INTELLIGENZA ARTIFICIALE E DIRITTO: UNA RIVOLUZIONE?

A CURA DI
ALESSANDRO PAJNO, FILIPPO DONATI E ANTONIO PERRUCCI

VOLUME I
DIRITTI FONDAMENTALI, DATI PERSONALI E REGOLAZIONE

SOCIETÀ EDITRICE IL MULINO

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

ISBN 978-88-15-29967-3

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

Capitolo primo

L'INTELLIGENZA ARTIFICIALE: DALLA RICERCA SCIENTIFICA ALLE SUE APPLICAZIONI. UNA INTRODUZIONE DI CONTESTO

1. *L'intelligenza artificiale: breve storia, terminologia e sfide*

L'intelligenza artificiale è oramai ovunque ed è entrata prepotentemente anche nelle nostre vite quotidiane. Ne sentiamo parlare in continuazione. Ne parliamo tutti, anche se forse talvolta un po' a sproposito. Negli ultimi anni, gli Stati Uniti e la Cina, per cui la tecnologia continua ad essere una forte chiave di attrito commerciale, hanno iniziato una corsa a due per il predominio nell'intelligenza artificiale. Cinque anni fa, nel luglio 2017, il Consiglio di Stato della Cina ha emesso un Piano di sviluppo dell'intelligenza artificiale di nuova generazione, con fortissimi investimenti (dell'ordine dei miliardi di dollari) e con l'obiettivo dichiarato di raggiungere la supremazia nell'intelligenza artificiale in pochi anni, entro il 2030. Dopo quasi due anni, nel febbraio 2019, anche il presidente degli Stati Uniti ha firmato un *executive order* per creare un programma denominato *The American AI Initiative*. Nel frattempo, altri Paesi, oltre a Cina e Stati Uniti, hanno elaborato una loro strategia nazionale per l'intelligenza artificiale. Anche l'Unione europea ha elaborato una strategia europea per l'intelligenza artificiale. La Commissione europea ha pubblicato il 21/4/2021 una Proposta di Regolamento sull'approccio europeo all'intelligenza artificiale[1], che propone il primo quadro giuridico europeo sull'intelligenza artificiale, e un

Questo capitolo è di Giuseppe F. Italiano, Stefano Civitarese Matteucci e Antonio Perrucci.

[1] *Proposal for a Regulation laying down harmonised rules on artificial intelligence* (COM[2021] 206 final).

nuovo Piano coordinato sull'intelligenza artificiale[2], che ha la finalità di rafforzare allo stesso tempo l'adozione dell'intelligenza artificiale e gli investimenti e l'innovazione nel settore in tutta l'Unione europea. Questo è il contesto in cui ci stiamo muovendo oggi, e che evidenzia ancora di più l'importanza dell'intelligenza artificiale nelle nostre vite e nelle nostre società.

Ma cosa è esattamente l'intelligenza artificiale? È difficile rispondere a questa domanda se non si segue la storia di questa disciplina. Come sostiene la scrittice Pamela McCorduck nel suo libro *Machines Who Think*[3], l'intelligenza artificiale ha in realtà radici molto profonde nella storia del genere umano: nata con l'antico desiderio di «fabbricare gli dei» (*forge the gods*), è un'idea che ha pervaso continuamente la storia della civiltà occidentale, anche se nel corso dei secoli è stata espressa sotto forme molto diverse ed eterogenee, come miti, storie o leggende. Nonostante il continuo interesse verso questo concetto affascinante, fu soltanto a partire dal 1940 che l'idea di costruire un oggetto «pensante» cominciò a sembrare più concreta e in qualche modo realizzabile: la disponibilità dei primi elaboratori elettronici riuscì a catalizzare l'interesse di scienziati provenienti da varie discipline, come la psicologia, la matematica, l'ingegneria, l'economia e le scienze politiche, sull'effettiva possibilità di costruire un cervello artificiale. In effetti, le nuove ricerche sulle macchine pensanti sviluppate in quegli anni nacquero dalla confluenza di idee provenienti da aree scientifiche molto diverse. I progressi della neurologia avevano mostrato che il cervello era costituito da una rete di neuroni che trasmettevano impulsi elettrochimici. Questa scoperta, combinata con i nuovi progressi scientifici delle teorie cibernetiche nel campo del controllo e della stabilità delle reti elettriche di Norbert Wiener, della teoria dell'informazione di Claude Shannon, e della teoria del calcolo di Alan Turing, condusse naturalmente a chiedersi se fosse possibile costruire un

[2] *Coordinated Plan on Artificial Intelligence 2021 Review* (COM[2021] 205 final).

[3] P. McCorduck, *Machines Who Think*, Natick, MA, 2004[2].

cervello elettronico. Qualche anno dopo, si cominciarono a costruire le prime architetture basate su neuroni artificiali, come ad esempio SNARC, costruita nel 1951 da un giovane Marvin Minsky insieme a Dean Edmonds, che furono i primi precursori delle moderne reti neurali artificiali.

Sin dai primi anni, la sfida dell'intelligenza artificiale sembrava quella di rendere i computer, le macchine, capaci di eseguire compiti tipici dell'intelligenza umana. Già nel 1950, Alan Turing pubblicò un lavoro pionieristico[4], in cui si studiava la possibilità teorica di costruire una macchina «pensante». L'articolo di Turing comincia con le testuali parole: «I propose to consider the question, "Can machines think?"».

Considerando la difficoltà di definire esattamente il significato di «pensante», Turing introdusse un test per valutare la capacità di una macchina di dimostrare un comportamento intelligente, inteso come un comportamento che non fosse facilmente distinguibile da quello di un essere umano. Questo test, che successivamente prese il nome di «Test di Turing», è stato ed è ancora oggi alla base di molti sviluppi nell'intelligenza artificiale. In particolare, nel definire il suo test, Turing propose di considerare il seguente scenario, che chiamò *Imitation Game*:

> Let us fix our attention on one particular digital computer C. Is it true that by modifying this computer to have an adequate storage, suitably increasing its speed of action, and providing it with an appropriate programme, C can be made to play satisfactorily the part of A in the imitation game, the part of B being taken by a man?

Più in dettaglio, Turing propose che un giudice (umano) valutasse una conversazione tra una macchina e un essere umano. Il test prevedeva che la conversazione fosse limitata esclusivamente al testo, così da non valutare la capacità della macchina di generare una conversazione audio. Come

[4] A. Turing, *Computing Machinery and Intelligence*, in «Mind», LIX, 1950, n. 236, pp. 433-460.

sosteneva Turing, in un tale esperimento si potrà dire che la macchina è in grado di superare il test se il giudice umano, pur essendo a conoscenza del fatto che uno dei due interlocutori è una macchina, non riuscirà a distinguere in maniera affidabile la macchina dall'essere umano.

Negli anni, l'idea che un computer riuscisse a superare il Test di Turing sembrava fuori dalla portata delle tecnologie del momento. Anzi, per qualche addetto ai lavori fino a poco tempo fa anche il termine intelligenza artificiale sembrava indicare il futuro: l'intelligenza artificiale era tutto quello che avremmo voluto fare, ma che ancora non eravamo in grado di fare. Poi sono arrivati alcuni cambiamenti epocali: nel 2011 Watson, un sistema costruito da IBM, ha vinto a Jeopardy, un famoso e complicato gioco a quiz televisivo americano. Nel 2012 Jeff Dean e Andrew Ng hanno pubblicato un articolo scientifico pionieristico in cui descrivevano, tra le altre cose, un sistema basato su reti neurali multilivello (*deep learning*) che era sorprendentemente esperto nel riconoscimento di immagini di gatti, un compito che sembrava fuori dalla portata delle macchine. Grazie ai progressi incredibili nel riconoscimento delle immagini, oggi abbiamo costruito i primi sistemi di riconoscimento facciale con un grado di affidabilità molto elevato, e che si trovano oramai ovunque, dai nostri telefoni alle reti sociali, fino al controllo dei passaporti negli aeroporti. Anche grazie al riconoscimento delle immagini, abbiamo cominciato ad avere i primi veicoli a guida autonoma, che devono essere innanzitutto in grado di «vedere» e di riconoscere altri veicoli e altri oggetti nelle loro vicinanze. Nel 2009 i veicoli a guida autonoma viaggiavano soltanto nel deserto. Soltanto pochissimi anni dopo hanno cominciato a frequentare le nostre strade.

