Università
Ca'Foscari
Venezia

Corso di Laurea Magistrale

In

Economia e gestione delle aziende

Tesi di Laurea

Un'intelligenza artificiale affidabile

Design etico e uso nella giustizia predittiva

Relatore
Ch. Prof. Francesco Rullani

Laureando
Gabriele Frasson
Matricola 858639

Anno Accademico
2020 / 2021

INDICE

Introduzione

L'idea dell'intera tesi, sebbene non completamente pertinente al mio ambito di studi, è nata dalla volontà di capire meglio il mondo digitale, sempre più in espansione, e la sua relazione con le persone.

In particolare l'intelligenza artificiale e gli algoritmi sono onnipresenti, regolando ormai moltissimi aspetti della nostra quotidianità senza che noi ce ne accorgiamo. Molte sono le decisioni che si basano su di essi. Basti pensare che dei ricercatori in Cina hanno riferito di aver creato un'intelligenza artificiale che può presentare un'accusa di reato con una precisione del 97%, andando di fatto a sostituire, in parte, il processo decisionale dei pubblici ministeri[1]. Oppure ancora, che all'Università di Guelph è stato progettato un ipernetowrk in grado di addestrare una rete neurale settandone i parametri[2], quindi un'IA che allena un'altra IA.

L'intelligenza artificiale è tanto presente e potente, quanto sconosciuta. È una tecnologia da considerare attentamente, possibilmente progettandola in modo che sia il più affidabile possibile. Questa valutazione non deve tenere conto solo della parte che riguarda gli aspetti concreti, infatti non può prescindere da un approccio etico poiché di mezzo ci sono gli interessi degli esseri umani.

Scopo di questo elaborato è proprio questo: analizzare l'intelligenza artificiale in quest'ottica valutandone le implicazioni e cercando di definire un framework di design etico.

Il tentativo di perseguire questo intento si esplicherà nel modo seguente.

Nel primo capitolo si cercherà di capire in cosa consiste l'intelligenza artificiale, cosa rappresenta e come si è arrivati a conoscerla per quella che è oggi.

[1] "Chinese Scientists Develop AI 'Prosecutor' That Can Press Its Own Charges | South China Morning Post," n.d., https://www.scmp.com/news/china/science/article/3160997/chinese-scientists-develop-ai-prosecutor-can-press-its-own.
[2] Chris Zhang, Mengye Ren, and Raquel Urtasun, "GRAPH HYPERNETWORKS FOR NEURAL ARCHITECTURE SEARCH" (2019): 17.

Con il secondo capitolo entriamo/si entrerà nel vivo della discussione: vedremo/si vedrà cosa bisogna considerare per la progettazione di un'intelligenza artificiale affidabile, che segua soprattutto un approccio etico, rispettoso dell'essere umano.

Nel terzo capitolo analizzeremo/verrà analizzato un ambito specifico in cui l'intelligenza artificiale viene applicata, quello della giustizia predittiva.

Il quarto e ultimo capitolo consiste in un'analisi dati sul risk assessment tool Savry. Verrà utilizzato un dataset riguardante i minorenni autori di reato in Catalogna. Tale analisi ha lo scopo di rendere concrete parte delle riflessioni espresse nei precedenti capitoli.

Capitolo 1

L'intelligenza artificiale: informazioni per una panoramica generale

1.1 Storia dell'intelligenza artificiale

Fin dall'antichità l'uomo è stato portato ad imitare una prerogativa del divino, cioè l'atto della creazione[3]. Nella mitologia Efesto, scagliato giù dall'Olimpo e diventato incapace di camminare viene aiutato in tale mansione da degli automi così come un essere artificiale era presente nel mito di Pigmalione e Galatea[4]. Passando per il mito del Golem di Praga e fino ad anni più recenti nella letteratura si può ricordare la creazione di un essere da parte del dottor Frankenstein.

Inoltre sempre fin dagli albori dell'umanità l'uomo si è chiesto come poter automatizzare le proprie azioni tramite macchine o strumenti. Si parte da 7000 anni fa quando si inventò il primo abaco che fu perfezionato prima da Blaise Pascal con la pascalina e poi da Gottfried Wilhelm[5]. Tutte queste prime macchine non erano programmabili e potevano svolgere una singola operazione[6].

Sicuramente infatti il campo dell'intelligenza artificiale si sviluppò in concomitanza con quello dei computer.

La progettazione dei primi strumenti automatizzati risale agli inizi dell'Ottocento ad opera di Charles Babbage. La sua macchina è strutturata come una serie di elementi meccanici e le istruzioni scritte sotto forma algoritmica su tavolette perforate. Per via degli elevati costi di produzione fu sviluppato solo un modellino non in scala reale. Gli studi di Babbage risulteranno comunque utili negli anni successivi nel capire l'esecuzione in caso di algoritmi complessi[7].

[3] Isaac Asimov and Laura Serra, *Io, robot* (Milano: Mondadori, 2018).
[4] Kevin Warwick, *Intelligenza Artificiale - Le basi* (Dario Flaccovio Editore, 2015).
[5] Amedeo Santosuosso, *Intelligenza artificiale e diritto* (Mondadori, 2020).
[6] Giovanni Sartor, *L'informatica giuridica e le tecnologie dell'informatica: corso d'informatica giuridica*, 2016, https://ebookcentral.proquest.com/lib/concordiaab-ebooks/detail.action?docID=4771207.
[7] Stefano Quintarelli et al., *Intelligenza artificiale* (Bollati Boringhieri, 2020).

Il termine compute machine (macchina computazionale) inizia a diffondersi negli anni '20 riferendosi a macchine che riescono a svolgere calcoli in modo efficace. Inizialmente queste sono solamente meccaniche, nei decenni a seguire diventeranno elettronico/digitale[8].

Nel 1936 Alan Mathison Turing concepisce l'idea di macchina digitale astratta in grado di eseguire algoritmi e composta da un nastro infinito, cioè la memoria sulla quale sono salvati vari simboli, e da uno scanner che può leggere e modificare tali simboli. È cosi che si inizia a pensare a macchine programmabili[9].

Nel 1943 McCulloch e Pitts nel loro lavoro descrissero i primi neuroni artificiale Questo fu uno dei primi risultati ottenuto nel campo della cibernetica, disciplina che si occupa dello studio di strumenti e tecnologie e loro applicazione secondo modelli di funzionamento tipici degli esseri viventi. Tale studio fu poi perfezionato nel 1958 da Rosenblatt che inventò il percettrone, un modello matematico di neuroni, ossia una rete neurale ispirata al funzionamento del cervello umano[10].

Altri contributi riconducibili sempre a Turing sono riscontrabili qualche anno dopo. Nel 1946, in una conferenza parlò per la prima volta di "macchine che possono imparare dall'esperienza", quindi in grado di modificare le proprie istruzioni da sole. Nel 1950 nell'articolo "Computing Machinery and Intelligence"[11] Turing si chiese se le macchine potessero essere dotate della facoltà di pensiero, introdusse perciò il noto test di Turing per verificare l'intelligenza di una macchina: un intervistatore, ponendo delle domande, deve cercare di distinguere il comportamento di una macchina da quello dell'essere umano.

Grazie a Turing il campo dell'AI ottenne grande considerazione in quegli anni.

Alcuni anni dopo Marvin Minsky e Dean Edmonds crearono il primo computer IA, gli studi di Pitts e McCulloch sui modelli neurali si rivelarono fondamentali. Tale computer

[8] Santosuosso, *Intelligenza artificiale e diritto*.
[9] Quintarelli et al., *Intelligenza artificiale*.
[10] Warwick, *Intelligenza Artificiale - Le basi*.
[11] A. M. Turing, "I.—COMPUTING MACHINERY AND INTELLIGENCE," *Mind* LIX.236 (1950): 433–60, https://doi.org/10.1093/mind/LIX.236.433.

chiamato SNARC, acronimo per Stochastic Neural Analog Reinforcement Computer, era basato su una rete neurale che simulava 40 neuroni[12].

Si ritiene che il termine intelligenza artificiale e tale disciplina nacquero durante un seminario tenutosi nell'estate del 1956 al Dartmouth College. Vi parteciparono i maggiori esponenti dell'informatica dell'epoca: John Mccarthy, Marvin Minsky, Claude Shannon e Nathaniel Rochester. In tale contesto non vennero fatte grosse scoperte o passi avanti ma venne ben definito un ambito disciplinare, quello dell'intelligenza artificiale. Si arrivò a ritenere che qualsiasi caratteristica dell'intelligenza possa essere trasposta e simulata da una macchina[13].

Negli anni Sessanta un autorevole contributo fu dato da Newell e Simon con il loro General Problem Solver, programma per simulare la capacità di problem solving degli esseri umani.

Nel decennio successivo non vi furono particolari avanzamenti nel settore. Bisogna considerare che la potenza di calcolo dei computer di allora non era di certo quella di oggi. Questo fu uno dei principali limiti, infatti i compiti che avrebbe dovuto risolvere l'intelligenza artificiale richiedevano la capacità di elaborare una mole considerevole di informazioni. La macchina elaborata da Ross Quillan per il linguaggio naturale poteva elaborare solo venti parole. Si può facilmente intuire che non fossero abbastanza per permettere ad una macchina di comunicare come un essere umano[14].

Negli anni Ottanta sorse nuovo interesse verso questa disciplina. In particolare vennero sviluppati i sistemi esperti, delle macchine progettate per far fronte a compiti molto specifici, spostandosi da sistemi bassati sulla formalizzazione dei meccanismi di ragionamento a sistemi knowledge-based[15]. Secondo la definizione di Jackson sono "sistemi computerizzati in grado di emulare l'attività di decisione di un essere umano esperto in un determinato settore"[16]. I sistemi esperti erano dominio-specifici, erano ossia dotati di conoscenza altamente specializzata, infatti si credeva che proprio questa fosse alla base di competenze umane rare che differenziassero gli esperti dai

[12] Warwick, *Intelligenza Artificiale - Le basi*.
[13] Santosuosso, *Intelligenza artificiale e diritto*.
[14] Warwick, *Intelligenza Artificiale - Le basi*.
[15] Jerry Kaplan, *Intelligenza artificiale. Guida al futuro prossimo* (Luiss University Press, 2018).
[16] Peter Jackson, *Introduction to Expert Systems*, 3rd ed., International Computer Science Series (Harlow, England ; Reading, Mass: Addison-Wesley, 1999).

principianti. Tali programmi erano composti da due componenti: una base di conoscenza rappresentata in forma simbolica tramite relazioni, regole e un motore di inferenza che serve per leggere e modificare i simboli. Il programmatore doveva essere anche un esperto della materia ed essere disponibile ad aggiungere nuova conoscenza qualora fosse stato necessario. Vennero ottenuti successi in campi quali la diagnosi, la progettazione, il monitoraggio, l'interpretazione di dati e la pianificazione. Ma la codifica manuale delle conoscenze e la forzatura nel ricorrere alle capacità umane sono stati evidenti limiti che, congiuntamente all'esplosione di approcci diversi, quali quelli adottati dai personal computer che erano anche più economici e pratici, hanno portato a ridurre gli investimenti riguardanti i sistemi esperti[17].

Per capire l'evoluzione che negli anni è avvenuta in questo campo basti pensare che nei primi anni '90 un computer/IA, Deep Blue, giocò a scacchi contro Garry Kasparov, granmaestro e campione del mondo in carica, e lo sconfisse[18].

Questo è stato possibile grazie ai progressi ottenuti in termini di potenza, memoria ed interconnettività.

A partire dagli anni Sessanta infatti vi fu un continuo miglioramento nel campo dell'elettronica. Grazie alla progressiva miniaturizzazione dei componenti hardware si svilupparono computer sempre più piccoli e veloci, aumentando esponenzialmente la velocità di calcolo e la capacità di memorizzazione dei dati. Questo seguiva quanto postulato da Moore nella sua prima legge: il numero di componenti elettronici che formano un chip sarebbe raddoppiato ogni anno, incrementando la potenza di calcolo[19]. La miniaturizzazione ha consentito inoltre di dotare la macchina di un numero sempre maggiore di sensori, ottenendo dati sempre più accurati e completi riguardanti qualsiasi aspetto del mondo circostante[20].

Negli anni '90 fecero la loro comparsa le Graphics Processing Unit, ossia i processori grafici che, proveniente dal mondo dei videogiochi, furono subito di grande aiuto per la loro capacità di eseguire processi complessi velocemente, più delle CPU fin lì utilizzate.

