprio dell'IA. Si affronteranno quindi casi come quello di Cambridge Analytica e Strava, dove, nel primo caso, i dati sensibili degli utenti di Facebook vennero usati per manipolare intenzionalmente gli utenti stessi come stabilito dalle sentenze, mentre nel secondo l'applicazione rivelava segreti militari, come il perimetro delle basi statunitensi, senza che fosse stata presa in considerazione questa eventualità. Tuttavia questo capitolo conclusivo non vuole solo evidenziare casi degeneri, ma portare esempi di come le più recenti norme in materia siano state rispettate da un'ampia maggioranza di aziende nel settore.

Capitolo 2

Intelligenza artificiale e sue declinazioni

2.1 Intelligenza artificiale: breve introduzione

L'intelligenza artificiale è, spesso, soprattutto dalle persone non esperte in materia, fraintesa con un insieme di tecniche incomprensibili, atte a relegare l'uomo in uno stato di inferiorità rispetto alla macchina [75]. Se è vero che questo rischio è presente, non si deve dimenticare che l'utilizzo di sistemi che si basano sull'intelligenza artificiale ha il notevole pregio di aiutare l'uomo nello svolgimento di mansioni rischiose o impossibili per lui, quindi deve essere quest'ultimo a supervisionare ed impedire che una tale eventualità si verifichi. Questa incomprensione è dovuta al fatto che è stata fin da subito ambigua la sua definizione.

2.1.1 Storia ed evoluzione

L'interesse e la ricerca in tale ambito si erano già destate da qualche anno, quando A. M. Turing, nell'articolo del 1948, dal titolo programmatico "Intelligent machinery" [96], annuncia di voler investigare se fosse possibile costruire una macchina che mostrasse un comportamento intelligente. Nel 1950 Turing stesso, vero padre fondatore non solo dell'IA, ma dell'informatica tutta, in un articolo [97], questa volta con un taglio più filosofico, propose un test, che ancora oggi porta il suo nome, per determinare se la macchina fosse dotata di una intelligenza propria (2.1).
Tuttavia non è stata questa importante personalità che ha introdotto il termine intelligenza artificiale; infatti si riferiva alla capacità da parte di una macchine di elaborare un comportamento autonomo con il termine *machine intelligence*. Fu l'americano J. McCarthy, che in una proposta del 1955 per un seminario che si sarebbe tenuto nel 1956, usò per primo questa definizione, affermando che ogni aspetto dell'apprendimento e ogni altra caratteristica dell'intelligenza possano essere descritti così precisamenteche una macchina possa simularli [60].
Dato che il problema si presentava troppo complesso nella sua interezza, si è deciso di suddividerlo in sottocategorie; in questo modo, per lungo tempo, la robotica rimase

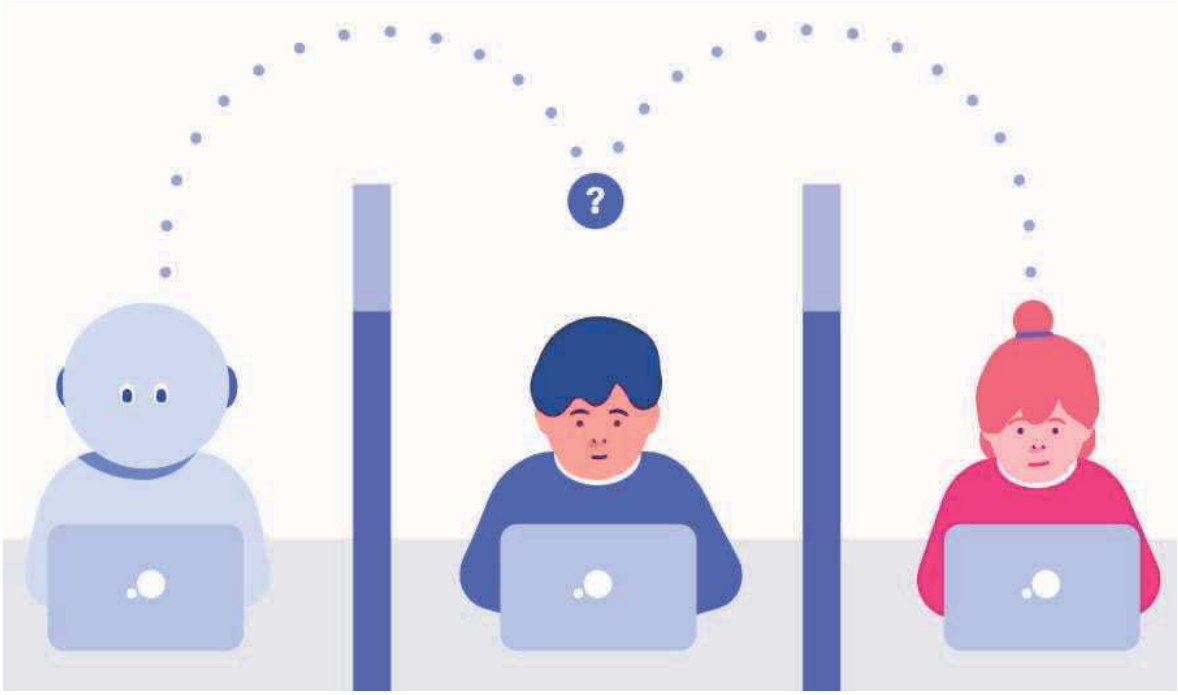

Figura 2.1: Il test di Turing: all'interrogante (figura centrale) sono sottoposte le risposte a uno stesso quesito da parte di un umano e un sistema di IA; se non riesce a determinare quale dei due sia la macchina, allora essa ha raggiunto un livello di intelligenza pari a quella umana © Elements of AI

slegata dall'intelligenza artificiale. Fino agli anni '60 si è cercato di creare un'IA generale, che permettesse di risolvere problemi di qualsiasi natura, ma tale sforzo fu vano, perché presto si capì che si doveva ridurre il campo di appartenenza del problema per aumentarne l'efficacia; l'epoca che va dal 1974 al 1980 costituì il primo inverno dell'IA, con risultati scarsi e conseguente diminuzione dei finanziamenti. Dunque dall'IA generale si passò all'IA stretta, che persegue la stessa ambizione, ma in un dominio molto più ristretto. Questo cambiamento comportò una rinnovata fioritura e un impiego massiccio dei sistemi intelligenti negli anni Ottanta, quando venivano chiamati sistemi esperti, perché derivavano la conoscenza dagli esperti del settore; molti ambiti trassero beneficio dall'introduzione di questo strumento: il settore della diagnosi medica, quello finanziario, quello della prospezione geologica. Le aspettative così alte per questo enorme progresso sociale ed economico furono però deluse dalla fragilità in cui si trovava all'epoca la tecnologia. Infatti il periodo compreso tra il 1988 e il 1993 coincide con il cosiddetto secondo inverno dell'intelligenza artificiale, poiché, essendo venuta meno la fiducia in tale ambito da parte della società, si ridussero i finanziamenti in questo settore. Tuttavia, proprio in quegli anni si continuava a fare ricerca in questo settore soprattutto sulle reti neurali fino ad arrivare alla nuova fusione con la robotica con la nascita della *cognitive robotics* [16].

Il fuoco della ricerca non si era ancora spento e covava sotto le braci, tant'è che dal '93 al 2011 si assistette a una nuova esplosione e un rinnovato entusiasmo nei confronti dell'intelligenza artificiale, dal momento che la tecnologia dell'hardware e l'ideazione di nuovi algoritmi portarono a successi in svariati settori; la guida automatica, i sistemi biometrici, il mondo dei videogiochi trassero profondo beneficio dall' applicazione di

sistemi basati sull'IA. Fino ad arrivare ai tempi attuali, dove la possibilità di avere dataset formati da quantità enormi di dati, i cosiddetti big data, ha comportato un ulteriore sviluppo e affinamento dell'intelligenza artificiale. L'evento fondamentale che ha evidenziato questo incredibile sviluppo nel campo della Computer Vision si è verificato nel 2012, quando AlexNet, una Convolutional neural network, ha vinto in maniera schiacciante un contest di classificazione di immagini [55].

Come si è visto, benché molti ricercatori, ammaliati dalla possibilità di creare effettivamente una macchina intelligente come l'essere umano, si siano profusi nell'avvicinarsi a questo obiettivo poco realistico nel corso dell'intera vita di questo settore, non si è riusciti a sviluppare un sistema in grado di operare al di fuori di uno specifico ambito, con specifici task da raggiungere. È dunque importante sottolineare come gli enormi progressi individuati in ambiti specifici non implichino passi in avanti verso una intelligenza generale che possa svolgere tutte le attività che un essere umano è in grado di fare [14]. Tuttavia, non essendo possibile fare previsioni sul futuro a lungo termine, in quanto la rapidità dell'innovazione tecnologica in questi ultimi anni ha sorpreso anchei più fiduciosi analisti, non è detto che in un qualche momento storico si possa giungere alla creazione di una macchina super-intelligente in grado di manipolare chi la utilizza. La domanda piena di preoccupazione, che in molti si chiedono, è se si sia raggiunta una singolarità tecnologica, cioè se si sia arrivati al punto che il progresso tecnologico ha superato la capacità di comprendere degli esseri umani moderni. Attualmente esistono computer capaci di *raw computing power* di molto superiore a quella del cervello umano, però ciò non significa avere un comportamento intelligente. Per questo motivo nel gennaio del 2015, molte personalità coinvolte a vario titolo nella ricerca e nell'attuazione di tali sistemi, come Hawkins, Musk, Etzioni, Rossi, LeCun, hanno firmato la lettera *"Research Priorities for Robust and Beneficial Artificial Intelligence: An Open Letter"*, in cui pongono l'accento sulla necessità del controllo dell'IA in modo che la macchina non prenda il sopravvento sull'uomo [81].

2.1.2 Base di conoscenza e ragionamento

Agli albori dell'intelligenza artificiale, ragionando sul sistema di apprendimento umano, si individuarono due processi da ricalcare nella macchina per permettere la risoluzione di problemi in autonomia: la conoscenza del mondo e il ragionamento che permette di derivare ulteriore conoscenza. Queste due questioni sono gli obiettivi dei sistemi basati sulla conoscenza, che possono essere implementati tramite approccio dichiarativo o procedurale. L'approccio procedurale è costituito dall'esplicita volontà di indirizzare un comportamento direttamente nel codice, mentre quello dichiarativo permette alla macchina di derivare le azioni tramite ragionamento, definendo una base di conoscenza (knowledge base KB) costituita dalla conoscenza di base e di una specifica istanza del problema, nonché dagli obiettivi dell'agente.

Fu quest'ultimo il *modus operandi* adottato per la simulazione del metodo d'apprendimento umano, perché permetteva l'acquisizione di nuova conoscenza a partire

da una KB effettuando il processo di ragionamento tramite un motore di inferenza. Nel caso degli agenti intelligenti la KB e il mondo reale sono messi in relazione da sensori e apprendimento, usando la logica. La logica, mutuata dalla filosofia, è lo studio delle condizioni secondo le quali un'argomentazione risulta corretta; tuttavia sorge un problema di non poco conto: il linguaggio naturale è ambiguo. Per questo motivo vennero creati linguaggi logici formali, attraverso la logica proposizionale e quella del primo ordine, formati da una sintassi rigida e una semantica che definisse i valori di verità di ogni formula in ogni possibile modello, applicabili in maniera non ambigua. Tali linguaggi hanno permesso lo sviluppo di motori di inferenza, che sono algoritmi di inferenza, cioè procedure derivanti formule a partire da premesse date.

Questi algoritmi, costituiti da regole di inferenza che rappresentano ogni singolo passo del ragionamento, la cui correttezza è dimostrabile facilmente, determinano se è possibile o meno aggiungere un'ulteriore formula alla base di conoscenza. Nel caso della logica proposizionale tali procedure sono molto dispendiose dal punto di vista computazionale, poiché questo linguaggio è poco espressivo e poco conciso. Per questo motivo si passò alla logica dei predicati, più ricca di sfumature e meno verbosa [66].

2.1.3 Problema della ricerca e del soddisfacimento dei vincoli

Una larga parte dei problemi di IA è legata alla ricerca, per cui sono state ideate diverse strategie che esplorano lo spazio degli stati, formato da uno stato iniziale, un insieme di azioni eseguibili dall'agente e un cammino, definito dal passaggio tra uno stato e il successivo per arrivare al goal. Un algoritmo in tale ambito, avendo come input un problema, restituisce una soluzione come sequenza di azioni, che possono essere eseguite in un secondo tempo. Dall'algoritmica si conosce che ad ogni algoritmo di ricerca è associato un albero di ricerca, il quale rappresenta tutti gli stati in cui la macchina può trovarsi, svolgendo una determinata azione. Vi sono alcune strategie per scorrere tutto l'albero, come *breadth first* e *depth first*, cioè per ampiezza e per profondità rispettivamente, ma gli agenti intelligenti possono sfruttare la conoscenza sul problema per evitare di dover visitare l'intero albero. Si hanno così le strategie informate o euristiche, con l'introduzione di funzioni di valutazione che danno una stima computazionale dello sforzo per raggiungere l'obiettivo. Questi metodi hanno diverse origini, da quelli più tecnici come l'algoritmo A* [54] o l'*Hill climbing* [95], alle cosiddette metaeuristiche come la *ANT colony optimization* [35], derivata dalla zoologia, e gli algoritmi genetici [45], che sfruttano le conoscenze in genetica, per esempio applicando il crossover per determinare i figli di nodi genitori in maniera che posseggano una parte del DNA di un genitore e l'altra dell'altro.

Cercare di risolvere il problema del motore di inferenza ragionando sulla ricerca non è l'unico approccio possibile per definire una strategia valida al fine di determinare le azioni che un sistema di IA deve intraprendere; dato che deve sottostare ad alcuni vincoli, nel tempo si sono sviluppati metodi che pongono l'accento su questo tipo di problemi: i *Constraint Satisfaction Problems* (CSP), definiti su un insieme di variabili

appartenenti a domini finiti e su un insieme di vincoli, che legano tra di loro le variabili. Questi problemi presentano diversi metodi di soluzione, tra cui quelli che usanoi vincoli a posteriori, come il *Generate and Test* e lo *Standard Backtracking* (2.2), e gli algoritmi di propagazione, che cercano di prevenire i fallimenti nell'ottenimento delgoal, piuttosto che effettuare costosi backtracking, cioè tornare indietro nell'esecuzionedelle istruzioni.

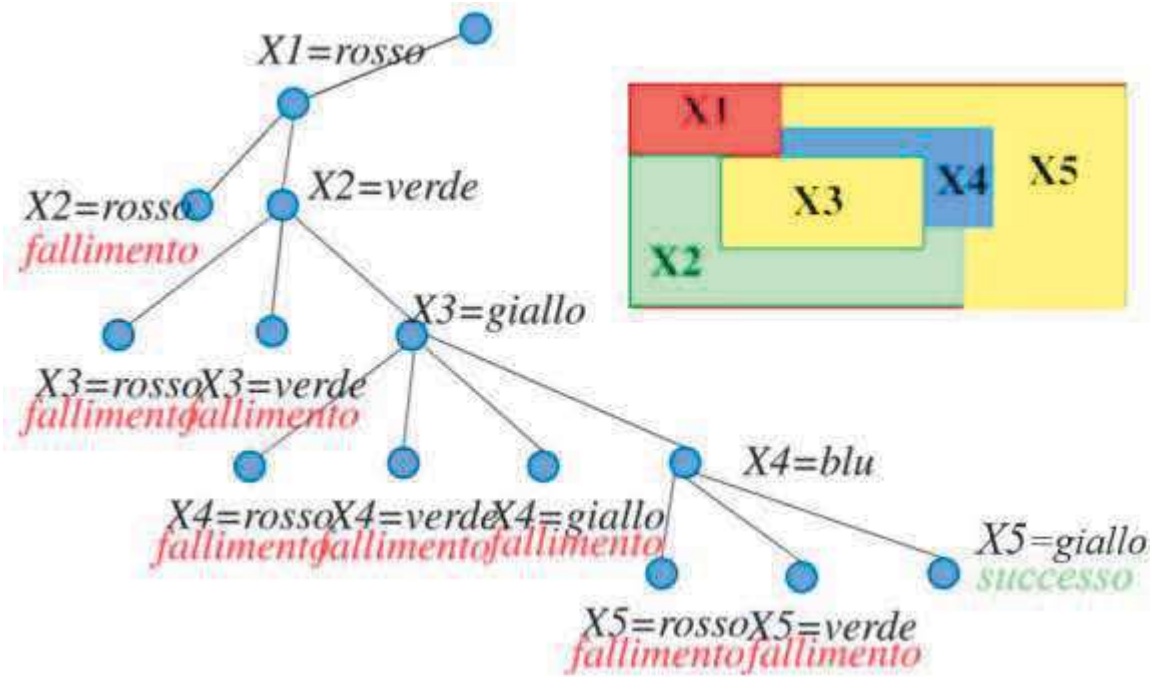

Figura 2.2: Esempio di Standard Backtracking per un problema di Map Coloring; come si può vedere vi sono numerosi fallimenti, prima di giungere al goal © Università di Bologna

Tecniche più performanti sono, dunque, il *Forward checking* (2.3), che elimina, dal dominio delle variabili non ancora istanziate, il valore appena assegnato alla variabile corrente e si controlla la compatibilità dei vincoli contenenti la variabile appena assegnata con le precedenti e le successive, e il *Look Ahead* (2.4), tecnica più completa della precedente, che, oltre ad effettuare i controlli del *Forward checking* , riduce i domini delle variabili non ancora istanziate, propagando anche le relazioni contenenti coppie di variabili non ancora istanziate. In questo caso le euristiche che vengono applicate sono di due categorie: euristiche per la selezione della variabile, che consiste nell'assegnare per prima la variabile con dominio di cardinalità minore (*first-fail*) o nello scegliere la variabile legata a più vincoli (*most-constrained principle*), o quelle per la selezione del valore, che consiste nello scegliere il valore con più alta probabilità di successo.

2.2 Stato dell'arte dell'intelligenza artificiale

Nei settant'anni di vita dell'intelligenza artificiale si sono susseguite molteplici tecniche [87] per modellare una macchina in modo che potesse apprendere e simulare un cervello

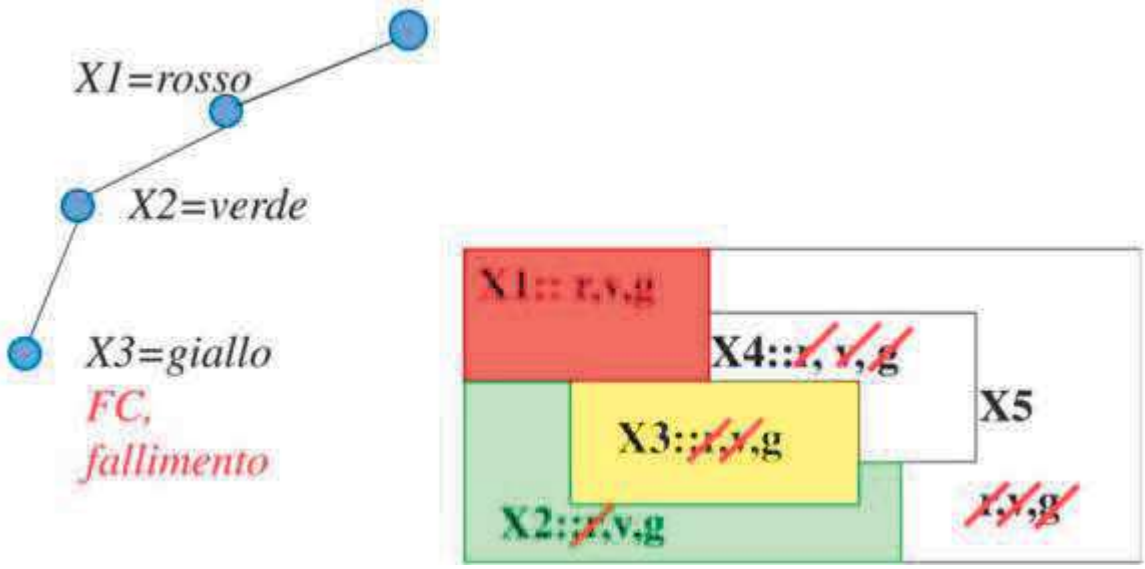

Figura 2.3: Esempio di Forward Checking per un problema di Map Coloring © Università di Bologna

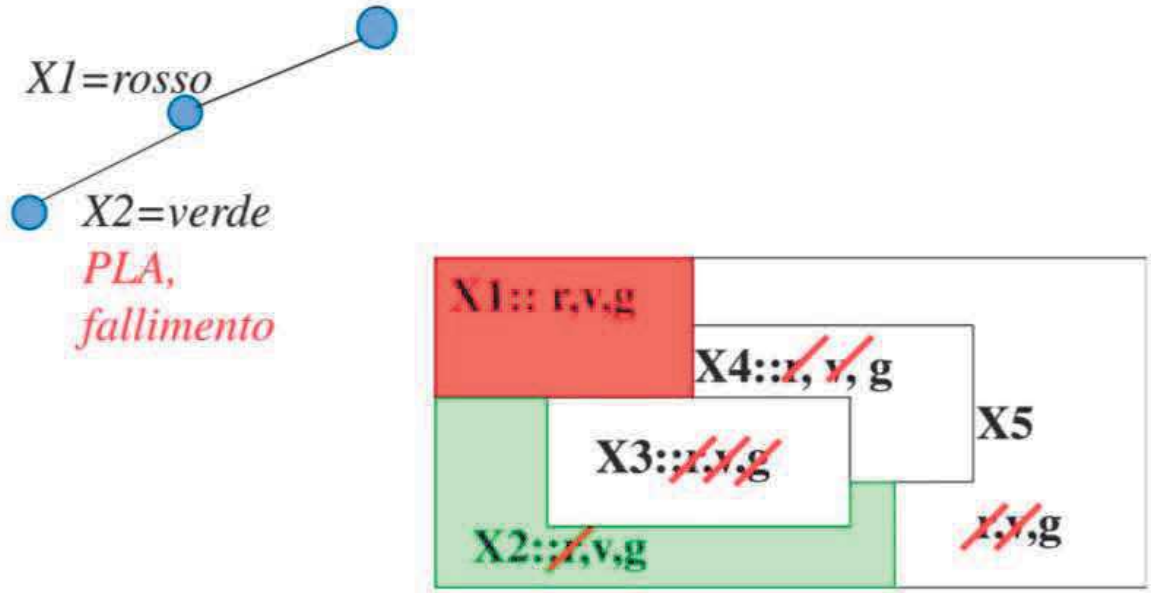

Figura 2.4: Esempio di Partial Look Ahead per un problema di Map Coloring © Università di Bologna

umano [37]. Attualmente, in ragione delle notevoli quantità di dati disponibili, grazie alla rapida e capillare diffusione dei dispositivi elettronici, si sono raggiunti obiettivi che hanno del fantascientifico in ambiti quali la risoluzione di immagini nella *computervision* [104], la telechirurgia [46] e la guida autonoma [32]. Per consentire una più ampia comprensione di come sia difficile giungere a queste incredibili innovazioni nei prossimi paragrafi si indagheranno le modalità attraverso le quali si sono sviluppati e si continuano ancora oggi a sviluppare sistemi capaci di apprendere e quindi agire in modo autonomo. Certamente, dato lo spazio ridotto in questo studio, non si esaurisce l'intera trattazione che ognuno dei sopraccitati argomenti richiederebbe, ma consente di capire quali siano le tecniche più diffuse attraverso cui l'elemento artificiale apprendee le sue più iconiche applicazioni.