Negli ultimi anni, poi, i progressi nell'intelligenza artificiale sono diventati ancora più impressionanti. Nel 2015 AlphaGo, sempre basato su *deep learning*, ha battuto un campione mondiale di Go, un gioco molto complicato. Nel febbraio 2019, a San Francisco si è svolto un *debate* (una gara di dibattito) tra Harish Natarajan, che ha il più grande numero di vittorie in *debate* al mondo, e una mac-

china costruita da IBM per questo scopo, Project Debater[5]. Anche se alla fine la giuria ha assegnato la vittoria a Harish Natarajan, se guardiamo il video di questo dibattito ci possiamo rendere conto che le capacità dialettiche e retoriche della macchina Project Debater sono veramente impressionanti. Nel 2020 OpenAI ha presentato GPT-3 (Generative Pre-training Transformer), un modello di linguaggio (*large language model*) basato su *deep learning* in grado di generare testi molto simili a quelli prodotti da noi esseri umani. Tanto per fare un esempio, abbiamo di recente sottoposto a GPT-3 il seguente testo:

> L'Intelligenza Artificiale (IA) sarà il maggiore protagonista tecnologico di questo secolo. Sono molti gli indicatori e le ricerche che non hanno dubbi in proposito. Secondo il World Economic Forum, entro il 2022 l'Intelligenza Artificiale creerà 133 milioni di nuovi posti di lavoro, chiaramente allo stesso tempo 75 milioni andranno persi, così potremo contare un saldo netto pari a ben 58 milioni di posti di lavoro aggiuntivi nei prossimi anni che riguarderanno in particolare i data analyst e i data scientist.

Abbiamo chiesto a GPT-3 di continuare l'argomento generando un paragrafo, e abbiamo ottenuto il seguente risultato:

> Inoltre, l'IA sarà la più grande rivoluzione della nostra vita quotidiana. Un esempio? Le macchine saranno in grado di comprendere e interagire con noi in una maniera che non ci lascerà mai soli, anche quando non possiamo accedere a servizi come chat, web o telefonia. Ma quanto tempo ci vorrà perché queste tecnologie si diffondano?

Dopo aver ascoltato il dibattito tra Harish Natarajan e Debater, o dopo aver letto i testi generati da GPT-3, si ha quasi la sensazione che le macchine stiano continuando la loro lenta, ma inarrestabile, marcia di avvicinamento al Test di Turing. All'improvviso, grazie anche a molti altri incredibili sviluppi, per noi l'intelligenza artificiale non è più il

[5] https://www.youtube.com/watch?v=-d4Uj9ViP9o.

futuro, ma è diventata il presente. Ognuno ha cominciato a chiamare intelligenza artificiale tutto quello che sta facendo. Abbiamo assistito quasi a una personificazione dell'intelligenza artificiale, tanto che spesso sentiamo dire: «L'intelligenza artificiale si comporta in questo modo, l'intelligenza artificiale fa così». Intelligenza artificiale oggi è un termine molto abusato, che significa più o meno tutto, e quindi forse rischia di essere un po' svuotato dai suoi significati.

Ma cosa è esattamente l'intelligenza artificiale? Bisogna riconoscere che il tema è molto complesso, ed è difficile darne una definizione precisa e pienamente condivisa, anche se tutti noi riconosciamo di utilizzare quotidianamente sistemi basati su intelligenza artificiale, ad esempio per comunicare con il nostro assistente vocale, per bloccare e-mail di spam, per utilizzare un motore di ricerca, per guardare un film o per ascoltare musica. Nel corso dei decenni sono state introdotte varie definizioni di intelligenza artificiale, e l'argomento è stato molto dibattuto nel corso degli anni. Persino l'Unione europea ha prodotto un rapporto di oltre 90 pagine per dare una definizione operativa di intelligenza artificiale![6] Senza avere la pretesa di occupare tanto spazio, potremmo dire molto più brevemente che l'intelligenza artificiale fa riferimento a sistemi che mostrano comportamenti intelligenti, attraverso un'analisi dell'ambiente in cui operano e intraprendendo, con un certo grado di autonomia, azioni mirate a raggiungere determinati obiettivi specifici. Questi sistemi possono essere basati soltanto su software, e agire quindi in una dimensione puramente digitale (ad esempio, assistenti vocali, software di analisi delle immagini, motori di ricerca, sistemi di riconoscimento vocale e facciale), oppure possono essere incorporati anche in dispositivi hardware, ovvero dispositivi fisici (ad esempio, robot evoluti, autoveicoli a guida autonoma, droni e applicazioni relative a *internet of things*).

[6] S. Samoili, M. Lopez Cobo, E. Gomez Gutierrez, G. De Prato, F. Martinez-Plumed e D. Delipetrev, *AI WATCH. Defining Artificial Intelligence*, EUR 30117 EN, Publications Office of the European Union, Luxembourg, 2020 (online), doi: 10.2760/382730 (online), JRC118163.

Forse la prima sfida nella definizione dell'intelligenza artificiale consiste proprio nel definire quali siano le caratteristiche di un comportamento intelligente. In primo luogo, possiamo pensare all'intelligenza come alla capacità di inferire corrispondenze e correlazioni in un insieme complesso di dati, e di stabilirne il contenuto informativo. In secondo luogo, possiamo pensare all'intelligenza come alla capacità di pianificare un comportamento o una strategia costituita da una successione di scelte e comportamenti, in modo coerente rispetto a un obiettivo prestabilito. In queste accezioni, l'intelligenza non si identifica affatto con l'autocoscienza o con la consapevolezza, che coinvolgono l'elaborazione di informazione a un livello di maggiore complessità. Non dobbiamo quindi assolutamente aspettarci che un sistema dotato di intelligenza artificiale possieda una mente e una consapevolezza artificiale, come molti film o libri sembrano suggerire al nostro immaginario collettivo.

Per elaborare meglio questi concetti, prendiamo in considerazione uno dei contributi più importanti nell'area dell'intelligenza artificiale degli ultimi anni: gli algoritmi di *machine learning*. Anche se sembrano una novità degli ultimi anni, in realtà si è cominciato a lavorare sul *machine learning* fin dagli anni '60, e la ricerca sul *machine learning* è fiorita in ambito accademico soprattutto agli inizi degli anni '90. Quasi trenta anni fa. Ma gli algoritmi di *machine learning* sono usciti dall'ambito accademico e sono entrati prepotentemente nelle nostre vite soltanto negli ultimi anni, grazie anche a due fattori importanti: la disponibilità di grandissime quantità di dati, e di computer sempre più potenti e in grado di elaborare velocemente queste grandissime quantità di dati. In realtà gli algoritmi di *machine learning* sono solo una piccola parte della disciplina dell'intelligenza artificiale. E anche il *deep learning*, per intenderci quello reso noto dal lavoro pionieristico di Dean e Ng, e poi utilizzato da sistemi molto potenti, come ad esempio AlphaGo o GPT-3, oggi è molto di moda. Ma non è che una delle tantissime direzioni del *machine learning* in cui abbiamo lavorato negli ultimi decenni. Non è neanche tanto chiaro se il *deep learning* sarà la tecnica più utile nel futuro immediato. Però negli ultimi

anni ha prodotto risultati strabilianti. Come riconoscere i gatti nelle fotografie, e tutto quello di importante che ne è derivato.

Per capire cosa sono gli algoritmi di *machine learning*, dobbiamo prima comprendere cosa sono gli algoritmi. Semplificando al massimo, potremmo dire che gli algoritmi sono sequenze di istruzioni, linee di codice, che possiamo scrivere e che possiamo leggere. Ad esempio, nel recente scandalo Dieselgate, lo scandalo delle emissioni in cui sono state coinvolte varie case automobilistiche, il problema risiedeva in un'applicazione software, che era in grado di riconoscere se il veicolo fosse stato sottoposto a un test di omologazione (e di conseguenza di ridurre le emissioni). In quel caso, è stato possibile prendere quel codice software, e l'algoritmo che ne era alla base, sezionarli ed esaminarli con attenzione, e capire esattamente cosa facessero e perché producessero un certo risultato durante i test di omologazione e un risultato molto diverso durante il normale funzionamento. In modo del tutto trasparente, spiegabile, interpretabile e riproducibile.

Gli algoritmi di *machine learning* agiscono però in modo molto diverso da algoritmi tradizionali. Infatti, gli algoritmi di *machine learning* sono in grado di migliorarsi automaticamente attraverso l'esperienza e l'utilizzo di dati. Per fare questo, costruiscono un modello basandosi su dati (*training data*) che utilizzano in una fase di addestramento, così da riuscire a effettuare predizioni senza essere stati esplicitamente programmati per questo scopo. Ad esempio, possiamo addestrare un algoritmo di *machine learning* con un numero opportuno di immagini di cani e di gatti, senza aver bisogno di definire quali sono le caratteristiche che riescono a distinguere i cani dai gatti. Dopo questa fase di addestramento, l'algoritmo sarà in grado di apprendere in autonomia quali siano le caratteristiche comuni alle immagini dei cani e quali siano le caratteristiche comuni alle immagini dei gatti, e sarà in grado di costruire un modello in grado di riconoscere, con precisione elevata, se una nuova immagine, che non gli è stata presentata nella fase di addestramento, contenga un cane oppure un gatto.