[17] Kaplan, *Intelligenza artificiale. Guida al futuro prossimo*.
[18] Warwick, *Intelligenza Artificiale - Le basi*.
[19] Sartor, *L'informatica giuridica e le tecnologie dell'informatica*.
[20] Alessandra Carleo, ed., *Decisione Robotica*, Percorsi. Diritto (Bologna: Il mulino, 2019).

Con l'avvento della Seconda guerra mondiale vi fu uno sviluppo della tecnologia particolarmente intenso, ovviamente a fini militare, e nello specifico tale sviluppo ottenne grossi risultati nei sistemi di trasmissione. Vennero piantati i semi per la futura nascita di sistemi di comunicazione senza fili come il Wi-Fi o il GPS e per internet. Infatti il World Wide Web ha consentito l'accesso ad una mole incredibile di dati e conoscenze semplificando la vita al campo dell'intelligenza artificiale.

Ai nostri giorni questa l'intelligenza artificiale è applicata in ogni dove, anche in ambiti per noi impensabili. Dal settore finanziario, a quello produttivo e militare l'IA ormai ha raggiunto performance impensabili per una normale persona[21].

1.2 Che cos'è l'intelligenza artificiale: una serie di definizioni

L'intelligenza artificiale è una disciplina che negli ultimi anni ha contribuito in modo significativo al mondo dell'informatica consentendo l'evoluzione di molti settori grazia alla sua versatilità

L'intelligenza artificiale è una disciplina influenzata dagli studi dei più svariati campi:

- ingegneria e informatica: sviluppo hardware dei calcolatori ma anche software dei programmi e algoritmi;
- psicologia: studi riguardanti la mente e il comportamento, la coscienza e l'apprendimento;
- medicina e neuroscienze: studi sul sistema nervoso e sul funzionamento dei neuroni e del cervello umano;
- matematica: la logica matematica e le teorie sulla probabilità sono importanti per quanto riguarda il processo decisionale;
- filosofia: logica formale ed epistemologia utili per la codificazione della conoscenza e per capire gli ambiti applicativi dell'IA;
- economia: teoria per la gestione di risorse scarse, ragionamenti sulle scelte dei consumatori, la teoria dei giochi con la quale si analizzano i comportamenti degli agenti;

[21] Sartor, *L'informatica giuridica e le tecnologie dell'informatica*. ; Warwick, *Intelligenza Artificiale - Le basi*.

- linguistica: gli studi sul linguaggio sono fondamentali per far processare al meglio all'IA tutti i dati di questo ambito[22].

Questo risulta essere solamente un breve accenno del contributo dei vari campi all'IA. Risulta assai chiaro come il tema Intelligenza artificiale sia interdisciplinare richiedendo le competenze e conoscenze di svariati professionisti.

Anche per quanto riguarda le definizioni è difficile trovarne una univoca. Dalla letteratura ne sono emerse varie, in ognuna vengono messe in evidenza aspetti diversi, evidenziando la complessità di questo fenomeno, tant'è che abbiamo visto come sia difficile definire cosa sia la stessa intelligenza.

Tra le molte definizioni sono emersi dei fattori comuni che concordano nel definire l'intelligenza artificiale come l'insieme di tecniche e studi che consentono di sviluppare e progettare algoritmi che permettano alle macchine di svolgere prestazioni che, per un osservatore comune, sembrerebbero appartenere ad un essere dotato di intelligenza umana[23].

Secondo l'Enciclopedia Britannica l'intelligenza artificale è "the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience"[24].

É interessante come venga nominato il termine generalizzare. Secondo l'opinione di Kaplan proprio in questo potrebbe risiedere l'essenza dell'intelligenza: la capacità di condurre generalizzazione corrette con il giusto tempismo e su una base di dati non illimitati. Seguendo questa logica tanto più sarebbe vasto il campo di applicazione e tanto più velocemente si potrebbe trarre conclusioni con informazioni minime, allora tanto maggiore sarebbe da considerarsi intelligente il compito svolto[25].

[22] Giuseppe Contissa, *Information Technology for the Law*, Informatica Giuridica. Serie Didattica 6 (Torino: G. Giappichelli, 2017).
[23] Francesco Amigoni, Viola Schiaffonati, and Marco Somalvico, "Intelligenza artificiale in 'Enciclopedia della Scienza e della Tecnica.'"
[24] J. Copeland, "Intelligenza Artificiale in 'Encyclopedia Britannica.'"
[25] Jerry Kaplan, *Intelligenza artificiale. Guida al futuro prossimo* (Luiss University Press, 2018).

L'intelligenza artificiale consiste quindi nella capacità di un computer di svolgere funzioni cognitive simili a quelle della mente umana come risolvere problemi, ragionare, comunicare, imparare, percepire, interagire, risolvere problemi, creare[26]. Questo secondo il Cambridge Dictionary, mentre l'English Oxford Living Dictionary fa degli esempi più concreti indicando la percezione visiva, il riconoscimento vocale, il processo decisionale e la traduzione tra le lingue quali compiti che richiedono intelligenza umana.

Ancora, la Commissione Europea nel suo White Paper on IA parla di intelligenza artificiale come un insieme di tecnologie che combinano dati, algoritmi e potenza di calcolo.

Vi sono poi altre definizioni che non associano l'intelligenza artificiale alle competenze umane, ma più alla percezione e interazione con l'ambiente circostante.

In particolare sono quattro le macroaree alle quali possiamo ricondurre tali abilità per creare un framework[27]:

- sentire: fare la possibilità ad una macchina di percepire il mondo circostante acquisendo ed elaborando immagini, suoni, parole, testi e altri dati;
- comprendere: consente alla macchina di comprendere le informazioni e i dati raccolti, applicandovi l'analytics e ricavandone predizioni;
- agire: consente alla macchina di compiere azioni nel mondo fisico o digitale sulla base della comprensione precedente
- imparare: permette alla macchina di ottimizzare in maniera continua le sue prestazioni[28].

Oltre a questa classificazione gli studiosi Perter Norvig e Stuart Russel[29] hanno individuato quattro campi o parametri derivanti da due diverse classificazioni. A seconda che l'interesse per l'IA riguardi il pensiero/ragionamento o il comportamento e che misurino il successo secondo prestazioni umane o razionalmente. Incrociando

[26] Alina Köchling and Marius Claus Wehner, "Discriminated by an Algorithm: A Systematic Review of Discrimination and Fairness by Algorithmic Decision-Making in the Context of HR Recruitment and HR Development," *Bus Res* 13.3 (2020): 795–848, https://doi.org/10.1007/s40685-020-00134-w.
[27] Massimo Morielli, Leonardo Galimberti, and Applied intelligence, 2018, "Intelligenza Artificiale: Istruzioni per L'uso," 2018, Accenture Applied intelligence, https://www.accenture.com/it-it/insights/artificial-intelligence/artificial-intelligence-explained-executives.
[28] Morielli, Galimberti, and Applied intelligence, 2018, "Intelligenza Artificiale: Istruzioni per L'uso."
[29] Stuart J. Russell, Peter Norvig, and Ernest Davis, *Artificial Intelligence: A Modern Approach*, 3rd ed., Prentice Hall Series in Artificial Intelligence (Upper Saddle River: Prentice Hall, 2010).

queste due classificazioni si ottengono quattro diverse tipologie di definizioni: 1) sistemi che agiscono come gli esseri umani, identificabile tramite il Test di Turing, 2) sistemi che agiscono razionalmente, quindi sono degli agenti razionali, 3) sistemi che pensano come gli esseri umani, che seguono un approccio derivante dalle scienze cognitive, 4) sistemi che pensano razionalmente, seguendo le "leggi del pensiero"[30].

Tutti i diversi approcci durante l'evoluzione dell'intelligenza artificiale sono stati affrontati. Poiché non si è ancora arrivati al punto in cui le macchine possano pensare e poiché è sicuramente utile che i computer ottengano risultati migliori di quelli che farebbe una persona, gli studi attualmente si concentrano soprattutto nel campo dei sistemi che agiscono in modo razionale[31].

Varie sono le interpretazioni che si possono dare all'IA. Considerando l'esistenza di vari punti di vista, sia per quanto riguarda il lato teorico che quello realizzativo, appare sensato considerare riduttivo il tentativo di classificarla o definirla in modo univoco. Questo anche in virtù del fatto che spesso viene utilizzata come metro di giudizio la capacità umana, già di gran lunga superata in alcuni contesti[32]. Un computer è sicuramente da considerare più intelligente di una persona in contesti limitati e definiti. I computer hanno prestazioni di gran lunga superiori per quanto riguarda la velocità di elaborazione, la precisione nei calcoli, la capacità di memorizzazione, la gestione di dati complessi e la possibilità di funzionare ininterrottamente. Si potrebbe considerare solamente modi diversi di definire l'intelligenza[33].

Seppur l'IA non sia considerabile come una scienza dura, la fisica o la biologia lo sono poiché le teorie possono essere confermate in modo oggettivo, è plausibile negli anni a venire lo diventi. Nonostante l'intelligenza artificiale non sia completamente una disciplina a sé stante l'impatto che già oggi ha su tutti noi non è discutibile[34].

[30] Russell, Norvig, and Davis, *Artificial Intelligence*.
[31] Gianluigi Ciacci and Giovanni Buonomo, *Profili di informatica giuridica*, 2. ed., CEDAM scienze giuridiche (Milano: Wolters Kluwer, 2021).
[32] Kaplan, *Intelligenza artificiale. Guida al futuro prossimo*.
[33] Roberto Marmo, *Algoritmi per l'intelligenza artificiale* (HOEPLI, 2020).
[34] Kaplan, *Intelligenza artificiale. Guida al futuro prossimo*.

1.3 Filosofia dell'IA

1.3.1 Concezione di intelligenza

L'intelligenza artificiale, semplificando, consiste nella scienza e nell'ingegneria che consentono di sviluppare macchine intelligenti[35]. Ma cos'è l'intelligenza?

L'intelligenza è quella caratteristica che contraddistingue l'essere umano, quindi prima di passare ad analizzare l'intelligenza artificiale vorrei soffermarmi brevemente sul concetto di intelligenza.

L'intelligenza è un'entità molto complessa e multiforme, ognuno ne ha la propria idea e in base a questa valuta di conseguenza gli altri. Spesso si è influenzati dal contesto culturale e sociale in cui si vive[36].

Uno degli studi più interessanti è quello del neuropsicologo Gardner che nel corso dei propri studi ha individuato vari tipi di intelligenza:

- corporeo-cinestetico: riguardante il controllo corporeo e il controllo sugli oggetti tramite esso;
- interpersonale: volta alla comprensione delle altre persone e delle loro decisioni;
- intrapersonale: volta alla comprensione di se stessi e dei propri stati d'animo;
- linguistico: capacità di padroneggiare una lingua consentendo un'accurata espressione;
- logico-matematico: capacità di analisi in modo logico-scientifico;
- creativo: arrivare alla creazione di nuovi schemi di pensiero, si può considerare come parte di questa anche l'intelligenza filosofico-esistenziale, ossia la capacità di ragionare in modo astratto;
- visuale-spaziale: intelligenza che riguarda la percezione dello spazio fisico[37].

[35] John McCarthy, "What Is Artificial Intelligence?," *Computer Science Department Stanford University* (2007): 15, http://www-formal.stanford.edu/jmc/.
[36] Warwick, *Intelligenza Artificiale - Le basi*.
[37] Howard Gardner and Ester Dornetti, *Cinque chiavi per il futuro* (Milano: Feltrinelli, 2015); Luca Massaron, *Intelligenza artificiale for dummies* (S.l.: HOEPLI, 2020).

Il cervello possiede una moltitudine di peculiarità che lo rendono capace di varie funzioni, le intelligenze che contraddistinguono l'essere umano e lo differenziano dalle altre specie sono quella creativa e quella intrapersonale.[38]

1.3.2 Intelligenza Artificiale debole e forte

Possono delinearsi due differenti tipologie di intelligenza artificiale: weak AI e strong AI. Della prima, quella debole, si caratterizzerebbero i sistemi capaci di imitare determinate funzionalità cognitive appartenenti all'uomo, ma senza giungere alle stesse capacità intellettuali, limitandosi dunque a replicarne i processi logici. La macchina infatti confronta casi simili, li elabora e impara la soluzione più razionale per risolverli, addestrando quindi la propria capacità di problem-solving, che migliora con l'ulteriore immissione di dati[39]. Un esempio è il gioco degli scacchi, in cui la macchina si dimostra in grado di intraprendere decisioni e svolgere ragionamenti simili a quelli umani.