2.2.1 *Machine learning* e apprendimento automatico

Uno degli ostacoli più difficili da aggirare nella costruzione di una macchina "pensante" è come permettere l'acquisizione di conoscenza a partire dai dati. Come visto per altre terminologie che afferiscono all'IA anche in questo caso la parola inglese *machine learning* presenta ambiguità, non tanto per la prima parte, quanto per la seconda, non essendo sempre identificabile in maniera certa cosa s'intenda per *learning*, per apprendimento. Questa poca chiarezza nel linguaggio naturale specchia alla perfezione come nemmeno per un essere intelligente come l'uomo possa esserci a volte la piena certezzadi comprendere ciò che si vuole dire; trasporre quindi le caratteristiche umane in una macchina risulta ancora più complesso in quanto due persone diverse per cultura ed etnia potrebbero dissentire sul significato di un termine e quindi programmare in modo diverso la risposta che deve fornire l'entità artificiale. "Acquisire conoscenze e abilitàe averle prontamente disponibili dalla memoria in modo da poter dare un senso a problemi e opportunità futuri" [12] può essere una valida definizione di apprendimento,come lo è "un cambiamento persistente nelle prestazioni [...] [che] deve avvenire come risultato dell'esperienza e dell'interazione dell'agente con il mondo" [39]. Entrambe queste definizioni sono valide e confermano quanto visto in precedenza in modi leggermente diversi. Nella prima si fa riferimento alla memoria di quanto appreso per agire anche a input differenti da quelli già visti in precedenza, mentre nella seconda si parla di mondo esterno, di esperienza e di come essa possa cambiare indelebilmente il mododi comportarsi. Per una macchina, che funziona su dati, a loro volta formati da innumerevoli bit, si presenta dunque la difficoltà di rappresentare il mondo che la circonda e determinare quale azione intraprendere in base alle conoscenze pregresse. Queste asperità hanno contribuito a definire quindi in modo decisamente differente il machine learning, per esempio come campo che studia l'abilità del programma di imparare senza essere esplicitamente programmato; questa definizione però risulta troppo generica e di vecchia data, risalendo al '59 [82], quando ancora si sperava in una intelligenza artificiale generale. Come sopra detto, essendosi delineata con più forza l'IA ristretta, ciò ha portato a una definizione più dettagliata e rigorosa nella quale si evidenzia come un programma possa apprendere dall'esperienza E rispetto a una classe di compiti specifici T e una misurazione di performance P, se la sua prestazione in compiti in T, misurato da P, migliora con l'esperienza E [61]. In questo caso quindi si fa riferimento a come l'esperienza aiuti l'agente in modo da aumentarne il livello prestazionale; in pratica si impara dall'esperienza, ma si vuole che, al contrario degli uomini che imparano anche dagli errori, il sistema sia infallibile. Ciò stride con il fatto che l'intera tecnologia basata su IA è sviluppata da esseri umani, quindi fallibili; si tratta dunque di un incapsulamento dell'errore, che all'utente non esperto potrebbe sembrare dovuto alla macchina, ma è nella maggior parte dei casi riconducibile a uno sbaglio umano. Gli errori che il programmatore può causare sono molteplici e possono avere connotazioni fortemente discriminatorie [36]; una delle cause principali di comportamenti sbagliatida parte di una IA è costituita dalla poca attenzione data alla composizione del dataset su

cui si intende addestrarla [33].

I dati sono muti senza l'uomo che ne conferisca un significato in modo che la macchina possa apprendere le azioni da svolgere prima con la sua supervisione e poi autonomamente. Per essere più precisi in questo campo si utilizza il termine pattern in luogo del termine dati, tant'è che una parte del machine learning, detta per l'appunto *pattern recognition*, si occupa di inventare nuove tecniche per distinguere le diverse tipologie di pattern [10]. Questa parola, pattern, presenta molteplici sfumature, che ancora una volta, se ce ne fosse bisogno, sottolineano l'ambiguità e l'intraducibilità dei concetti legati a quest'ambito della tecnologia che abbracciano più significati contemporaneamente. Infatti all'interno di questo termine si possono inserire tipi di dato anche molto diversi tra di loro e complessi, quindi costituiti da altri dati, quali un'impronta digitale, un segnale sonoro, un volto, un carattere scritto a mano. Per mettere ordine in questo guazzabuglio di pattern li si è categorizzati: ci sono dunque quelli numerici, che sono rappresentabili come vettori numerici nello spazio multidimensionale, poi quelli categorici, associati a caratteristiche qualitative, quelli sequenziali, che permettono relazioni spazio-temporali, infine ci sono una buona parte di pattern strutturati che non possono essere facilmente inquadrati, di solito organizzati in strutture complesse come alberi e grafi. Molte strutture naturali quali le proteine e il DNA permettono facili trasposizioni numeriche sfruttabili per l'addestramento di sistemi in grado di processare notevoli quantità di catene amminoacidiche per scoprire anomalie o nuove cure mediche [76]. Un esempio di questo genere che tutti noi abbiamo sotto gli occhi, ma che non viene mai troppo riconosciuto per i sentimenti contrastanti che l'uso dell'intelligenza artificiale suscita, è stata l'individuazione della proteina spike del retrovirus del Covid-19, che ha permesso di capire il funzionamento di questa infezione e costituito un fondamentale passo verso la messa in produzione di un vaccino efficace [99].

Per avvicinarsi ancora un poco al cuore del problema dell'apprendimento, cruciale per comprendere in maniera più nitida il funzionamento dell'intelligenza artificiale, si deve capire come l'elemento artificiale possa apprendere da un insieme di pattern. Infatti, i pattern dati alla macchina permettono di effettuare classificazioni, che li mappano in differenti classi a seconda delle loro caratteristiche (*features*), regressione, che assegna loro un valore continuo per consentire stime in uno spazio bidimensionale, clustering, che individua gruppi di pattern con proprietà simili. Dunque, come si è appena visto, il problema della rappresentazione dei pattern passa per le loro features e il loro apprendimento automatico tramite i *row data*, punto centrale sviluppato dalle tecniche di *deep learning*, che saranno trattate in seguito.

Per questo motivo l'insieme di dati su cui il sistema andrà ad operare è suddiviso in tre sottoinsiemi che prendono il nome di Training Set, Test Set e Validation Set. Il primo è l'insieme di dati su cui si addestrerà l'intelligenza artificiale prima del suo completamento; esso può essere già etichettato, in modo che i suoi dati appartengano già a una determinata classe, oppure può non contenere nessun sottoinsieme, lasciando al sistema la possibilità di addestrarsi. Esiste un terzo approccio tra apprendimento supervisionato e non supervisionato, che prende il nome di semi-supervisionato, ove il

training set è etichettato solo in parte ed etichettarlo tutto richiederebbe un notevole sforzo economico, perché il lavoro di classificazione a priori deve essere svolto da un esperto umano del settore in cui successivamente l'IA andrà ad operare. Un esempio di apprendimento semi-supervisionato è costituito dall'*active learning*, che, mentre è in esecuzione l'algoritmo, ad ogni sua iterazione sceglie i pattern più utili per il sistema e li etichetta. Non è l'unico tipo di apprendimento inusuale, infatti vi è anche l'approccio multilabel, che a ogni pattern si associano più classi, come nel caso di un articolo di giornale che può discettare di più ambiti, trattando il medesimo argomento.
Il Test Set è l'insieme di pattern su cui si valutano le prestazioni finali del sistema; solitamente non si utilizza solo un metodo per minimizzare l'errore in output, dati determinati input, perché in questo modo si potrebbe avere un sistema molto efficiente da un punto di vista computazionale, ma che non rispetta i principi di adeguatezza richiesti per l'applicazione che se ne vuol fare.

La natura dell'apprendimento, come visto dalla sua definizione, fa molta leva sull'esperienza; quindi sorge spontanea la domanda su come è possibile dotare l'elemento artificiale di un bagaglio di conoscenza che consenta di sviluppare il cosiddetto *learning by doing*. Questo non può avvenire con un apprendimento *batch*, cioè tutto ciò che la macchina può imparare è dato una volta per tutte, solo con una sessione di training iniziale senza permettere di apprendere ulteriormente. Né tantomeno può avvenire usando un approccio incrementale, addestrando ogni tanto la macchina con altre sessioni oltre a quella iniziale. La risposta più ovvia è quella di implementare la possibilità di imparare artificialmente ricalcando il processo naturale, che consente all'essere biologico di apprendere tramite l'interazione continua con l'ambiente che lo circonda formato tanto da oggetti inanimati quanto da suoi simili. Come visto in altri campi, risulta però moto più complesso simulare la realtà attraverso il digitale; quindi, risulta molto più difficile degli altri due approcci sopra menzionati, perché comporta una commistione tra approccio supervisionato e non supervisionato. Per modellare in modo più vicino alla percezione umana il problema e permettere l'apprendimento di un comportamento in base alle esperienze passate si è sviluppato negli anni il *reinforcement learning*, che associa al risultato positivo dell'azione compiuta dalla macchina una ricompensa, mentre lo penalizza se commette un errore. La robotica, attraverso queste tecniche, ha fatto notevoli progressi, perché questo approccio ha consentito, in ambiti ristretti, l'ottimizzazione implicita o esplicita dei parametri per poi agire sulla loro base, come la minimizzazione dell'errore e la massimizzazione del risultato. Questi non sono gli unici parametri sulla cui base valutare se il sistema di intelligenza artificiale lavora correttamente e in modo efficace; nel caso di un problema di classificazione binaria si utilizzano le frequenze di pattern collocati giustamente nelle classi di appartenenza e quelle di pattern erroneamente classificati, dando così origine alla *confusion matrix* (2.5).

Molti di questi metodi di apprendimento automatico richiedono di impostare parametri fondamentali, come il numero di neuroni di una CNN che sono allocati prima di partire con il learning: sono gli iperparametri. Proprio in virtù del loro ruolo essen-

	Positivo previsto	Negativo previsto	Formule
Positivo effettivo	True Positive (TP)	False Negative (FN)	Sensitività $\frac{TP}{(TP+FN)}$
Negativo effettivo	False Positive (FP)	True Negative (TN)	Specificità $\frac{TN}{(TN+FP)}$
Formule	Precisione $\frac{TP}{(TP+FP)}$	Valore negativo previsto $\frac{TN}{(TN+FN)}$	Accuratezza $\frac{TP+TN}{(TP+Tn+FP+FN)}$

Figura 2.5: Matrice di confusione, tramite cui si valutano le prestazioni della classificazione binaria

ziale per l'intero algoritmo è loro dedicato il Validation Set visto in precedenza,per permettere alla macchina di tararli in modo adeguato. Se tale insieme di dati è rappresentativo del Test Set, allora si deve massimizzarne l'accuratezza; se i gradi di libertà del classificatore sono eccessivi si raggiungerà una elevata accuratezza nel Training Set, ma non sul Validation Set, dando luogo ad *overfitting*, che potrebbe farne deperire rapidamente l'accuratezza. Dunque risulta buona regola cominciare con pochi gradi di libertà, controllabili tramite iperparametri, ed aumentarli ad ogni iterazione dell'algoritmo, monitorando l'accuratezza sul Training e sul Validation Set.

Nel corso della sua vita il machine learning si è dotato di diverse pratiche per arrivare alle più moderne e sviluppate. Quindi, partendo dallo studio della statistica, si è ideato un approccio Bayesiano, basato per l'appunto sul conosciuto teorema di Bayes, che ha dato vita a un classificatore molto efficiente sul piano teorico se sono note le distribuzioni iniziali [93]. Tuttavia nella pratica non sono disponibili le densità di probabilità condizionali, quindi si sono trovate due soluzioni, seguendo un approccio parametrico o non parametrico [68]. Nel primo caso, adatto per training set di dimensioni non elevate, si assumono ipotesi sulla forma delle distribuzioni e si apprendono i parametri fondamentali dal Training Set, mentre nel secondo si apprendono le distribuzioni direttamente da tale insieme di dati.

2.2.2 *Deep learning* e reti neurali

Le reti neurali prendono nome e spunto da quella che risulta una modellizzazione, e quindi una semplificazione, del neurone biologico. L'idea di applicare le conoscenze mediche alla realizzazione di una intelligenza artificiale è stata una delle prime intuizioni della ricerca nel 1943 addirittura prima del conio del termine IA ed esemplifica in modo inequivocabile la notevole fiducia che risiedeva in questo ambito. Tuttavia, come evidenziato in 2.1.1, la teoria abbisognava di una tecnologia che solo in anni moltosuccessivi, negli anni '80, ha potuto dare uno sviluppo pratico soddisfacente, quando si sono sviluppati algoritmi di training efficaci. In principio, nel 1956, Rosenblatt, psicologo statunitense, schematizzò

dunque una rete di neuroni artificiali, ognuno con i propri dendriti, le proprie connessioni sinaptiche e il proprio assone, costituita da un solo livello di input e uno di output di un neurone: era nato il percettrone, la prima rete neurale [78]. Ognuno dei neuroni di input è collegato al neurone di output e scambia con esso informazioni pesate in modo differente per dare, appunto, maggiore peso ai neuroni più importanti. All'interno del neurone che riceve queste informazioni tramitei rami pesati è presente una funzione di attivazione che determina quale sarà il suo output, solitamente si usa la funzione sigmoide, ma nel caso del percettrone si adoperò una funzione di attivazione lineare a scalino, molto limitante perché adatta ad apprendere solo mapping lineari. Negli anni ovviamente le reti neurali hanno aggiunto non solo neuroni nei livelli di input e di output, ma si sono creati livelli intermedi, andandoa creare le cosiddette *Multilayer perceptron* (MLP), permettendo di ottenere reti *feed-forward*, le cui connessioni collegano i neuroni di un livello precedente esclusivamentecon quello successivo, senza che siano possibili connessioni all'indietro. Accanto a questa tipologia si è sviluppata la sua duale, la rete *recurrent*, che quindi permette anche connessioni di *feedback*, ma risulta più complessa e meno utilizzata della precedente perché complica notevolmente il flusso di informazioni e l'addestramento, in quanto richiede di considerare il comportamento di uno stesso neurone in più istanti. L'output che queste reti consentono di ottenere corrisponde ad un mapping, che può consisterein un'etichetta da assegnare ai pattern di input se si tratta di classificazione o in un valore corretto della variabile dipendente corrispondente alle variabili indipendenti di input se si vuole effettuare una regressione [85].

Procedendo negli anni con lo studio delle dinamiche di funzionamento del cervello umano, si è potuto applicare quanto scoperto anche nell'ambito dell'intelligenza artificiale; da questa ampia contaminazione tra tecnologia e medicina, si sono sviluppate le *deep neural network*, reti neurali profonde nel senso che sono costituite da almeno quattro livelli, uno di input, uno di output e almeno due livelli intermedi. Ciò ricalca in modo pedissequo, per quanto conosciamo del nostro cervello, l'elaborazione che deve compiere tale organo per riconoscere un'immagine tramite il sistema visivo; come nelcaso della vista, però, non è necessario un numero di livelli esageratamente alto per ottenere reti neurali dall'elevata efficienza, dal momento che si è sperimentato comeaumentare il numero di livelli superiore ai 50 aumenta le prestazioni di poco a discapito dell'efficienza. Allo stesso modo non è possibile aumentare a dismisura il numero di neuroni e quello dei pesi, perché appesantirebbero di molto il training e, in particolare, i meccanismi di feedback e di feed-forward [1].

Avendo sempre come principale differenza la modalità di apprendimento, anche le reti neurali si dividono in modelli con training supervisionato, non supervisionato, ricorrenti. Molto conosciuta è una particolare rete neurale con training supervisionato, la CNN (*Convolutional Neural Network*), introdotta da LeCun nel 1998, grazie alla quale si sono ottenuti risultati notevoli in problemi di piccole dimensioni [56]. Le performance della CNN sono agevolate dal processing locale, che permette la sola connessione locale del neurone al livello precedente, riducendo il numero di connessioni, e dai pesi condivi-

si tra più neuroni dello stesso livello, riducendo il numero di pesi. Inoltre, la convoluzione, da cui tale rete neurale prende il nome, rappresenta una notevole miglioria soprattutto nel riconoscimento di immagini, perché, tramite un filtro digitale, consente la riduzione del numero di neuroni utilizzati.
Tuttavia l'ambito delle CNN ha conosciuto un aumento delle dimensioni di applicabilità e uno sviluppo particolarmente accelerato a partire dal 2012, anno in cui la CNN AlexNet ha battuto in modo schiacciante le altre intelligenze artificiali in un contestdi Image Detection [80], battendo il secondo classificato del 10% nella probabilità di errore nella classificazione di immagini; negli anni seguenti in cui tale competizione èstata attiva, cioè fino al 2017, si è arrivati a un errore di classificazione del 2,3% circa,che è di molto inferiore all'errore umano pari al 5% [89]. Questo risultato sorprendente si è ottenuto grazie a un addestramento su una quantità enorme di dati etichettati, i Big Data, che hanno permesso la distinzione di milioni di immagini in migliaia di classi diverse, grazie all'elevata potenza di calcolo delle GPU (*Graphic Power Unit*), che hanno abbassato da mesi a giorni il tempo di training, e grazie all'applicazione della funzione di attivazione Relu. Si potrebbe pensare che comunque anche un training digiorni possa essere dispendioso e sicuramente lo è, ma, una volta effettuato, il tempo richiesto per la classificazione di nuovi pattern risulta molto veloce. Per abbattere anche il tempo di training iniziale si può ricorrere al *Transfer Learning*, che consente il riutilizzo delle features e delle reti, le quali risultano pre-addrestate (*pre-trained*). Le applicazioni di queste particolari tecnologie non si limitano all'ambito della classificazione di immagini, ma presentano una molteplicità di possibili settori in cui operare, quali il *natural language processing* [64], la guida autonoma [20], la diagnosi medica
[31] e la bioinformatica [105].

2.2.3 Visione futura

È indubbio che la nostra società si sta sviluppando in modo da dare ampia ragione a quello che Luciano Floridi prospetta nel suo libro "La quarta rivoluzione: come l'infosfera sta cambiando il mondo" [43], in cui la separazione tra ciò che è virtuale e ciò che è reale non è più così marcata, per il fatto che la prima si sta avvicinando sempre più alla seconda, creando un connubio con essa e modificandola in modo importante. Notevole spinta a questo cambiamento la sta dando l'applicazione di sistemi di intelligenza artificiale che in questi ultimi anni hanno permesso un'elaborazione grafica e digitale di gran lunga superiore a tutti gli altri tipi di tecnologia, applicando nuovi algoritmi a partire da enormi agglomerati di dati. Tutto ciò ha spinto anche osservatori estraneia tematiche scientifiche a interessarsi dell'IA, come i giornali più diffusi; di questi fa parte il *The Guardian*, che in un articolo pubblicato lo scorso febbraio ha riportato l'eccezionale ruolo che questa tecnologia ricopre nella salvaguardia delle specie animali a rischio [48].

Attraverso delle telecamere disseminate alle entrate del parco nazionale di Kafue, in Zambia, si sono potuti localizzare, tramite Image Detection, bracconieri, che, col favore

delle tenebre, cercavano di introdursi illegalmente in esso tramite l'uso di imbarcazioni. Non è l'unica applicazione utile alla salvaguardia del pianeta che viene affrontata grazie ad applicazioni di intelligenza artificiale: per esempio, si monitora il livello di acqua dei bacini idrici del Brasile, cosa impensabile da compiere senza l'ausilio di elaborazioni satellitari delle immagini per la natura selvaggia che ricopre la maggior parte del paese.

Tutte queste applicazioni di IA potrebbero far pensare a una loro limitazione all'ambito specifico del riconoscimento di immagini, ma non è il solo settore dove l'IA contribuisce al miglioramento della società in cui siamo immersi. La capacità di ragionamento ha permesso a sistemi basati su tale tecnologia di battere in giochi sempre più complicati gli esseri umani più dotati: si era partiti con gli scacchi [15] e si è arrivati, notizia di pochi mesi fa, a Gran Turismo [7]. Questo avvenimento potrebbe essere ritenuto marginale, ma non bisogna sottovalutarlo, perché, trattandosi di un gioco di corse di automobili, potrebbe trarne vantaggio anche lo sviluppo di auto a guida autonoma, grande tematica che vede le case automobilistiche effettuare test continuativi pur di lanciare sul mercato la prima macchina che si guida da sola.

Da segnalare che in questi ultimi anni abbiamo potuto assistere a un fatto che ha dell'incredibile: l'individuazione della proteina spike causa del Covid-19 e la produzione praticamente simultanea di un vaccino adatto a difenderci da questo virus [99]. Anche in questo caso il merito è soprattutto dell'IA, che ha consentito alle industrie farmaceutiche di ridurre notevolmente il periodo di prova del vaccino, effettuando il test su poche persone volontarie e consentendo al mondo intero una rapida somministrazione.

Altro ambito in cui si è potuta apprezzare l'applicazione dell'intelligenza artificiale è l'archeologia; in questo caso si sono adoperate reti neurali per trovare nuovi siti archeologici tanto terrestri, quanto subacquei, ricostruire testi frammentari e ceramiche spezzate, dopo averle addestrate su reperti già rinvenuti e classificati [57]. Altri algoritmi sono stati sviluppati per studiare i resti umani, determinandone il sesso a partire dal cranio, l'altezza tramite la lunghezza delle ossa. Non solo per la ricerca archeologica sul campo sono stati sviluppati strumenti basati sull'IA, ma anche per consentire di progettare guide-robot interattive per i musei.