Questo tipo di apprendimento, descritto nell'esempio precedente, è noto come *machine learning supervisionato*. In questo caso l'algoritmo necessita di dati di training che siano opportunamente etichettati, cioè contengano l'esplicita informazione della categoria a cui appartiene il dato utilizzato nell'apprendimento, ovvero se all'algoritmo sia stata sottoposta, durante la fase di apprendimento, l'immagine di un cane o di un gatto. A partire da questi dati di addestramento, che contengono sia gli input (le immagini) che l'output desiderato (se raffiguri un cane o un gatto), algoritmi di *machine learning* supervisionato costruiscono un modello che verrà poi utilizzato nella fase di predizione. Diremo che un algoritmo che migliora l'accuratezza delle sue predizioni nel tempo avrà imparato ad eseguire un particolare compito (nell'esempio, riconoscere immagini di cani da immagini di gatti).

L'apprendimento supervisionato non è l'unico possibile, ed esistono anche algoritmi di *machine learning* che utilizzano apprendimento *non supervisionato*. Algoritmi di questo tipo prendono in ingresso dati che contengono soltanto gli input, senza alcuna specifica degli output desiderati, e tipicamente hanno l'obiettivo di apprendere una particolare struttura dei dati, come ad esempio i gruppi o i cluster in cui tali dati possono essere organizzati. Apprendono da dati di ingresso che non sono quindi etichettati, classificati o già rappresentati in categorie. Questo fanno, ad esempio, aziende che sono interessate a segmentare i propri clienti, o piattaforme come Netflix e Spotify, interessate a utilizzare algoritmi di raccomandazione basati su criteri di similarità dei propri utenti, per suggerire film o musica di particolare gradimento per i loro utenti.

Una terza tipologia di apprendimento è quella del *reinforcement learning* (apprendimento con rinforzo), che consente di risolvere problemi che richiedono di elaborare strategie particolarmente sofisticate. Questa tecnica, infatti, richiede un'esplorazione dell'ambiente che calcola le azioni più promettenti per raggiungere un particolare obiettivo, con alta probabilità di successo. Il *reinforcement learning* può essere impiegato quando il modello dell'ambiente in cui

opera l'agente di intelligenza artificiale non è conosciuto a priori (come, ad esempio, il movimento di robot in ambienti sconosciuti), oppure se tale modellazione richiede una quantità di risorse eccessiva (come, ad esempio, prevedere le conseguenze di tutte le possibili mosse di una partita a scacchi). Uno degli elementi costitutivi del *reinforcement learning* è una funzione di premio (*reward*) che consente di valutare le azioni stabilite a ogni passo dall'algoritmo. Questo sistema premiale è basato sulla valutazione sia di quanto è soddisfacente la nuova situazione in cui ci si viene a trovare per aver seguito l'indicazione dell'agente, sia della bontà dell'azione che è stata stabilita per arrivare a questa situazione a partire da quella di partenza. Identificare la funzione di premio è un compito in genere non banale. Ad esempio, nel gioco degli scacchi, si potrebbe pensare di premiare un agente di intelligenza artificiale ogni volta che cattura un pezzo avversario. Questo premio è solo apparentemente idoneo a vincere la partita, perché così facendo l'agente non cercherà di dare scacco matto ma finirà per prolungare al massimo la partita per mangiare quanti più pezzi possibili, incrementando quindi la probabilità di perdere pur di mangiare molti pezzi e fare molti punti. L'ambiente in cui addestrare l'agente può essere simulato come nel caso di una scacchiera, di un simulatore di volo, di un videogioco, oppure reale, come una stanza con oggetti in cui far muovere un robot. Durante l'addestramento l'agente sperimenta dapprima azioni casuali (esplorazione) che fanno accrescere la sua conoscenza dell'ambiente. In seguito, alternerà in modo opportuno l'esplorazione (*exploration*) con lo sfruttamento delle conoscenze acquisite durante le precedenti esplorazioni (*exploitation*). Grazie alla funzione di premio, verranno «rinforzate» le decisioni più promettenti, ovvero che hanno portato a conseguire un punteggio maggiore secondo quanto stabilito dalla funzione premio. In questo senso, un algoritmo basato su *reinforcement learning* impara grazie agli errori commessi, e al fatto di essere lasciato a esplorare liberamente l'ambiente circostante.

Alla base di molti algoritmi di *machine learning* ci sono *reti neurali artificiali* (*artificial neural networks*), note

semplicemente anche come reti neurali, e che hanno tratto ispirazione dalle reti neurali che si trovano all'interno del cervello biologico. Proprio le reti neurali artificiali derivanti dai lavori pionieristici di Minsky ed Edmonds del 1951, di cui abbiamo trattato precedentemente. Una rete neurale artificiale è costituita da una collezione di nodi, chiamati neuroni artificiali, interconnessi con una struttura di rete, e che rappresentano un modello molto semplificato dei neuroni che si trovano in un cervello biologico. Ogni connessione, così come le sinapsi di un cervello biologico, può trasmettere un segnale ad altri neuroni, a grandi linee come descritto nel seguito. Un neurone artificiale riceve un segnale, lo elabora e lo trasmette agli altri neuroni a cui è connesso. Il segnale di una connessione è un numero reale, e un neurone artificiale calcola una opportuna funzione non lineare della somma dei suoi input. Neuroni artificiali e connessione hanno un peso che viene calibrato opportunamente durante il processo di apprendimento: questo peso rafforza o indebolisce la «forza» del segnale della connessione e contribuisce a definire l'architettura della rete neurale. Tipicamente, i neuroni di una rete neurale artificiale sono organizzati su molti strati, dove strati diversi della rete effettuano diverse trasformazioni dell'input. Ultimata la fase di apprendimento, tutti i pesi delle connessioni risultano definiti e la rete neurale viene quindi completamente determinata. Durante il funzionamento di una rete neurale, i segnali attraversano la rete neurale dal primo strato (strato di input) fino all'ultimo strato (strato di output), subendo tutte le trasformazioni indicate dai neuroni e dalle connessioni della rete.

Quando si parla di *deep learning*, che abbiamo già evidenziato in sistemi come AlphaGo e GPT-3, ci si riferisce tipicamente a reti neurali artificiali a moltissimi strati, e per questo denominate *profonde*. Architetture basate su *deep learning* sono state applicate in moltissimi settori, inclusi visione artificiale, riconoscimento del parlato, elaborazione del linguaggio naturale, traduzioni automatiche, riconoscimento di immagini e bioinformatica, dove hanno prodotto risultati confrontabili e spesso superiori a quelli degli esperti umani.

Da quanto appena descritto, risulta evidente che gli algoritmi di *machine learning* sono molto diversi dagli algoritmi tradizionali, perché in generale il loro obiettivo è quello di generare modelli che siano in grado di consentire determinate operazioni, come ad esempio effettuare predizioni di particolare accuratezza. A differenza degli algoritmi tradizionali, gli algoritmi di *machine learning* spesso non sono trasparenti, spiegabili o interpretabili. Soprattutto se utilizzano reti neurali artificiali e tecniche di *deep learning*. Infatti, una delle differenze principali è che un algoritmo di *machine learning* usa del codice per consentire a un sistema automatico di imparare e di creare un modello del problema che si vuole risolvere. Il punto fondamentale è che il modello creato da un algoritmo di *deep learning* non è facilmente dissezionabile, non è decomponibile come le linee di codice di algoritmi tradizionali. Non possiamo dissezionare un algoritmo di *deep learning* per capire cosa è successo, perché ha raggiunto una determinata conclusione, perché è stato sollecitato in un certo modo un neurone artificiale. Possiamo provare a osservare dall'esterno, a fare delle domande, interrogare l'algoritmo e analizzare le sue risposte. Ma questo è un processo lungo e non sempre fa capire perché è stata raggiunta una certa conclusione. Il meccanismo diventa complessivamente meno chiaro, meno trasparente e più difficilmente spiegabile. Per questo gli algoritmi di *machine learning* sono come una scatola chiusa, una *black box*, che non può essere aperta facilmente. Non è realmente chiaro, in tutti i dettagli, perché una rete neurale riconosca una determinata immagine. Si costruisce un suo modello, attribuisce pesi ai suoi neuroni e alle sue connessioni, in base ai dati su cui è stata precedentemente addestrata. E questa è proprio la parte più delicata di tutto il processo: gli algoritmi di *machine learning* usano i dati per imparare, cercano di trovare segnali nascosti all'interno dell'enorme quantità di dati su cui lavorano per creare il loro modello (da cosa possiamo capire che c'è un gatto o un cane nell'immagine? Da cosa possiamo capire che c'è un pedone che sta attraversando la strada? O comprendere la segnaletica stradale?). Gli algoritmi usati per creare questi

modelli sono molto sofisticati. Come diceva lo statistico George Box, «Tutti i modelli sono sbagliati, ma qualcuno è utile» (*All models are wrong, but some are useful*). Gli algoritmi di *machine learning* fanno esattamente questo: costruiscono modelli complicati, modelli che possono anche essere sbagliati, ma che sono fondamentalmente utili e che funzionano molto bene in pratica.