Della seconda invece si caratterizzerebbero i sistemi in grado di giungere a sviluppare una propria sapienza, o addirittura coscienza di sé, senza la necessità di imitare modelli di ragionamento e di pensiero simili a quelli umani[40]. Si tratterebbe quindi di un'intelligenza indipendente da quella umana, e c'è chi teme che a lungo andare diventerebbe anche superiore. Chi sposa questa distinzione, teorizzata per la prima volta dal filosofo americano John Searle nel 1980, ritiene che l'intelligenza artificiale forte sia un modello a cui non si giungerà nell'immediato, ma solo nel lungo periodo[41].

La distinzione potrebbe essere riassunta dicendo che esistono due tipi di intelligenza artificiale: macchine che sono davvero intelligenti e macchine che agiscono come se fossero tali[42].

Va precisato come nello studio dell'intelligenza artificiale si debba mirare a un approccio scevro dai pregiudizi che ricercano un costante confronto tra l'intelligenza artificiale e

[38] Yuval Noah Harari and Giuseppe Bernardi, *Sapiens. Da animali a dèi: breve storia dell'umanità* (Milano: Bompiani, 2020).
[39] Cabirio Cautela et al., "The Impact of Artificial Intelligence on Design Thinking Practice: Insights from the Ecosystem of Startups," *Strategic Design Research Journal* 12.1 (2019): 114–34, https://doi.org/10.4013/sdrj.2019.121.08.
[40] Sartor, *L'informatica giuridica e le tecnologie dell'informatica*.
[41] Marmo, *Algoritmi per l'intelligenza artificiale*.
[42] Kaplan, *Intelligenza artificiale. Guida al futuro prossimo*.

quella umana. Con la prima non si tenta necessariamente di imitare la seconda, anche perché non si potrebbe mai riuscire nella perfetta simulazione dell'intelligenza umana partendo da un computer. Un altro pregiudizio comune è credere che i computer pensino in modo meccanicistico e deciso a priori, e che invece gli umani pensino in modo casuale. Il ragionamento umano appare casuale perché si fatica a comprenderne a pieno il funzionamento, che nasce dagli impulsi cerebrali. Chi sostiene la teoria dell'intelligenza artificiale forte crede che questa possa comportarsi allo stesso modo di un cervello umano, ritenendo non significative, o per lo meno non preponderanti, le questioni relative al rapporto della mente con il corpo, alle esperienze di vita tipicamente umane, e al concetto di coscienza in una veste più spirituale che materialista. Chi invece propende verso il non ritenere valida questa teoria tiene in considerazione questi altri aspetti (non scientificamente misurabili) e ne fa discendere concetti importanti come l'impossibilità del libero arbitrio, ritenendo che una macchina intelligente non abbia un'effettiva libertà di scelta. I detrattori della teoria tendono ad accostare le due intelligenze, sottolineando i limiti di quella artificiale rispetto a quella umana, incappando talvolta nei pregiudizi di cui sopra. Tuttavia, anche la posizione che vede l'IA come debole ha un forte punto di vista umanocentrico, poiché si basa sull'assunto che l'intelligenza artificiale tenda all'imitazione di quella umana, ritenendo quest'ultima il modello a cui aspirare.

Un terzo e ulteriore punto di vista è detto "IA razionale", e considera l'intelligenza come un concetto che ricomprende al suo interno un insieme più ampio di caratteristiche, che afferiscono non solo agli umani, ma anche alle macchine e agli animali.

Searle aveva ideato un esperimento mentale detto "la Stanza Cinese" con cui voleva dimostrare che per quanto una macchina appaia intelligente non sarà mai dotata di coscienza e comprensione. Un computer che "impara" il cinese tramite i caratteri cinesi che ha acquisito tramite input è in grado di produrne altri come output, ma non ha una vera capacità di comprensione e dunque, secondo la teoria di Searle, non ha un pensiero. L'attività del computer potrebbe essere paragonata a quella di una persona in carne ed ossa che riceve dei caratteri cinesi, li elabora seguendo le informazioni che gli sono state fornite, e produce altri caratteri cinesi, pur non capendo nulla di cinese. Searle infatti

sosteneva che la coscienza è frutto di processi neuronali prettamente tipici dell'attività cerebrale umana, a cui una macchina non potrà mai giungere[43].

La semiotica, che studia l'uso dei simboli per il ragionamento e la comunicazione, fa una distinzione tra la sintassi, che ricomprende le regole con cui i simboli vengono organizzati e manipolati, e la semantica, che concerne il significato dei simboli stessi. La sintassi è semplice da capire, la semantica no[44].

C'è chi sostiene che l'esperimento della stanza cinese sia confutabile se si considera ad esempio il caso in cui, invece del cinese, la persona umana debba avere a che fare con il linguaggio macchina. Il protagonista dell'esperimento anche in questo caso seguirebbe le istruzioni senza imparare il linguaggio, ma quel linguaggio, che per lui non significa nulla, è invece comprensibile alla macchina, e ciò dimostrerebbe che in certe situazioni è la macchina ad avere una coscienza, e non l'uomo[45]. Dopotutto, si potrebbe dire che le idee elaborate dal cervello e i byte elaborati dal computer rappresentano entrambe informazioni in forma simbolica (un esempio sono i segnali nervosi che arrivano agli occhi), che vengono processate e infine condivise con l'esterno. Allo stato attuale dell'arte, posto che non si può naturalmente ritenere che umani e macchine si equivalgono, non è stata individuata alcuna dimostrazione per cui i processi decisionali di uomini e macchine obbediscano a diversi principi[46].

Esiste un'ulteriore classificazione che aggiunge alla IA debole e alla IA forte (in questo modello dette "IA ristretta e IA generale") la IA superintelligente, in grado di sviluppare abilità creative e scientifiche, ma anche sociali ed emotive[47].

L'argomento è certamente dibattuto e ricco di implicazioni filosofiche che rimarranno, almeno per ora, irrisolte.

[43] Warwick, *Intelligenza Artificiale - Le basi*.
[44] Kaplan, *Intelligenza artificiale. Guida al futuro prossimo*.
[45] Kevin Warwick, *Intelligenza Artificiale - Le basi* (Dario Flaccovio Editore, 2015).
[46] Maddalena Castellani and Beppe Carrella, *Blockchain: guida pratica tecnico giuridica all'uso* (Firenze: goWare, 2019).
[47] Francesco Corea, *Artificial Intelligence and Exponential Technologies: Business Models Evolution and New Investment Opportunities* (New York, NY: Springer Berlin Heidelberg, 2017).

1.4 Funzionamento

Esistono due tipi di approcci all'intelligenza artificiale:

- top down: è caratterizzato dal fatto che gli stati mentali sono identificati con rappresentazioni di tipo simbolico all'interno di un sistema simbolico-fisico, il ragionamento è frutto di una manipolazione dei simboli. È un approccio efficace quando è necessario affrontare in tempo limitato compiti ben definiti per cui si dispone già di un insieme di regole. Questa viene anche chiamata impostazione funzionale o "comportamentista" e ritiene hardware e l'intelligenza siano separati.
- Bottom up: detto anche approccio connessionista, costruisce modalità di ragionamento più complesse partendo da reti di neuroni artificiali che simulano i neuroni cerebrali. In questo caso il ragionamento è il risultato dell'interconnessione di una molteplicità di semplici unità computazionali, come una sorta di cervello in grado di adattarsi e imparare nel tempo. Tale impostazione viene chiamata anche "strutturale" o "costruttivista".

Gli approcci simbolici sono più comprensibili perché trasparenti, ma non sono in grado di gestire con la stessa facilità degli approcci connessionisti le realtà più incerte e incomplete. Negli ultimi tempi questi ultimi si sono sviluppati raggiungendo risultati piuttosto rilevanti[48].

[48] Warwick, *Intelligenza Artificiale - Le basi*; Massaron, *Intelligenza artificiale for dummies*.

1.4.1 Machine Learning e Deep Learning

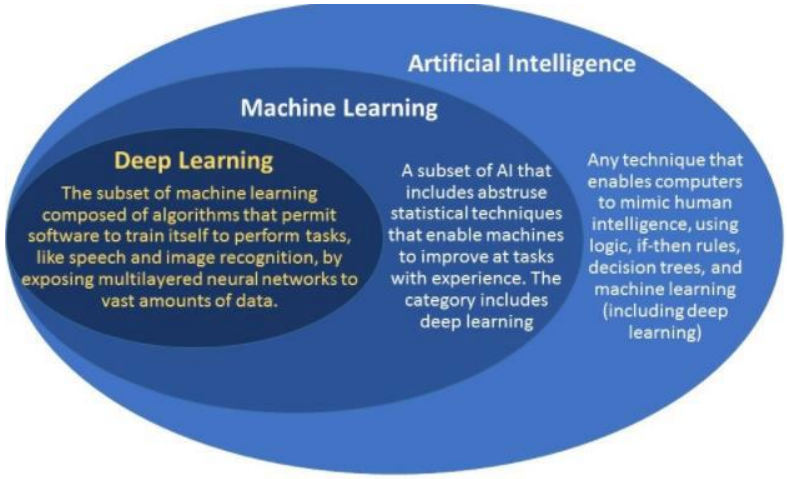


Figura 1. Relazione tra IA, ML, DL[49]

Il Machine Learning, o apprendimento automatico, è definito come "il concetto secondo cui un programma per computer può imparare e adattarsi a nuovi dati senza interferenze umane"[50], dunque è un sottoinsieme dell'IA che si serve dei computer per stimare statisticamente una funzione complessa: si serve di un insieme di tecniche che consentono alle macchine di imparare dai dati che ricevono, estraendone dei pattern, per poi prendere decisioni o fare previsioni su di essi[51].

Caratterizzante il Machine Learning è il tipo di apprendimento, che si classifica secondo tre differenti modelli:

- L'apprendimento con supervisione didattica (Supervised Learning): all'IA vengono fornite delle etichette (soluzioni) dall'addestratore, in modo che possa scoprire le relazioni tra input ed output conosciuti ed elaborare un modello predittivo. Dando al software nozioni specifiche questo è in grado di attingere al suo database di esperienze per risolvere nuovi problemi, simili a quelli conosciuti. Questa tipologia di apprendimento è adatta ai casi in cui i dati di input sono conosciuti, e naturalmente più sono i dati disponibili, più la capacità della macchina di imparare sarà accurata. Le principali applicazioni sono:

[49] AndreaBacciu2018, "Una Panoramica Introduttiva Su Deep Learning e Machine Learning," *DeepLearningItalia*, 25 September 2017, https://www.deeplearningitalia.com/una-panoramica-introduttiva-su-deep-learning-e-machine-learning/.
[50] Rushdi Shams, "Developing Machine Learning Products Better and Faster at Startups," *IEEE Eng. Manag. Rev.* 46.3 (2018): 36–39, https://doi.org/10.1109/EMR.2018.2870669.
[51] Tra le quali le reti neurali artificiali, la teoria dei sistemi dinamici, il riconoscimento di pattern, il data mining, il filtraggio adattivo, la statistica computazionale, l'elaborazione delle immagini, ecc.

    - classificazione: questa tecnica raggruppa gli input in categorie, è usata quando i dati possono essere suddivisi in specifiche categorie, ed è utilizzata soprattutto nel campo del riconoscimento delle immagini e del riconoscimento vocale;
    - regressione: questa studia la relazione tra due o più variabili indipendenti l'una dall'altra.
- L'apprendimento senza supervisione didattica (Unsupervised Learning): i dati di input non vengono codificati, la macchina quindi non può attingere ad esempi. L'apprendimento avviene tramite l'analisi dei risultati, dunque l'IA tramite gli output è in grado di esaminare i risultati dei compiti che è chiamata a svolgere, senza l'ausilio delle etichette, e crea dei propri raggruppamenti. Non ci sono risposte corrette, conosciute a priori, a cui giungere. Il software quindi ha più libertà di scelta, dovendo organizzare le informazioni in modo intelligente senza un ausilio esterno. La tipologia di apprendimento non supervisionato più conosciuta è il "clustering", con il quale dati con caratteristiche simili tra loro vengono raggruppati dal software. Un'altra tipologia è l'associazione, che serve a scoprire regole che descrivono porzioni di dati elevate tramite la determinazione di associazioni, correlazioni che il software individua in una base di dati che gli è stata fornita (si pensi allo studio del comportamento di acquisto di un prodotto, che viene effettuato tramite questa tecnica).
- L'apprendimento per rinforzo (Reinforcement Learning): il software riceve informazioni che si trovano a metà strada tra quelle dell'apprendimento con supervisione e quello senza supervisione, non vengono indicati esempi che indicano il giusto output per un determinato input, ma le informazioni vengono fornite con lo scopo di indicare la correttezza di un'azione. La macchina quindi viene premiata (ecco perché "rinforzo") quando raggiunge gli obiettivi ed è invitata a trovare la soluzione corretta, diventando così capace di distinguere le azioni corrette da quelle errate tramite un meccanismo cosiddetto di "*try and error*", in cui le etichette vengono fornite dopo l'azione. Nel caso in cui vi siano più "hidden units" si parla di Deep Learning, che, come si vedrà a breve, è una sotto-tipologia di Machine Learning che si caratterizza da un apprendimento automatico da parte del software che si articola in più livelli: non solo quindi i tre

livelli di base costituiti da input layer, hidden layer e output layer, ma uno numero molto più elevato di hidden leyer, che talvolta arriva anche a 150.