Se è vero quindi che il nostro tempo ha tratto notevole conforto dalle più svariate applicazioni dell'IA, si deve sempre tenere conto dell'altra faccia della medaglia: si sviluppano continuamente nuove armi con l'ausilio di questa prodigiosa tecnologia [84]. Purtroppo, il progresso umano si misura anche in armamenti e la situazione ucraina ne è lo specchio atroce [103]; in questo scenario sono stati rilevati anche attacchi guidati da sistemi autonomi, ma non è detto che questo possa essere l'unico conflitto in cui si adopereranno se non se ne vieterà o limiterà l'uso [30]. Anche in questo caso viene sottolineata la centralità dell'essere umano, ultimo e unico vero arbitro delle conseguenze a cui le sue scellerate azioni possono portare; si è sviluppato un pensiero secondo cui tali sistemi autonomi debbano sempre essere soggetti a un controllo umano significativo, pur tuttavia avendo contorni sfumati: si parte dal dare espressamente ordini all'arma verso quale obbiettivo si deve dirigere l'ordigno o il proiettile di cui è dotata per arrivare a un controllo del processo decisionale per cui l'arma autonoma ha

intrapreso una determinata decisione [92]. Quest'ultima azione è ad ogni modo molto irrealistica, se non irrealizzabile, data l'opacità di sistemi dotati di milioni di parametri pressoché impossibili da processare da una mente umana, come spiegato nei paragrafi precedenti. Inoltre in casi di crimini di guerra sarebbe molto difficile determinarne la responsabilità, qualora fossero stati compiuti da un'arma *loitering* quale un drone che autonomamente scegliesse il suo obbiettivo ed agisse senza una supervisione o finanche un avvallo umano; in questo modo si interromperebbe la catena di responsabilità che in casi analoghi viene attribuita dal soldato semplice fino al suo superiore più alto in grado che ha dato l'ordine causa del crimine[92].

Dunque non ci si può più esimere dal prendere una posizione forte ed inequivocabile in grado di cessare o perlomeno limitare la produzione e la ricerca in questo ambito tramite una convenzione internazionale per evitare che tali sistemi creino una seria minaccia per la sopravvivenza stessa del genere umano, se, per esempio, fossero usati come controllori di missili con testata nucleare [92]. Da apprezzare la proposta avanzata dalla Croce Rossa Internazionale [30], che ha sollevato preoccupata la questione, chiedendo a tutti i paesi del mondo di vietare l'utilizzo delle armi autonome letali e quelle imprevedibili, termine che però suona ambiguo, dal momento che ogni sistema di IA può essere ritenuto imprevedibile per un osservatore umano, in quanto può avere comportamenti non decifrabili dall'uomo.

Tuttavia, questo non è l'unico dilemma etico che solleva l'intelligenza artificiale. Per esempio, l'impronta antropica che il nostro pianeta deve sopportare è anche dovuta ai reiterati addestramenti che si fanno dell'IA sempre in continua espansione [88]; in questo caso il dilemma è come ridurre le emissioni di anidride carbonica dovuta a questo fatto senza pregiudicare il progresso e i benefici che l'intelligenza artificiale può portare. Non di meno interessante ed allarmante è il fatto che sono aumentati in maniera massiccia i rischi collegati all'uso di dati personali per l'addestramento, dal momento che le nostre vite sono sempre più pervase da queste tecnologie, caratterizzate da una molto spiccata avidità di dati [83]. Il prosieguo dell'attuale trattazione avrà come focus quest'ultimo ambito così delicato e importante per il nostro mondo iperstorico, analizzando la proposta di regolamentazione europea e quella italiana che sono state avanzate per consentire la creazione di una IA affidabile e antropocentrica.

Capitolo 3

Diritto e tecnologia

3.1 Legge e tecnologia

Le innovazioni tecnologiche hanno suscitato e suscitano in molti casi sentimenti contrastanti, da una parte un entusiasmo sfrenato, dall'altra una paura irrazionale [75]. Il primo di questi stati emotivi si manifesta nelle persone sempre attente all'ultima novità e disposte a pagare qualsiasi cifra pur di accaparrarsi un bene di cui, magari, non avrebbero neanche bisogno. Il secondo, invece, è ben esemplificato dal Frankenstein di Mary Shelley, in cui l'atmosfera gotica contribuisce alla diffidenza verso la tecnica, vista come una sorta di sortilegio. Benché siano passati più di due secoli dalla pubblicazione di questo romanzo, rappresenta ancora nella mentalità odierna il paradigma dello scienziato pazzo, che cerca di manipolare a suo piacimento il mondo che lo circonda e la natura [4]. Fuor di metafora gli informatici più visionari sono visti come autentici guru, che tramite le loro invenzioni consentono di manipolare non solo la natura, ma lo stesso essere umano. Proprio per evitare questi eccessi si è cercato e si sta cercando di regolamentare il settore tecnologico in tutto il mondo, ma specialmente in Europa, dove è più presente un'attenzione ad una legislazione incentrata sull'essere umano [24].

3.1.1 La rapida evoluzione tecnologica ha impattato contro la lentezza della legiferazione? Legge e IA sono inconciliabili?

Si ha sovente l'idea che l'incontro tra legge e tecnologia debba essere necessariamente uno scontro tra una forza irresistibile e un'altra inamovibile. Come ben spiegato dalla professoressa Julie E. Cohen, questa visione è affetta da due errori di notevole importanza. Da un lato si ha la convinzione che la tecnologia sia una forza irresistibile, che il legislatore non riuscirà mai a governare. Tuttavia, coloro che hanno questa visione, forse per scarsa informazione, ignorano che *"le tecnologie dell'informazione sono altamente configurabili, e la loro alta configurabilità permette molteplici spunti da parte di terzi interessati e con buone risorse per delineare il loro sviluppo"* [21]. Il secondo errore è costituito dall'idea che la legge sia un'entità immobile, ma come ribadito più volte dagli studiosi di diritto, essa sta già rispondendo in modo reattivo ai cambiamenti anche molto repentini apportati dalla tecnologia, cercando sempre di sviluppare una

legislazione agile e puntuale. A comprova di questa affermazione si può vedere quante proposte siano state avanzate, approvate e messe in pratica in tema di regolamentazione di strumenti tecnologici in questi ultimi anni.
Il diritto si trova in una situazione simile all'epoca compresa tra il tardo diciottesimo e la metà del secolo scorso, quando l'industrializzazione sempre più crescente è stata una fonte di conflitto, che ha plasmato l'azione legislativa in modo massiccio, permettendo agli stati di ammodernarsi. Quindi, seguendo la stessa procedura, oggigiorno *"la proprietà delle risorse dell'età dell'informazione e la responsabilità per i danni dell'età dell'informazione sono diventate fonti pervasive di conflitto"*, sviluppando nuove rivendicazioni circa i diritti e le responsabilità che conseguono dall'utilizzo notevole delle moderne tecnologie. Proprio per questo motivo, si assiste all'emergere di istituzioni legali atte alla regolazione del problema, ma la loro forma e la loro sostanza rimangono indeterminate e nebulose [21].

Da qualche tempo a questa parte l'economia politica dei paesi sviluppati ha intrapreso una trasformazione dal capitalismo industriale a quello informazionale. Come indicato da Manuel Castells [17], il capitalismo è detto informazionale quando costituisce una modalità di produzione con l'informazionalismo come strumento di sviluppo. Con informazionalismo s'intende un "modello di sviluppo in cui la fonte principale della produttività è la capacità qualitativa di ottimizzare la combinazione e l'impiego di lavoro, capitale e risorse naturali sulla base della conoscenza e dell'informazione". Dunque in alcuni casi la fonte di guadagno principale, su cui l'attuale capitalismo fa affidamento, non è più legata in prima battuta alla manifattura, ma è orientata principalmentealla produzione, all'accumulazione e al processo dell'informazione. In altri casi, purnon costituendo il fine ultimo dell'azienda, le tecnologie legate all'informazione hanno trasformato profondamente anche l'attività industriale tradizionale. Da queste considerazioni si deduce che urge una nuova o più efficace regolamentazione della tecnologia e, in particolare, del suo prodotto che forse più degli altri ha contribuito ad allarmare la società civile, vale a dire l'intelligenza artificiale. Come ben puntualizzato dalgruppo di esperti che ha prodotto l'ultimo documento per il governo italiano in materia Proposte per una strategia italiana per l'intelligenza artificiale[62] *"scrivere regole puntuali per tecnologie come l'IA fotografandone l'evoluzione attuale equivale a scrivere sulla sabbia: alla prima ondata tecnologica, è necessario ricominciare da capo"*. Quindi, l'azione da sviluppare nel campo deve essere agile e lungimirante, definendo alcuni principi cardine, sufficientemente invarianti nel medio periodo. *"Dal "governo" si passa dunque alla governance della tecnologia, che implica strumenti flessibili e adattabili al contesto, nonché una costante cooperazione tra istituzioni, università e società civile"*.

A livello europeo si è passati da una riflessione sull'etica dell'IA ad un più ampio dibattito sull'affidabilità della stessa [24], basata sulla necessità di garantire il rispetto dei diritti fondamentali dell'individuo e lo sviluppo sostenibile per un miglioramento dell'intera società, perché ci si è adoperati per soluzioni che salvaguardino i valori europei, ma al contempo siano agili e multi-stakeholder.

3.1.2 Questione sulla regolazione dell'IA: l'inefficacia delle tre leggi della robotica di Asimov

Riamane da specificare una questione importante: a chi deve essere riferita la responsabilità in caso di violazioni delle leggi riguardanti l'IA?
Per rispondere a questa fondamentale domanda si cita soventemente l'idea che aveva a riguardo Isaac Asimov [3], padre fondatore di quella fantascienza ricca di droidi, cyborg e robot che hanno in seguito popolato il panorama della letteratura e del cinema. Asimov, scrittore dalle solide conoscenze scientifiche, essendo stato anche un insigne biochimico, può essere considerato il Jules Verne del XX secolo, perché, come è vero che le mirabolanti invenzioni immaginate dall'autore francese sono state effettivamente seguite dalla loro realizzazione, così le idee avanzate da Asimov sono state rivalutate e tuttora sono parte integrante della ricerca soprattutto in intelligenza artificiale. C'è da fare tuttavia un distinguo chiaro per evitare fraintendimenti: lo scrittore in questione ha immaginato i robot dotati di un cervello positronico, in grado di attribuire agli automi un livello empatico, provando emozioni e sentimenti del tutto simili a quelli degli esseri umani. Questo tipo di cervello è ancora una chimera impossibile da realizzare e forse rimarrà sempre tale, date le scarse conoscenze in materia di cervello umano, che trovano ancora difficoltà a spiegare con chiarezza scientifica sufficiente l'origine delle emozioni e degli stati d'animo. La sua idea che in questo caso si prende in esame e che più ha influenzato il dibattito sulla IA è costituita dalla sua visione della regolazione dell'uso dei robot e degli obblighi cui questi devono attenersi. Nello specifico Asimov ha espresso una propria posizione riguardo la questione della responsabilità, in cui tende a privilegiare il lato macchina. Infatti nelle sue famose tre leggi della robotica non viene nominato mai l'umano, se non come parte lesa:

1. *Un robot non può recar danno a un essere umano né può permettere che, a causa del suo mancato intervento, un essere umano riceva danno*;
2. *Un robot deve obbedire agli ordini impartiti dagli esseri umani, purché tali ordini non vadano in contrasto alla Prima Legge*;
3. *Un robot deve proteggere la propria esistenza, purché la salvaguardia di essa non contrasti con la Prima o con la Seconda Legge* [3].

Come si può intuire già ad una prima lettura, in queste tre leggi si suppongono esseri artificiali dotati di una autonomia decisionale paragonabile a quella umana e capaci di discernere tra azioni benevole o nocive, cioè ciò che costituisce uno dei dilemmi più dibattuti in sede filosofica ed etica anche per l'uomo moderno. Questo fatto è dovuto alla questione sopra riportata del cervello positronico e della conseguente vicinanza caratteriale del robot asimoviano all'essere umano. L'approccio seguito da Asimov cerca di superare quello che lui chiamava complesso di Frankenstein, secondo cui i robot sono intrinsecamente minacciosi o malvagi, per aumentare la fiducia del pubblicoverso questi artefatti. Quello che però risulta chiaro dalle tre leggi della robotica è la loro natura vaga e indefinita che porta molti dei personaggi delle storie dello scrittore

a scontrarsi con esse e a fornire diverse interpretazioni delle stesse. Attraverso questo costante conflitto si arriva a formulare una nuova legge fondamentale, la legge zero, la cui retro-compatibilità deve essere assicurata dalle altre tre leggi: *Un robot non può recare danno all'umanità, né può permettere che, a causa del proprio mancato intervento, l'umanità riceva danno* [3]. Ovviamente si sta parlando di un frutto della fantasia dell'autore e un espediente letterario che rivela il suo notevole acume, ma potrebbe essere realmente preso in considerazione nella stesura di un corpus giuridico che regoli l'ambito dell'intelligenza artificiale? Possono certamente essere prese come spunto queste leggi nate dalla penna di Asimov, ma devono essere fatte modifiche sostanziali al loro impianto, perché si deve porre al centro della salvaguardia giuridica l'essere umano e non il robot, mero strumento frutto della sua tecnica per quanto avanzata essa possa essere. Quindi esse dovranno essere modellate attorno alla figura tanto dello sviluppatore di agenti intelligenti e degli algoritmi su cui si basano, quanto della persona che, in un secondo momento, li adopererà. Infatti la società che si sta andando a delineare è caratterizzata da un uso pervasivo di tali strumenti tecnologici che, con un felice termine coniato da J. Balkin, si sta trasformando in una "società algoritmica" [4].

3.1.3 Leggi per delineare una società algoritmica: approccio *bottom-up* e *top-down*

La presunzione alla base della società algoritmica risiede nello sfruttamento dei dati attraverso algoritmi per governare e migliorare la società. L'ambizione sottesa a questa definizione è l'onniscienza, sapere e quindi predire tutto, un'ambizione antica quanto l'uomo. Da questa aspirazione derivano potenziali rischi, che vanno dai danni fisici, alle violazioni della privacy, dalla discriminazione alla manipolazione. Tuttavia, il punto più critico è costituito dalla asimmetria di potere informazionale che si viene a determinare tra utilizzatori dell'IA e loro programmatori o esperti del settore, derivante dalle maggiori conoscenze nel campo di questi ultimi. Per evitare che questi significativi problemi possano comportare un notevole intralcio allo sviluppo di questo tipo di società, Balkin ha introdotto, sulla falsa riga delle tre leggi di Asimov, tre principi per regolarne la vita:

1. *Per quanto riguarda i clienti e gli utenti finali, gli utilizzatori dell'algoritmo sono fiduciari informazionali*;

2. *Rispetto a coloro che non sono clienti e utenti finali, gli utilizzatori degli algoritmi hanno compiti pubblici. Se sono governi, ciò deriva dalla loro natura di governi. Se sono attori privati, lo sono i loro affari interessati da un interesse pubblico*;

3. *Il dovere pubblico centrale degli utilizzatori dell'algoritmo è di evitare di esternalizzare i costi (danni) delle loro operazioni*.

Da questi requisiti derivano gli obblighi di trasparenza, interpretabilità, giusto processo e responsabilità che caratterizzano i temi più dibattuti presso la comunità scientifica e giuridica [4].

Il primo principio si basa sulla figura del fiduciario che ha due obblighi fondamentali: il dovere della cura e quello della lealtà. Il primo dovere trova la sua applicazione nell'evitare di danneggiare il cliente col proprio operato, mentre il secondo nell'evitare conflitti di interessi con i propri clienti. Ciò accade nelle professioni di medici e avvocati, cui il paziente o cliente affida la propria salute o la propria difesa in ambito giuridico. Nella società algoritmica si sono venuti a creare altri fiduciari, molto più potenti dei precedenti, che presentano alcune affinità, quale la asimmetria della conoscenza tra fiduciario e cliente, ma anche differenze, perché vi afferiscono molti più utenti: le piattaforme digitali. Mentre il fiduciario tradizionale è tenuto dalla propria deontologia professionale al segreto e al riserbo sulle condizioni di salute del paziente o sulle confessioni del proprio cliente, senza conseguire un ritorno economico effettivo, i dati personali sono monetizzabili facilmente dalle compagnie come Facebook, Google *et similia*. In secondo luogo, i social network e i motori di ricerca hanno interesse che l'utente riveli quanto più possibile sulla sua personalità in modo da personalizzare il suo habitat virtuale, differentemente dalla discrezione cui si attengono i medici o gli avvocati. La terza differenza fondamentale tra fiduciari tradizionali e informazionali sta nel servizio atteso dall'utente: un paziente non si aspetta solo di non essere danneggiato dal proprio dottore, ma anche di ricevere consigli e avvertimenti su potenzialirischi. Le persone non si aspettano obblighi di assistenza così completi dai loro motori di ricerca e siti di social media.

Dunque, benché i colossi del web non debbano sottostare alle stringenti regole di deontologia professionale dei professionisti, non dovrebbero allo stesso tempo pretendere di offrire un sistema sicuro e rispettoso della privacy e, in un secondo momento, agire in modo discriminatorio o addirittura manipolatorio nei confronti dei loro utenti finali o rivendere i loro dati personali ad altre aziende dall'operato poco chiaro. La seconda legge discende direttamente dalla prima: dato che i governi sono fiduciari nei confronti delle persone che governano, tanto più se utilizzano sistemi informatici e algoritmi sono fiduciari informazionali nei confronti delle persone che governano. Il principio espresso dalla prima legge non esaurisce il problema, perché non tutte le compagnie private che operano nel mondo digitale e usano algoritmi, agenti IA o robot sono fiduciari informazionali. Soprattutto nell'ambito della robotica i danni che un malfunzionamento può arrecare possono coinvolgere anche persone che non sono utenti e non hanno relazioni contrattuali con tali imprese. Da ciò deriva una responsabilità pubblica dell'azienda che opera nel settore, quando si avvalgono dell'utilizzo di robot o sistemi di IA.

Per spiegare la ragione della terza legge, Balkin suggerisce un'analogia con i "fastidi" pubblici, quali il fumo, i rumori e l'inquinamento. Tradizionalmente questi erano problemi legati all'utilizzo di proprietà reali, ma quest'idea si è ampliata a comprendere anche un'ampia gamma di danni. Un "fastidio" pubblico può recare danni per un indefinito numero di persone. Certamente i danni legati all'uso fraudolento o improprio

di algoritmi non ricalcano i "fastidi", come tradizionalmente sono intesi, ma presenta notevoli somiglianze. Per esempio, non si può asserire che un algoritmo sia malevolo a priori, trattandosi di uno strumento, quindi programmabile da un essere umano; è dunque assimilabile all'inquinamento acustico, che dipende in prima battuta dall'utilizzo improprio di uno stereo, che, di per sé, non avrebbe queste intenzioni. Una seconda ragione del parallelismo è costituita dalla portata del danno algoritmico, che, come il danno dovuto all'inquinamento, ha differenti gradi di gravità. Un terzo argomento che accomuna i due problemi è dato dall'insieme di attori che entrambi devono averein modo da ottenere un problema significativo: come per l'inquinamento ambientale servono più automezzi, così "i danni della società algoritmica derivano da processi decisionali e da giudizi cumulativi da parte di un'ampia gamma di attori pubblici e privati". Balkin continua argomentando che il problema centrale oggigiorno risiede non tanto nella discriminazione intenzionale, quanto nei danni cumulativi all'identità, perché la società algoritmica aumenta a dismisura la rapidità, la portata e la pervasività della categorizzazione, della classificazione e della decisione, in modo da influenzare la vita degli esseri umani.

Visti quindi gli espedienti teorici sulla creazione di una società algoritmica più proporzionata, l'ultimo punto da chiarire risulta essere come sviluppare un sistema di IA in grado di seguire linee etiche conformi alla società in cui è sviluppato. In letteratura si distinguono due approcci principali, cui si affianca una terza via intermedia tra le altre due, detta per l'appunto ibrida [2] [40]. La prima che si è sviluppata è costituita dall'approccio *top-down*, in cui un codice comportamentale fondamentale è inserito all'interno dell'agente artificiale e ad esso si attiene scrupolosamente alla ricerca di quale azione sia la più corretta da prendere. Per corredare un sistema di intelligenza artificiale di queste basi etiche si possono seguire varie strade, come dotarlo di basi morali fondanti la società in cui si trovano, per esempio i Dieci Comandamenti, la regola aurea o le tre leggi della robotica, ma le più condivise risultano essere le seguenti due: una utilitaristica, in cui si cercano di massimizzare i benefici e minimizzare i danni, mentre l'altra segue teorie deontologiche, che si concentrano sui motivi dell'azione e richiedono agli agenti di rispettare doveri e diritti specifici, per esempio l'imperativo categorico kantiano. Entrambe le soluzioni, tuttavia, oltre a sollevare problemi enormi dal punto di vista computazionale, sollevano anche il problema secondo cui il sistema sia in grado di processare i dati, seguendo la base morale immessa, in tempo reale in modo da dare una risposta in tempi ragionevoli. Questo non è solo un dilemma per un'entità artificiale, ma anche per l'essere umano, perché le sue azioni non sono solo dettate da un insieme di regole morali cui si attiene, ma anche dalla contingenza degli avvenimenti [2].

Dati questi ostacoli molto difficili da superare, si prende in esame l'approccio complementare al *top-down*: il cosiddetto approccio *bottom-up*. In questo caso il comportamento morale non è dato a priori, codificandolo direttamente, ma segue la teoria del *learning-by-doing*, per cui l'agente artificiale apprende dall'ambiente esterno quale sia il comportamento corretto, imparando anche dagli errori [19]. Tuttavia, anche questo

approccio ha i suoi limiti; infatti, non è semplice far evolvere un sistema con queste caratteristiche né è scontato che la macchina apprenda solo comportamenti corretti (potrebbe addirittura violare alcuni principi riconosciuti, ma il più delle volte disattesi dagli utenti umani [40]).

Entrambi gli approcci incarnano differenti aspetti che costituiscono ciò che solitamente viene inteso come una spiccata sensibilità morale. Questo fatto evidenzia come sia necessario un approccio ibrido, in modo da fondere le caratteristiche; questa fusione però risulta complessa, poiché richiede di intrecciare diverse filosofie ed architetture [2]. Forse in un futuro, prossimo o lontano che sia, la potenza di calcolo sempre più sviluppata potrà aiutare la risoluzione di questa importante questione.

3.2 Regolamento generale sulla protezione dei dati (GDPR)

La questione della regolazione della tecnologia ed in particolare dell'intelligenza artificiale passa anche attraverso un'attenta analisi delle sue implicazioni nella sfera piùintima degli individui. Come rilevato da molti studiosi [43] [9] [106] [98], l'esperienza in rete ha acquisito sempre più importanza nella vita delle persone, fino, in taluni casi, a plasmare la realtà nella quale l'internauta si ritrova. Dunque è naturale che i dati personali forniti dall'utente durante la navigazione siano da intendersi effettivamente parte dell'identità della persona, data la crescita del tempo impiegato nel web (3.1), come esplicitato dal Trattato di Lisbona.