Nonostante questi limiti, gli algoritmi di *machine learning* hanno avuto un impatto incredibile in molte discipline, nelle aziende e persino nelle nostre vite quotidiane. Le loro applicazioni sono incredibili, e in un numero di settori troppo lungo per essere persino elencato. In tutti questi settori, hanno risolto in modo nuovo molti problemi difficili, che prima non si sapevano affatto risolvere. Ma hanno anche creato problemi completamente nuovi, e anche loro di non facile risoluzione.

Come è avvenuto, ad esempio, nell'economia e nella finanza. Secondo stime recenti, in questo settore le *FinTech companies* stanno conquistando velocemente quote molto importanti del mercato totale. Abbiamo già molti strumenti basati su *machine learning* che hanno una enorme diffusione nel mondo finanziario, come ad esempio i *robo-advisor*, che sono degli assistenti finanziari digitali, le piattaforme digitali di valutazione dei rischi (*risk assessment*), gli strumenti per la gestione del portafoglio di investimenti, per l'affidabilità creditizia (*credit score*) delle società e delle persone, per la rilevazione di frodi, abbiamo anche piattaforme di *algorithmic trading* che analizzano velocemente enormi quantità di dati ed effettuano in maniera automatica acquisti e vendite di titoli e azioni, e molto altro. Questo sviluppo rapido, oltre ad aprire nuovi mercati e nuove opportunità, ha anche sollevato qualche preoccupazione. Innanzitutto, gli algoritmi di *machine learning*, per loro natura, costruiscono i loro modelli su grandi quantità di dati, che sono ovviamente dati storici, dati del passato. Il fatto che siano dati storici può ovviamente rafforzare *bias*, discriminazioni e pregiudizi storici, ad esempio nel concedere mutui o prestiti. E il fatto che abbiamo bisogno di una grandissima quantità di dati genera ovviamente anche problemi complessi relativi alla

privacy e alla proprietà di questi dati. Un altro aspetto molto importante è relativo ai profili etici degli algoritmi. Man mano che gli algoritmi assumono responsabilità importanti, come ad esempio eseguire transazioni finanziarie, influenzare decisioni di affidabilità creditizia, oppure guidare veicoli, sembra importante poter spiegare agli utenti perché è stata presa una certa decisione, capire come poter assicurare comportamenti etici nell'interesse degli utenti. Ma questo non è affatto banale. Proprio perché se è un algoritmo di *machine learning*, non è chiaro perché abbia raggiunto quella decisione. Tutti i problemi evidenziati, come pregiudizi, *bias*, privacy, responsabilità degli algoritmi, sono problemi molto complicati. Problemi che non sono soltanto tecnologici, che non sono soltanto economici, che non sono soltanto tipici delle discipline sociali, e che non possono essere affrontati con approcci tradizionali. Problemi che richiedono sempre più una stretta collaborazione tra esperti provenienti da discipline completamente diverse, che devono confrontarsi e lavorare insieme.

Oltre che nei settori industriali, l'intelligenza artificiale sta entrando sempre più prepotentemente anche nelle nostre vite quotidiane, suggerendoci film da guardare, musica da ascoltare, beni da acquistare, oppure fornendoci assistenti vocali sempre più sofisticati, come ad esempio Alexa, Siri, Cortana o Google Assistant. Queste tecnologie, che fino a pochi anni fa erano disponibili soltanto nei laboratori di ricerca più avanzati, sono oramai a disposizione di tutti. Ma anche questo solleva nuovi problemi. Se da un lato questo progresso esponenziale non può che suscitare comprensibili entusiasmi, perché con l'intelligenza artificiale sono alla portata di tutti cose che soltanto pochi anni fa sembravano inimmaginabili, d'altro canto sembrerebbe opportuno avviare fin da oggi riflessioni serie, e scevre da pregiudizi, sull'impatto sociale dell'intelligenza artificiale e su come continuare ad ottenerne benefici minimizzandone i rischi. Rischi che includono, tra gli altri, anche la diffusione di immagini o video *deep-fake*, in grado di danneggiare la reputazione delle persone; l'utilizzo di *bot* per manipolare l'opinione pubblica; trasferimenti di pregiudizi e *bias* tipicamente umani negli

algoritmi di intelligenza artificiale; sistemi di riconoscimento che possono risultare invasivi della privacy.

Vale forse la pena sottolineare che molti dei problemi evidenziati non sono relativi soltanto all'intelligenza artificiale, ma in generale sono forse indicativi dei nostri rapporti con le tecnologie informatiche. Per fare un esempio, consideriamo un caso concreto, accaduto pochi anni fa a Boston. Riguarda il servizio degli *school bus*, i bus navetta utilizzati per accompagnare gli studenti nelle scuole pubbliche. A Boston, vengono spesi ogni anno oltre 120 milioni di dollari per fornire questo servizio a più di 25.000 studenti di oltre 200 scuole pubbliche con un totale di 650 *school bus*. Organizzare bene questo servizio è un problema complicatissimo. Bisogna decidere quanti *school bus* utilizzare, che giro far fare a ogni *school bus*, decidere tutte le fermate in modo che ognuna delle fermate sia raggiungibile a piedi, e garantire che tutti gli studenti arrivino a scuola prima dell'inizio delle lezioni. A questi vincoli se ne aggiungono moltissimi altri. Non ultimi, nel caso di una grande città come Boston, i problemi di traffico che noi tutti conosciamo. In poche parole, trovare una buona soluzione a questo problema è veramente un incubo. Trovare una soluzione ottima è praticamente impossibile. Non lo può fare un essere umano, ma è anche molto complicato per un algoritmo. Nel 2017 il sistema delle scuole pubbliche di Boston (Boston Public Schools) avviò una collaborazione con Dimitris Bertsekas, uno dei più noti scienziati di ricerca operativa della Sloan School of Management di MIT, con l'obiettivo di ridurre di 50 unità il numero di *school bus* utilizzati, per un risparmio stimato di circa 5 milioni di dollari l'anno (per una media di circa 25.000 dollari a scuola). Sembrava un successo annunciato, avrebbe sicuramente migliorato l'efficienza di un servizio pubblico, e consentito di investire le risorse risparmiate in progetti per il miglioramento della qualità dell'istruzione. La soluzione individuata non utilizzava assolutamente approcci basati su intelligenza artificiale, ma algoritmi classici di ottimizzazione. Tutti si aspettavano che avrebbe dimostrato ancora una volta l'incredibile impatto degli algoritmi e delle tecnologie informatiche sul miglio-

ramento della società. Ma purtroppo non è andata esattamente così. Appena è stato annunciato il nuovo orario degli *school bus*, alcune famiglie lo hanno trovato inaccettabile e hanno protestato veementemente, addirittura marciando verso la City Hall. Questa protesta ha costretto il sistema delle scuole pubbliche di Boston a ripristinare velocemente l'orario precedente, facendo sprecare tutto il lavoro fatto, tutte le risorse finanziarie e il tempo investito nel progetto. Questo è successo a Boston, uno degli *hub* tecnologici del Paese più avanzato al mondo. Un fallimento degli algoritmi? Non sembrerebbe così. La soluzione tecnologica, gli algoritmi, hanno probabilmente ottenuto la migliore soluzione possibile. Ma il problema non poteva essere affrontato soltanto con strumenti tecnologici. Chi ha protestato era una minoranza, e lo ha fatto perché ha visto che il nuovo orario avrebbe peggiorato la propria situazione particolare, senza considerare che forse la situazione complessiva, il *social welfare*, il benessere sociale, come si chiama in teoria dei giochi, sarebbe migliorato, e di molto. Chi ha protestato ha criticato un algoritmo, che anche se non basato su tecniche di intelligenza artificiale, è stato visto come una scatola chiusa, che in modo non trasparente prendeva decisioni sulla vita personale dei cittadini. L'algoritmo in realtà aveva l'obiettivo di ottenere il miglior compromesso possibile tra le varie esigenze, tra i costi del servizio per la collettività e gli interessi e i vincoli dei singoli cittadini, cercando di non penalizzare soprattutto le famiglie più deboli, per cui i cambiamenti negli orari dei propri figli avrebbero generato maggiori criticità. E su questo, l'algoritmo ha funzionato molto bene. Ciò che probabilmente non ha funzionato è il modo in cui i *policy maker* hanno pensato di utilizzare un algoritmo per cambiare alcune *policy* che avevano effetti immediati sulla vita delle persone. Più che altro, è stato un fallimento del processo politico con cui Boston ha deciso di affrontare il problema, che non è tecnologico, e che non può essere affidato esclusivamente a un algoritmo. Ancora una volta, questo ci insegna che diventa sempre più cruciale la collaborazione concreta di esperti in informatica, che progettano e comprendono come funzionano gli algoritmi,

e di esperti di *policy*, che comprendono come pesare i vari *trade-off* per il benessere sociale, e le caratteristiche fondamentali delle moderne democrazie.