Si può concludere con il constatare che il Machine Learning è quel metodo che allena l'IA, il Deep Learning invece simula il cervello umano, ed è più precisamente un sub-metodo del Machine Learning che usa "reti neurali profonde" ("*Deep Neural work*"), cioè reti composte da molti strati e da nuovi algoritmi per il pre-processamento dei dati. A differenza del cervello biologico, in cui ciascun neurone può connettersi a qualunque altro (seppur con limiti fisici) e alla cui struttura si ispirano le reti neurali, queste hanno un numero limitato di strati e la propagazione dell'informazione segue una direzione prestabilita. Alla macchina non vengono fornite informazioni, ma vengono apprese tramite algoritmi di calcolo statistico atti a comprendere il funzionamento del cervello biologico e le modalità con cui questo interpreta input di base.

Il Deep Learning viene utilizzato per l'apprendimento supervisionato e per quello non supervisionato, si caratterizza per essere composto da due fasi:

- Fase di formazione: consiste nel trasformare i livelli di base che hanno dati grezzi in livelli superiori, la rete paragona i dati ricevuti e ne impara le caratteristiche e confronta il risultato ottenuto con quello che il set di esempi indicava come corretto. Se il risultato è diverso la rete si corregge cambiando la propria configurazione dei pesi sulle connessioni. Tramite l'algoritmo di *backpropagation*, infatti, la rete assimila una gran moltitudine di esempi svolti, è in grado di capire quanto le sue risposte si discostano dagli esempi e autocorregge il proprio output. La rete compie poi il processo inverso, dalla rete output verso quella di input (ecco spiegato "*backpropagation*") e dopo una serie di questi cicli diventa capace di dare risposte sempre più corrette.
- Test/Training: riguarda l'allenamento del modello, tramite dati di natura differente rispetto a quelli di input.

Il meccanismo non funziona con lo schema "*if…then*", ma la rete è in grado di trovare informazioni nuove elaborando i dati conosciuti con un'elaborazione di tipo statistico in modo da renderli coerenti con l'obiettivo da raggiungere. Tali reti infatti sono funzionali a quei casi in cui non si conoscono tutti i dati, quando è necessaria un'attività "integrativa", caratteristica tipica del cervello biologico.

È bene aprire una parentesi sulle reti neurali, a cui si ispira e su cui si basa il Deep Learning, per comprenderne il funzionamento.

Il neurone costituisce la cellula fondamentale di un cervello biologico, ha la grandezza che va da due a trenta micrometri di diametro, e conta fino a diecimila connessioni, tramite sinapsi, con altri neuroni. Di per sé i neuroni hanno una struttura piuttosto semplice, ciò che garantisce la potenza del cervello umano è la complessità delle connessioni che essi formano. Quando una persona impara, le connessioni si rafforzano o si indeboliscono a seconda della probabilità che l'individuo si comporterà o meno in un certo modo: se infatti costui opera in maniera "corretta" rispetto a un evento i percorsi neuronali coinvolti in quella decisione tenderanno a rafforzarsi, in modo che sia più alta la probabilità di rispondere allo stesso modo in una situazione simile. Ispirati a questi principi, vengono create le reti neurali artificiali (*Artificial Neural Network* – ANN), naturalmente senza la pretesa di ricalcare perfettamente il funzionamento di un cervello biologico[52]. Del cervello biologico, infatti, si conosce molto bene quali regioni di esso sono coinvolte nelle più variegate attività, ma molto poco di come i neuroni si connettano tra loro. Generalmente, gli studi condotti dai ricercatori di intelligenza artificiale si concentrano sul comportamento dei singoli neuroni per poi capirne e studiarne le connessioni[53]. Capire il funzionamento del cervello umano è una tappa fondamentale per lo sviluppo dell' IA forte[54].

Le unità neurali artificiali sono costituite da piccole unità di processamento dati, i cosiddetti neuroni artificiali, in grado di svolgere attività elementari come assumere decisioni in base a regole, inviare risposte agli altri neuroni e ricevere da questi altre informazioni. Una rete artificiale ha una struttura molto più semplice rispetto a quello di un cervello biologico, i neuroni artificiali sono dei numeri il cui valore va da 0 a 1 e sono detti "neuroni" per convenienza, in quanto sono organizzati in una modalità che ricorda, seppur in modo semplicistico, quella di una rete neurale biologica.

In una rete neurale artificiale i neuroni sono organizzati in diversi livelli, ognuno connesso solo al livello superiore e a quello inferiore. Il livello più basso riceve input dall'esterno e quelli più alti (detti "livelli nascosti") ricevono input solo dai neuroni

[52] Warwick, *Intelligenza Artificiale - Le basi*.
[53] Castellani and Carrella, *Blockchain*.
[54] Santosuosso, *Intelligenza artificiale e diritto*.

sottostanti. I neuroni ricevono dati in input, li processano, e inviano le informazioni ai nodi successivi. In questo modo, ripetendo questo ciclo di input-elaborazione-output, il software diventa capace di generalizzare e quindi fornire output corretti che si riferiscono ad input che non fanno parte del training set.

Queste reti neurali sono molto complesse, tuttavia non sono in grado di "imparare", ma sono spiccatamente in gamba nel trovare le correlazioni, ricordandosi le loro vecchie strategie di fronte a input nuovi. È lecito chiedersi se l'essere umano, in fondo, non faccia poi la stessa cosa, o se invece si ponga su uno scalino superiore rispetto alle reti neurali artificiali nel suo modo di imparare e interagire con l'esterno[55].

Tornando ora al Deep Learning, si osserva il suo essere caratterizzato da tre abilità:

- La generalizzabilità: si tratta della precisione con cui la macchina fa una stima su determinati dati che non sono stati ancora formulati
- La "trainability": la rapidità con cui un framework è in grado di lavorare
- L'espressività: il parametro che individua la capacità della macchina di valutare le stime generali

Si caratterizza inoltre per elevati livelli di prestazioni, i modelli si rivelano infatti in grado di essere applicati ad una grande eterogeneità di problemi con ottimi risultati.

Per addestrare un modello di Deep Learning servono training set notevolmente ampi, e questo li rende adatti ad affrontare le sfide derivanti dai Big Data[56]. Questo fa sì che sia necessario una quantità di tempo elevata per la creazione della rete, ecco perché accade spesso che le reti siano sviluppate parallelamente tra loro, tramite il partizionamento del livello su più schede GPU (*Graphic Processing Unit*)[57].

1.5 Stato dell'arte

Lo studio dell'intelligenza artificiale è sempre crescente ed è corroborato dalla consapevolezza della numerosa varietà di ambiti in cui può essere applicato: forme più o

[55] Castellani and Carrella, *Blockchain*.
[56] Kaplan, *Intelligenza artificiale. Guida al futuro prossimo*.
[57] Corea, *Artificial Intelligence and Exponential Technologies*.

meno raffinate di IA si trovano infatti nello sviluppo di assistenti vocali, di sistemi di teleassistenza, di veicoli a guida autonoma, nell'ambito militare, in quello della ricerca, dell'istruzione, della cura personale, della medicina, del consumo energetico, dell'ambiente, e in molti altri.

Rispetto alla rivoluzione informatica la rivoluzione che verrà innescata dallo sviluppo dell'intelligenza artificiale e della miniaturizzazione degli automi sarà più profonda e pervasiva, potendo i robot occupare sempre più spazi fisici, dalle nostre case alle industrie, dagli aerei alla chirurgia. Naturalmente un cambiamento di tale portata trascinerà con sé l'esigenza di risposte di natura giuridica, che come spesso accade di norma seguono lo sviluppo tecnologico, ma il diritto ha un ruolo fondamentale in quanto contribuisce a disegnare le modalità d'uso delle nuove tecnologie.

L'IA ha sviluppato abilità che da sempre erano considerate caratteristiche inerenti alla sola natura umana, come ad esempio l'abilità nel gioco degli scacchi. Tuttavia le macchine ancora non hanno sviluppato una completa abilità nel ragionamento del buon senso o nel comprendere il linguaggio naturale. Secondo il professor Joshua Greene, che insegna psicologia ad Harvard, la capacità che resta ancora prettamente umana è quella di immaginare, considerare qualunque tipo di idea e saper distinguere ciò che è vero da ciò che è finzione[58].

Fin dagli albori uno dei timori più radicati è quello che l'IA un giorno diventi autonoma e superi l'intelligenza umana, ma al momento non è una questione che possa destare un'oggettiva preoccupazione, perché per lo sviluppo dell'apprendimento automatico servirebbe una quantità molto ampia di dati che non può prescindere dall'apporto umano. Ulteriormente, la potenza di calcolo delle macchine attuali non è paragonabile a quella umana, e per quanto riguarda i computer quantistici, che potrebbero segnare una effettiva rivoluzione, bisognerà aspettare ancora molto tempo affinché siano pronti all'uso quotidiano.

L'intelligenza artificiale ha comunque già un notevole stuolo di ambiti applicativi, , viene utilizzata praticamente in tutti i domini sociali, elenchiamo qualche esempio. In campo medico e sanitario l'impatto e il ruolo di tale tecnologia è formidabile, per esempio per la diagnosi. In ambito lavorativo, volendo citare solo pochi casi lampanti, viene usata nel

[58] Santosuosso, *Intelligenza artificiale e diritto*.

marketing per meglio analizzare il comportamento dei consumatori ed è di supporto nel supply chain management per ottimizzare la catena di approvvigionamento e distribuzione. Nel mondo del banking e della pubblica sicurezza la gestione del rischio viene spesso affidata agli algoritmi. Anche nel campo della ricerca ed istruzione, così come quello dell'intrattenimento le applicazioni sono veramente infinite [59].

Un approccio affermatosi negli ultimi anni è il cosiddetto "*IA for good*", che spinge sull'idea della neutralità della tecnologia, e che questa possa essere utilizzata per risolvere numerose questioni. L'IA può essere infatti un ottimo strumento per promuovere le potenzialità dell'essere umano e creare quindi opportunità, permettendo all'uomo di vivere in modo più intelligente, oppure può anche essere sottoutilizzato, o ancor peggio può essere abusato. Naturalmente una tecnologia così potenzialmente dirompente comporta rischi proporzionati ai benefici. Inquadrare lo sviluppo dell'IA in una logica di responsabilità è un passo essenziale per una società equa[60].

[59] Consoft Sistemi, "L'intelligenza Artificiale al Servizio Dell'uomo," 2019, https://www.cospe.org/wp-content/uploads/2019/07/03_dossier_INTELLIGENZA-ARTIFICIALE_080719-1.pdf; Patrick Lin, Keith Abney, and George Bekey, "Robot Ethics: Mapping the Issues for a Mechanized World," *Artificial Intelligence* 175.5–6 (2011): 942–49, https://doi.org/10.1016/j.artint.2010.11.026; Christopher Burr, Mariarosaria Taddeo, and Luciano Floridi, "The Ethics of Digital Well-Being: A Thematic Review," *Sci Eng Ethics* 26.4 (2020): 2313–43, https://doi.org/10.1007/s11948-020-00175-8.

[60] Luciano Floridi et al., "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," *Minds & Machines* 28.4 (2018): 689–707, https://doi.org/10.1007/s11023-018-9482-5.

Capitolo 2

Un'intelligenza artificiale affidabile per le persone

La diffusione degli algoritmi nella società ha incrementato la necessità di rispondere a questioni fondamentali. Le informazioni e le tecnologie riguardanti le comunicazioni hanno permeato ormai totalmente il nostro vivere quotidiano. Basti pensare che già nel 2011 la popolazione presente in internet ha raggiunto i 2,1 miliardi di persone, ossia il 30 % della popolazione mondiale. Nello stesso anno le ore spese online sono state 35 miliardi, la maggior parte di queste dedicata alle interazioni[61]. Nella realtà di questo mondo i dati la fanno da padrona e sono utilizzati dagli algoritmi per estrarre informazioni. In questo modo si possono prendere decisioni informate e anche completamente automatizzate. Le potenzialità che ne possono derivare sono enormi, ma ovviamente il dibattito pubblico è aperto[62].