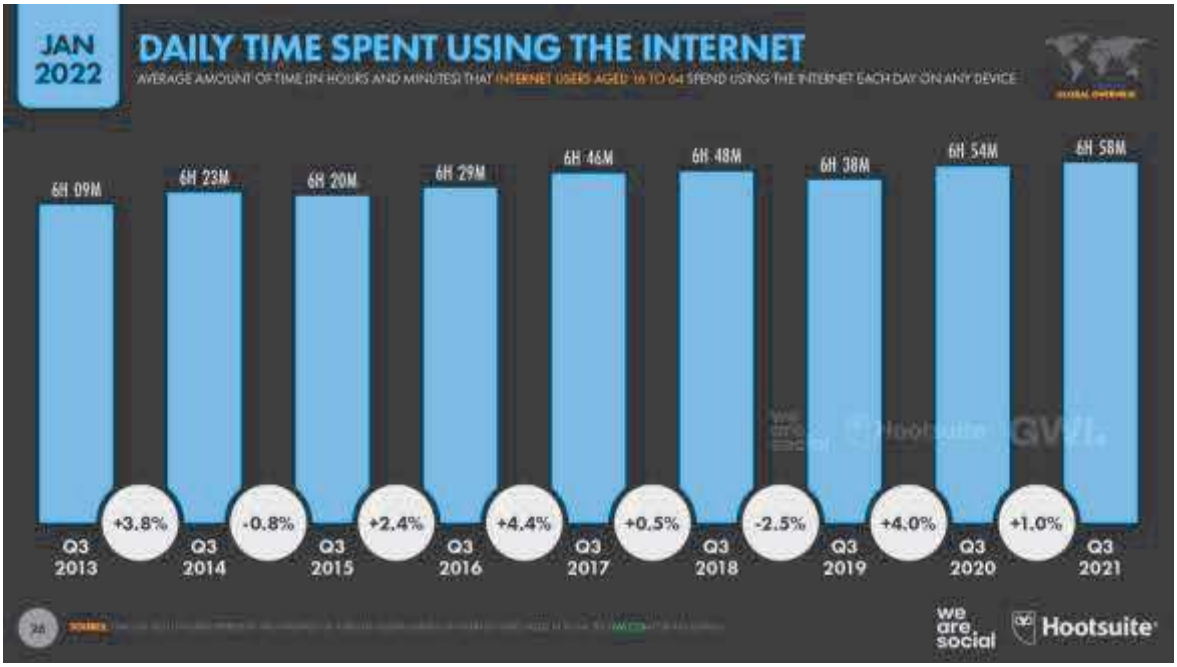

Figura 3.1: Tempo impiegato online al giorno in media © we are social

Ma tutto questo ragionamento non deve far perdere di vista il fine ultimo della qui presente trattazione: come questo traffico online può essere collegato con l'IA? Come visto nel precedente capitolo i sistemi di intelligenza artificiale richiedono una quantità

abnorme di dati per apprendere come estrarre conoscenza dagli stessi e quale migliore dataset dei big data offerti dagli utenti del web? Tuttavia non è l'unico aspetto che lega i dati personali all'IA su cui è bene vigilare; infatti i dati personali possono essere acquisiti da un sistema di intelligenza artificiale per perfezionare il proprio compito durante il suo utilizzo, comportando un rischio per l'integrità dell'identità dell'utente, ma non solo, i dati sensibili potrebbero essere utilizzati per effettuare discriminazioni immotivate o accentuare quelle già in atto [36]. Dunque solo con un approfondimento sul delicato tema dei dati personali, si può veramente comprendere come la società algoritmica abbia necessità di tutelare la loro integrità. La comunità europea ha già predisposto un regolamento in materia nel 2016, entrato in vigore nel 2018, conosciuto come GDPR [71], che risponde prontamente a questo bisogno impellente. Nei prossimi paragrafi si daranno quindi alcuni spunti su questo tema vitale per la società odierna.

3.2.1 Diritto alla protezione dei dati: un diritto fondamentale

Risulta molto utile effettuare una precisazione: la protezione dei dati, infatti, è spesso usata come sinonimo di diritto alla privacy, mentre sono due concetti che, pur afferendo alla stessa tematica, quella dei dati personali, hanno diverse sfumature. Quando si parla del primo si intende *"un sistema di trattamento degli stessi che identifica direttamente o indirettamente una persona. La sua definizione accoglie, oltre al principio di riservatezza, quelli inerenti alla disponibilità ed all'integrità dei dati personali"* [77]. Invece la privacy ha assunto un significato diverso fin dalle sue origini in terra americana nel 1890 come *"diritto ad essere lasciato solo"*, da cui deriva la definizione europea leggermente differente: la privacy *"fa riferimento al diritto alla riservatezza delle informazioni personali e della propria vita privata [. . .] Usiamo il termine privacy quando vogliamo rappresentare uno spazio personale che gli sconosciuti non possono oltrepassare"*. Dunque, in ultima analisi, *"mentre la privacy è stata costruita come un dispositivo "escludente" ovvero come uno strumento per allontanare lo sguardo indesiderato, la protezione dei dati personali mette al centro la persona in riferimento ai suoi dati perché questi costituiscono un'identità"*.

La protezione dei dati è un tema che sta interessando in modo sempre più massiccio la tecnologia e l'uso che se ne fa, dal momento che mondo reale e virtuale si stanno sempre più fondendo, andando a formare la cosiddetta onlife, come ben spiegato daLuciano Floridi, noto filosofo italiano, che ha teorizzato come nelle società attuali, definite iperstoriche (3.2), in cui le informazioni sono alla base dell'economia, non si riesca più a distinguere in quale realtà ci si trovi, *online* o *offline* [43].

Significativa è l'importanza che si è voluta dare alla tutela dei dati personali nel territorio dell'Unione Europea tramite due riferimenti espliciti nei documenti che compongono il più ampio Trattato di Lisbona. Si tratta di due articoli che discendono direttamente dalla Direttiva 95/46/CE [23], ormai non più in essere, essendo stata abrogata dall'art. 94 del Regolamento Generale per la Protezione dei Dati (GDPR) [71]. Con questa Direttiva l'Unione Europea, all'epoca Comunità Europea, con un'opera

Figura 3.2: La società iperstorica, nuova frontiera dell'umanità

legislativa lungimirante, trovandosi ancora agli albori della rete e lontano dal volume odierno di dati scambiati tramite essa, riconosceva come diritto della persona fisica quello alla vita privata, con riguardo al trattamento dei dati personali. Questa convinzione venne dunque ribadita sia all'articolo 8 della Carta dei diritti fondamentali dell'Unione Europea [70], sia all'articolo 16 nel Trattato sul funzionamento dell'Unione Europea [69], entrambi elementi costitutivi del Trattato di Lisbona, ratificato nel 2007 e approvato da tutti i parlamenti nazionali degli stati facenti parte dell'Unione nei due anni successivi. Da sottolineare la formulazione dell'articolo 8 citato, che scinde la vita privata, il cui rispetto è garantito dall'articolo 7, dalla protezione dei dati di carattere personale, a testimoniare come nei dodici anni intercorsi tra la Direttiva e la Carta l'evoluzione tecnologica abbia spinto l'ente legislatore a considerarli due ambiti separati. Inoltre questo fatto storico risulta ancora più palese nell'aver inserito la protezione dei dati sensibili nei diritti fondamentali riconosciuti alla persona fisica, cioè un diritto personalistico, che ad essa appartiene [73]. Nel solco di questa legislazione il Parlamento Europeo e il Consiglio dell'Unione Europea hanno adottato il GDPR, ribadendo fin dal primo Considerando che la protezione delle persone fisiche con riguardo al trattamento dei dati di carattere personale è un diritto fondamentale. Partendo da questo assunto si chiarisce come il trattamento dovrebbe essere al servizio dell'uomo e il diritto alla protezione dei dati non è fine a sé stesso, ma deve essere considerato alla luce della sua funzione sociale e contemperato con altri diritti fondamentali. Si argomenta anche come questo regolamento costituisca un superamento della Direttiva 95/46/CE, perché essa non ha consentito di evitare la frammentazione dell'applicazione della protezione dei dati personali nel territorio dell'Unione, né ha eliminato l'incertezza giuridica o la

percezione, largamente diffusa nel pubblico, che in particolare le operazioni online comportino rischi per la protezione delle persone fisiche, come riportato nel Considerando 9. Il GDPR consente anche il rispetto di una procedura già in atto dal 1995, che, sulla scia del Trattato di Schengen, in cui si garantisce la libera circolazione dei cittadini europei in tutti gli stati facenti parte dell'Unione, garantisce la libera circolazione dei dati all'interno dell'Unione per promuovere lo sviluppo dell'economia digitale, consolidando la fiducia delle persone fisiche e rafforzando la certezza giuridica e operativa delle regole. Il GDPR ha valenza transnazionale e comunitaria anche nei confronti dei trattamenti che riguardano cittadini dell'Unione, ma con titolari siti al di fuori dell'Unione.

Tuttavia il GDPR non solo preserva o perfeziona alcune tecniche già in atto nella Direttiva, ma aggiunge anche qualche innovazione come il diritto delle persone al controllo sui dati che le riguardano. In questo caso si supera significativamente la concezione dei dati personali come proprietà della persona che li ha generati, attuata seguendo il principio dell'autodeterminazione informativa di matrice tedesca, derivata dal concetto della tutela della dignità della persona e della sua identità [73], come era avvenuto nella Direttiva del '95. È avvenuto giustamente questo cambio di prospettiva, proprio in ragione del cambiamento economico-sociale verificatosi negli ultimi anni, in cui la struttura della società digitale consente a più soggetti tra loro connessi di trattare dati personali forniti dall'interessato per un determinato servizio, andando a creare una sorta di catena di trattamenti e, quindi, di responsabilità. Invero è che, nell'indirizzo di una visione proprietaria del dato, si offre una nuova possibilità all'interessato, vale a dire di tenere traccia dei dati riferiti alla propria persona, introducendo il diritto alla portabilità dei dati, come esplicitato nell'art. 20. La norma permette a un utente, qualora abbia dato il suo consenso o abbia stipulato un contratto, di ricevere i propri dati personali in un formato leggibile da un dispositivo automatico, per effettuare un eventuale cambio di trattamento in modo da *"ottenere la trasmissione diretta dei dati personali da un titolare del trattamento all'altro, se tecnicamente fattibile"*.

I due diritti sopra riportati, cioè il diritto alla portabilità e quello al controllo sui dati non risultano essere gli unici di cui un utente si può avvalere, ma sono previsti anche altri diritti a sua tutela: il diritto di accesso, il diritto di rettifica, il diritto alla cancellazione, meglio noto come diritto all'oblio, il diritto di limitazione dei trattamenti, il diritto all'opposizione e, molto importante in materia di IA, il diritto connesso alle decisioni automatizzate, che costituisce una novità rispetto alla legislatura precedente.

Il diritto di accesso permette all'interessato di sapere non solo se i suoi dati sono sottoposti al trattamento, ma anche quali siano le finalità di tale trattamento, quali siano le categorie di dati prese in esame, se siano destinati a terzi, quali siano le origini dei dati non raccolti presso il titolare e le garanzie nel caso in cui fossero destinati a paesi esterni all'Unione.

Il diritto alla rettifica, come si può intuire dal nome, permette all'interessato di chiedere correzioni o integrazioni al titolare del trattamento, in una sorta di collaborazione con esso, per evitare i danni che potrebbero scaturire dal trattamento di dati errati.

Il diritto alla cancellazione consente di apprezzare in maniera più chiara l'innovazione apportata dal GDPR: infatti, l'art. 17 impone che il titolare cancelli dal trattamentoi dati riferiti all'utente che ne ha richiesto l'interruzione, compresi quelli resi pubblici dal titolare, eliminando ogni link, copia o riproduzione. Ciò risulta molto complessoda attuare alla luce di quanto visto per il diritto al controllo sui dati, poiché si è venuta a creare una stratificazione di trattamenti, che obbliga il titolare cui si è rivolto l'interessato a farsi tramite della richiesta di cancellazione nei confronti di chiunque stia trattando i suoi dati personali.

Il diritto di limitazione dei trattamenti consente all'interessato di contestare l'esattezza dei dati personali; qualora però il titolare non riscontrasse questo errore, egli sarebbe libero dall'effettuare azioni correttive.

Il diritto di opposizione ha nel termine stesso la sua definizione, ma si deve chiarire quando è lecito opporsi a un trattamento: nel caso in cui il trattamento venga effettuato senza il consenso dell'interessato oppure sulla base del legittimo interesse del titolare, come nel caso del marketing diretto. Quindi tale diritto si fonda principalmente sul controllo dei dati da parte dell'interessato.

L'ultimo dei diritti sopra riportati riguarda il diritto di conoscere il procedimento logico che ha portato alla decisione automatica corrente, con particolare attenzione alla profilazione delle persone. Questo diritto è già stato sviluppato nell'art. 15 della Direttiva, ma assume un rilievo molto maggiore nel GDPR, perché sottolinea la crescita di trattamenti automatizzati che è avvenuta nel ventennio che intercorre tra Direttivae Regolamento. La profilazione sopra menzionata è un buon esempio di questa rapida evoluzione: consiste nell'aggiunta di personalità ai dati già disponibili su un utente, che nel corso degli anni è divenuta una tecnica sempre più diffusa. Inoltre tale procedura è di notevole interesse per l'applicazione dell'intelligenza artificiale, perché i procedimenti decisionali automatici messi in pratica dall'IA potrebbero agevolarla in modo molto spinto, costituendo una possibile violazione delle norme vigenti.

Per evitare queste gravi ripercussioni si può agire in modo da garantire la pseudonimizzazione, più semplice, o l'anonimizzazione, più complessa dei dati personali. L'anonimizzazione dei dati personali dell'utente non rientra nelle pratiche regolate dal GDPR, poiché tale termine si riferisce a dati che non si riescono più a collegare con certezza al cittadino che li ha generati. Tuttavia raggiungere la completa anonimizzazione è molto difficile, se non impossibile, infatti, il più delle volte, si hanno alcuni elementi che, anche in assenza delle generalità del cittadino, permettono di riferire ildato a uno specifico individuo. Invece per pseudonimizzazione s'intende *"il trattamento dei dati personali in modo tale che i dati personali non possano più essere attribuiti a un interessato specifico senza l'utilizzo di informazioni aggiuntive, a condizione che tali informazioni aggiuntive siano conservate separatamente e soggette a misure tecniche e organizzative intese a garantire che tali dati personali non siano attribuiti a una persona fisica identificata o identificabile"*, come esposto nell'art. 4(5), in cui vengono definiti vari termini utili per il prosieguo del regolamento. Esemplificando la differenza tra i due approcci, la crittografia è un metodo di pseudonimizzazione e non di anonimizzazione, come spesso

creduto, poiché le trasformazioni dei dati dovute agli algoritmi sono reversibili tramite la decrittografia e le chiavi segrete che consentonodi decodificare il messaggio possono essere ritenute come informazioni aggiuntive che permettono comunque l'identificazione. Come riportato nel considerando 26 quindi *"i dati sottoposti a pseudonimizzazione [. . .] dovrebbero essere considerati informazioni su una persona fisica identificabile"* [58].

La pseudonimizzazione, la cui finalità non è tanto l'aumento della sicurezza dei dati, quanto la diminuzione del rischio derivante dal trattamento, non è l'unico strumento possibile per arginare possibili usi fraudolenti o non voluti dei dati personali; altre tecniche possibili sono la *privacy by design*, la *privacy by default* e la minimizzazione dei dati. Le due tipologie di privacy sono indicate nell'articolo 25, dove si fa riferimento alla protezione dei dati fin dalla progettazione (by design) e per impostazione predefinita (by default); esse sono metodologie differenti, ma al contempo si sovrappongono: la prima, che il titolare deve sempre tenere in considerazione dall'inizio del trattamento alla sua fine, può variare nel tempo per essere calibrata a seconda dei rischi identificati,mentre la seconda riguarda essenzialmente la fase di progettazione dei trattamenti in modo che siano trattati, per impostazione predefinita, solo i dati personali necessari per ogni specifica finalità del trattamento [73]. La privacy by default opera in basead un automatismo inserito nel trattamento stesso e scatta alla sua messa in atto;essa quindi costituisce un valido supporto per garantire la tutela dei dati in tutti quei sistemi, come l'intelligenza artificiale, in cui l'automatismo ne è la parte preponderante. La minimizzazione dei dati è invece un principio più generale, cui si fa riferimento all'art. 5, paragrafo 1, lettera c), che impone una restrizione ai dati trattabili ai soli dati adeguati, pertinenti e limitati a quanto necessario rispetto alle finalità per le quali sono trattati. Questo principio deve essere garantito nel passaggio tra un titolare e l'altro, da una fase all'altra, inoltre deve essere garantito anche nei casi in cui sia la macchina stessa ad effettuare il trattamento, con supervisione parziale o nulla da parte dell'uomo.

Nel GDPR non è mai menzionata l'IA, ma molte delle direttive in esso contenute sono rilevanti per i sistemi che la utilizzano. Infatti, l'applicazione completa dell'intelligenza artificiale comporta rischi che cozzano con i principi di protezione dei dati, perché questi ultimi apportano significative riduzioni nell'azione decisionale automatica e permettono di utilizzare solo in parte i dati per l'addestramento del sistema tramite la cosiddetta minimizzazione dei dati, che ne comporta una diminuzione della personalità. Inoltre il regolamento proibisce di utilizzare i dati raccolti per altri fini diversi dallo scopo cui si è dato il consenso, tranne in alcuni casi particolari. Questo concetto è evidenziato nell'articolo 5(1b) dove si afferma che *"un ulteriore trattamento dei dati personali a fini di archiviazione nel pubblico interesse, di ricerca scientifica o storica o a fini statistici non è [. . .] considerato incompatibile con le finalità iniziali"*. Secondo l'articolo 9 vige il divieto di trattare dati personali *"che rivelino l'origine razziale o etnica, le opinioni politiche, le convinzioni religiose o filosofiche, o l'appartenenza sindacale, nonché trattare dati genetici, dati biometrici intesi a identificare in*

modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale della persona", ma il paragrafo successivo, costituito da ben dieci punti, permette di aggirare questo divieto adducendo le più svariate argomentazioni: dalla pubblicazione del diretto interessato di informazioni riguardanti gli ambiti sopra definiti, all'avvenuto consenso al trattamento di questi dati, dalla salvaguardia del cittadino, alla medicina preventiva e motivi di interesse pubblico in ambito sanitario. Dunque, benché la profilazione sia in linea di massima vietata, e di conseguenza anche la re-identificazione dei dati sottoposti a pseudonimizzazione, ove non esplicitamente dichiarata al titolare del trattamento, vi sono ampie possibilità di eccezione. Non solo in fase di addestramento può essere messa a rischio la protezione dei dati sensibili, ma anche se i sistemi che li sfruttano permettono ai proprietari, agli altri utenti o alle organizzazioni statali, di determinarli in maniera indiretta tramite il loro utilizzo.

Il Regolamento, come la Direttiva, inserisce norme stringenti e chiare anche per i titolari del trattamento, che sono tenuti a fornire informazioni facilmente intelligibili agli utenti, in modo che le sue finalità non risultino opache e le violazioni siano immediatamente riferite all'interessato stesso. Inoltre il GDPR istituisce nuovi organismi di controllo e obbliga coloro che si impegneranno a fornire un servizio riguardante i dati personali degli utenti a redigere una valutazione d'impatto del sistema in questione.

3.2.2 Dati personali: valutazione d'impatto e autorità di controllo

L'intero Regolamento si basa sulla valutazione preliminare da parte del titolare del rischio che il trattamento che si vuole attuare comporta per l'interessato, come riportato all'art. 24. Il titolare deve tener *"conto della natura, dell'ambito di applicazione, del contesto e delle finalità del trattamento, nonché dei rischi aventi probabilità e gravità diverse per i diritti e le libertà delle persone fisiche"* e sulla base di queste considerazioni deve adottare misure tecniche e organizzative adeguate. Questi obblighi non sonoda adempiere una volta per tutte prima o all'inizio del trattamento, ma si devono applicare costantemente durante tutta l'applicazione di un sistema che implica anche dati personali, quindi significa che l'azione regolamentare del GDPR non è da intendersicome una foto, ma come un video in continuo movimento.

Una volta ritenuto che tale trattamento è conforme agli standard e costituisce un livello di rischio accettabile per il contraente, anche attraverso tecniche di privacy bydesign o by default e minimizzazione dei dati, si deve effettuare un'ulteriore valutazione, detta d'impatto, prevista dall'art. 35. Non in tutti i casi è obbligatorio adempiere a questa prescrizione, solo se l'utilizzo di nuove tecnologie nel trattamento comporta un rischio elevato per i diritti e le libertà della persona fisica. Per valutare l'effettivo livello di rischio è consigliabile consultarsi con il *Data Protection Officer* (DPO), figura centrale nel sistema di relazioni che intercorre tra autorità e titolari del trattamento per rendere effettiva la tutela dei diritti delle persone fisiche, con riguardo alla tutela dei dati personali, nonché alla loro libera circolazione. Questa figura è obbligatoria non solo per tutto il settore pubblico, quindi per autorità e organismi pubblici, ma anche per il set-

tore privato, se i trattamenti in questione rientrano in categorie specifiche riportate nei punti b) e c) del primo paragrafo dell'art. 37. È arduo trovare trattamenti operanti nell'ambito dell'IA che non prendano in considerazione entrambi gli scenari, vale a dire, nel caso b), se i trattamenti comportano un monitoraggio su vasta scala di dati riferiti agli interessati, mentre nel caso c), se i trattamenti vengono effettuati su larga scala dei dati sensibili riportati all'art 9.1 o 10, anche senza un controllo co- stante e continuo. Quindi è molto probabile che il DPO sia indispensabile per tutte le aziende che prendono in considerazione dati sensibili e intelligenza artificiale. Il DPO è una funzione che deve essere svolta da una personalità dotata di specifiche competenze giuridiche, ma anche di conoscenze tecniche nell'ambito dei trattamenti svolti all'interno dell'impresa o organizzazione cui afferisce. In questo modo può fare da ponte tra Autorità di controllo e interessati, supportando il ruolo del titolare del trattamento, come riportato dall'art. 39, in cui si riassumono i principali compiti di questa figura: informare e fornire consulenza al titolare o al responsabile, sorvegliare l'osservanza del Regolamento e delle altre leggi dell'Unione e degli Stati, fornire un parere sulla valutazione d'impatto, assicurare la collaborazione con l'Autorità di controllo e fungere da punto di raccordo tra quest'ultima e il titolare o responsabile. In questi compiti sifa riferimento al titolare o al responsabile, perché non in tutti i casi il titolare svolge anche il ruolo di responsabile del trattamento; i due ruoli si sovrappongono se l'azienda ha pochi dipendenti e i suoi trattamenti coinvolgono pochi interessati, altrimenti è auspicabile assumere un responsabile che possa dirimere le questioni più complesse. Le violazioni al GDPR hanno sempre come referente principale il titolare o responsabile, mai il DPO, perché egli ha il compito di supporto, ma limitata capacità decisionale.