Cosa ci aspetta e cosa possiamo fare nell'immediato futuro? Alan Turing ha concluso il suo famoso articolo *Computing machinery and intelligence* del 1950, in cui tra l'altro ha definito *The Imitation Game*, il già citato Test di Turing, con la famosa frase: «We can only see a short distance ahead, but we can see plenty there that needs to be done» (Possiamo vedere soltanto una piccola distanza davanti a noi, ma possiamo vedere che ci sono moltissime cose da fare). A 70 anni di distanza è cambiato quasi tutto. Le tecnologie hanno completamente rivoluzionato la nostra società, il nostro modo di lavorare e il nostro modo di vivere. Ma l'affermazione di Turing sembra ancora molto attuale. Anche oggi riusciamo a vedere soltanto cosa succederà a breve, due o tre anni. Ma anche basandoci su questo limitatissimo orizzonte temporale, possiamo vedere che c'è ancora moltissimo lavoro da fare. Si sta creando rapidamente un nuovo rapporto tra noi e le macchine, con una diversa distribuzione dei ruoli, e che richiede interazioni e forme di collaborazione profondamente diverse rispetto al passato. Con le macchine che sono in grado di eseguire compiti sempre più sofisticati e «intelligenti», ma che qualche volta possono interferire e produrre nuovi problemi e nuovi conflitti con le azioni degli esseri umani, come i recenti disastri aerei dei Boeing 737 Max ci hanno purtroppo insegnato. Ma nel contesto attuale disegnare questo nuovo rapporto tra esseri umani e macchine non è per niente facile. Anche perché le tecnologie digitali hanno una velocità impressionante. Le tecnologie di ieri, come ad esempio la tv, la radio, l'elettricità, l'automobile hanno impiegato più di 50 anni per raggiungere i 50 milioni di utenti. Ci hanno concesso tutto il tempo per abituarci alle loro innovazioni, per avere nuove regole sul loro utilizzo, e per organizzare le nostre vite e le nostre società di conseguenza. Oggi, le tecnologie digitali irrompono molto più velocemente, e non ci danno affatto il tempo per organizzarci e per abituarci alle loro dirompenti innovazioni. Un esempio evidente di questa

velocità viene dalle reti sociali: Twitter ha impiegato meno di tre anni per raggiungere i 50 milioni di utenti, Facebook e Instagram meno di due anni, Tik Tok meno di sei mesi. In questo scenario di trasformazioni profonde caratterizzate da un'incredibile rapidità, è molto importante creare le condizioni perché esperti e scienziati di varie discipline riescano a lavorare tutti insieme sulle nuove sfide che l'intelligenza artificiale sta creando, soprattutto sugli aspetti etici, di responsabilità, di discriminazione, di trasparenza, di equità, di organizzazione del lavoro e di governo nella nostra società.

2. *I problemi di un inquadramento giuridico dell'IA con particolare riferimento alle decisioni automatizzate delle autorità pubbliche*

Entro le coordinate delineate nel precedente paragrafo, il ruolo del diritto investe due aspetti fondamentali. Il primo riguarda la regolazione dell'uso che i privati, specialmente le imprese, fanno dell'IA. Questa è per certi versi la parte più difficile, anche in ragione di quanto osservato alla fine del precedente paragrafo. La vertiginosa avanzata della «società algoritmica» guidata da colossi informatici operanti su scala globale rende oltremodo complessa l'opera di irreggimentazione di attività che si sono tutte sviluppate in assenza di regole specifiche. La proposta della Commissione EU del 21/4/2021 di «Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale)» mira ad assicurare «il buon funzionamento del mercato interno per i sistemi di intelligenza artificiale… nel contesto del quale tanto i benefici quanto i rischi legati all'intelligenza artificiale siano adeguatamente affrontati a livello dell'Unione».

Il secondo aspetto, assai meno sviluppato del primo (soprattutto considerando il nostro Paese) riguarda l'uso degli algoritmi da parte delle istituzioni pubbliche nell'esercizio delle loro funzioni.

Entrambi questi aspetti sono ampiamente sviluppati nei vari capitoli dei volumi che raccolgono la presente ricerca. In queste pagine introduttive, ci limitiamo ad alcune notazioni di inquadramento generale, specialmente con riguardo al secondo profilo.

In primo luogo, appare opportuno ribadire l'importanza di chiarire a cosa precisamente ci si riferisca quando si parla di regolazione dell'IA. È normale pensare che la questione dell'automazione decisionale presenti aspetti più sensibili quando la decisione in questione sia il provvedimento di un'autorità pubblica. Non vi è dubbio che la decisione di una banca di concedere o meno un prestito – che si fondi su profilazione e algoritmi – possa avere il medesimo impatto sulla vita delle persone di un algoritmo che non conceda un beneficio fiscale o una misura di assistenza sociale. Ma negli ordinamenti come il nostro le decisioni dei privati non sono funzionalizzate, non rispondono, vale a dire, a quell'insieme di regole, sostanziali e formali, che mirano a rendere le decisioni delle autorità pubbliche verificabili e giustificabili dinanzi alla collettività. È, pertanto, evidente che quanto osservato in precedenza sulla difficoltà di spiegare e interpretare come un algoritmo di *machine learning* giunga a una determinata conclusione ponga problemi da un certo punto di vista più delicati nell'area delle decisioni pubbliche[7].

Non a caso l'art. 22 del GDPR subordina le decisioni automatizzate delle amministrazioni pubbliche fondate sulla profilazione di dati al requisito della loro autorizzazione da parte di una disposizione di legge. Prima di tornare, alla fine di questo paragrafo, su questo aspetto, relativo al fondamento legale di una decisione automatizzata, occorre notare che non tutte le decisioni automatizzate cadono sotto

[7] Il progresso dell'IA, peraltro, pone il problema di come definire il perimetro dei decisori pubblici. Si potrebbe, infatti, opinare che nella sostanza le Big Tech siano poteri (privati ma) sostanzialmente «pubblici», come tali da assoggettare alle stesse regole dei poteri pubblici statali. Su questo tema vedi di recente L. Ammannati, *I «signori» nell'era dell'algoritmo*, in «Diritto Pubblico», 2021, pp. 381-413.

l'ombrello concettuale dell'IA in senso stretto. O almeno, questo dipende da cosa stipuliamo essere IA. Nella proposta di legge sull'intelligenza artificiale dell'UE, per esempio, si accoglie una accezione onnicomprensiva di IA.

L'art. 3 definisce «sistema di intelligenza artificiale» (sistema di IA)

> un software sviluppato con una o più delle tecniche e degli approcci elencati nell'allegato I, che può, per una determinata serie di obiettivi definiti dall'uomo, generare output quali contenuti, previsioni, raccomandazioni o decisioni che influenzano gli ambienti con cui interagiscono.

A sua volta l'Allegato I individua tre approcci, distinguendo innanzitutto tra quelli di apprendimento automatico e quelli basati su logica e conoscenza. I primi includono l'apprendimento supervisionato, l'apprendimento non supervisionato e l'apprendimento per rinforzo, con utilizzo di un'ampia gamma di metodi, tra cui l'apprendimento profondo (*deep learning*). I secondi includono la rappresentazione della conoscenza, la programmazione induttiva (logica), le basi di conoscenze, i motori inferenziali e deduttivi, il ragionamento (simbolico) e i sistemi esperti. In terzo luogo, vengono elencati gli approcci statistici, la stima bayesiana, i metodi di ricerca e ottimizzazione.

Le tecniche maggiormente utilizzate nel settore pubblico sono ancora quelle riconducibili al secondo approccio, che si riferisce essenzialmente a quella che è stata chiamata la prima ondata di IA. Essa è anche definita *symbolic*[8] a indicare i simboli memorizzati e le associate regole «se-allora» (*if-then*), utilizzate da un software per determinare inferenze decisionali. Tali regole consentono di operare su problemi circoscritti, mancando tali sistemi di capacità di apprendimento. Come è stato osservato, con una metafora che è divenuta familiare tra gli addetti ai lavori, questo genere di

[8] J. Launchbury, *A DARPA perspective on artificial intelligence*, in https://www.darpa.mil/attachments/AIFull.pdf (ultimo accesso: 20/1/2020).

IA mantiene «the human in the loop»[9]. Mentre tali sistemi cosiddetti *expert rule-based* vengono impiegati, soprattutto al di fuori dell'Italia, da molti anni[10], le forme più evolute sono assai meno sviluppate nel settore pubblico rispetto a quello privato[11].