Come in seguito all'avvento di qualsiasi tecnologia, la costante diffusione dell'intelligenza artificiale ha visto contrapporsi due differenti correnti di pensiero: quella degli apocalittici e quella degli integrati. Gli apocalittici hanno il timore che l'umanità diventerà succube di tale tecnologia, che l'intera vita sarà condizionata dall'IA che ruberà loro il lavoro, deciderà per loro e ne violerà la privacy. Gli integrati invece hanno una linea di pensiero opposta, credono che l'avvento dell'IA sia inevitabile, ma che questo possa essere di grande aiuto all'essere umano, sgravandolo dei compiti più difficili e amplificando le sue migliori qualità[63]. I due punti di vista sono estremi, ognuno ha la propria peculiarità, ma devono essere entrambi compresi per rispondere adeguatamente a questa sfida etica e far sì che l'innovazione possa continuare senza particolari intoppi. L'IA è ormai utilizzata in qualsiasi campo, da quello sociale, a quello sanitario, ma anche giudiziario, la crescente automazione del processo decisionale impone la necessità di riflettere adeguatamente sui risvolti etici e di governance[64].

[61] Mariarosaria Taddeo, "Cyber Security and Individual Rights, Striking the Right Balance," *Philos. Technol.* 26.4 (2013): 353–56, https://doi.org/10.1007/s13347-013-0140-9.
[62] S. C. Olhede and P. J. Wolfe, "The Growing Ubiquity of Algorithms in Society: Implications, Impacts and Innovations," *Phil. Trans. R. Soc. A.* 376.2128 (2018): 20170364, https://doi.org/10.1098/rsta.2017.0364.
[63] AGID (agenzia per l'Italia digitale), "Libro Bianco Sull'Intelligenza Artificiale al Servizio Del Cittadino," 2018, https://www.agid.gov.it/it/argomenti/intelligenza-artificiale.
[64] AGID (agenzia per l'Italia digitale), "Libro Bianco Sull'Intelligenza Artificiale al Servizio Del Cittadino."

È quindi auspicabile arrivare alla creazione di un'IA affidabile. Per questo possiamo considerare l'intelligenza artificiale come composta da tre componenti, delle macroaree che vanno analizzate e definite, se vogliamo raggiungere il nostro obbiettivo.
L'intelligenza artificiale deve essere[65]:

- lecita: non deve essere contraria a legge e regolamenti in vigore;
- etica: deve essere rispettosa di principi etici e valori globalmente riconosciuti;
- robusta: deve essere robusta da una prospettiva tecnica, ma anche sociale.

L'intelligenza artificiale può avere un enorme impatto positivo sulla società, nell'aiutare ad eliminare le disuguaglianze e nella lotta climatica. Per massimizzare i benefici e minimizzare i rischi che ne possono conseguire è bene impostare un sistema di intelligenza artificiale che sia human-centric, al servizio quindi dell'umanità, aumentandone il benessere e le libertà[66]. La fiducia è un prerequisito, ossia un elemento essenziale in questo gioco[67]. Essa facilita le relazioni all'interno di un sistema i cui agenti possono essere umani, artificiali o ibridi. Più le tecnologie si evolvono e più noi ci affidiamo ad esse, sfruttandone il valore[68]. Il motivo per il quale ci si si affida è quello di ottenere un qualche vantaggio, è una scelta conveniente con la quale si delega un qualsiasi compito. La fiducia è alla base di una società e così come regola le interazioni tra le persone, così dovrebbe valere anche nel mondo "digitale" [69].

Anche secondo Floridi, quando si parla di digitale, bisogna considerare tre diversi punti di vista: la governance del digitale, l'etica del digitale e il regolamento del digitale. Queste tre prospettive, che ben si allineano con quelle appena viste, sono complementari tra loro, ma anche interdipendenti.

[65] AI HLEG (High-Level Expert Group on Artificial Intelligence), "Ethics Guidelines for Trustworthy AI," 8 April 2019, Commissione Europea.
[66] AI HLEG (High-Level Expert Group on Artificial Intelligence), "Ethics Guidelines for Trustworthy AI."
[67] Commissione Europea, "AI Ethics Communication," 8 April 2019.
[68] Mariarosaria Taddeo, "Trusting Digital Technologies Correctly," *Minds & Machines* 27.4 (2017): 565–68, https://doi.org/10.1007/s11023-017-9450-5.
[69] Mariarosaria Taddeo, "Modelling Trust in Artificial Agents, A First Step Toward the Analysis of e-Trust," *Minds & Machines* 20.2 (2010): 243–57, https://doi.org/10.1007/s11023-010-9201-3.

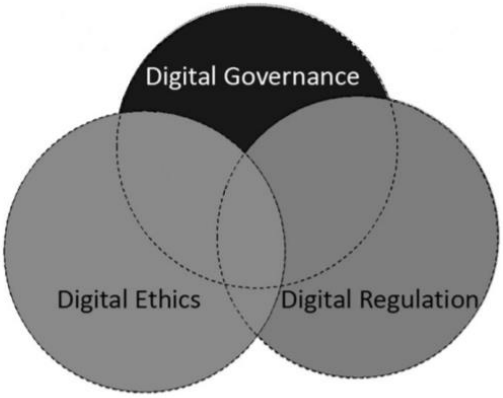


Figura 1. Approcci nella gestione del digitale[70]

Se fin qui la governance poteva essere più ampliamente intesa, ora assume un significato più specifico. Per governance si intendono le attività con cui si attuano le politiche, tutte quelle pratiche utili per un adeguato sviluppo. Sono comprese anche le linee guida e le raccomandazioni attraverso cui le istituzioni governative regolano il comportamento degli interessati[71]. Nel successivo paragrafo proprio di questo ci occuperemo.

L'etica digitale è quella branca dell'etica che si occupa di dati e informazioni, algoritmi e pratiche quali la programmazione e gli standard, e che plasma la regolamentazione e la governance[72].

Nel settore di riferimento un quadro normativo non è ancora venuto alla luce, e al momento, come vedremo, si fa affidamento alla governance e ai principi morali. La governance può indicare una giusta direzione da seguire, il regolamento indica ciò che è illegale e ciò che è legale, mentre l'etica fornisce informazioni circa le migliori azioni da compiere per ottenere una società migliore. Andremo ora quindi a vedere più nel dettaglio queste prospettive, tentando il più possibile di operare una distinzione, ma rimanendo coscienti che a volte le componenti possono essere interrelate.

[70] Luciano Floridi, "Soft Ethics and the Governance of the Digital," *Philos. Technol.* 31.1 (2018): 1–8, https://doi.org/10.1007/s13347-018-0303-9.
[71] Floridi, "Soft Ethics and the Governance of the Digital."
[72] Roberto Pasca di Magliano, "Etica e innovazione nella governance pubblica" Policy Paper.1 (2021): 16; Floridi, "Soft Ethics and the Governance of the Digital."

2.1 Governance e indirizzo strategico internazionale

Il coinvolgimento delle autorità pubbliche nella governance di questo campo ha un costo, ma ormai è diventato inevitabile. Analizzare il trade-off tra libertà e regolamentazioni è fondamentale, in quanto in questo caso il potere delle autorità non servirebbe a limitare le libertà individuali, ma a tutelarle. Il "potere della legge" serve infatti a proteggere diritti come la privacy, l'anonimato, la trasparenza, poiché le tecnologie stanno dando nuova forma alla vita per come era conosciuta finora[73]. Poiché gli algoritmi sono un importante mezzo di regolamentazione ed esercizio del potere la loro governance è essenziale[74]. Il rapporto tra algoritmi, leggi e diritti umani è sicuramente complesso, le principali domande alle quali si sta tutt'oggi cercando di trovare una risposta riguardano il livello di governance che le istituzioni dovrebbero esercitare, cioè fino a che punto questo deve estendersi e in che forme. Diamo per assodato che gli algoritmi ormai governino le nostre vite e che una corretta regolamentazione aumenterebbe la percezione di controllo e sicurezza che le persone necessitano, se fosse eccessivo il rischio sarebbe quello di soffocare possibili sviluppi e benefici futuri dell'implementazione dell'intelligenza artificiale.

Fuori da specifici settori l'approccio della legge, e più in generale delle regolamentazioni, è quello di permettere la libera innovazione, ma quando si parla di intelligenza artificiale vi sono alcune persone talmente tanto preoccupate dalla possibilità di possibili danni che chiedono a gran voce l'istituzione di un'autorità di regolamentazione centrale. I motivi a svantaggio di tale soluzione riguardano i rischi da mitigare che, essendo ancora sconosciuti o inesistenti, non sono controllabili da un ente. Inoltre, almeno tendenzialmente, i legislatori non hanno successo nella legislazione quando questa riguarda il futuro, spesso a causa della loro inesperienza in materia. Infine, i campi di applicazione attuali, per non nominare il potenziale futuro, dell'intelligenza artificiale sono talmente ampi che una regolamentazione completa sarebbe complicata, dovendo regolare vari aspetti[75]. Una strategia ottimale sarebbe quella di affrontare il problema in maniera incrementale. Con il progressivo sviluppo e la conseguente comparsa di rischi,

[73] Taddeo, "Cyber Security and Individual Rights, Striking the Right Balance."
[74] Mariarosaria Taddeo, "The Struggle Between Liberties and Authorities in the Information Age," *Sci Eng Ethics* 21.5 (2015): 1125–38, https://doi.org/10.1007/s11948-014-9586-0.
[75] Chris Reed, "How Should We Regulate Artificial Intelligence?," *Phil. Trans. R. Soc. A.* 376.2128 (2018): 20170360, https://doi.org/10.1098/rsta.2017.0360.

se necessario sarà pensato un regolamento specifico. Quello che si può fare è continuare nella ricerca di tali rischi. Ad oggi le minacce maggiori riguardano i diritti fondamentali, infatti le decisioni prese dall'intelligenza artificiale non si basano sulla legge stessa, ma sui dati, aprendo a possibili discriminazioni per motivi di sesso, razza, religione. Si potrebbe ovviare al problema richiedendo che tali strumenti, potenzialmente dannosi, diano spiegazioni circa le ragioni sulla base delle quali hanno preso le loro decisioni[76]. Questo ovviamente non è sempre possibile o facile, poiché spesso gli algoritmi operano come black-box e la spiegabilità è uno dei problemi che scaturisce.

Delle linee guida sono quindi necessarie e le organizzazioni statali e mondiali hanno già da anni iniziato a muovere i primi passi in questa direzione, fornendo al momento solo una loro visione, neppure completa. Analizzeremo brevemente il panorama internazionale e la strategia che diversi soggetti stanno improntando.

Il contributo che l'IA può dare alla società è indubbio. Se si vuole far sì che l'intelligenza artificiale abbia la possibilità di diventare un bene pubblico, è necessario indirizzare il settore con una vision e una strategia ben chiare. Negli ultimi anni questo ruolo è stato svolto dal settore privato e accademico, ma per svelare tutto il potenziale e condividere benefici ed opportunità un regolamento sarebbe auspicabile[77]. Il quadro che andremo a delineare riguarderà quindi l'orientamento socio-politico, ma anche l'approccio strategico, che questi Paesi cercano di perseguire, disegnando le relazioni e l'interesse dei vari stakeholder[78].

Unione Europea

Nel 2016 l'Unione Europea ha emanato un regolamento per la trattazione dei dati, il GDPR. Nello stesso anno è stato rilasciato un documento intitolato "Civil Law Rules on Robotics" che contiene un mix di leggi e principi con il quale si cerca di indirizzare tutti i paesi membri lungo un approccio comune. Tale documento contiene anche molti riferimenti etici e sociali che analizzeremo meglio in seguito, ma risulta non essere

[76] Reed, "How Should We Regulate Artificial Intelligence?"
[77] Huw Roberts et al., "Achieving a 'Good AI Society': Comparing the Aims and Progress of the EU and the US," *Sci Eng Ethics* 27.6 (2021): 68, https://doi.org/10.1007/s11948-021-00340-7.
[78] Corinne Cath et al., "Artificial Intelligence and the 'Good Society': The US, EU, and UK Approach," *Sci Eng Ethics* (2017), https://doi.org/10.1007/s11948-017-9901-7.

sufficiente. Nel 2018 25 Paesi hanno firmato la Declaration of Cooperation on Artifcial Intelligence ed è stato formato un gruppo di esperti: l'High-Level Expert Group on AI (HLEG) il quale ha pubblicato delle linee guida[79] e alle quali gli studiosi hanno dedicato particolare attenzione. L'approccio dell'Unione Europea è stato reso noto tramite una comunicazione da parte della Commissione e consiste nel:

- promuovere lo sviluppo pubblico e privato della tecnologia e dell'industria;
- preparare il mercato e gli attori che occupano della formazione a questo cambiamento;
- definire un adeguato framework etico e legale.