L'Autorità di controllo costituisce un altro attore sul palcoscenico del Regolamento, quasi a costituire il *deus ex machina* dell'intero regolamento, alle cui norme ci si deve attenere per vedere riconosciuti come sicuri i propri sistemi. Infatti, l'Autorità non solo ha il compito di sorvegliare la corretta attuazione del GDPR, ma soprattutto detienealtri poteri di vitale importanza nella prospettiva dell'evoluzione dei sistemi di IA: promuovere la consapevolezza dei titolari o responsabili riguardo gli obblighi loro imposti, collaborare con le altre autorità di controllo in modo da costituire una sinergia positiva, anche svolgendo indagini sulla base di informazioni ricevute da autorità analoghe, sorvegliare gli sviluppi di tecnologie innovative, qualora fossero incidenti sulla protezione dei dati personali. Il ruolo fondamentale dell'Autorità è evidente anche nell'obbligo di redigere e rendere pubblici due elenchi di fondamentale importanza tanto per i titolari, quanto per gli interessati: *"un elenco delle tipologie di trattamenti soggetti al requisito di una valutazione d'impatto sulla protezione dei dati"*, ai sensi dell'art.35, e *"un elenco in relazione al requisito di una valutazione d'impatto sulla protezione dei dati ai sensi dell'articolo 35, paragrafo 4"*, ai sensi dell'art. 57, paragrafo 1, lettera k. L'Autorità non impone solamente obblighi, ma soddisfa anche alla necessità di introdurre buone prassi, come approvare idonei codici di condotta stilati da associazioni o organizzazioni rappresentanti le categorie di titolari o responsabili dei trattamenti e creare meccanismi di certificazione per rendere un trattamento approvato, tenere registri delle violazioni e definire criteri per l'accreditamento

di organismi per il controllo dei codici di condotta. Dunque l'attività di controllo dell'Autorità deve essere costante e continuativa, in modo da dare avvertimenti o ammonimenti, affinché i titolari adeguino le loro policies al Regolamento, indicandone le modalità e i tempi. Come visto il Regolamento prevede l'auspicata stesura, quindi non obbligatoria, di un codice di condotta che preveda di assicurare la puntuale applicazione del GDPR negli ambiti di: trattamento corretto e trasparente dei dati, tutela dei legittimi interessi, modalità di raccolta adottate dei dati personali, pseudonimizzazione dei dati personali, informazione fornita al pubblico e agliinteressati, esercizio dei diritti degli interessati, tutela del minore, misure relative alla responsabilità del titolare, notifica all'Autorità di controllo delle eventuali violazioni, modalità del trasferimento dei dati verso Paesi terzi o organizzazioni internazionali, procedure stragiudiziali per comporre controversie tra titolari e interessati.

Una volta stilato questo codice, il monitoraggio dello stesso deve essere garantito dall'Autorità o anche da un organismo terzo rispetto alla stessa, che deve possedere un adeguato livello tecnico e deve essere accreditato dall'Autorità competente. Per poter operare, questo organismo di controllo ha l'obbligo di essere indipendente e competente, l'obbligo di istituire procedure adeguate a valutare l'ammissibilità dei titolari o responsabili ad applicare il codice, nonché procedure adeguate a gestire eventuali reclami relativi a violazioni del codice; sempre nell'ambito dell'indipendenza da determinati titolari, tale organismo deve poter dimostrare all'Autorità l'assenza di conflitti d'interesse nello svolgimento delle proprie funzioni.

I codici di condotta costituiscono uno strumento di salvaguardia dei dati personali di estrema rilevanza nell'ambito dell'intelligenza artificiale e nello specifico della catena di trattamenti che la caratterizza. Infatti un codice di condotta comune tra i vari titolari dei trattamenti consente di avere una linea comune d'intervento che permetta una attuazione uniforme del Regolamento nell'ambito specifico. Proprio questa sua alta specificità dà la possibilità che sia possibile governare efficacemente ogni futura realizzazione di sistemi dotati di intelligenza artificiale, perché, tramite la flessibilità del meccanismo sopra riportato, in cui il pubblico, l'ente regolatore e il privato lavorano in sinergia, questo diviene un potente mezzo, che può essere sempre adattato alle esigenzedel momento.

Infine, se il trattamento dei dati personali è approvato dall'Autorità, allora è caldamente consigliato dal Regolamento un sistema di certificazione della protezione dei dati attraverso sigilli e marchi validi all'interno dell'UE e consegnati esclusivamente in sede europea. La differenza tra Certificazione e Codice di condotta risiede nel fatto che ilprimo caso rappresenta il frutto dell'adesione volontaria dei titolari o dei responsabili, mentre il secondo caso è il frutto di un'iniziativa di un gruppo o di un'associazione. L'apporto di sigilli e marchi rappresenta il tassello fondamentale del rapporto tra titolare e Autorità di controllo; ovviamente, non essendo obbligatoria, tale ulteriore norma potrebbe essere trascurata, senonché nel GDPR si incentivano Stati, Autorità di controllo e lo stesso Comitato europeo per la protezione dei dati a promuoverne l'utilizzo. Per confermare l'importanza di questo meccanismo si deve notare che gli standard per l'attribuzione di sigilli e marchi

è una prerogativa della Commissione europea, perché tali metodi di certificazione hanno valenza europea.

L'insieme di queste norme e prerogative cui si deve attenere il titolare del trattamento, costituivano già prima dell'approvazione di altri documenti in materia di intelligenza artificiale una base su cui fare affidamento per la conciliazione della tutela dei dati personali con tali sistemi; il Parlamento europeo e la Commissione hanno però optato per rendere ancora più certa l'applicazione del diritto in questo specifico ambito in modo da prevenire futuri conflitti dovuti a trattamenti indiscriminati di dati personali effettuati usando tecniche ancora impensabili per lo stato dell'arte attuale. Proprio per ovviare a un futuro dominato dalla automazione, anche pericolosa nei confronti dei dati sensibili degli utenti, sono state avanzate proposte originali, come quella di dotare il sistema automatico di uno status giuridico, che permetta di attribuire in modo certo la colpa e la responsabilità in caso di danni o lesioni anche al sistema stesso.

3.2.3 Verso una responsabilità robotica?

Come visto nel corso dell'intero capitolo la responsabilità di un trattamento errato o discriminatorio non è di facile risoluzione soprattutto se avviene automaticamente, perché risulta molto difficile attribuirlo senza una efficace verifica degli strumenti utilizzati. Infatti molti sistemi di IA, che lavorano con una supervisione minima o nulla da parte dell'uomo, lavorano come *black boxes*, costituendo un ostacolo di non poco conto alla risoluzione di questo problema. Inoltre risulta molto complessa la gestione della responsabilità anche per un altro fattore, più volte citato in questo studio:la presenza di catene di trattamenti tra loro interconnessi, affidati solitamente a diversi titolari. Si è dunque pensato, per ovviare a questo problema fondamentale, di eliminare questa indeterminatezza inserendo una cosiddetta responsabilità del robot.In questo caso si indaga se il titolare del trattamento possa nominare o meno come responsabile, seguendo l'art. 28, il sistema dotato di autonomia decisionale, dato che la sua figura si caratterizza per il fatto di trattare i dati per conto del titolare. Inquesto caso, come nel caso del responsabile umano, è sempre il titolare che risponde in ultima battuta alle istruzioni impartite al responsabile artificiale. Risulta però difficile immaginare un'intelligenza artificiale che agisca senza che il suo programmatore mantenga le responsabilità derivanti dalle azioni poste in essere, almeno per quanto riguarda le violazioni e i danni provocati. È inoltre da vedere con molta attenzione come il dispositivo automatico non possiede uno status giuridico effettivo, risultando quindi difficilmente imputabile di un trattamento discriminatorio, anche nel caso di una nomina a responsabile dello stesso.

È per questo motivo che in molti paesi, anche esterni all'UE, come il Giappone [59] e gli Stati Uniti, si sono date direttive tali da ritenere i robot, anche con capacità decisionali elevate, prodotti e, per questo motivo, la responsabilità di eventuali danni a cose o persone sono direttamente riconducibili alle aziende produttrici [8]. Uno degli ambiti in cui l'applicazione di questo criterio è più visibile corrisponde a quello medico, in cui

il paziente è operato dal chirurgo, coadiuvato da un robot chirurgo. Negli ultimi anni si sono avute molteplici cause intentate da vari pazienti nei confronti di alcune compagnie che offrivano questi servizi in ospedali statunitensi, ma risulta essere molto complessa non solo l'attribuzione della colpa, ma anche dimostrare il rapporto di causa-effetto che una mancata o errata manovra del braccio meccanico abbia potuto impattare sulle condizioni di salute derivate dall'operazione; infatti, molti di questi dibattimenti hanno avuto esito negativo per chi denunciava proprio per mancanza di chiarezza dell'effettiva causa dello stato di salute attuale del paziente. In questo specifico contesto il danno è causato da azioni o reazioni del robot in un contesto di interazione con l'uomo, ma il danno può anche derivare da un inadempimento contrattuale o da difetti di produzione. Per i robot, se il loro comportamento è impostato dal produttore, si può ancora riferire la responsabilità allo stesso produttore, mentre risulta più complicata l'attribuzione nel caso di un robot dotato di capacità di apprendimento. Per esempio, si esamina il caso in cui una macchina a guida autonoma, impostata dall'utente per effettuare un determinato percorso, decida di cambiarlo autonomamente, perché magari c'è meno traffico o è la strada più breve, e provochi un incidente per il cattivo stato di manutenzione del manto stradale [6]. In questo caso l'attribuzione della responsabilità potrebbe essere riferita al produttore, che ha previsto un algoritmo di auto-apprendimento, o al programmatore, che ha implementato l'algoritmo stesso senza controllare le condizioni fisiche del tragitto, o il Comune, che ha l'obbligo di effettuare la manutenzione stradale. Proprio per via di questa indeterminatezza, le azioni legislative disponibili sono molteplici: dall'assicurazione obbligatoria per questi sistemi all'istituzione di un numero d'immatricolazione individuale, come suggerito dal Parlamento europeo, agli approcci più estremi, come il divieto di commercializzazione di sistemi di IA o l'immissione nel mercato degli stessi senza regole [6].

Ovviamente sono state sviluppate teorie che applicano tipi di responsabilità già presenti sia in ambito civile che penale. Il problema più significativo, attorno al quale si sviluppano altri punti critici, consiste nella definizione giuridica di un sistema di intelligenza artificiale: secondo alcuni analisti può essere assimilabile a un prodotto, mentre secondo altri potrebbe assumere la qualifica di agenti. Il primo caso, già discusso sopra, presenta notevoli difficoltà di applicabilità giuridica, in quanto risulta complessa l'attribuzione di responsabilità per prodotti dotati di un livello di autonomia molto elevato. Il secondo caso prende in considerazione diversi scenari: la responsabilità genitoriale, che presenta una non semplice assimilazione tra automi ed esseri pensanti, la responsabilità per il danno provocato da animali, che prospetta un'equiparazione tra animali e automi, e la responsabilità per l'esercizio di attività pericolose, che, essendo più generica, consente una applicabilità più adatta all'IA. Significativo è il parallelismo che solitamente è preso a paradigma tra animali ed esseri artificiali; infatti l'animale, pur se addestrato, presenta un grado di imprevedibilità dovuto alla sua natura, proprio come la macchina autonoma, che potrebbe deviare il proprio comportamento da quanto programmato dal suo creatore. Tale equiparazione è però affetta da più errori, comeil grado di imprevedibilità delle macchine, che è di molto ridotto rispetto a quello cagio-

nato dagli animali, fatto dovuto alla natura artificiale delle prime, e la possibilità di programmare anticipatamente limitazioni dei sistemi, qualora costituissero un pericolo per l'uomo.

Per quanto riguarda la responsabilità penale, secondo lo studioso israeliano G. Hallevy, assumendo che le macchine siano dotate di una personalità giuridica, vi sono tre distinti paradigmi. Il primo di questi si basa sulla riconduzione delle forme di intelligenza artificiale alla categoria degli agenti innocenti [51]. In questo caso, i responsabili dei danni derivati dal trattamento sono individuati nel programmatore del sistema e nell'utente, analogamente ai tutori delle persone affette da degradazione o menomazione mentale. Il programmatore potrebbe realizzare un robot con scopi illeciti, mentre l'utente potrebbe usare lo stesso robot, non creato per fini dolosi, in modo fraudolento. Il secondo modello individuato consiste nell'assenza di intenzioni atte a commettere un reato da parte del programmatore o dell'utente, ma per negligenza degli stessi, non previste in fase di progettazione o utilizzo. In questo caso alla responsabilità del cliente o del programmatore potrebbe essere aggiunta una responsabilità del sistema, se esso stesso non agisce come mero strumento, ma si comporta come un'entità pienamente intelligente. Il terzo scenario, forse il più innovativo e originale, si basa sulla condizione di considerare l'agente intelligente come *mens rea* e *actus res*, cioè essere sia la mente del reato, sia l'esecutore materiale. In questo modo il sistema di intelligenza artificiale potrebbe essere imputato di responsabilità penale. Ciò però comporterebbe uno stravolgimento del significato della pena, che, finora applicata solo ad esseri umani, prevede una diminuzione della propria libertà in modo da evitare la reiterazione del crimine, con il rischio di forzarne l'applicazione anche a categorie non rientranti di fatto in quelle canoniche. Tale ultimo paradigma è quindi viziato da un fondamentale fraintendimento, cioè sul fatto che le caratteristiche umane differiscono in modo sostanziale da quelle presentate da sistemi di intelligenza artificiale [6].

La risposta quindi alla domanda del titolo di questo paragrafo è negativa, anche memori delle considerazioni fatte nel paragrafo 1.3, in cui si cambia la prospettiva delle tre leggi della robotica, dall'agente artificiale a quello umano, poiché quest'ultimo può essere realmente indipendente, mentre l'altro risulta comunque succube delle decisioni umane. Per concludere, è quindi impensabile un futuro apocalittico in cui il genere umano possa essere minacciato dalle macchine da lui costruite [4], perché basta prendere precauzioni adeguate in modo che la tecnologia, anche la più avanzata, quale l'intelligenza artificiale, possa essere sempre programmata e sorvegliata dall'uomo, per esempio, senza il bisogno di incaricare un sistema di IA come responsabile di un trattamento che potrebbe rivelarsi, in alcuni casi, dannoso.

Capitolo 4

Governance dell'intelligenza artificiale

4.1 Giurisdizione europea

Come visto nei precedenti capitoli, l'intelligenza artificiale costituisce un notevole ostacolo per l'ente legislativo, date le sue tecniche sempre innovative e la sua alta dosedi imprevedibilità. Inoltre, l'immaginazione collettiva risulta affetta da pregiudizi che tendono ad aumentare la diffidenza in un'entità artificiale dotata di una così ampia autonomia, dotata anche della capacità di elaborare dati personali degli utenti, il cherende ancora più complesso il quadro in cui andrà ad operare un'efficiente azione legislativa. Per evitare una eccessiva libertà nel settore, l'Unione Europea ha stilato una Proposta di Regolazione, che non pregiudicasse la crescita in un ambito così ricco di speranze ed opportunità economiche. Probabilmente sull'onda del boom di utilizzo e quindi commercializzazione di prodotti tecnologici, l'UE ha cercato di sviluppare uno strumento efficace che potesse tutelare da una parte i consumatori e, dall'altra, potesse predisporre un terreno sicuro per la ricerca anche in ambito di IA, considerata come prossima ad un utilizzo capillare, anche nella vita quotidiana, benché i dati forniti dall'Unione abbiano registrato un utilizzo ancora limitato di questi sistemi, almeno in ambito lavorativo fino al 2020 [41], come si può vedere in figura 4.1.

4.1.1 Unione Europea e intelligenza artificiale: un cammino che parte da lontano

Nel 2017 tramite due diversi comunicati, uno della Commissione e uno del Consiglio Europeo, le autorità di Bruxelles hanno preso in considerazione la creazione di una nuova legislazione dell'intelligenza artificiale; queste prime fasi sono state seguite da un'ulteriore comunicazione dal titolo programmatico *"L'intelligenza artificiale per l'Europa"*, approvato dalla Commissione nell'aprile del 2018 [27]. In questo documento si è posto l'accento sugli aspetti economici, dall'ambito della ricerca, a quello della concorrenza, agli aiuti per le piccole e medie imprese, ma non si è approfondita solo questa questione, perché si aveva ben chiaro che uno sviluppo repentino di questa specifica tecnologia avrebbe comportato cambiamenti socioeconomici molto impattanti. Seguendo questa previsione, rivelatasi corretta anche per l'insorgere prepotente della pandemia da Covid-19, la Commis-

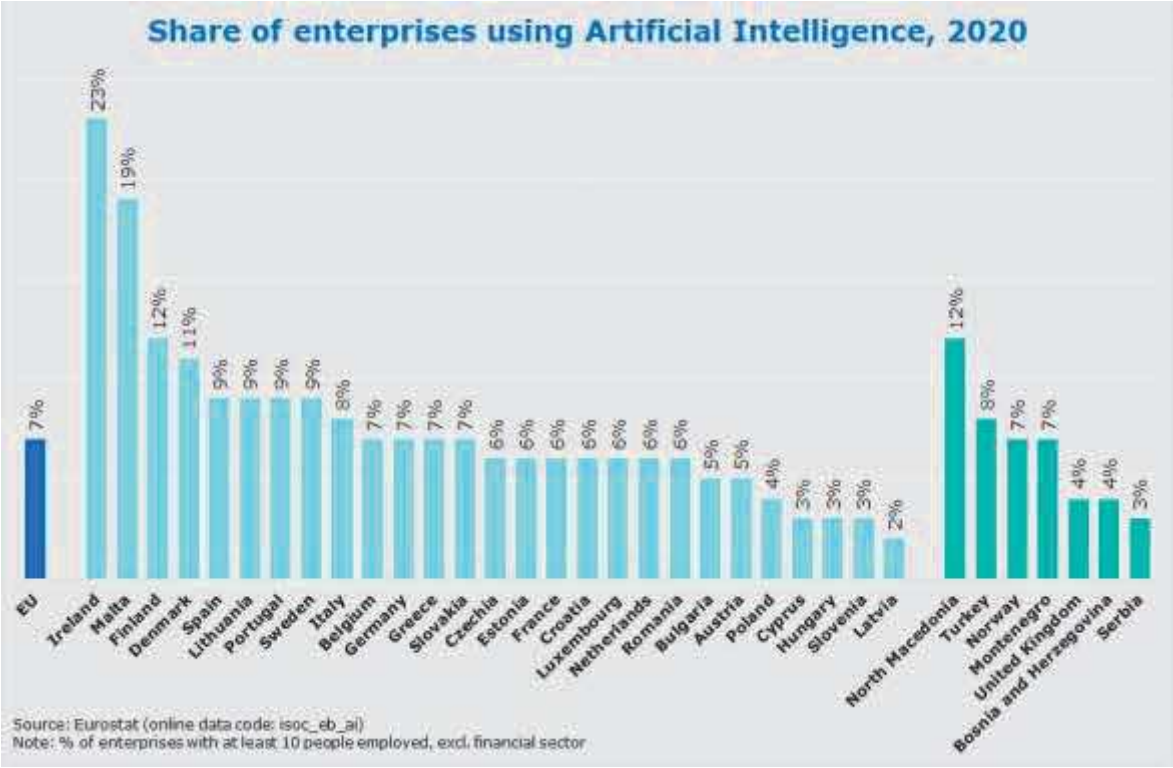

Figura 4.1: Il tasso di utilizzo dell'intelligenza artificiale nelle imprese europee © Eurostat

sione indicava come imprescindibile la creazione di professionalità con più competenze in IA, attraverso incentivi ai poli di ricerca e alle università. Questi ultimi non sono solo destinatari di provvedimenti atti a favorirne il lavoro, ma costituiscono anche, sempre secondo questa comunicazione, veri e propri stakeholder, che, in sinergia con l'UE, gli enti dei consumatori e le aziende produttrici, alimenteranno un dibattito, da cui prendere spunto per un'azione legislativa idonea. È proprio in questa comunicazione che si palesa per la prima volta un intento legislativo dell'UE per consentire un quadro etico e giuridico adeguato quando sono utilizzati sistemi di IA. Ancor più importante è il richiamo contenuto all'interno del documento al GDPR,a stabilire uno stretto contatto tra un futuro regolamento e la tutela dei dati personali. La frase che conclude il documento, oltre a risuonare molto altisonante, racchiude bene quale è lo spirito dell'iniziativa: *"insieme, potremo mettere la forza dell'IA al servizio del progresso umano"* [27]. Da sottolineare, anche in questo caso, l'affermazione più volte ribadita in sede europea dell'intelligenza artificiale come mezzo di cui un essere umano può servirsi e che, quindi, non risulta essere un ente giuridico cui imputare responsabilità o un'autonomia tale da renderla capace di azioni dannose nei confronti dell'uomo, senza che sia esso stesso ad averne dato la facoltà tramite la sua errata o fraudolenta programmazione.

Seguendo questo graduale percorso, l'UE ha pubblicato nel 2020 un libro bianco, documento che contiene proposte di azione comunitaria nel settore specifico dell'IA, che consentisse di fornire una base per un regolamento più completo in materia e il cui sottotitolo evidenzia in modo ancora più marcato quale direzione si vuole prendere: l'eccellenza e la fiducia sono i due fari verso cui puntare, due concetti che tendono a

mostrare ancora una volta la volontà di tutelare l'utente, salvaguardando al contempo la competitività europea [22]. Per non perdere la bussola della trattazione, che vuole sempre riconoscere la necessità di consentire un equo trattamento dei dati sensibili anche in ambienti di difficile previsione come l'intelligenza artificiale, si individua, anche in questo documento prodromico alla Proposta avanzata nel 2021, una sezione detta *"Rischi per i diritti fondamentali, e la protezione dei dati personali e della privacy e la non discriminazione"*, che si basa su una ricerca del 2017 del Consiglio d'Europa, che svela come un utilizzo non regolato dall'IA potrebbe comportare rischi importanti per i cittadini.