In un importante rapporto del febbraio 2020 commissionato dalla Administrative Conference of the United States[12], tra i risultati dell'indagine condotta da scienziati informatici dell'Università di Stanford su ben 142 agenzie federali emerge che l'amministrazione degli Stati Uniti solo nel 12% dei casi impiega tecniche che possono considerarsi sofisticate. I casi di decisioni pubbliche automatizzate di cui si è discusso in Italia negli ultimi anni sono anch'essi riconducibili all'IA «per simboli». Ci si riferisce in particolare ad alcune controversie dinanzi alla giustizia amministrativa originate dall'uso di un algoritmo da parte

[9] P. Boucher, *How artificial intelligence works*, in «European Parliamentary Research Service», March 2019, in https://www.europarl.europa.eu/at-your-service/files/be-heard/religious-and-non-confessional-dialogue/events/en-20190319-how-artificial-intelligence-works.pdf.

[10] Una delle prime indagini sistematiche relative alle decisioni automatizzate nell'amministrazione pubblica fu condotta dall'Australian Administrative Review Council nel 2004 (*Automated Assistance in Administrative Decision-Making*, in https://www.ag.gov.au/LegalSystem/AdministrativeLaw/Documents/publications/report-46.pdf). In appendice al rapporto si dà conto di un elevato numero di agenzie e dipartimenti in vari settori che utilizzavano tecniche di automazione. Tra queste prevalentemente *rule-based* ma anche di auto-apprendimento, incluse le tecniche chiamate reti neurali (*neural networks*). NetRisk, impiegata dall'Australian Taxation Office, un profilatore del rischio di insolvenza, è un esempio dell'ultimo tipo. NetRisk usa la situazione degli utenti e il loro comportamento passato per suggerire cosa fare in caso di inadempimento.

[11] Del resto, quasi sempre questi sistemi sono stati elaborati dapprima per operare nel settore privato. Vedi K. Yeung, *Algorithmic regulation: A critical interrogation*, in «Regulation & Governance», 2018, n. 12, pp. 505-523.

[12] D. Freeman Engstrom, D.E. Ho, C.M. Sharkey e M. Cuéllar, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*, in https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf.

del ministero dell'Istruzione per gestire le procedure di mobilità dei docenti[13].

Detto questo, occorre osservare che i confini tra l'IA tradizionale e quella più evoluta (secondo le varie tecniche di apprendimento descritte in precedenza) nella prassi sono sfumati. Alcuni studi mostrano come si assista in diversi casi a una combinazione tra queste diverse tecniche che rende problematico stabilire la misura della intelligibilità e della riferibilità a una scelta umana dell'esito dell'algoritmo. In un lavoro che prende in esame quattro casi di studio – ripetizione di indebiti benefici assistenziali in Australia, calcolo del rischio di recidiva negli Stati Uniti, aiuti finanziari agli studenti in Svezia, calcolo della «reputazione sociale» in Cina[14] – emerge che in nessuno di questi gli esseri umani sono completamente estranei al processo decisionale. Anche nel contesto dell'apprendimento automatico supervisionato, riscontrabile in alcuni dei casi indicati (Stati Uniti e Cina), sono esseri umani che decidono quali processi automatizzare e quali tecniche utilizzare, oltre a identificare i dati o le regole che costituiscono la base per le inferenze.

Ciononostante, quando il sistema fa perno sull'applicazione di regole elaborate sulla base di risultati empirici raccolti da dati storici (attraverso statistiche o apprendimento automatico) il modo attraverso cui la regola viene inferita può risultare opaco. Il tipo più problematico di opacità è quello suddetto relativo alla difficoltà di comprendere l'azione di complessi meccanismi di apprendimento, che si aggiunge agli altri due tipi di opacità – che possono riguardare anche l'IA tradizionale – identificati da Jenna Burrel[15]. Questi ultimi, che sembrano maggiormente ge-

[13] TAR Lazio-Roma, Sezione III-bis, 10/9/2018, n. 9227; Consiglio di Stato, VI, 4/2/2020, n. 881.

[14] M. Zalnieriute, L. Bennett Moses e G. Williams, *The Rule of Law and Automation of Government Decision-Making*, in «The Modern Law Review», 2019, n. 82, p. 433.

[15] J. Burrell, *How the Machine «Thinks»: Understanding Opacity in Machine Learning Algorithms*, in «Big Data & Society», 2016, n. 3, p. 1.

stibili mediante una regolamentazione giuridica, sono la «segretezza intenzionale» e «l'ignoranza tecnologica». Il primo si ha quando le tecniche impiegate sono coperte da segreto commerciale o di Stato, o quando i dati utilizzati contengono informazioni personali che non possono essere divulgate a causa delle leggi sulla privacy o sulla protezione dei dati. L'ultimo tipo di opacità riguarda per così dire un problema di competenza, nel senso che l'algoritmo opera secondo regole oggettivamente trasparenti, ma la loro comprensione richiede un livello di conoscenze specialistiche che le persone comuni non hanno. Naturalmente, anche questo genere di opacità può dare luogo a inconvenienti molto seri e comunque chiama in causa la questione della fiducia negli «esperti». Rendere effettivamente trasparenti questo tipo di decisioni implica, infatti, instaurare meccanismi attraverso cui, per esempio, tecnici (indipendenti) fungano da «mediatori-divulgatori» tra la macchina e i cittadini.

Una soluzione di carattere generale per tentare di affrontare il problema dell'opacità (dei tre tipi) è quello – per esempio previsto nel sistema svedese di Student Welfare – di riferire la responsabilità (*accountability*) non a ogni singola decisione ma all'intero processo. In questo caso, vale a dire, vi è un funzionario che è comunque tenuto a dare conto della decisione automatizzata qualora questa venga contestata. Il che significa che questi dovrà fornire una propria giustificazione della decisione non necessariamente consistente nello spiegare il processo compiuto dall'algoritmo. In questo modo, è come se le decisioni algoritmiche fossero valide solo *prima facie*.

Oltre a trasparenza e responsabilità, altri due «valori» che entrano in gioco nel dibattito sulle decisioni automatizzate sono la prevedibilità e l'uguaglianza. Questi sono normalmente i valori che l'impiego dell'IA si pensa dovrebbe favorire. Nell'applicare un insieme di regole predefinite – per esempio l'attribuzione di un beneficio in base al reddito – un algoritmo è certamente più «affidabile» e imparziale di un essere umano. Ancora una volta, però, non sempre può dirsi lo stesso se le regole da applicare dipendono da più

complessi meccanismi di apprendimento dell'algoritmo. Per esempio, è molto difficile prevedere le decisioni di Compas (Correctional Offender Management Profiling for Alternative Sanctions), utilizzato dai giudici in alcuni stati degli Stati Uniti per dedurre da dati storici quali imputati condannati presentino il maggior rischio di recidiva, in particolare laddove vi sia il rischio di violenza. Coloro che hanno sviluppato l'algoritmo non sapevano necessariamente in anticipo quali criteri sarebbero stati trovati per correlare, da soli o in combinazione, certi comportamenti. È molto probabile che le regole per l'assegnazione dei «punteggi» agli individui siano state ottenute attraverso un processo di apprendimento automatico da un ampio set di dati che registrano le caratteristiche di chi ha comportamenti recidivi storici. Certe caratteristiche, come la razza o il genere, presentano elevati margini di opinabilità e riguardano una materia fortemente sensibile dal punto di vista dei diritti umani. Compas, per esempio, era stato «addestrato» a tenere conto delle differenze di genere. Nel famoso caso *Loomis*[16] la Corte suprema del Wisconsin ritenne che un trattamento differenziato tra uomini e donne non ledesse il diritto al giusto processo dell'imputato a non essere giudicato in base al sesso, poiché avendo uomini e donne tassi di recidiva diversi, ignorare il genere «fornirebbe risultati meno accurati». Fino a che punto, però, questo tipo di approccio è compatibile con il principio di eguaglianza?

Su questo genere di questioni un tradizionale elemento di discussione riguarda l'assunzione di decisioni discrezionali. Da una parte, vale a dire, il ricorso all'IA può essere visto come un modo per innalzare la prevedibilità delle decisioni e quindi contenere la discrezionalità dei funzionari pubblici, nella misura in cui si ritenga che questo sia desiderabile. Dall'altra parte, però, si ritiene che i casi in cui la legge conferisca a un'autorità pubblica il compito di prendere una decisione sulla base di standard vaghi, implicando ponderazione tra diversi possibili

[16] State of Wisconsin *vs* Loomis 881 N.W.2d 749 (Wis. 2016).

corsi d'azione, siano da sottrarre al campo di operazione delle decisioni automatizzate. Questa, per esempio, è la scelta operata dalla Germania, nella legge generale sul procedimento amministrativo (paragrafo 35), modificata nel 2017 per introdurre la previsione del provvedimento amministrativo completamente automatizzato (*Vollständig automatisierter Erlass eines Verwaltungsaktes*). Essa prevede che «un atto amministrativo può essere adottato interamente da una macchina, purché ciò sia consentito dalla legge e non vi sia discrezionalità né margine di valutazione». Si tratta di un orientamento non certo isolato. Per esempio, al centro del ricordato rapporto australiano vi è la distinzione tra macchine che «prendono una decisione» e «aiutano qualcuno a prendere una decisione». La prima ipotesi dovrebbe essere limitata a quelle decisioni che non comportano elementi discrezionali, mentre quando il «sistema esperto» assiste un funzionario nel prendere una decisione discrezionale, questo dovrebbe essere concepito in modo che il decisore non sia tenuto a, o indebitamente influenzato da, qualsiasi risultato specifico.