Recentemente la stessa Commissione ha proposto un approccio di regolamentazione basato sul rischio[80]: inaccettabile, ad alto rischio, rischio limitato e rischio minimo/assente. Categorizzando quindi le varie attività secondo questo principio, i requisiti che si dovrebbero rispettare cambierebbero da caso a caso diventando più stringenti, fino a proibire quei sistemi il cui rischio è inaccettabile. La politica domestica inizialmente considerava solo la robotica, ora comprende una più ampia serie di discipline. Grande importanza viene riservata ai diritti individuali con il fine di promuovere il benessere collettivo, e in questa sfumatura forse si differenzia l'approccio europeo da quello americano. Anche a livello internazionale l'Unione Europea promuove la cooperazione e la creazione di misure che possano diventare degli standard globali[81].

Che l'Europa propenda per un approccio regolamentativo forte lo si può vedere anche con l'adozione del General Data Protection Regulation (GDPR), nel 2016, un vero e proprio regolamento per la collezione, la conservazione e l'uso delle informazioni personali. In particolare, il GDPR contiene specifiche per la trasparenza e l'equità, obblighi per la responsabilità, specifiche legali per il processamento dei dati, il diritto degli individui per opporsi[82]. La portata del GDPR è globale, poiché riguarda qualsiasi azienda che processa dati di residenti in Europa. Lo scopo è stato dunque quello di

[79] AI HLEG (High-Level Expert Group on Artificial Intelligence), "Ethics Guidelines for Trustworthy AI."
[80] Europena Commission, "Europe Fit for the Digital Age: Artificial Intelligence," 2021, https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682.
[81] Roberts et al., "Achieving a 'Good AI Society.'"
[82] Christina Blacklaws, "Algorithms: Transparency and Accountability," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2128 (2018): 20170351, https://doi.org/10.1098/rsta.2017.0351.

settare degli standard. Con regolamentazioni più stringenti anche in ambito IA l'Europa cerca quindi recuperare la propria sovranità digitale.

Stati Uniti d'America

Il White Office of Science and Technology nel 2016 ha rilasciato dei report per la prima volta specifici sull'intelligenza artificiale: 'Preparing for the Future of Artificial Intelligence[83]' e il 'National Artificial Intelligence Research and Development Strategic Plan[84]' (aggiornato poi nel 2019). I report definiscono il possibile ruolo del governo nel facilitare il progresso in questo campo e nel tentare di porre dei limiti regolatori minimi che fungano da guida per gli investimenti in ricerca e sviluppo. Secondo Cath, applicare i framework esistenti a questi nuovi problemi è inadeguato[85]. L'amministrazione Trump ha proseguito con una politica completamente liberale fino al 2020, quando sono state proposte delle linee guida che si basano su tre principi: l'eccesso di regolamenti limitanti, il crescente ingaggio del pubblico e la promozione della fiducia (trustworthy). Recentemente questa visione che mira a far primeggiare gli Stati Uniti nel campo dell' R&D è stata tradotta in legge. Se da un punto di vista domestico il governo ha deciso in sostanza di non intromettersi e lasciare la più totale libertà, da quello internazionale la situazione cambia. Gli USA hanno promosso un ambiente internazionale aperto così da favorire le proprie industrie, questa visione liberale comprende sfumature mercantili poiché l'intento è quello di generare un certo controllo[86]. Rispetto alla visione Europea c'è sicuramente una minor preoccupazione per i possibili rischi, comunque nominati tra le linee guida, ma sicuramente una forte attenzione a non farsi scappare questa ghiotta possibilità[87]. Il governo degli Stati Uniti è quindi completamente orientato all'innovazione, con applicazioni anche per il bene pubblico, e alla crescita economica.

[83] National Science and Technology Council, "Preparing for the Future of AI," October 2016, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.
[84] National Science and Technology Council, "The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update" (2019): 50.
[85] Cath et al., "Artificial Intelligence and the 'Good Society.'"
[86] Roberts et al., "Achieving a 'Good AI Society.'"
[87] Cath et al., "Artificial Intelligence and the 'Good Society.'"

Questo però ovviamente andrà a favorire il settore industriale e le applicazioni commerciali[88].

Regno Unito

L'House of Commons' Science and Technology Committee, sempre nel 2016, ha rilasciato il report britannico sull'IA. Si può dedurre dai toni una certa urgenza insieme all'esortazione ad attivarsi per individuare potenziali applicazioni così come i possibili rischi. Viene suggerita la creazione di una commissione con l'obiettivo di aiutare a sviluppare un quadro normativo. Gli aspetti da tenere in considerazione non sarebbero solo quelli legali, ma anche sociali ed etici. In questo senso tale commissione avrebbe come obiettivo quello di dare ascolto e avvicinare gli interessi dei vari stakeholder[89]. Viene data molta importanza all'ambito accademico, che ha lo scopo primario di svolgere ricerca e ridurre il gap di conoscenze con il mondo industriale. Il Regno Unito è stato in passato un grande polo di ricerca, soprattutto in ambito IA, e a Londra vi è infatti una start-up, comprata da Google, che proprio di questo si occupa. Parte delle preoccupazioni sono dovute alla Brexit e agli effetti negativi che ha sui fondi per la ricerca. Proprio per questo il Regno Unito, così come gli Stati Uniti, promuovono una regolamentazione leggera, per non penalizzare l'innovazione ulteriormente[90].

I vari report hanno molti punti in comune, riguardano soprattutto la governane dell'intelligenza artificiale e trattano solo in modo superficiale la tematica della normativa. Tutte le visioni sottolineano l'importanza del dialogo a livello internazionale. Gli Stati Uniti solo nel maggio del 2020 sono entrati nel GPAI, Global Partnership on AI, prima erano timorosi che l'attenzione incanalata sulle questioni etiche potesse frenare l'innovazione. Questo blocco di alleanze, confermato anche dall' adesione ai principi etici dell'Organizzazione per la cooperazione e lo sviluppo economico, l'OECD, si trova principalmente a dover fronteggiare la minaccia della Cina. Gli Stati Uniti hanno

[88] Roberts et al., "Achieving a 'Good AI Society.'"
[89] D Majumdar and H K Chattopadhyay, "AI and Human Rights: From Business and Policy Perspectives" (n.d.): 10.
[90] Cath et al., "Artificial Intelligence and the 'Good Society.'"

ricominciato a puntare sulla collaborazione internazionale in particolare dall'inaugurazione di Biden come presidente.

Questa spinta alla collaborazione non deve intendersi come un completo allineamento di intenti da parte dell'Unione Europea e degli Stati Uniti. Poiché i principi etici forniscono solo piccole indicazioni circa lo sviluppo dell'intelligenza artificiale, gli approcci ideologici possono rimanere diversi, con entrambe le superpotenze che stanno cercando di esportare il proprio. Il desiderio di sovranità digitale dell'Ue risulta essere in netto contrasto con la politica completamente liberale che gli Usa hanno per favorire le loro immense imprese. Questo significa che sarà poco chiaro l'utilizzo che si farà dell'intelligenza artificiale, non tanto se l'IA dovrà essere affidabile o meno[91]. Sicuramente l'approccio dell'UE è eticamente superiore, con l'obiettivo di creare quella che viene definita una "good AI society". Tale volontà deve trovare anche riscontro in un approccio pratico e non rimanere solo su un piano teorico.

E l'Italia? Nel 2021 con l'approvazione del Consiglio dei ministri ha adottato il Programma Strategico per l'Intelligenza artificiale per gli anni 2022-2024. Tale programma è in linea con la strategia europea e prevede una serie di politiche da adottare per rafforzare questo settore in Italia. Anche in questo caso le aree prioritarie di intervento sono lo sviluppo di competenza tramite le università, la ricerca e le applicazioni sia nel settore industriale che in quello della pubblica amministrazione. Un gruppo di lavoro permanente lavorerà fianco a fianco con la Commissione per la Transizione Digitale[92].

Una riflessione finale si può fare sul fatto che le sfere operative per qualsiasi Paese rimangono tre, e sono quella corporativa, il settore pubblico e quello accademico, ognuno con i propri compiti o funzioni. La confusione nasce quando questi attori vanno ad interagire in un'altra sfera o quando vengono adottate pratiche e codici etici di

[91] Roberts et al., "Achieving a 'Good AI Society.'"
[92] "Artificial Intelligence: Italy Launches National Strategy," *Connessioni - Bridging Worlds*, 7 December 2021, https://www.connessioni.biz/en/artificial-intelligence-italy-launches-national-strategy/.

un'altra sfera. Il digitale ha avvicinato queste sfere eliminando le barriere e quindi rendendo meno definiti i confini operativi[93].

2.2 Etica

Dopo aver visto brevemente le strategie a livello nazionale e sentito come queste principalmente siano orientate a un rispetto dei principi etici, vediamo in cosa questi consistono.

La rivoluzione digitale ha trasformato le nostre vite e le tecnologie sottostanti sono onnipresenti. Adottare un approccio etico per quanto concerne la tecnologia dell'intelligenza artificiale è utile per mitigare i danni e al tempo stesso incorporarne i benefici. Questo è un duplice vantaggio dovuto al valore sociale che l'IA è in grado di abilitare[94]. La società verso cui ci stiamo dirigendo sarà sempre maggiormente dominata dagli algoritmi, ritenendo che questi siano potenti e infallibili[95]. È necessario considerare forzatamente le ripercussioni di tali tecnologie su morale ed etica, e per non incorrere in una dittatura basata su asimmetrie informative la politica deve continuare sulla strada intrapresa[96]. La conformità alla legge, in questo campo non ancora completamente formatasi, è necessaria, ma non sufficiente perché il ruolo dell'etica è altrettanto fondamentale.

In questi anni è sempre più frequente il fenomeno della filantropia riguardante i dati, ossia la donazione di dati da parte di aziende private. Compito delle organizzazioni internazionali è quello di creare le infrastrutture adeguate a coordinare e gestire tali dati[97]. La governance di tali dati pone le medesime sfide e problemi etici che andremo ad analizzare in seguito. Ma la rivoluzione informatica ha svelato come comportamenti

---

[93] Peter Grindrod, "Beyond Privacy and Exposure: Ethical Issues within Citizen-Facing Analytics," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160132, https://doi.org/10.1098/rsta.2016.0132.

[94] Luciano Floridi et al., "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," *Minds & Machines* 28.4 (2018): 689–707, https://doi.org/10.1007/s11023-018-9482-5.

[95] Marcello Pistilli, "Etica e algoritmi. Verso l'algoretica," *innovationgym*, 21 October 2019, https://www.innovationgym.org/en/etica-e-algoritmi-verso-lalgoretica/.

[96] Pistilli, "Etica e algoritmi. Verso l'algoretica."

[97] Mariarosaria Taddeo, "Data Philanthropy and Individual Rights," *Minds & Machines* 27.1 (2017): 1–5, https://doi.org/10.1007/s11023-017-9429-2.

moralmente buoni siano il risultato di valori morali e di una infrastruttura etica[98]. Alcuni principi "infraetici" – neologismo introdotto da Floridi col quale ci si riferisce ad un framework non completamente etico che facilita le decisioni morali - alla base delle relazioni umane sono la fiducia, il rispetto e la lealtà. Questi sono elementi che non garantiscono un comportamento moralmente buono, ma che lo promuovono, aiutando quindi le persone di una società a raggiungere i propri obiettivi. Trasponendo tale analisi alla società dell'informazione matura in cui viviamo, possiamo trarne che principi infraetici sono la fiducia, la sicurezza, la trasparenza e la filantropia dei dati stessi. La filantropia dei dati permette di trasferire conoscenza e comprensione, migliorando la governance e promuovendo una società più aperta e pluralista[99]. Questi principi infraetici sono onnipresenti nella letteratura, spesso proposti all'interno di framework diversi, noi li andremo ora ad analizzare poiché in ogni caso rappresentano dei componenti essenziali dell'intelligenza artificiale all'interno della società.