In questo libro bianco si approfondiscono alcune tematiche fondamentali già presenti negli altri atti sviluppati in precedenza, come l'incentivo ad aumentare le competenze in intelligenza artificiale, il partenariato dell'Unione col settore privato, l'aiuto alle piccole-medie imprese, al fine di costruire un ecosistema di fiducia tra cittadini e nuove tecnologie; allo stesso tempo si inseriscono dettagli tecnici sui possibili meccanismi e principi da seguire per un regolamento che sia il più adeguato possibile ad una realtà in continuo cambiamento come l'IA. In questo quadro si inseriscono la trasparenza dei trattamenti, in modo da soddisfare gli obblighi di mantenere informati gli utenti sui sistemi ad alto rischio, la robustezza, ottenuta tramite uno sviluppo responsabile e una valutazione ex ante dei possibili rischi derivanti dall'utilizzo del sistema, la sorveglianza umana, termine fondamentale per consentire il rispetto degli altri principi; proprio per rendere un sistema di IA spiegabile, nei limiti del possibile, si è deciso di obbligare già in questo abbozzo di regolamento il titolare del sistema stesso a mantenere datie registri relativi alla programmazione dell'algoritmo, nello specifico, alla sua fase di addestramento. Data la natura delle tecnologie facenti uso dell'intelligenza artificiale, caratterizzata il più delle volte da una continua crescita del bagaglio di conoscenze anche durante il suo utilizzo, si è voluto chiarificare che le valutazioni sulla liceità dei trattamenti debbano essere ripetute per tutta la durata dei sistemi; per svolgere queste valutazioni si prendono in considerazione strutture di sostegno esterne alle aziende operanti nel settore per compiere queste onerose e dispendiose valutazioni, anche attraverso i poli dell'innovazione digitale presenti nel territorio dell'Unione.

Una questione che riveste notevole importanza, solo accennata nel libro bianco, è la presenza di riferimenti continui a sistemi di intelligenza artificiale ad alto rischio,che suggerisce un approccio basato sul rischio e una conseguente scala di rischiositàdi tali sistemi, come indicato nella sezione C *"Ambito di applicazione di un futuro quadro normativo dell'UE"*: *"Tale approccio basato sul rischio [. . .] richiede che siano definiti criteri chiari per distinguere tra le diverse applicazioni di IA, in particolare perstabilire se tali applicazioni siano o meno "ad alto rischio""*[22]. La Commissione ha individuato alcuni criteri da valutare per ogni sistema posto in essere: se l'IA è usatain un settore in cui si prevedono rischi significativi ed è usata in modo tale da poter generare rischi significativi.

4.1.2 Proposta europea di regolamento sull'intelligenza artificiale

Il documento prodotto nel 2021 è dunque solo l'ultimo in ordine di tempo che l'Unione Europea ha varato in materia di intelligenza artificiale. Tramite le molteplici iniziative sopra menzionate si è giunti all'attuale *"Proposta di Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale"* [25]. I portatori di interesse (stakeholders) hanno richiesto di definire, oltre al termine intelligenza artificiale, anche una scala di rischio, memori della sua definizione assai approssimativa data nel libro bianco, cui si possa fare riferimento nel caso di assunzione di un sistema basato su tale tecnologia. Anche in questo caso si tratta di una proposta che deve essere presa in considerazione non solo dal Parlamento Europeo edagli altri organi europei deputati, ma anche e soprattutto dai produttori di tali sistemi, dalle associazioni dei consumatori, dalle università e dai poli di ricerca. Dunque è ragionevole aspettarsi che da questo dibattito scaturiscano modifiche anche sostanziali all'impianto normativo così come è stato ideato.

È sempre importante ricordare che con questo atto, come indicato anche in tutte le precedenti iniziative, l'UE vuole promuovere un'intelligenza artificiale affidabile e per raggiungere questo scopo ha adottato un approccio equilibrato, che non pregiudichi la crescita nel settore e, al contempo, si basi sui valori e sui diritti fondamentali della persona. Quindi, si è voluto evidenziare come l'IA non sia un prodotto tecnologico come gli altri, ma abbia anche implicazioni umane ed etiche di non poco conto. L'atteggiamento di dialogo dimostrato dall'Unione ha portato già alcuni frutti, come aver capito che lo strumento per governare questa tecnologia deve essere flessibile, dato il rapido sviluppo cui è soggetta. Ciò ha indirizzato la Proposta verso una governance dell'intelligenza artificiale, che non deve essere confusa con un ammorbidimento dei termini della manovra, ma è altresì da intendersi un modo di regolare il settore senza appesantirne la ricerca e l'applicazione. In ambito europeo si è scelto di seguire un approccio proporzionato basato sul rischio, costituendo un regolamento che determinasse quali sono i pericoli cui un cittadino può andare incontro, prevedendo e prevenendo, nei limiti del possibile, future criticità che l'uso pervasivo dell'IA potrà causare.

La Proposta ha un carattere orizzontale, cioè non riguarda l'intelligenza artificiale applicata in un'unica filiera, bensì ad ogni ramo dell'economia; perciò essa richiede una completa coerenza con la normativa vigente nell'Unione, in modo da garantirnela retro-compatibilità con la Carta dei diritti fondamentali dell'Unione europea e il diritto derivato dell'UE in vigore in materia di protezione dei dati, il GDPR. Anche in questo caso, come nel GDPR, si è voluto sottolineare come il regolamento sia a servizio del cittadino inteso come persona, quindi rimarcando la centralità dell'essere umano anche nel processo tecnologico in modo da promuovere i principi etici di cui l'Unione si fa garante. Tutto questo sforzo per predisporre una normativa utile al cittadino e alle imprese o agli stati che vogliono adottare sistemi basati sulla intelligenza artificiale intende fornire una certezza del diritto, rimuovendo gli ostacoli alla fornitura transfrontaliera di tali sistemi e accrescendo la fiducia della comunità intera.

Per stilare questa proposta si sono seguiti quattro principi etici in modo da fornire un'IA antropocentrica. Il primo e fondamentale principio adottato è il rispetto dell'autonomia umana, secondo cui gli umani che interagiscono con un sistema di intelligenza artificiale devono essere in grado di mantenere una piena e completa autodeterminazione. La seconda direttiva presa in considerazione è stata la prevenzione di un danno fisico o psicologico nei confronti dell'uomo; come visto, la manipolazione può comportare un'alterazione della psiche, mentre un danno materiale può essere causato da una macchina a guida autonoma o da un robot industriale fuori controllo. Si è cercato di modellare il regolamento sull'imparzialità che tali sistemi dovrebbero consentire, sia dal punto di vista sostanziale, garantendo un'equa distribuzione di costi e benefici e anche l'assenza di pregiudizi e stigmatizzazioni ingiustificate, sia da quello procedurale, consentendo la possibilità di contestare e richiedere un risarcimento, qualora dovessero verificarsi discriminazioni non tollerabili. Infine, l'ultimo principio etico sul quale si basa l'impianto giuridico è rendere possibile una spiegazione del comportamento di una IA, offrendo una sufficiente trasparenza algoritmica, che consenta di determinare quali sono state le cause dell'azione, magari dannosa, verificatasi. Per consentire il rispetto di questo particolare concetto, sono stati inseriti non solo obblighi di prova del sistema ex ante, prima di essere inserito nel mercato, ma anche rigidi controlli ex post in modo che, in caso di violazioni dei diritti fondamentali, tra cui il diritto alla protezione dei dati di carattere personale, sia possibile un ricorso efficace da parte delle persone lese. Partendo dal rispetto di queste fondamentali basi etiche, il regolamento si articola in dodici titoli, che determinano gli ambiti di applicazione, una scala di rischi, obblighi di trasparenza, sostegni all'innovazione, strumenti di governance e codici di condotta. Soffermandoci sull'approccio basato sul rischio, se ne sono definiti quattro gradi, da inaccettabile a minimo.

4.1.3 Rischio inaccettabile: quando un'intelligenza artificiale non può essere commercializzata?

Si ha un rischio inaccettabile qualora i sistemi che fanno uso di intelligenza artificiale fossero qualificati come *"chiara minaccia per la sicurezza, i mezzi di sussistenza e i diritti delle persone"*, attraverso la manipolazione per aggirare il libero arbitrio dell'essere umano o attraverso l'applicazione di un punteggio sociale. In questo caso l'Unione ha pensato di vietare tali applicazioni, fornendo solo qualche rara eccezione alla loro messain opera. Di queste pratiche si occupa il titolo II, composto dal solo articolo 5, che pone l'accento sull'uso dell'IA che potrebbe provocare una distorsione del comportamentodi una persona non consapevole del trattamento o appartenente a categorie con latenti vulnerabilità in modo che provochi danni psicologici o fisici ad un altro individuo.

Inoltre, è ritenuta vietata l'assegnazione di un *social score*, che classifica le persone fisiche in base alle proprie azioni o al proprio comportamento sociale noto o previsto. In questo caso sono due i fattori che ne causano il divieto presi in considerazione: potrebbe consentire un trattamento basato su pregiudizi o sfavorevole di persone fisiche o interi

gruppi sociali, non collegati direttamente ai contesti in cui sono stati generatii dati originali, oppure, in uno scenario ancora più cupo in cui vi sarebbe uno stato repressivo, potrebbe consentire un trattamento basato su pregiudizi o sfavorevole di persone fisiche o interi gruppi sociali ingiustificato o sproporzionato rispetto alla gravitàdelle loro azioni. Questo paragrafo c) dell'art. 5 costituisce il più chiaro riferimento a quella pratica in Europa e nel mondo occidentale unanimemente deplorata, che consistenella continua ingerenza dello stato nelle attività dei cittadini, fornendo una continua valutazione delle loro azioni, come avviene in Cina. Bisogna ammettere che, datala lontananza culturale e geografica della Repubblica Popolare Cinese, si tende ad ingigantire molte delle pratiche svolte dal suo governo autoritario, come il punteggio sociale personale, su cui si basa l'accesso ai mezzi pubblici, alla sanità o la richiestadi un prestito. È allo stesso tempo molto difficile comprendere appieno la situazione, data la poca trasparenza che avvolge le autorità statali. A partire dal 2020 si è riacceso l'interesse internazionale sull'argomento, dal momento che il partito nazionale ha posto come obiettivo del quattordicesimo piano quinquennale l'utilizzo efficiente dei big dataper modernizzare l'azione governativa [49]. Ulteriore risalto all'annosa questione è stato dato anche durante questo ormai lungo periodo caratterizzato dalla pandemia da Covid-19, in cui molte regioni cinesi hanno imposto restrizioni severe e adattato il *SocialCredit System* (SoCS) in modo da tracciare e sanzionare le violazioni alle stesse e, in un secondo momento, in modo da incentivare il ritorno al posto di lavoro in sicurezza [38]. Un aspetto risulta evidente: per come la società occidentale si è sviluppata, con un alto rispetto dell'autonomia individuale, discendente dal concetto cristiano di liberoarbitrio, e il conseguente rispetto della sfera privata, l'attribuzione di un social score costituisce qualcosa, per come dichiarato dall'UE, di inaccettabile.

È importante soffermarsi su come lo stato cinese riesca a monitorare in maniera capillare l'attività dei propri cittadini; ovviamente si utilizzano banche dati per controllare le transazioni di danaro da un conto a un altro, ma anche un elemento che avvicina di più alla distopica visione orwelliana: le onnipresenti telecamere, che permettono il riconoscimento in tempo reale degli autori delle trasgressioni, anche le più trascurabili. Sembra una riproposizione orientale dell'iconica frase di "1984": *"Big Brother is watching you"* [67]. Non è un caso che proprio l'uso di sistemi di identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico a fini di attività di contrasto costituisce un'altra pratica non tollerata. Ovviamente queste pratiche sono vietate nella misura in cui ledono la libertà dell'individuo e soprattutto offrono achi le promuove e ne ha il controllo ingerenze inaccettabili nella sua sfera privata. In particolare, i dati personali vengono utilizzati come punto di partenza per la messa in pratica di tali metodi per sfruttare le vulnerabilità di un certo gruppo di persone, in base alle loro ricerche sul web, ai loro comportamenti o addirittura in base alle loro caratteristiche somatiche. Vi sono però tre scenari in cui questo obbligo si fa meno stringente:

1. *"Ricerca mirata di potenziali vittime specifiche di reato"*;

2. *"Prevenzione di una minaccia specifica, sostanziale e imminente [...] o di un attacco terroristico"*;
3. *"Il rilevamento, la localizzazione, l'identificazione o l'azione penale nei confronti di un autore o un sospettato di un reato"*.

La prima deroga probabilmente sarà la più utilizzata, dal momento che potrà essere messa in pratica per casi di rapimenti anche di minori o allontanamenti volontari. Per aumentare la sicurezza in alcune città italiane ed europee era stata avanzata l'ipotesidi installare telecamere con identificazione biometrica remota in tempo reale per ogni cittadino che transitasse in luoghi pubblici [11]; tale soluzione è deprecata, ma, se vi sono state minacce consistenti, mirate a una specifica persona o all'intera popolazione, come ammesso nel secondo scenario, vi è la possibilità di attuarla. Da sottolineare il carattere temporaneo per cui queste misure sono possibili, comprensibile soprattutto dall'uso del termine imminente, perché il continuo processo delle azioni attraverso l'identificazione biometrica potrebbe non solo impattare negativamente sull'opinione pubblica, ma anche costituire un affronto alla libertà degli esseri umani. La terza deroga consentita è speculare alla prima: se nella prima si ricercava la vittima, in questo caso si cerca il colpevole o il sospettato di un reato; in questo modo non viene meno a quella intenzione di rafforzare la sicurezza dei cittadini anche tramite nuove tecnologie, ma la si delimita per processare i volti di colpevoli o presunti tali entro un ragionevole insieme di individui.

Come sottolineato più avanti nello stesso art. 5, la tutela dei cittadini e dei loro dati sensibili è rispettata da condizioni necessarie, specificatamente temporali, geografiche e personali limitate, ma allo stesso tempo proporzionate in relazione all'uso. Queste limitazioni sono anche garantite dall'autorizzazione preventiva rilasciata da un'autorità giudiziaria o amministrativa indipendente dallo Stato in cui avviene l'uso; è specificata anche la possibilità di approvare in tutto o in parte l'identificazione biometrica remota "in tempo reale" in spazi accessibili al pubblico ai fini di attività di contrasto che deve essere decisa dal diritto nazionale di uno Stato membro, indipendentemente dagli altri facenti parte dell'Unione. Questa ultima precisazione potrebbe sembrare superflua o addirittura controproducente, dal momento che sembrerebbe incentivare la frammentazione della normativa, ma non si deve dimenticare che questo documento costituisce solo una bozza di quello che verrà preso in considerazione e approvato in sede comunitaria e, in questo modo, si incentiva uno sviluppo autonomo di leggi che potrebbero aiutare la stesura finale.

4.1.4 Rischio alto, limitato o minimo: classificazione e obblighi

Per definire un sistema di intelligenza artificiale ad alto rischio è necessario si verifichino entrambe le seguenti condizioni: costituisce un componente di sicurezza di un prodotto o è un prodotto regolato in altre direttive precedenti, che si possono consultare nell'allegato II della Proposta, e se risulta *"soggetto a una valutazione della conformità da parte di terzi ai fini dell'immissione sul mercato o della messa in servizio di tale*

prodotto", come riportato nel paragrafo 1 dell'art. 6. L'allegato II della Proposta cerca di armonizzare la legislazione con precedenti atti dell'UE e prende in considerazione, nella prima sezione, svariati settori dell'economia, come la robotica, la nautica, la giocattoleria, la sicurezza in ambienti esplosivi o sugli ascensori, la protezione individuale, l'ambito medico. La seconda sezione arricchisce i riferimenti a sistemi di IA ad alto potenziale di rischiosità con quelli legati al trasporto sia esso aereo, veicolare o ferroviario, perché costituisce un segmento dove deve essere garantita la massima sicurezza in modo da evitare possibili sciagure. L'articolo 7 consente alla Commissione l'aggiunta di sistemi che potrebbero differire da quelli specificati nell'allegato III, in cui si precisa che sono da considerarsi ad alto rischio i sistemi usati nelle seguenti categorie:

1. Identificazione biometrica "in tempo reale" e "a posteriori" delle persone fisiche;
2. Gestione e funzionamento delle infrastrutture critiche, come acqua, gas, riscaldamento ed elettricità;
3. Istruzione e formazione professionale, per esempio per determinare l'accesso o l'esclusione di un soggetto da un istituto;
4. Occupazione e gestione dei lavoratori, anche nel caso di lavoratori autonomi;
5. Accesso a sistemi pubblici o privati essenziali;
6. Attività di contrasto, ambito già sviluppato nei sistemi di IA vietati, ma qui indicato prendendo in considerazione altre tecniche, come poligrafi o profilazione delle persone fisiche nel corso di un'indagine;
7. Gestione della migrazione, dell'asilo e del controllo delle frontiere;
8. Amministrazione della giustizia.

In tutti questi ambiti i sistemi di IA, prima di essere immessi nel mercato, dovranno rispettare rigorosi obblighi, cui possono essere ammesse deroghe solo in casi eccezionali. Come indicato nel secondo paragrafo dell'articolo 7, la Commissione, per valutare l'effettiva rischiosità del sistema, deve tenere conto di molteplici criteri, come le sue finalità, la misura in cui sarà usato, la portata di un potenziale danno per un utente o per un individuo coinvolto, pur non essendo l'utilizzatore materiale del sistema. Tuttavia, oltre a sottostare a questo rigido controllo, che ne verificherà la conformità ai requisiti, si deve anche porre in essere un efficace sistema di gestione dei rischi, eseguito nell'intero ciclo di vita dell'IA in questione, come riportato all'art. 9.

Per tornare all'argomento principe della trattazione ivi esposta, le criticità che solleva l'intelligenza artificiale in questi contesti sono strettamente legate ai dati personali e alla loro governance, come evidenziato nell'articolo 10, dove si fa esplicito riferimento al GDPR, che vieta il trattamento di quei dati rivelatori di origine razziale o etnica, opinioni politiche, convinzioni religiose o filosofiche, o appartenenza sindacale, nonché il trattamento di dati genetici, dati biometrici intesi a identificare in modo univoco una persona fisica, dati relativi alla salute o alla vita sessuale o all'orientamento sessuale

della persona. Come sopra indicato questo divieto categorico viene smorzato dal secondo comma in cui si paventano dieci scenari possibili per utilizzare comunque questi dati sensibili. Nel medesimo comma dell'articolo 10 si cita un articolo del regolamento UE 2016/680, che ammette delle eccezioni al divieto imposto dal GDPR, autorizzando il trattamento dei dati sopra descritti solo se strettamente necessario e sotto garanzie adeguate per i diritti e la libertà dell'interessato [72]. Per i sistemi di IA ad alto rischio si richiede il rispetto di una serie di norme molto stringenti: si deve stilare una documentazione tecnica dettagliata e aggiornata, si devono conservare le registrazioni automatiche dell'utilizzo in modo da garantire trasparenza in caso di lesioni, si devono fornire informazioni sul fornitore e sul sistema agli utenti, si deve garantire la supervisione da parte di persone fisiche durante il periodo di attività. Inoltre, come indicato dall'articolo 15, lo sviluppo di questo tipo di sistemi deve seguire i dettami dell'accuratezza, della robustezza, tramite, per esempio, backup dei dati, e della cybersicurezza, prevenendo potenziali attacchi volti a manipolare il training set, come nel caso di *data poisoning*, o difetti del modello. Dopo aver fornito il sistema di queste proprietà, il fornitore del servizio che adotta l'intelligenza artificiale deve sottoporlo ad una serie di verifiche da parte degli organi competenti, i quali, alla fine, se è tutto in regola, appongono la marcatura CE di tecnologia conforme alle norme vigenti nell'Unione. Come nel caso del GDPR è richiesta una fondamentale cooperazione tra fabbricanti di tali sistemi o loro rappresentanti autorizzati e autorità competenti, per costituire un ecosistema di fiducia che comporti una serena convivenza con l'intelligenza artificiale. Diversamente da molti altri regolamenti, la Proposta introduce obblighi per coloro che usufruiscono di sistemi di IA ad alto rischio: essi si devono attenere scrupolosamente alle regole date assieme al sistema stesso, monitorandone il funzionamento e conservandone i log generati automaticamente, vale a dire le registrazioni sequenziali e cronologiche delle operazioni effettuate.

Si abbassa la potenziale rischiosità dei sistemi di IA a limitata, qualora si esigano determinati obblighi di trasparenza; per esempio, deve essere richiesta un'autorizzazione a procedere ad un'operazione, rendendo consapevole l'utente del servizio di avere a che fare con una chatbot. Anche nel caso di sistemi usati per il riconoscimento delle emozioni o per la caratterizzazione biometrica, così come nel caso del *"deep fake"* [100, 50], in cui si alterano immagini o video in modo che somiglino notevolmente a persone, luoghi o altre entità esistenti, l'utente deve essere informato e deve informare le altre persone fisiche che li utilizzano del fatto che sono sottoposte a un provvedimento del genere, come riportato nell'articolo 52 della proposta. Determinati sistemi di IA, come quello sopra riportato non richiedono una stringente normativa come nei casi precedentemente esposti, ma vanno incontro comunque all'obbligo di trasparenza: nel caso di un sistema in grado di riconoscere le emozioni o in grado di effettuare una categorizzazione biometrica si deve informare la persona su quale sia il trattamento cui viene sottoposta.

Infine, il rischio è classificato come minimo per i sistemi che rappresentano un rischio nullo o limitato per la sicurezza dei cittadini, come per i filtri antispam basati sull'IA.

4.1.5 Sostegno all'innovazione, governance e autorità preposte

Prima dell'immissione nel mercato di qualsiasi sistema innovativo che sfrutta l'intelligenza artificiale, sono forniti dalle autorità preposte degli Stati membri o dal Garante europeo della protezione dei dati spazi di sperimentazione normativa, che consistono in ambienti controllati che permettono uno sviluppo sicuro per un periodo di tempo limitato di tali innovazioni. La sicurezza in questi spazi deve essere garantita non solo nei riguardi delle persone fisiche, ma anche nel rispetto della protezione dei dati personali degli utenti, su cui devono vigilare le autorità nazionali competenti, come sottolineato dal secondo paragrafo dell'articolo 53. La questione è approfondita in modo dettagliato nell'articolo appena successivo, il 54, nel quale si afferma che i dati personali raccolti legalmente per altre finalità possono essere usati nello spazio di sperimentazione solo sotto determinate condizioni. Prima e fondamentale eccezione è dovuta alla salvaguardia di un interesse pubblico rilevante, come la salute, la sicurezza o la protezione dell'ambiente. Altre norme da rispettare qualora si voglia agire su questo tipo di dati sono la messa a punto di meccanismi di monitoraggio efficaci o la separazione dei *personal data* supervisionata dallo stesso individuo cui tali dati si riferiscono dagli altri dati usati. È fatto inoltre divieto di trasmissione, trasferimento o addirittura consultazione di dati personali da parte di terzi esterni al progetto di sviluppo, che si occuperà della cancellazione degli stessi una volta terminata la partecipazione allo spazio di sperimentazione e fornirà risultati e descrizione completa del processo e della logica alla base dell'addestramento, delle prove e della convalida del sistema di IA, nonché una breve sintesi del progetto alle autorità preposte. L'ultimo articolo che fa parte del Titolo V "Misure a sostegno dell'innovazione", il 55, permette ai fornitori di piccole dimensioni e alle start-up un accesso prioritario allo spazio di ricerca europeo.