I giudici amministrativi italiani, pur presentando sinora un atteggiamento ambivalente nei confronti delle decisioni algoritmiche, sembrano invece non escludere in linea di principio che queste possano includere anche elementi discrezionali.

Tutta questa discussione può anche essere riguardata dal punto di vista del principio di legalità, inteso come predeterminazione normativa da parte del potere di assumere decisioni automatizzate. Secondo l'impostazione della legge tedesca sopra menzionata – ma anche quella dell'Australian Administrative Council nel rapporto citato – un provvedimento può essere completamente automatizzato solo a condizione che vi sia una disposizione di legge che conferisca il relativo potere. Alla stessa logica si ispira il GDPR come accennato, seppure dalla diversa angolazione della protezione dei dati personali. Questa impostazione si contrappone all'idea che l'impiego degli algoritmi sarebbe da ricondurre a un fatto organizzativo, quindi a scelte autonome delle burocrazie professionali. Anche rispetto

a questo profilo – per chi scrive decisivo[17] – può essere ritenuto rilevante il passaggio da semplici «sistemi esperti» a sistemi più complessi e la notata difficoltà di distinguere precisamente tra i due. Come è stato osservato[18], adottare l'uno o l'altro tipo di IA comporta in ultima istanza la necessità di accettare un *trade-off* tra svantaggi e vantaggi concreti per la comunità. Nel sopra ricordato recente rapporto statunitense *Government by Algorithm* si osserva che l'imposizione di vincoli alle scelte del modello, a esempio limitando il numero di caratteristiche dei dati o vietando approcci di modellazione più sofisticati, compromette la potenza di analisi dello strumento e, quindi, la sua utilità. Per esempio, i vantaggi in certi campi di sistemi avanzati di apprendimento automatico che elaborano in modo massiccio dati personali, potrebbero essere ottenibili solo accettando una nozione di trasparenza e *accountability* meno esigente che nel caso delle decisioni umane. A chi spetta, in un ordinamento democratico, effettuare tale valutazione?

La rilevanza di tale questione può essere anche apprezzata dal punto di vista empirico. In effetti, vi è un problema di «meta-trasparenza» relativo alla conoscenza dell'uso che le autorità pubbliche fanno dell'IA. Uno dei sottogruppi della ricerca Astrid ha tra i suoi obiettivi quello di cominciare a colmare la lacuna conoscitiva sull'uso degli algoritmi nella prassi delle istituzioni pubbliche italiane e sulla percezione da parte dei funzionari pubblici di quali siano le condizioni – legali e fattuali – che rendano o meno possibile tale uso. Il rapporto *Government by Algorithm* parte proprio dalla premessa che, al netto del grande dibattito accademico sull'IA, si sa ancora poco su come le pubbliche amministrazioni la stiano utilizzando, ne acquisiscano la disponibilità o ne controllino l'uso.

[17] S. Civitarese Matteucci, *Public Administration Algorithm Decision-Making and the Rule of Law*, in «European Public Law», 27, 2021, pp. 103-130.

[18] Zalnieriute, Bennett Moses e Williams, *The Rule of Law and Automation of Government Decision-Making*, cit.

3. *L'intelligenza artificiale: gli impatti sul sistema economico e la strategia italiana*

3.1. *Il ruolo pervasivo dell'intelligenza artificiale sui sistemi economici*

I sistemi di intelligenza artificiale svolgono ormai un ruolo cruciale a sostegno della competitività e dello sviluppo di numerosi settori produttivi, ed assumono una valenza strategica anche in comparti «extra-economici», quali la difesa e la sicurezza. Con riguardo ai primi, l'IA, coniugata ai *big data*, ai progressi nel supercalcolo ed alla diffusione della robotica, rappresenta infatti una straordinaria opportunità per ridurre i costi operativi ed accrescere l'efficienza dei processi produttivi, nonché per migliorare la qualità dei prodotti. A tale proposito, si è parlato di una «quarta rivoluzione industriale».

Vi sono numerose ricerche ed analisi che attestano la portata della «rivoluzione» dell'IA applicata al mondo della produzione: l'industria manifatturiera, senz'altro, ma anche i servizi e l'agricoltura. Tra i Paesi più avanzati nell'introduzione di sistemi di IA nei processi produttivi, spiccano quelli europei: Germania, Francia e Regno Unito che sopravanzano sia gli Stati Uniti che la Cina. L'Italia, dal canto suo, deve ancora impegnarsi molto per la diffusione dell'IA, nonostante i notevoli progressi degli ultimi anni. Quindi, l'Europa, o almeno i principali Paesi membri, sembrano poter svolgere un ruolo di primo piano nello sviluppo e nella diffusione dei sistemi di IA: sotto il profilo industriale e della produzione, oltre che nel campo delle regole, di cui si è trattato nei paragrafi precedenti.

I più significativi impatti dell'IA, nonché i primi che hanno raggiunto una dimensione rilevante, hanno interessato i mercati finanziari, cui è dedicata la Parte terza del volume III (*Intelligenza artificiale e FinTech*), cui si rinvia.

All'interno del sistema del credito, due mercati che registrano impatti significativi dell'IA sono quelli dei servizi di pagamento e del credito alla clientela *retail*, oggetto peraltro di una ricerca in corso presso Astrid. In questi mercati,

come in altri simili, diviene cruciale valutare e comparare le opportunità offerte dall'IA, in termini di guadagni di efficienza, personalizzazione dei servizi, miglioramento della *customer experience*, riduzione delle frodi, con taluni «rischi», riconducibili alla privacy ed alla *data governance*, nonché ad aspetti quali la trasparenza, l'*accountability*, la *fairness* e la non discriminazione, senza dimenticare i profili etici.

Tra gli altri comparti produttivi in cui l'intelligenza artificiale ha concorso al mutamento dei modelli di business e della stessa organizzazione delle imprese, vanno senz'altro menzionati l'industria dei trasporti, la farmaceutica, il commercio al dettaglio, in cui l'obiettivo principale ha riguardato il miglioramento della *customer experience*, ed il settore televisivo (soprattutto, a seguito della crescita impetuosa dello *streaming*)[19].

Non si può concludere questa veloce rassegna dei riflessi dell'IA sui sistemi economici, senza menzionare il caso particolare del mercato del lavoro, dove l'IA, assieme alle altre tecnologie digitali, è destinata a produrre effetti di estremo rilievo, sotto diversi profili. Da un lato, in termini di livelli occupazionali, con riguardo cioè al saldo tra posti di lavoro creati e posti di lavoro eliminati dall'introduzione e dalla diffusione dell'IA. Dall'altro lato, in relazione alla struttura delle qualifiche professionali ed alle competenze necessarie per interagire efficacemente con i sistemi di IA. Infine, per i conseguenti riflessi sulla struttura delle retribuzioni.

Al di là di questi cenni sintetici, l'analisi degli impatti dell'IA sui sistemi economici continua ad essere oggetto di studi e ricerche, a livello internazionale ed anche nel nostro Paese. In questo contesto, si colloca l'attività del Laboratorio sull'ecosistema digitale (LED), avviato dalla Fondazione Astrid nel settembre 2019: oltre alla ricerca che costituisce l'oggetto di questi volumi, altre iniziative

[19] Al riguardo, LED ha in corso una ricerca sulle prospettive del settore televisivo, che affronta l'impatto delle tecnologie digitali – tra cui l'intelligenza artificiale – sia dal punto di vista infrastrutturale, sia con riguardo alle modalità di fruizione dei contenuti audiovisivi.

stanno affrontando il tema dell'IA, all'interno dell'ecosistema digitale[20].

3.2. *Il «Programma strategico intelligenza artificiale 2022-2024»*

Si è accennato in precedenza che il nostro Paese non è all'avanguardia nell'applicazione di sistemi di IA nel mondo della produzione e, più in generale, nell'economia e nella società. Come è stato notato da diversi esperti della materia, ad esempio dall'Osservatorio sull'intelligenza artificiale della Bocconi, in Italia ci sarebbe il potenziale per svolgere un ruolo di primo piano nel campo dell'IA: vantiamo posizioni di rilievo nella preparazione dei talenti e nella ricerca, che tuttavia risulta molto frammentata, diverse nicchie di specializzazione, una apprezzabile capacità brevettuale, limitatamente ad alcune applicazioni, una significativa crescita degli investimenti da parte delle imprese[21]. Ciò che serve è una strategia adeguata, un approccio di sistema, che coinvolga pubblico e privato.