Volendo precisare quanto detto sopra, la società dell'informazione è ormai superata, poiché la società odierna, e la stessa intelligenza artificiale, si basano sui dati. L'etica dei dati ha come obiettivo quello di valutare i problemi morali derivanti dalla trattazione dei dati, degli algoritmi e dalle pratiche correlate. Questo cambiamento di prospettiva non è nuovo, infatti il livello di astrazione (LoA) affrontato dall'etica è variato spesso nel corso degli anni in seguito al rapido e diffuso progresso tecnologico. Nel campo dell'etica i primi studi avevano un LoA in cui l'uomo ne era il protagonista, si è poi passati a porgere maggior attenzione ai computer, ossia ai mezzi, per arrivare alle informazioni, cioè il contenuto, fino ai giorni nostri in cui il focus è incentrato sui dati[100]. Quest'ultimo cambiamento secondo Floridi è più semantico che concettuale, evidenziando però come l'attenzione sia stata spostata su problemi etici riguardanti la raccolta e la gestione di grandi quantità di dati. Seppur i dati, gli algoritmi e le pratiche, tra cui la programmazione e l'innovazione, siano campi di ricerca distinti tra loro, sono altrettanto sovrapponibili secondo uno specifico punto di vista, cioè quello etico. Sarebbe impossibile parlare di elementi quali la fiducia, la trasparenza, la privacy poiché

[98] Luciano Floridi, *The 4th Revolution: How the Infosphere Is Reshaping Human Reality*, First edition. (New York ; Oxford: Oxford University Press, 2014).
[99] Taddeo, "Data Philanthropy and Individual Rights."
[100] Luciano Floridi and Mariarosaria Taddeo, "What Is Data Ethics?," *Phil. Trans. R. Soc. A.* 374.2083 (2016): 20160360, https://doi.org/10.1098/rsta.2016.0360.

direttamente interconnessi a tutti e tre i diversi campi[101]. Seppur il livello di astrazione sia cambiato è importante notare come questi temi siano rimasti rilevanti, li andremo quindi a osservare più nel dettaglio, facendo riferimento nello specifico all'intelligenza artificiale, pur ben sapendo che quanto detto ha valenza anche per i dati.

L'etica va nella stessa direzione della legge, alcuni elementi sono codificati tramite essa come accennato nel precedente paragrafo, mentre altri sono lasciati alla libera morale[102].

2.2.1 Diritti fondamentali

L'intelligenza artificiale dovrebbe rispettare gli stessi diritti alla base dell'essere umano, ossia i diritti fondamentali. Se ci si dovesse chiedere quale possa essere il ruolo dei diritti umani nella governance dell'AI, ormai sono vari gli studi che hanno confermato come questi possano mitigare i rischi dell'intelligenza artificiale. I diritti umani possono guidare nel design fornendo le fondamenta sulle quali creare. Qualsiasi progetto tecnologico non dovrebbe prescindere dal chiedersi quale potrebbe essere il suo impatto sugli esseri umani[103]. I diritti umani trovano anche espressione nella legge, cosa che ne sottolinea l'importanza.

La fonte dei diritti fondamentali è internazionale e risiede nella Dichiarazione Universale dei Diritti Umani del 1948 (UDHR), questi poi vengono ribaditi con il Patto internazionale di New York sui diritti civili e politici del 1966 (ICCPR) e il Patto internazionale sui diritti economici, sociali e culturali sempre del 1966 (ICESCR)[104]. I diritti fondamentali non hanno confini, poiché vengono ripresi da tutti gli Stati o comunità.

Alcuni dei diritti e principi previsti e protetti da queste fonti sono[105]:

---

[101] Floridi and Taddeo, "What Is Data Ethics?"
[102] Cat Drew, "Data Science Ethics in Government," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160119, https://doi.org/10.1098/rsta.2016.0119.
[103] Mark Latonero and Aaina Agarwal, "Human Rights Impact Assessments for AI: Learning from Facebook's Failure in Myanmar," Carr Center for Human Rights Policy (2021): 18.
[104] Effy Vayena and John Tasioulas, "The Dynamics of Big Data and Human Rights: The Case of Scientific Research," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160129, https://doi.org/10.1098/rsta.2016.0129.
[105] Maria Stefania Cataleta, "The Fragility of Human Rights Facing AI," *Humane Artificial Intelligence* Working Paper No.02 (n.d.): 33.

- principio della dignità umana, con il quale si intende il valore intrinseco di ogni individuo in quanto essere umano, il diritto al lavoro e ad un adeguato tenore di vita possono essere qui ricompresi;
- il diritto alla privacy e alla protezione dei dati (in campo europeo esiste il GDPR), che se di per sé è già ritenuto un diritto fondamentale, con l'avvento delle moderne tecnologie lo è ancora di più;
- il diritto alla libertà in generale e in particolare alla libertà di espressione e di movimento, in tale diritto si può ritrovare anche il diritto alla partecipazione politica e all'autodeterminazione[106].

Tali valori etico-sociali potrebbero essere classificati a seconda del livello a cui fanno riferimento, da quello individuale, a quello sociale e globale[107].

2.2.2 Principi etici

Numerose sono state le organizzazioni che hanno analizzato le implicazioni etiche dell'intelligenza artificiale. I ricercatori di AI5People nel corso del loro studio hanno cercato di considerare tutti i più importanti principi presenti in letteratura per dedurne quelli più rilevanti. Osservando i 47 principi totali e confrontandoli è apparso ci fossero delle somiglianze, in particolare con i quattro principi, pilastri nella bioetica: beneficenza, non maleficenza, autonomia, giustizia. Questi quattro principi ben si adattano alle sfide etiche create dall'intelligenza artificiale, ma vi è la necessità di introdurre un quinto principio, la spiegabilità[108].

Anche secondo le linee guida espresse dal gruppo di esperti della Commissione Europea (High-Level Expert Group on Artificial Intelligence) i principi etici da rispettare sono: autonomia umana, prevenzione del danno, giustizia/equità, spiegabilità. Tra i principi espressi nei due casi vi è una perfetta sovrapposizione, se non fosse che in quest'ultimo

[106] Majumdar and Chattopadhyay, "AI and Human Rights: From Business and Policy Perspectives"; Stefano Quintarelli et al., "AI: profili etici Una prospettiva etica sull'Intelligenza Artificiale: princìpi, diritti e raccomandazioni."3 (n.d.): 22.
[107] Quintarelli et al., "AI: profili etici Una prospettiva etica sull'Intelligenza Artificiale: princìpi, diritti e raccomandazioni."
[108] Teresa Scantamburlo, "Non-Empirical Problems in Fair Machine Learning," *Ethics Inf Technol* 23.4 (2021): 703–12, https://doi.org/10.1007/s10676-021-09608-9.

manca il principio della beneficenza, esplicato in diverso modo, ma comunque non ignorato[109].

Cerchiamo di vedere in cosa sussistono tali principi.

- Beneficenza: si intende la promozione del benessere di tutte le creature senzienti. L'intelligenza artificiale deve essere quindi organizzata con priorità il benessere umano, preservandone la dignità e la prosperità. In tale principio possiamo inserire anche la necessità che l'IA sia human-centric. Alcuni studiosi hanno generalizzato tale principio, portandolo ad uno step superiore con il concetto più ampio di sostenibilità[110].

- Non maleficenza: sebbene sembri semplicemente l'opposto del principio precedente, non si intende semplicemente "non fare del male". I rischi negativi derivanti da una negligente gestione dell'intelligenza artificiale esistono e non si devono ignorare. Vanno tutelati la dignità umana, che è il fondamento dei diritti, così come l'integrità fisica e mentale. Uno dei principali problemi consiste nelle violazioni di privacy, ritenuta da molti un diritto individuale fondamentale, un altro problema si manifesta nei casi in cui si possano venire a creare delle asimmetrie informative. Bisogna prestare attenzione anche per quanto riguarda il campo di applicazione dell'IA e della capacità di cui la si dota. Appare evidente come possa essere considerata pericolosa non tanto la stessa tecnologia, ma le persone che la sviluppano. In caso di danni e responsabilità la questione è tra le più dibattute. Quando si tratta di robotica ed intelligenza artificiale spesso la sicurezza è un tema che coinvolge il software e il design. Gli stessi programmatori o computer sciences sono semplicemente delle persone e come chiunque possono sbagliare, vista la complessità dei programmi è plausibile si celino degli errori o delle vulnerabilità tra le migliaia di righe di codice che devono scrivere[111].

- Autonomia: consiste nell'idea che alle persone spetti il diritto di prendere decisioni da sole. Con l'uso dell'intelligenza artificiale cediamo parte del nostro potere

[109] AI HLEG (High-Level Expert Group on Artificial Intelligence), "Ethics Guidelines for Trustworthy AI."
[110] Floridi et al., "AI4People—An Ethical Framework for a Good AI Society."
[111] Patrick Lin, Keith Abney, and George Bekey, "Robot Ethics: Mapping the Issues for a Mechanized World," *Artificial Intelligence* 175.5–6 (2011): 942–49, https://doi.org/10.1016/j.artint.2010.11.026.

decisionale alle macchine. Bisogna cecare di instaurare un equilibrio tra questi due aspetti, in quanto l'adozione di sistemi intelligenti non dovrebbe privare l'essere umano della propria libertà decisionale e del controllo. Va limitata invece la capacità decisionale delle macchine, proteggendo il valore umano che risiede nella scelta umana, mantenendo il potere di scegliere quali decisioni prendere. Va ponderata la migliore allocazione possibile di funzioni tra l'essere umano e le macchine.

- Giustizia: rappresenta l'ultimo principio della bioetica. I termini inglesi coi quali ci si riferisce sono due: justice e fairness, sono sinonimi, ma hanno sfumature di significato lievemente differenti. Il progresso nel campo dell'IA dovrebbe essere di aiuto nell'eliminazione delle discriminazioni, di qualsiasi tipo esse siano, poiché il beneficio, che ha come scopo lo sviluppo dell'intelligenza artificiale, deve essere condiviso. Per giustizia si intende quindi equità, tradotto a mio avviso in miglior modo col termine fairness[112]. Nell'allenamento di questi sistemi si deve considerare con solidarietà il rispetto degli interessi di tutti quanti. Tale concetto ha infatti sia una dimensione sostanziale che una procedurale: con la prima si intende l'impegno nella distribuzione dei benefici e dei costi, eliminando discriminazione e bias ingiusti, promuovendo eque opportunità; la dimensione procedurale implica di prevedere la possibilità di contestare e porre rimedio alle soluzioni prese dall'intelligenza artificiale[113].

In questo principio rientra il diritto alla non-discriminazione che troviamo all'art 21 della Carta dei Diritti Fondamentali dell'UE. La discriminazione è connaturata all'uso degli algoritmi, la stessa categorizzazione può far nascere dei bias[114]. I big data sono un campione della società e la società stessa presenta disuguaglianza, se questi dati vengono utilizzati per allenare gli algoritmi, le decisioni "obiettive" che ne deriveranno saranno macchiate dalle stesse discriminazioni.

La distorsione delle decisioni è un problema chiave e può avere diverse cause: può essere dovuto alla raccolta di un campione troppo piccolo per essere rappresentativo,

[112] Jessica Morley et al., "Ethics as a Service: A Pragmatic Operationalisation of AI Ethics," *Minds & Machines* 31.2 (2021): 239–56, https://doi.org/10.1007/s11023-021-09563-w.
[113] Floridi et al., "AI4People—An Ethical Framework for a Good AI Society," 4.
[114] Bryce Goodman and Seth Flaxman, "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation,'" *AIMag* 38.3 (2017): 50–57, https://doi.org/10.1609/aimag.v38i3.2741.

gli stessi dati potrebbero essere distorti oppure i dati riflettono una discriminazione presente nella società[115].

L'implemento di appropriate misure tecniche e organizzative previene questi effetti, come suggerito al paragrafo numero 71 del recital del GDPR[116]. Bisogna prestare particolare attenzione ai dati sensibili quali dati personali rilevanti origine etniche o razziali, orientamento politico e religioso, dati genetici e riguardanti la salute.

Un altro tipo di bias, riguardante l'incertezza, accade quando un gruppo è sottorappresentato nel campione destinato al training dell'algoritmo e quando l'algoritmo è avverso al rischio, preferendo basare le proprie previsioni su dati più certi. In sostanza l'algoritmo favorirebbe i gruppi meglio rappresentati nei dati di training poiché vi sarebbe meno incertezza legata a queste predizioni[117].

- Spiegabilità: il manipolo di persone che progetta le IA è esiguo, mentre quelle che le subisce o ne beneficia è enorme. Temi come l'intelligibility e l'accountability, la responsabilità, risultano essere fondamentali per mantenere alta la fiducia degli utilizzatori verso questa tecnologia. Per ovviare alla necessità di comprensione delle decisioni bisogna capire il funzionamento dell'intelligenza artificiale, questo però è spesso invisibile e incomprensibile come nel caso di algoritmi blackbox, a volte anche per gli esperti. Questo principio viene spesso espresso con il termine trasparenza ed integra i precedenti elencati. Infatti per capire se, semplificando, stiamo facendo del bene o del male, la comprensione è un elemento di prim'ordine[118].