Si arriva così al Titolo VI, che istituisce il Comitato Europeo per l'Intelligenza Artificiale e fornisce regole certe per la designazione delle autorità nazionali competenti in materia di IA. Il Comitato è di fondamentale rilevanza, perché è istituito per permettere una efficace cooperazione tra autorità nazionali e Commissione, contribuendo in questo modo a determinare gli orientamenti legislativi e il rispetto del regolamento. Significativo è il coinvolgimento nella struttura del Comitato, presieduto dalla Commissione di Garante europeo della protezione dei dati, assieme ovviamente al capo o alto funzionario di livello equivalente delle autorità nazionali, a sottolineare il carattere inscindibile dell'intelligenza artificiale dal trattamento equo dei dati personali. L'autorità nazionale, una per ciascuno stato, deve basare la propria competenza in mododa salvaguardare l'obiettività e l'imparzialità dei loro compiti e attività, come indicato nel primo paragrafo dell'art. 59.

Nell'art. 60 viene anche istituita una banca dati europea per i sistemi di intelligenza artificiale indipendenti ad alto rischio, che contiene dati personali delle persone fisiche responsabili della registrazione del sistema stesso, quali nome e dati di contatto, solo nella misura necessaria per il trattamento in conformità al regolamento. Questa banca dati permette più agevolmente di contattare il responsabile in caso di violazioni o

pratiche poco chiare, ma permette altresì di individuare con certezza a chi si devono rivolgere le autorità che vigilano sull'IA. Ciò è reso possibile anche da un piano di monitoraggio successivo all'immissione sul mercato che deve essere stilato dai fornitori nel caso di sistemi ad alto rischio, ex art. 61. Come accade nel GDPR, l'avvenuta violazione degli obblighi previsti deve essere tempestivamente segnalata alle autorità preposte, qualora sia stato accertato un nesso causale o una sua alta probabilità tra il sistema e il danno. Alla luce della protezione dei dati, risulta di notevole importanza la possibilità per le autorità di vigilanza di accedere ai dataset di addestramento, convalida e prova usati dal fornitore, per consentire una visione più chiara possibile del panorama entro cui funziona il sistema; questo controllo si estende a qualsiasi documento mantenuto stando alle norme della Proposta nel caso di sistemi di IA ad alto rischio. Proprio come avvenuto nel GDPR, anche in questo documento si fa molto affidamento nella creazione di marchi, ma soprattutto nell'adesione volontaria a codici di condotta elaborati spontaneamente da singoli fornitori o da organizzazioni che li rappresentano.

Nel decimo titolo, che va dall'articolo 70 al 74 compresi, il regolamento passa a definire le sanzioni previste, seguendo il principio di riservatezza delle indagini, e, nello specifico, come il Garante europeo della protezione dei dati può infliggere sanzioni amministrative pecuniarie agli enti presi in considerazione dalla proposta, fermo restando il diritto a difendersi delle parti interessate, consentendo la tutela dei propri dati personali o dei segreti aziendali che questi procedimenti potrebbero svelare.

Dunque, sintetizzando la proposta avanzata dall'Unione, essa garantisce in ogni possibile scenario, perfino nel caso in cui si ravvisassero scorrettezze da parte dei fornitori o da parte degli organi di controllo, di salvaguardare l'integrità e la riservatezza dei dati personali forniti. In questo modo si è cercato di instillare anche nella parte di pubblico più dubbiosa, soprattutto in coloro che ne vedono l'eccessiva invadenza nel privato, la fiducia verso questa forma di tecnologia così avanzata da temere una sua espansione incontrollata e deleteria. Infine, l'UE ha cercato di imbrigliare l'IA in modo da costituire una base giuridica certa anche per i fornitori, soprattutto quelli interni, ma anche ponendo un freno a quei sistemi esterni all'Unione ritenuti inaccettabili per la libertà del cittadino; così si è garantito anche un ambiente sicuro per lo sviluppo e la ricerca in questo ambito. Si vedrà col tempo e le possibili modifiche, che saranno apportate in sede comunitaria o nei singoli stati, l'efficacia della attuale proposta e la sua effettiva attuabilità. Il documento stilato ha avuto già il pregio di indirizzare gli stati membri a prendere in considerazione una legislazione conforme alle linee guida individuate, come nel caso dell'Italia, che nel novembre del 2021 ha redatto un *"Programma strategico per l'Intelligenza Artificiale"* con orizzonte il 2024.

4.2 Programma italiano

Come in sede comunitaria, così anche in terra italiana questo Programma risulta essere solo l'ultimo provvedimento in ordine di tempo in materia di intelligenza artificiale; infatti, nel luglio 2020 uscì un documento del Ministero dello Sviluppo Economico

dal titolo "Proposte per una Strategia italiana per l'intelligenza artificiale", citato già all'inizio del capitolo 3. In questo lungo e dettagliato documento si sviluppano i temi cari alla visione europea, come lo sviluppo sostenibile, la salvaguardia del pianeta, un'IA antropocentrica, i possibili rischi di tali sistemi, un ecosistema affidabile e competitivo; per questo motivo è stato preso come base per stilare il Programma, più pragmatico e specifico della Strategia. È ribadito inoltre un concetto caro a tutto il mondo della ricerca scientifica odierna: *"come tutte le tecnologie general purpose, l'IA non è buona o cattiva in sé: dipende dall'uso che se ne fa"* [62]. Per questo motivo si suddividono le cause di un danno o di un pregiudizio in volontarie e involontarie. Mentre nel primo caso lo scopo fraudolento è chiaro, nel secondo può derivare da un uso maldestro dell'IA, che comporta un'amplificazione del bias già esistente, specialmente se si utilizzano dataset non rappresentativi, come si vedrà nel capitolo 5.

Tale Programma, approvato dal Consiglio dei ministri il 24 novembre del 2021, ha evidenziato innanzitutto come il panorama italiano sia caratterizzato da quattro tipologie di attori: la ricerca, l'istruzione e la formazione, le infrastrutture e le comunità. La prima è svolta in maniera preponderante nei laboratori di ricerca delle università e nei centri di ricerca pubblici e in fondazioni di ricerca dedicate [28]. L'istruzione e la formazione avvengono all'interno delle università che forniscono più di 200 curricula in IA. In Italia sono presenti diverse infrastrutture di ricerca di alto livello, fornite, per esempio, dal CINECA o dal CNR, che permettono a scienziati di qualsiasi disciplina esperimenti, anche complessi, richiedenti alta capacità di calcolo. Molti dei ricercatori italiani attivi nell'ambito dell'intelligenza artificiale partecipano alle più prestigiose reti di ricerca internazionale sull'IA. Benché l'Italia formi un'eccellente comunità di ricerca riconosciuta a livello internazionale, si constata una frammentazione della ricerca, data dalla bassa integrazione interdisciplinare nei laboratori e dalla mancanza di scala e massa critica; ciò comporta anche una insufficiente attrazione di talenti, favorita anche dall'assenza di una strategia che attiri i ricercatori stranieri. La conseguenza diretta di questi problemi è costituita dalla limitata capacità brevettuale. Inoltre, è da sottolineare come in questo settore si riscontra un problema cronico della ricerca ingegneristica in Italia, cioè una minore presenza di donne, meno del 20%, contro il 50% delle STEMin generale [28]. In confronto alle altre nazioni europee economicamente simili all'Italia il nostro paese investe poco in ricerca, sviluppo, formazione e nel settore privatosi effettuano meno investimenti nell'intelligenza artificiale. Dunque, dal momento che molte tecnologie ruoteranno attorno all'applicazione efficace e sempre più spinta di IA, urgono provvedimenti, come riportato dal Programma, in grado di rendere l'Italia un paese competitivo in questo fondamentale ambito. Proprio in quest'ottica si inserisce la volontà di stabilire un quadro giuridico basato su valori e principi etici che non vadano a ostacolare lo sviluppo nel settore. Come richiesto dall'Unione, anche l'intelligenza artificiale italiana vuole essere antropocentrica, affidabile e sostenibile dal puntodi vista sia economico, sia ambientale. Faro di questo rinnovamento sarà la pubblica amministrazione, che porrà in atto sistemi per l'attenuazione o l'eliminazione dei rischi che comporta l'IA e ne sfrutterà gli enormi vantaggi che offre, seguendo il mantra di

"governare l'IA e governare con l'IA" [28].

Il grande numero di settori presi in considerazione dal Programma, da quello finanziario a quello agroalimentare, dalla salute alla cultura, sottolinea la trasversalità dell'utilizzo che si può e si potrà fare di questa tecnologia, se supportata in modo oculato dallo stato, promuovendo corsi e carriere in materie scientifiche (STEM), rafforzando le competenze della PA in materia e sostenendo la ricerca. Un'altra proposta avanzata dal documento approvato dal Consiglio riguarda una piattaforma di dati e software condivisa da tutti i protagonisti della ricerca, che garantisca la protezione dei dati personali, per tutelare la proprietà intellettuale nazionale e favorire tutte quelle aziende che intendono investire in questa tecnologia. La PA dovrà garantire protocolli di conformità rispetto alle normative nazionali ed europee, in modo da permetterela riutilizzabilità dei nuovi sistemi rilasciati dalle diverse amministrazioni nel rispetto delle regole per la protezione dei dati personali. Nel settore pubblico l'Italia è pronta ad affrontare sfide come la sua digitalizzazione e la sua modernizzazione, la tutela del territorio e delle risorse idriche, la manutenzione stradale, la telemedicina, l'innovazione e la digitalizzazione della sanità. Tutti questi ambiti riguardano infrastrutture critiche, che rientrano nei sistemi di IA ad alto rischio, come indicato dalla proposta europea; quindi, andrà disposta in sede parlamentare una discussione per conformare questa spinta all'innovazione con le direttive comunitarie.

Le trasformazioni digitali, quale è l'intelligenza artificiale, coinvolgono l'economia intera, avendo ripercussioni in molti aspetti della società, perciò, non possono essere gestite in isolamento. Per questo motivo il Programma ha istituito un gruppo di lavoro permanente in materia all'interno del Comitato Interministeriale per la Transizione Digitale, che richiede l'attiva partecipazione di diversi ministeri: il Ministero dell'Università e della Ricerca, il Ministero dello Sviluppo Economico e il Ministro per l'Innovazione Tecnologica e la Transizione Digitale.

Capitolo 5

Casi di studio

Tutte le misure che l'Unione Europea e l'Italia stanno intraprendendo per cercare di mettere ordine nella babele di provvedimenti normativi che caratterizza l'ambito tecnologico e l'IA nello specifico, non sono altro che risposte a questioni aperte presso l'opinione pubblica. In questi ultimi anni sono stati diversi i casi di violazione dei dati personali, ripresi dai più importanti media del mondo, che hanno suscitato interrogativi e perfino scatenato inchieste nazionali; sicuramente il più famoso per quantità di utenti raggirati e clamore mediatico è quello di Cambridge Analytica [86].

5.1 Cambridge Analytica

La qui presente trattazione non vuole addentrarsi né nella cronaca della vicenda, né tantomeno nelle vicende giudiziarie che hanno portato alla chiusura di questa azienda di consulenza nel 2018, ma vuole mettere in luce come ha agito in modo fraudolento per manipolare l'espressione di voto negli Stati Uniti d'America al fine di aver forse influenzato le elezioni presidenziali del 2016. L'entità del danno provocato dalla consapevole violazione dei diritti dei cittadini americani non è facilmente calcolabile, ma rimane inequivocabile che in alcuni stati, soprattutto in quelli più in bilico, questa azione abbia potuto seriamente compromettere la validità del voto. Le elezioni statunitensi non furono l'unico caso a finire sotto la lente d'ingrandimento, ma un'altra votazione ne risultò con molta probabilità influenzata, anche se in maniera minore, per via della quantità minore di abitanti inglesi rispetto agli americani: il referendum sulla Brexit. Ovviamente, non essendoci dati certi in questo caso, si approfondirà più nel dettaglio l'azione fraudolenta mesa in pratica oltreoceano.

Quello che avvenne in terra americana si rivelò essere una sistematica violazione della privacy degli utenti, atta a ottenere una profilazione quanto mai accurata, in modo da influenzarne la libertà di esprimere un voto senza essere condizionato da terzi, in questo caso dalle nuove tecnologie e dai loro creatori. Poco prima delle elezioni venne proposto a milioni di possibili votanti americani un dettagliato test della personalità per determinare quale partito fossero più inclini a votare. Il test, effettuato da circa 320000 elettori, per essere remunerato tra i 2 e i 5 $, richiedeva di accedere al proprio profilo

Facebook ed era, secondo l'azienda di consulenza, sottoposto per scopi accademici.Fino a questo punto nulla di male, ma già l'accesso alla propria pagina personale dovevafar pensare sulle reali intenzioni della "ricerca accademica" [83]. Infatti, semplicemente con questa richiesta, Cambridge Analytica ha effettuato la profilazione dell'utente, usando i suoi dati pseudonimizzati come fossero dati personali contenuti in una banca dati interna all'azienda. Giunti a questo punto è legittimo chiedersi se 320mila persone potessero influire sull'elezione di un presidente di una nazione come gli USA di quasi 330 milioni di abitanti; sicuramente il numero è risibile al confronto, se non si tenesse conto che tra i dati così acquisiti vi erano anche le amicizie virtuali richieste e accettate nella piattaforma digitale: in questo modo Cambridge Analytica ha potuto condizionare tra i 30 e i 50 milioni di elettori. Vista sotto quest'ottica la violazione assume proporzioni macroscopiche, fornendo la tesi della possibile illegittimità delle elezioni federali. Quindi l'azienda in questione usò l'insieme di dati dei cittadini che sierano sottoposti al test come training set per costruire un modello in grado di definirele preferenze politiche dei loro amici e di altre persone. In questo modo si è riusciti ad individuare quali elettori fossero incerti nell'indicazione del proprio voto, soprattutto negli stati con poco scarto secondo i sondaggi, e, di conseguenza, ad influenzarne il comportamento elettorale tramite pubblicità o altri messaggi persuasivi 5.1.

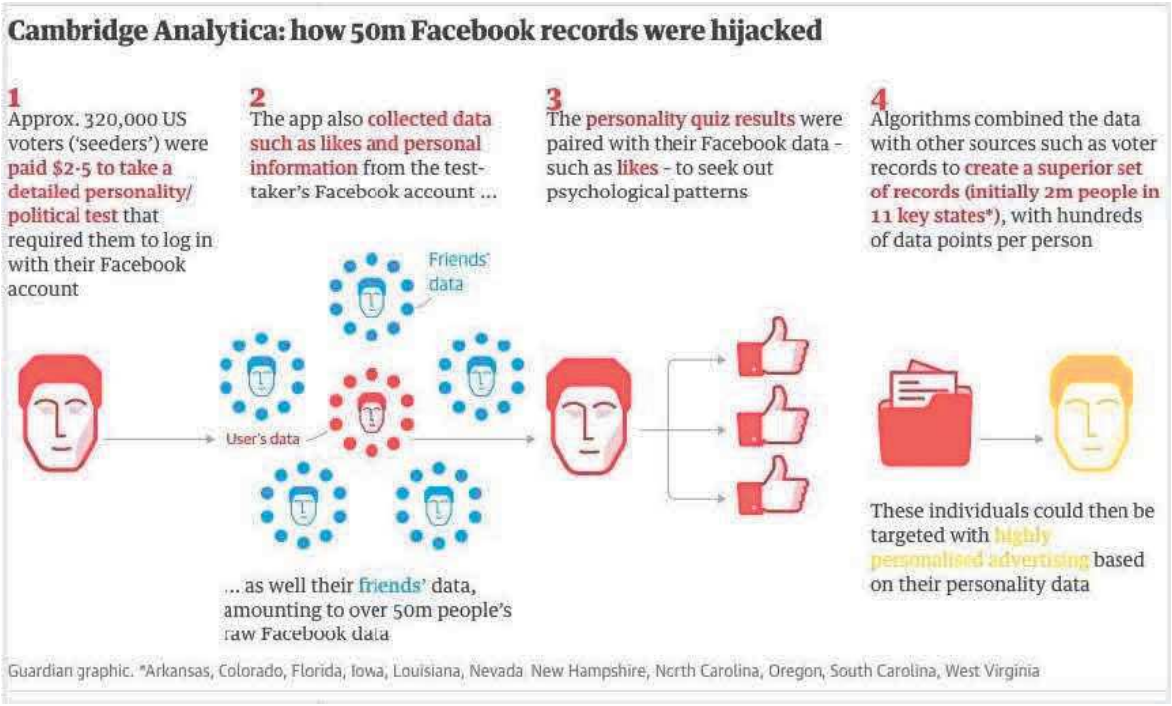

Figura 5.1: Meccanismo di profilazione degli utenti nel caso Cambridge Analytica © Giovanni Sartor

Proprio dalle rivelazioni dovute alla scoperta e alla condanna di questa manipolazione si è partiti a modellare la proposta europea, come il divieto di usare dati per scopi diversi da quelli cui si è dato il consenso oppure l'uso di algoritmi basati sull'IA atti alla profilazione degli individui (in quest'ultimo caso si violerebbe anche il GDPR, in quegli anni ancora in fase di stesura definitiva).

Lo scandalo di Cambridge Analytica rappresenta un esempio in cui gli autori del

reato erano consci dell'illegalità del loro operato, sfruttando anche buchi normativi dovuti all'innovativo sistema di IA usato per profilazione. Un ulteriore caso in cui, invece, il software non era stato sviluppato con scopi fraudolenti, ma non considerando alcuni aspetti delicati come la geolocalizzazione in tempo reale e condivisa con altri utenti, è costituito dall'applicazione Strava.

5.2 Strava

Seguendo l'onda della vita ai tempi dei social, venne rilasciata sul mercato questa app di fitness, che permette di salvare i propri percorsi effettuati in bici o di jogging per poi condividerne la mappa con gli altri utenti. In Europa non ha riscosso tanto successo quanto negli USA, dove spopolò, anche all'interno degli ambienti militari,e continua a essere scaricata. Strava, nata nel 2009, è stata al centro di un caso abbastanza seguito dalla stampa nel 2018, quando un ricercatore australiano esperto in sicurezza internazionale, Nathan Ruser, si è accorto, leggendo un post su un blog di cartografia, come sulla *heatmap*, una mappa di calore, che l'applicazione metteva a disposizione degli utenti per segnalare quali fossero i percorsi più consigliati e utilizzatinella zona vicino a dove ci si trovava, comparissero percorsi che combaciavano con le planimetrie delle basi militari statunitensi in luoghi di conflitto, come l'Afghanistan, laSiria e lo Yemen [42]. Benché solitamente i perimetri delle basi fossero ben conosciuti,in questo modo si potevano trarre conclusioni su quali fossero le basi più usate e i percorsi effettuati dai soldati all'interno delle stesse. In questo modo si sarebbero potute effettuare incursioni che permettessero un risultato eccellente anche con pochi mezzi e ciò avrebbe comportato un rischio molto elevato dal punto di vista strategico, inzone, come quelle in questione, di conflitto guidato soprattutto da attacchi terroristici che hanno per loro natura caratteristiche di limitatezza di forze e obiettivi mirati.

Si andò quindi a cercare quale fosse la causa del problema e si capì come la heatmap fosse l'elaborazione di dati di migliaia di utenti, in questo caso soldati, che percorrevano il perimetro delle basi per allenamento con la geolocalizzazione del proprio smartphone attiva. Ovviamente le impostazioni di Strava permettevano già allora di disattivare esplicitamente la raccolta di dati per le mappe di calore, impostando "zone di privacy" in località predefinite dall'utente. Probabilmente per negligenza da parte dei militari stessi, quindi, sono state rese di dominio pubblico queste importanti informazioni. Dunque, sono state presi seri provvedimenti per evitare nuove dimenticanze come queste, decidendo di vietare l'uso di questa app ai soldati.

Come visto, gli sviluppatori dell'applicazione non ne avevano previsto un largo utilizzo da parte di utenti appartenenti all'esercito, ma l'unica possibile contromisura non è dipesa da loro. Infatti, a differenza di Cambridge Analytica, che dopo lo scandalo è stata chiusa, Strava continua a essere scaricata e utilizzata, perché considerata sicura e utile. Benché sia stata garantita la pseudonimizzazione delle informazioni apprese dagli utenti [91], senza una effettiva intrusione nella sfera privata degli utenti, comunque si sono verificati spiacevoli inconvenienti come quello sopra descritto e questa situazione

evidenzia come la tecnologia possa garantire delle comodità assolutamente apprezzabili, ma al contempo possa, anche nel caso più banale, creare non pochi problemi, se usata in modo superficiale.

5.3 Caso positivo

Agli onori della cronaca sono balzati casi come quelli precedentemente esposti, che hanno rivelato falle in questo o quel settore [90, 74], ma nella maggior parte dei casi avviene una attenta applicazione dei protocolli in materia di dataset di dati personali e loro utilizzo in materia di IA. Certamente creano molta meno audience e, quindi, vengono trattati in ambienti più specialistici rispetto alle violazioni più evidenti, ma costituiscono una prova tangibile di conciliazione tra dati personali e intelligenza artificiale.

Un esempio di questa tipologia è l'accordo stipulato tra il CIRPA (Centro Interdipartimentale per la ricerca in diritto, economia e management della Pubblica Amministrazione dell'Università degli studi di Salerno) e So.Re.Sa s.p.a. (Società Regionale per la Sanità per la regione Campania), che ha avuto come finalità lo sviluppo di un modello di analisi e controllo del percorso di nascita, in modo di supportare le politiche sanitarie per l'appropriato ricorso al parto cesareo, tramite un sistema integrato di *data management* e strumenti di *mobile health*, in grado di gestire la diversificata mole di dati legata all'evento nascita. Questo provvedimento si è reso necessario, dal momentoche la percentuale di parti cesarei in tale regione risulta molto più elevata di quella nazionale, eccedendo la quantità di tali operazioni raccomandata dall'OMS [94].