Al riguardo, nel luglio del 2020, il ministero per lo Sviluppo economico ha pubblicato un voluminoso rapporto dal titolo *Proposte per una strategia italiana per l'intelligenza artificiale*, sulla base del lavoro di un apposito Gruppo di esperti istituito dallo stesso ministero. La necessità di definire una strategia nazionale sull'IA derivava da una richiesta in tal senso rivolta dalla Commissione europea ai

[20] A tal fine, si ricorda che LED si occupa di esaminare i mercati legati alla «produzione» di infrastrutture, beni e servizi digitali (il *cloud computing*, la telefonia 5G, i cavi sottomarini), nonché di valutare gli impatti delle tecnologie digitali su alcuni mercati «utilizzatori» (i servizi postali, i settori verticali collegati alle reti 5G, l'industria dell'audiovisivo, i sistemi di pagamento ed il credito al consumo, il mercato del lavoro).

[21] Dai dati dell'Osservatorio Artificial Intelligence della School of Management del Politecnico di Milano, si apprende che il mercato italiano dell'intelligenza artificiale ha registrato un incremento del 27% nel 2021, per un valore di 380 milioni di euro, raddoppiato in soli due anni.

Paesi membri, in concomitanza con la presentazione della strategia europea per l'IA[22].

Pochi mesi prima della pubblicazione della proposta di strategia italiana, era stato pubblicato il *Libro bianco sull'intelligenza artificiale*[23], con il quale l'Europa comincia a delineare un quadro di intervento non limitato al solo versante regolamentare[24].

Attualmente, come anticipato nel primo paragrafo, le iniziative dell'Europa in materia di IA riguardano certamente il campo delle regole, ossia la Proposta di Regolamento pubblicata nell'aprile 2021[25], ma anche la «politica industriale», come indicato dal nuovo Piano coordinato sull'intelligenza artificiale[26], finalizzato appunto a rafforzare contestualmente l'adozione dell'IA e gli investimenti nel settore in tutta l'Unione europea.

Questo approccio, che coniuga regolamentazione e «politica industriale», si ritrova anche nel recente *Programma strategico intelligenza artificiale 2022-2024*, pubblicato dal governo italiano a fine novembre 2021[27].

Il documento si articola in tre sezioni: la prima riguarda il contesto, ossia la posizione competitiva dell'Italia in materia di IA; la seconda stabilisce i principi guida, gli obiettivi ed i settori prioritari per un programma strategico nazionale; la terza, infine, individua tre aree di intervento fondamentali (competenze, ricerca, applicazioni), nonché le relative politiche d'intervento.

[22] *L'intelligenza artificiale per l'Europa*, COM(2018) 237 finale.

[23] *Libro bianco sull'intelligenza artificiale. Un approccio europeo all'eccellenza e alla fiducia*, 19/2/2020.

[24] Sui temi della disciplina dell'IA, si rinvia al contributo di V. Falce (Parte terza del volume III).

[25] 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, aprile 2021.

[26] Nuovo Piano coordinato sull'intelligenza artificiale 2021, che aggiorna il precedente Piano, presentato nel dicembre 2018, alcuni mesi dopo la pubblicazione della Strategia europea su IA.

[27] *Programma strategico intelligenza artificiale 2022-2024*, a cura dei ministeri dell'Università e della ricerca, dello Sviluppo economico e del ministro per l'Innovazione tecnologica e la transizione digitale, 24/11/2021.

Ai fini che qui interessano, ci si limita solo ad alcuni richiami al documento, riservando ad altra parte della ricerca più approfondite valutazioni circa le indicazioni di *policy* avanzate dal governo italiano con riguardo all'IA.

La prima considerazione riguarda il terreno definitorio, oggetto prevalente di questo capitolo. L'esordio del Rapporto inquadra l'intelligenza artificiale come il complesso di «modelli digitali, algoritmi e tecnologie che riproducono la percezione, il ragionamento l'interazione e l'apprendimento»[28]. Come si può constatare, una definizione alquanto differente da quelle indicate nel paragrafo 1, in particolare rispetto a quella proposta a pagina 48 («l'intelligenza artificiale fa riferimento a sistemi che mostrano comportamenti intelligenti, operando attraverso un'analisi dell'ambiente in cui operano e intraprendendo, con un certo grado di autonomia, azioni mirate a raggiungere determinati obiettivi specifici»).

In altre parti del Rapporto, si individua un ecosistema dell'intelligenza artificiale, facendo ricorso ad un termine – ecosistema – alquanto usato (forse, abusato) negli ultimi tempi. In questo ecosistema IA, rientrano quattro tipologie di «operatori»/«soggetti»: la comunità scientifica; i centri di trasferimento tecnologico; i fornitori di tecnologie e soluzioni; gli utenti privati e pubblici. In effetti, questa rappresentazione dell'ecosistema IA richiama molto da vicino quella proposta per l'ecosistema digitale[29], evidenziando un aspetto di assoluto rilievo, diffusamente segnalato nel primo paragrafo: ossia, l'interazione tra IA, *big data*, algoritmi, supercalcolo.

Una seconda considerazione rimanda a quello che – ad avviso di chi scrive – rappresenta il più importante apporto di questo Rapporto: la focalizzazione sulle misure di intervento, ossia sul versante della «politica industriale di contesto». Si tratta di ben 24 *policies* da adottare nel prossimo triennio,

[28] *Programma strategico intelligenza artificiale 2022-2024*, cit., p. 2.

[29] Per una sintetica rappresentazione del concetto di «ecosistema digitale», sia consentito rinviare a A. Perrucci, *Dai Big Data all'Ecosistema Digitale: dinamiche tecnologiche e di mercato e ruolo delle politiche pubbliche*, in «Analisi giuridica dell'economia», 1, 2019.

con riferimento alle tre aree di intervento fondamentali: competenze (e talenti); ricerca; applicazioni (sia per la pubblica amministrazione che per le imprese private).

Queste misure possono essere definite «di contesto» perché, almeno al livello attuale di declinazione, sembrano confinate ad una fase pre-competitiva: infatti, nonostante siano individuati i settori prioritari – ben undici, peraltro – gli interventi si riferiscono a misure che riguardano i fattori produttivi in senso lato (capitale umano, ricerca di base ed applicata, innovazione ed adozione tecnologica, credito d'imposta, promozione start-up).

In definitiva, l'intervento pubblico in materia di IA sembra seguire la strada più promettente e corretta: si promuovono interventi che affiancano ed accompagnano le iniziative dei privati, senza sostituirsi ad essi, evitando un effetto di *crowding out* degli investimenti privati che – invece – vanno incentivati, magari riducendone il rischio, e sempre nel rispetto delle norme a tutela della concorrenza.

Da ultimo, una considerazione circa la relazione tra il complesso di queste misure, che probabilmente avranno bisogno di una successiva articolazione per i settori prioritari, e la posizione dell'Italia in relazione alla Proposta di Regolamento sull'IA presentata dalla Commissione europea. È senz'altro positivo che anche il nostro Paese, come sta facendo l'Europa, superi un approccio circoscritto alle misure di carattere regolamentare, ossia la «specializzazione» nel campo delle regole, che ha fatto parlare di un «effetto Bruxelles»[30]. Tuttavia, regolazione e «politica industriale» sono due leve che vanno considerate assieme, in una visione di sistema: assolutamente necessaria nel caso dell'IA, come – più in generale – per la grande sfida della trasformazione digitale che la nuova Commissione europea ha posto al centro del proprio programma, assieme al tema della transizione ecologica.

L'auspicio, pertanto, è che presto si definisca una visione di sistema per il digitale, a partire dalle iniziative previste al

[30] A. Bradford, *The Brussels Effect: How the European Union Rules the World*, Oxford, 2020.

riguardo dal Piano nazionale di ripresa e resilienza: misure senza dubbio importanti per sostenere la transizione digitale della nostra economia, ma che – necessariamente – sono state costruite in una logica *bottom up*, dati i tempi ristretti di identificazione dei progetti e delle riforme, senza quindi la possibilità di disegnare una impostazione di sistema.

Le considerazioni svolte dal ministro Colao, in occasione dell'audizione del 9/3/2022 sulla legge per l'intelligenza artificiale, sembrano andare in questa direzione, con una chiara indicazione:

> Se il regolamento europeo prova a definire una cornice normativa per l'ecosistema IA, noi con la Strategia nazionale approvata a novembre vogliamo dare maggior impeto al nostro sistema nazionale. Quindi non è uno scopo legislativo ma esecutivo e di governo.