In questo caso le parti in causa hanno messo in pratica il protocollo GDPR, ancora non in vigore nel 2017, anno della stipula, ma in via di attuazione a partire dal 2018, facendo riferimento all'articolo 9 che impone il divieto di trattamento di dati personali, tranne per alcune categorie specifiche; di queste tipologie fanno parte i dati resi manifestamente pubblici dall'interessata e, più comune in questo caso, i dati sanitari che hanno come scopo un miglioramento per la collettività. Sorge spontaneo il dubbio su quando una donna possa rendere manifestamente pubblico il parto cesareo: il fatto è che non corrisponde a una scelta prettamente privata operata dalla partoriente, ma auna scelta che può essere programmata ed essere espletata a suo piacimento in forma pubblica dalla stessa. Tale scelta, se non motivata da urgenze cliniche, quindi consente di evitare la richiesta di un ulteriore consenso ad operare sui suoi dati da parte della donna. Un ulteriore ambito in cui il processo di questi dati sanitari sono collezionabili è quello della ricerca e della statistica, che permettono, in questo specifico caso, di determinare l'eziologia di questo malfunzionamento della sanità pubblica [94]. Quindi lo Stato può prelevare questi dati anche in assenza di consenso da parte della paziente, in modo da consentire la ricerca in un ambito di vitale importanza per la società. Inoltre si considera anche il primo comma dell'articolo 89, per cui vige la minimizzazione e pseudonimizzazione dei dati collezionati, che possono essere trasferiti ad enti o istituti di ricerca solo se privati della loro personalità, in modo da renderli non identificabili con

questa o quella paziente. Trovandosi in suolo italiano, tale ricerca ha dovuto sottostare anche al "Codice in materia di protezione dei dati personali" del 2003, entrato in vigore nel 2004, che ha subito un ammodernamento fondamentale per integrare la legislazione nazionale con il GDPR, le "Regole deontologiche per trattamenti a fini statistici o di ricerca scientifica". Tali regole devono essere applicate in questo caso, avendo tre aspetti direttamente ricollegabili al primo articolo di questo provvedimento, dato che si tratta di un fenomeno collettivo, che richiede l'utilizzo di atti o documenti, quali le cartelle cliniche da parte di un incaricato da parte dell'ente di ricerca, come un privato. I dati processati in questo specifico caso possono comprendere anche dati genetici, che hanno restrizioni importanti nel trattamento, dato l'alto contenuto personale e la possibile discriminazione che ne potrebbe derivare; infatti gli unici ambiti in cui possono essere trattati e, nello specifico, in cui l'IA può processarli sono legati alla sperimentazione clinica o alla ricerca scientifica, atte a migliorare o tutelare la salute dell'interessata, di terzi o della collettività. Infine si è pronti ad indagare questa collezione di dati algoritmicamente tramite l'intelligenza artificiale, avendo assolto tutti gli obblighi del caso.

Questo esempio evidenzia come, anche prima della stesura di un esaustivo regolamento in materia, gli altri protocolli permettessero già di applicare le norme in modo coerente per tutelare i dati personali delle persone coinvolte. Tuttavia si è resa indispensabile una proposta, come quella individuata dall'Unione Europea, che ha l'obiettivo di costituire una base giuridica certa per le applicazioni in materia di *data protection* e intelligenza artificiale. Inoltre si può rilevare che, in casi specifici e limitati come quello qui esposto, la tutela della privacy individuale può retrocedere nell'interesse di un bene comune più ampio che la società tutta possa beneficiare.

Capitolo 6

Conclusione

Le nuove tecnologie sembrano lanciare una sfida difficilmente affrontabile con l'ausilio solo di impianti legislativi legati a un mondo passato caratterizzato da una limitata capacità di calcolo e di autonomia delle macchine, ma, data la rapidità del loro sviluppo e della loro commercializzazione, l'ente legiferante deve promuovere l'introduzione di un diritto sempre al passo con il ritmo, a volte frenetico, del mondo dell'hi-tech. La pandemia da Covid-19 ha aumentato una tendenza già in precedenza significativa: l'uso delle innovazioni più recenti da parte di quasi tutta la popolazione, che ha favorito l'uso di mezzi all'avanguardia anche nella produzione, come nel caso di agenti robotici dotati di autonomia all'interno delle fabbriche [41]. Quest'ultimo ambito ha sollevatonon poche criticità in fatto di responsabilità e di una sua equa assegnazione in caso di danno, ma non solo, perché viene a delinearsi una possibile disoccupazione tecnologica che potrebbe coinvolgere i lavori meno qualificati e più ripetitivi, con la conseguente scomparsa del ceto medio. Come visto quindi non è semplice trovare una regolazione equilibrata che non proibisca in modo categorico l'uso dell'intelligenza artificiale, mane consenta un suo sviluppo in serenità.

La comunità europea e nazionale sta cercando, con un'opera legislativa lungimirante, di accrescere la fiducia della società nell'intelligenza artificiale, senza al contempo ostacolarne ulteriori sviluppi. Certamente è apprezzabile l'attenzione rivolta alla tutela dei dati personali anche nell'utilizzo di tali sistemi, vero diritto da salvaguardare in questa realtà permeata di tecnologie sempre più sofisticate e difficilmente intelligibili. Un altro punto chiave fondamentale della Proposta europea, come anche del Programma italiano, consiste nell'apertura a un dialogo costruttivo con associazioni di consumatori, università, poli di ricerca, aziende operanti nel settore, sia multinazionali, sia piccole-medie imprese che indirizzino il diritto, garantendo un approccio equilibrato, che non pregiudichi una futura auspicata leadership europea nel settore. Questo dialogo potrebbe degenerare nel momento in cui le poche aziende al vertice e di fatto reggenti il monopolio nell'industria tecnologica avessero un peso più determinante nel dibattito e venissero prese come interlocutori privilegiati, magari sviluppando un regolamento conforme alle loro aspettative, in modo da avvantaggiarle. Si tratta però, allo stesso tempo, di non imporre una linea troppo dura verso le multinazionali,

che potrebbero scegliere di annullare i finanziamenti nelle aziende europee a fronte di obblighi insormontabili o molto difficili da adempiere. Dunque il lavoro che attende l'Unione è di tipo funambolico, come se si fosse sulla lama di un rasoio, ma la Proposta rappresenta un buon punto di partenza per assicurare la certezza del diritto in modo che rassicuri tutti i portatori di interesse, coinvolti in prima persona a fornire i loro pareri e suggerimenti per adottare un regolamento che tuteli il cittadino e il produttore al tempo stesso. La strada è tracciata, ora serve un lavoro di coordinamento delle operazioni che consenta un, se non eccellente, almeno accettabile compromesso dell'opera di regolazione, dove la centralità dei dati sensibili, come parte indivisibile della persona umana, non sia messa in discussione.

Bibliografia

[1] S. Albawi, T. A. Mohammed, e S. Al-Zawi. Understanding of a convolutional neural network. In *2017 international conference on engineering and technology (ICET)*, pages 1–6. Ieee, 2017.

[2] C. Allen, I. Smit, e W. Wallach. Artificial morality: Top-down, bottom-up, and hybrid approaches. *Ethics and information technology*, 7(3):149–155, 2005.

[3] I. Asimov. *Io, robot*. Edizioni Mondadori, 2018.

[4] J. M. Balkin. The three lawsof robotics in the age of big data. *Ohio St. LJ*, 78:1217, 2017.

[5] M. Ball. The metaverse: What it is, where to find it, and who will build it. *https://www.matthewball.vc/all/themetaverse*, 13 Gennaio 2020.

[6] M. Bassini, L. Liguori, e O. Pollicino. Sistemi di intelligenza artificiale, responsabilità e accountability. verso nuovi paradigmi? Giappichelli Editore, 2018.

[7] V. Bates Ramirez. Sony's Racing AI Just Beat The World's Best Gran Turismo Drivers. *https: // singularityhub. com/ 2022/ 02/ 10/ sonys-ai-beat-the-worlds-best-gran-turismo-drivers/*, 10 Febbraio 2022.

[8] A. Bertolini. Robots as products: the case for a realistic analysis of robotic applications and liability rules. *Law, innovation and technology*, 5(2):214–247, 2013.

[9] R. Biolcati. La vita online degli adolescenti: tra sperimentazione e rischio. *Psicologia clinica dello sviluppo*, 14(2):267–298, 2010.

[10] C. M. Bishop e N. M. Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[11] M. Borgobello. Il riconoscimento facciale vive tra noi (e ci osserva): ecco i rischi privacy. *https://www.agendadigitale.eu/sicurezza/privacy/il-riconoscimento-facciale-nelle-nostre-strade-ecco-i-rischi-privacy/*, 22 Settembre 2021.

[12] P. Brown et al. The science of successful learning, 2014.

[13] E. Brynjolfsson e T. Mitchell. What can machine learning do? Workforce implications. *Science*, 358(6370):1530–1534, 2017.

[14] L. Butterfield. Leading academics reveal: what are we getting wrong about AI, 2018.

[15] M. Campbell, A. J. Hoane Jr, e F. Hsu. Deep blue. *Artificial intelligence*, 134(1-2):57–83, 2002.

[16] L. Carlucci. Intelligenza artificiale, robotica e macchine intelligenti parte 1. *https://www.lincei.it/it/videoteca/05032021-intelligenza-artificiale-robotica-e-macchine-intelligenti-parte-1*, 5 Marzo 2021.

[17] M. Castells. *Volgere di millennio*. EGEA spa, 2014.

[18] E. Cau. I segreti del metaverso, nessuno sa niente del prossimo, mastodontico progetto di Zuckerberg. *Il Foglio*, 6 Novembre 2021.

[19] F. Chesani et al. A game-based competition as instrument for teaching artificial intelligence. In *Conference of the Italian Association for Artificial Intelligence*, pages 72–84. Springer, 2017.

[20] S. O. Chishti et al. Self-driving cars using CNN and Q-learning. In *2018 IEEE 21st International Multi-Topic Conference (INMIC)*, pages 1–7. IEEE, 2018.

[21] J. E. Cohen. *Between truth and power*. Oxford University Press, 2019.

[22] Commissione Europea. Libro bianco sull'intelligenza artificiale- un approccio europeo alla eccellenza e alla fiducia, 19 Febbraio 2020.

[23] Commissione Europea. Direttiva 95/46/ce del parlamento europeo e del consiglio, del 24 ottobre 1995, relativa alla tutela delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati, 2018.

[24] Commissione Europea. *Ethics guidelines for trustworthy AI*. Publications Office, 2019.

[25] Commissione Europea. Proposta di regolamento del parlamento europeo e del consiglio che stabilisce regole armonizzate sull'intelligenza artificiale. *https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206*, 21 Aprile 2021.

[26] Commissione Europea. Un'europa pronta per l'era digitale: la commissione propone nuove regole e azioni per l'eccellenza e la fiducia nell'intelligenza artificiale. *https://italy.representation.ec.europa.eu/notizie-ed-eventi/notizie/uneuropa-pronta-lera-digitale-la-commissione-propone-nuove-regole-e_it*, 21 Aprile 2021.

[27] Commissione Europea. L'intelligenza artificiale per l'Europa, 25 Aprile 2018.

[28] Consiglio dei ministri. Programma strategico, intelligenza artificiale 2022-2024. *https://assets.innovazione.gov.it/1637937177-programma-strategico-iaweb-2.pdf*, 24 Novembre 2021.

[29] F. Corea, C. G. Ferrauto, F. Fossa, A. Loreggia et al. *Intelligenza artificiale. Cos'è davvero, come funziona, che effetti avrà*. Bollati Boringhieri, 2020.

[30] Croce Rossa Internazionale. Icrc position on autonomous weapon systems. *https://www.icrc.org/en/document/icrc-position-autonomous-weapon-systems*,

12 Maggio 2021.

[31] S. Dabeer, M. M. Khan, e S. Islam. Cancer diagnosis in histopathological image: CNN based approach. *Informatics in Medicine Unlocked*, 16:100231, 2019.

[32] M. Daily et al. Self-driving cars. *Computer*, 50(12):18–23, 2017.

[33] E. Davis. AI amusements: the tragic tale of Tay the chatbot. *AI Matters*, 2(4):20–24, 2016.

[34] DLAPiper. Il futuro regolamento della tecnologia. *https://www.dlapiper.com/~/media/files/insights/publications/2021/05/futuro-regolamento-della-tecnologia-booklet.pdf*, 2021.

[35] M. Dorigo, M. Birattari, e T. Stutzle. Ant colony optimization. *IEEE computational intelligence magazine*, 1(4):28–39, 2006.

[36] J. Dressel e H. Farid. The accuracy, fairness, and limits of predicting recidivism. *Science advances*, 4(1):eaao5580, 2018.

[37] H. L. Dreyfus e S. E. Dreyfus. Making a mind versus modelling the brain: Artificial intelligence back at the branchpoint. In *Understanding the Artificial: On the future shape of artificial intelligence*, pages 33–54. Springer, 1991.

[38] K. Drinhausen e V. Brussee. China's social credit system in 2021. *From fragmentation towards integration*, page 12, 2021.

[39] M. P. Driscoll. *Psychology of learning for instruction*. Allyn & Bacon, 1994.

[40] A. Etzioni e O. Etzioni. Incorporating ethics into artificial intelligence. *The Journal of Ethics*, 21(4):403–418, 2017.

[41] Eurostat. Artificial intelligence in EU enterprises. *https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn 20210413-1*, 13 Aprile 2021.

[42] M. I. R. Fenwick, M. Hittle, e O. White. Fitness app Strava lights up staff at military bases. *BBCJournal Archive*, 2018.

[43] L. Floridi. *La quarta rivoluzione: come l'infosfera sta trasformando il mondo*. Raffaello Cortina Editore, 2017.

[44] L. Floridi. *Il verde e il blu: Idee ingenue per migliorare la politica*. Raffaello Cortina Editore, 2020.

[45] S. Forrest. Genetic algorithms. *ACM Computing Surveys (CSUR)*, 28(1):77–80, 1996.

[46] A. Frisoli. I nuovi utilizzi della robotica in medicina: scenari e ultime frontiere. *https://www.agendadigitale.eu/sanita/i-nuovi-utilizzi-della-telerobotica-scenari-e-ultime-frontiere/*, 28 Gennaio 2022.

[47] J. Furman e R. Seamans. AI and the economy. *Innovation policy and the economy*, 19(1):161–191, 2019.

[48] G. Green. Five ways ai is saving wildlife – from counting chimps to locating whales. *https://www. theguardian. com/environment/2022/feb/21/five-ways-ai-is-saving-wildlife-from-counting-chimps-to-locating-whales-aoe*, 21 Febbraio 2022.

[49] N. Grünberg e V. Brussee. China's 14th five-year plan – strengthening the domestic base to become a superpower. *https://merics.org/en/short-analysis/chinas-14th-five-year-plan-strengthening-domestic-base-become-superpower*, 09 Aprile 2021.

[50] D. Güera e E. J. Delp. Deepfake video detection using recurrent neural networks. In *2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS)*, pages 1–6. IEEE, 2018.

[51] G. Hallevy. I, robot-I, criminal: When science fiction becomes reality: Legal liability of AI robots committing criminal offenses. *Syracuse Sci. & Tech. L. Rep.*, page 1, 2010.

[52] AI HLEG. A definition of ai: Main capabilities and scientific disciplines. *https: // ec. europa. eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf*, 2019.

[53] A. K. Inkulu et al. Challenges and opportunities in human robot collaboration context of industry 4.0-a state of the art review. *Industrial Robot: the international journal of robotics research and application*, 2021.

[54] S. V. Konakalla. A star algorithm, 2014.

[55] A. Krizhevsky, I. Sutskever, e G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[56] Y. LeCun et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[57] L. Mantovan e L. Nanni. The computerization of archaeology: survey on artificial intelligence techniques. *SN Computer Science*, 1(5):1–32, 2020.

[58] M. Martorana. Dati personali: anonimizzazione e pseudonimizzazione. *https://www.altalex. com/documents/news/2021/06/08/dati-personali-anonimizzazione-e-pseudonimizzazione*, 8 Giugno 2021.

[59] H. Matsuzaki e G. Lindemann. The autonomy-safety-paradox of service robotics in Europe and Japan: a comparative analysis. *AI & society*, 31(4):501–517, 2016.

[60] J. McCarthy, M. L. Minsky, N. Rochester, e C. E. Shannon. Dartmouth artificial intelligence project proposal. *http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth. html*, 1955.

[61] R. S. Michalski, J. G. Carbonell, e T. M. Mitchell. *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, 2013.

[62] Mise. Proposte per una strategia italiana per l'intelligenza artificiale, 2019.

[63] B. Mondal. Artificial intelligence: state of the art. *Recent Trends and Advances in Artificial Intelligence and Internet of Things*, pages 389–425, 2020.

[64] P. M. Nadkarni, L. Ohno-Machado, e W. W. Chapman. Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, 18(5):544–551, 2011.

[65] J. Nida-Rumelin e N. Weidenfeld. *Umanesimo digitale: Un'etica per l'epoca dell'Intelligenza Artificiale*. Franco Angeli, 2019.

[66] P. Norvig e S. J. Russell. Intelligenza artificiale. *Un approccio moderno*, 2010.

[67] G Orwell. *1984*. Mondadori, 2004.

[68] H. Park, N. Kim, e J. Lee. Parametric models and non-parametric machine learning models for predicting option prices: Empirical comparison study over kospi 200 index options. *Expert Systems with Applications*, 41(11):5227–5237, 2014.

[69] Parlamento Europeo e Consiglio Europeo. Trattato sul funzionamento dell'unione europea,13 Dicembre 2007.

[70] Parlamento Europeo e Consiglio Europeo. Carta dei diritti fondamentali dell'unione europea, 2000.

[71] Parlamento Europeo e Consiglio Europeo. Regolamento generale sulla protezione dei dati (GDPR), 2016.

[72] Parlamento Europeo e Consiglio Europeo. Direttiva (UE) 2016/680 del parlamento europeo edel consiglio, 27 Aprile 2016.

[73] F. Pizzetti et al. *Intelligenza artificiale, protezione dei dati personali e regolazione*, volume 6. G Giappichelli Editore, 2018.

[74] W. Nicholson Price e I. G. Cohen. Privacy in the age of medical big data. *Nature medicine*, 25(1):37–43, 2019.

[75] M. Przepiórkowski. Technological revolution–irrational fears and justified dangers. *Młoda Humanistyka*, 17(2), 2021.

[76] Z. Qin et al. Artificial intelligence method to design and fold alpha-helical structural proteins from the primary amino acid sequence. 36:100652, 2020.

[77] L. Rendina. Privacy vs protezione dati personali: attenti alla differenza, ne vadella nostra identità. *https: // www. agendadigitale. eu/ sicurezza/ privacy/privacy-e-protezione-dati-personali-cosa-sono-quali-differenze-cosa-e-cambiato-col-gdpr/*, 2019.

[78] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[79] F Rossi. *Il confine del futuro: Possiamo fidarci dell'intelligenza artificiale?* Feltrinelli Editore, 2019.

[80] O. Russakovsky et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

[81] S. Russell, D. Dewey, e M. Tegmark. Research priorities for robust and beneficial artificial intelligence. *Ai Magazine*, 36(4):105–114, 2015.

[82] A. L. Samuel. Eight-move opening utilizing generalization learning (see appendix b, game g-43.1 some studies in machine learning using the game of checkers). *IBM Journal*, pages 210–229, 1959.

[83] G. Sartor e F. Lagioia. The impact of the general data protection regulation (GDPR) on artificial intelligence. *European Parliamentary Research Service*, pages 1–100, 2020.

[84] F. Sauer. *Lethal autonomous weapons systems*. Routledge, 2021.

[85] F. Scarselli et al. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80, 2008.

[86] C. O. Schneble, B. S. Elger, e D. Shaw. The Cambridge Analytica affair and internet-mediated research. *EMBO reports*, 19(8):e46579, 2018.

[87] M. Schuld, I. Sinayskiy, e F. Petruccione. The quest for a quantum neural network. *QuantumInformation Processing*, 13(11):2567–2586, 2014.

[88] Schwartz et al. Green AI. *Communications of the ACM*, 63(12):54–63, 2020.

[89] B. Siyah. Imagenet winning CNN architectures (ILSVRC). *https://www.kaggle.com/discussions/ getting-started/ 149448* , 2020.

[90] M. Smith e S. Miller. The ethical application of biometric facial recognition technology. *Ai &Society*, 37(1):167–175, 2022.

[91] Strava. Informativa sulla privacy di Strava, 7 Luglio 2021.

[92] G. Tamburrini. *Etica delle macchine: dilemmi morali per robotica e intelligenza artificiale*. Carocci editore, 2020.

[93] M. E. Tipping. Bayesian inference: An introduction to principles and practice in machinelearning. In *Summer School on Machine Learning*, pages 41–62. Springer, 2003.

[94] R. Trezza. Diritto e intelligenza artificiale: etica, privacy, responsabilità, decisione. *Diritto eintelligenza artificiale*, pages 29–48, 2020.

[95] I. Tsamardinos, L. E. Brown, e C. F. Aliferis. The max-min hill-climbing Bayesian networkstructure learning algorithm. *Machine learning*, 65(1):31–78, 2006.

[96] A. M. Turing. Intelligent machinery, 1948.

[97] A. M. Turing e J. Haugeland. Computing machinery and intelligence. *The Turing Test: VerbalBehavior as the Hallmark of Intelligence*, pages 29–56, 1950.

[98] J. M. Twenge. *iGen: Why today's super-connected kids are growing up less rebellious, more tole- rant, less happy–and completely unprepared for adulthood–and what that means for the rest ofus*. Simon and Schuster, 2017.

[99] E. Waltz. AI takes its best shot: What AI can-and can't-do in the race for a Coronavirus vaccine-[vaccine]. *IEEE Spectrum*, 57(10):24–67, 2020.

[100] M. Westerlund. The emergence of deepfake technology: A review. *Technology Innovation Management Review*, 9(11), 2019.

[101] A. Winfield. Ethical standards in robotics and ai. *Nature Electronics*, 2(2):46–48, 2019.

[102] Z. Ye et al. Tackling environmental challenges in pollution controls using artificial intelligence: A review. *Science of the Total Environment*, 699:134279, 2020.

[103] S. Zeitchik. The future of warfare could be a lot more grisly than Ukraine. *https://www. washingtonpost.com/technology/2022/03/11/autonomous-weapons-geneva-un/*, 11 Marzo 2022.

[104] Y. Zhang et al. Graphic recognition information processing technology based on artificial intelligence algorithm. In *Innovative Computing*, pages 521–527. Springer, 2022.

[105] Q. Zhu et al. Gram-CNN: a deep learning approach with local context for named entityrecognition in biomedical text. *Bioinformatics*, 34(9):1547–1554, 2018.

[106] G. Ziccardi. *Internet, controllo e libertà Trasparenza, sorveglianza e segreto nell'era tecnologica*.Raffaello Cortina Editore, 2015.