Il diritto alla spiegazione è richiamato dagli articoli 13 e 13 del GDPR nei quali viene esplicitato il diritto ad ottenere informazioni significative circa la logica coinvolta[119]. Capire come spiegare une decisione algoritmica comporta una differente difficoltà a seconda dell'algoritmo utilizzato. Vi è una sostanziale differenza in trasparenza tra l'utilizzo di una regressione o l'utilizzo del machine learning, in quanto il secondo è più probabile lavori a scatola chiusa, come una "black box". Una delle possibili

[115] Olhede and Wolfe, "The Growing Ubiquity of Algorithms in Society."
[116] Parliament and Council of the European Union, "General Data Protection Regulation," 2016, https://gdpr-info.eu/recitals/.
[117] Goodman and Flaxman, "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation.'"
[118] Floridi et al., "AI4People—An Ethical Framework for a Good AI Society," 4.
[119] Goodman and Flaxman, "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation.'"

soluzioni consiste nello sviluppo di algoritmi che indichino l'influenza degli input nell'output da loro ottenuto, questo consentirebbe di capire almeno superficialmente la logica sottostante.

Questi principi riflettono le formalità e gli obblighi richiesti anche da un punto di vista legale, non stiamo trattando di una questione puramente etica, ma definendo il framework entro il quale l'intelligenza artificiale e i suoi operatori possono agire. Alcuni principi possono apparire come in conflitto. Analizzeremo in seguito in dettaglio il caso della "polizia predittiva", l'uso dell'IA per ridurre il crimine e in cui i principi di non maleficenza e autonomia sono in conflitto. Infatti per aumentare il benessere collettivo, vengono violate le libertà personali di alcuni individui. Spetta al dibattito pubblico e chi di competenza decidere come disegnare e plasmare l'intelligenza artificiale per trovare la soluzione migliore a questo trade-off.

I problemi che sorgono, quali bias, trasparenza e fairness vedremo come essere devono essere affrontati e risolti nel prossimo paragrafo con l'introduzione di alcuni requisiti.

2.3 Design di un'IA, dai principi alla pratica: i requisiti fondamentali

Analizzando l'intelligenza artificiale da un punto di vista etico è facile cadere nei formalismi. È infatti difficile fornire risposte pratiche quando si parla di trasparenza, spiegabilità, giustizia, oltre che poterne sottolineare l'importanza. Bisogna allontanarsi da un approccio orientato alla soluzione e considerarne uno orientato al processo, vanno disegnati i confini nel quale gli attori possono operare traducendo i diritti fondamentali in requisiti[120]. I principi etici appena visti sono quindi tradotti in requisiti che permettano l'implementazione di un'intelligenza artificiale affidabile, ponendo freno anche ai relativi problemi. Tali requisiti sono stati esposti dalla Commissione Europea nel report sulle linee guida per l'intelligenza artificiale e creano un framework completo di tutti gli aspetti da affrontare. Con un'opportuna modellazione sono validi per tutti gli stakeholder, dagli sviluppatori agli utenti finali[121].

[120] Evgeni Aizenberg and Jeroen van den Hoven, "Designing for Human Rights in AI," *Big Data & Society* 7.2 (2020): 205395172094956, https://doi.org/10.1177/2053951720949566.
[121] Commissione Europea, "AI Ethics Communication."

I sette requisiti chiave sono: a) human agency e supervisione, b) robustezza tecnica e sicurezza, c) privacy e governance dei dati, d) trasparenza, e) diversità, assenza di discriminazione e fairness, f) accountability, g) benessere ambientale e sociale,[122].

Lo scopo di questi requisiti è quello di ottenere una intelligenza artificiale affidabile e socialmente utile, ossia un bene sociale (AI4SG). Tale la definizione riportata da Floridi™: "the design, development, and deployment of AI systems in ways that prevent, mitigate or resolve problems adversely affecting human life and/or the wellbeing of the natural world, and/or enable socially preferable and/or environmentally sustainable developments[123]". I requisiti sono coerenti con i principi etici, ma più pragmatici, rappresentano le basi su cui fondare questo bene sociale e aiutano in un'efficace protezione dei diritti umani.

Andremo ora a vederli più nello specifico, su alcuni ci soffermeremo particolarmente. Anche in questo caso cercheremo di affrontarli separatamente, anche se vi sono forti co-dipendenze. Inoltre l'ordine di rappresentazione non è rivelatore della maggior importanza di uno rispetto ad altri.

a) Human agency e supervisione

I sistemi di intelligenza artificiale devono essere di supporto al processo decisionale e all'autonomia umana, rispettando i diritti fondamentali. Le persone dovrebbero essere dotate della conoscenza necessaria e della possibilità, prima di tutto, di capire questi sistemi e in secondo luogo di interagirci qualora dovesse essere necessario contestare i risultati dell'IA[124]. Per facilitare questo l'intelligenza artificiale dovrebbe essere disegnata affinché il suo contributo sia limitato o, meglio, contestualizzato a specifiche aree. L'intervento dell'IA non deve invadere l'autonomia degli utenti. Una tecnologia particolarmente invadente potrebbe essere rifiutata dagli stessi utenti poiché non compresa. Il contributo non deve quindi sovrastare il giudizio dell'utente, ma essergli di supporto. Una soluzione ottimale consisterebbe nel rendere gli stessi utenti partecipi della progettazione, andando incontro alle problematiche viste[125].

[122] AI HLEG (High-Level Expert Group on Artificial Intelligence), "Ethics Guidelines for Trustworthy AI."
[123] Luciano Floridi et al., "How to Design AI for Social Good: Seven Essential Factors," *Sci Eng Ethics* 26.3 (2020): 1771–96, https://doi.org/10.1007/s11948-020-00213-5.
[124] AI HLEG (High-Level Expert Group on Artificial Intelligence), "Ethics Guidelines for Trustworthy AI."
[125] AI HLEG (High-Level Expert Group on Artificial Intelligence), "Ethics Guidelines for Trustworthy AI."

Dovrebbe rimanere in capo agli individui la possibilità di curare il capitale semantico, ossia attribuire alle cose o ai concetti il significato da loro designato per quella specifica situazione. Secondo Floridi è importante definire i compiti che spettano alle macchine, e la semanticizzazione non è uno di questi. Basti pensare al concetto giuridico di violazione. L'IA sarebbe allenata sui casi preesistenti e in un contesto sociale entrerebbero in gioco anche componenti soggettive ed emozioni, campo in cui le IA hanno difficoltà. Anche in questo caso il segreto sta nel trovare il giusto equilibrio e progettare i sistemi affinché possano svolgere i compiti nei quali eccellano[126].

b) Robustezza tecnica e sicurezza

Con tale requisito si richiama il principio della prevenzione dei danni. L'IA deve essere appunto sviluppata con un approccio preventivo per minimizzare i rischi. Questi possono essere causati dalla presenza di vulnerabilità esterne, quindi attacchi informatici, piuttosto che da errori interni. Maggiore è il rischio che il sistema comporta, maggiore dovrà anche essere la sicurezza, soprattutto per i rischi connessi ai diritti delle persone.

Temi di assoluta rilevanza che possiamo far rientrare in questo requisito sono quello dell'accuratezza, dell'affidabilità e della riproducibilità. Il primo consiste nella precisione e correttezza con cui l'intelligenza artificiale svolge le proprie previsioni. Il livello richiesto varia dal campo di applicazione, ma deve in ogni caso essere indicato e compreso. L'affidabilità e la riproducibilità sono due elementi interdipendenti, in particolare qualora un qualsiasi caso non potesse essere replicato diventerebbe di conseguenza anche meno affidabile[127].

La falsificabilità è un elemento cruciale per aumentare la fiducia verso i sistemi di IA. Il termine potrebbe trarre in inganno, ma indica la possibilità di verificare in modo pratico alcune condizioni alla base dell'intelligenza artificiale. La fase di testing dovrebbe avvenire in modo incrementale e, qualora possibile, in condizioni reali. Tramite una tale

[126] Floridi et al., "How to Design AI for Social Good."
[127] Commissione Europea, "AI Ethics Communication."

procedura il controllo sulla sicurezza è maggiore. Ad esempio in Germania vi sono zone deregolamentate in cui è possibile testare i veicoli a guida autonoma[128].

Errori dovuti alla manipolazione dei dati possono avere due fonti principali: i dati di input o la scelta stessa di comprendere variabili non significative. Gli indicatori scelti dovrebbero essere collegati da un nesso di causalità, in caso contrario il rischio è quello di prendere decisioni su variabile che appunto non sono fondamentali.

c) Privacy e governance dei dati

Sulla privacy numerosi sono gli autori che si sono spesi, in quanto ritenuto un elemento prioritario per la sicurezza e la dignità umana. In tale contesto risulta essenziale il consenso, così da rendere consapevoli gli utenti circa i possibili usi.

L'integrità dei dati, oltre ad essere fondamentale per la tutela della privacy, lo è al fine valutativo e prestazionale. I dati, come abbiamo visto, sono alla base dell'intelligenza artificiale, una loro corretta gestione è fondamentale[129].

Non sempre l'accesso ai dati è assicurato, infatti uno dei motivi per cui oggi i sistemi di intelligenza artificiale stanno ottenendo il loro successo è proprio perché i dati sono più accessibili. L'accesso ai dati implica però che in caso di dati sensibili questo debba essere regolato[130] e in tal senso abbiamo visto le tutele previste dal GDPR.

Secondo le teorie economiche la privacy è da considerare un bene intermedio, poiché la volontà degli individui circa la riservatezza dei propri dati dipende dall'effetto di tali dati su risultati futuri, ossia dai benefici che un utente pensa di ottenere, come la personalizzazione dei prodotti, dal concedere i propri dati. Alcune problematiche relative alla privacy da considerare sono:

- la persistenza dei dati: i dati una volta creati potrebbero perfino sopravvivere alla nostra morte poiché il costo di archiviazione è assai ridotto, dati inutilizzabili oggi potrebbero trovare una qualche funzione un domani, rivelando il loro potere

[128] Floridi et al., "How to Design AI for Social Good."
[129] Pistilli, "Etica e algoritmi. Verso l'algoretica."
[130] AI HLEG (High-Level Expert Group on Artificial Intelligence), "Ethics Guidelines for Trustworthy AI."

predittivo. Si noti che i dati creati non sono solo quelli per cui diamo il consenso, ma ci sono tutta una serie di dati che creiamo involontariamente;

- il riutilizzo dei dati: i dati sono beni a fecondità ripetuta e potrebbero avere infinite funzionalità nel futuro, maggiore la persistenza e minore sarà la sicurezza sul come verranno utilizzati. Secondo Tucker è proprio nelle correlazioni impreviste sui dati che si basa il più alto potenziale distorsivo, portando a conseguenze negative per la privacy;
- lo spillover dei dati: la creazione di dati potrebbe catturare più informazioni di quel che pensiamo e avere conseguenze anche su altri soggetti. Ad esempio la raccolta di dati genetici potrebbe avere ripercussioni sulla privacy dei familiari poiché dotati di un corredo pressoché identico, foto e video spesso riprendono individui che non hanno dato alcuna autorizzazione per l'uso di tali dati, gli stessi algoritmi potrebbero creare spillover sottoforma di discriminazione a partire dai dati[131].

La creazione, raccolta ed utilizzo dei dati sono pratiche radicate nella nostra società e col tempo, a causa di tecnologie sempre più invasive, aspettative e comprensione circa la privacy stanno cambiando. La privacy è un concetto normativo e consiste in un vero e proprio diritto alla base della dignità umana. Per una corretta tutela e per prevenire illeciti, le informazioni devono essere corrette, anonime e contestualizzate. Ma la privacy è anche un concetto tecnico, alcune delle prassi più diffuse prevedono l'anonimizzazione in modo da evitare una re-identificazione dei soggetti, il mantenimento della sicurezza del significato e l'applicazione di modelli formali come quello differenziale che definisce uno standard per la gestione del rischio[132]. Vista la sua complessità e il significato multiforme la privacy è giustamente definita come un concetto sostanzialmente contestato da Mulligan[133]. Far coesistere la doppia natura della privacy richiede una comprensione adeguata di entrambe e la risoluzione di alcuni conflitti come una corretta interpretazione delle norme che possono variare a seconda del settore[134]. La privacy

[131] Catherine Tucker and Gans, "Privacy, Algorithms, and Artificial Intelligence," *The Economics of Artificial Intelligence: An Agenda*, National Bureau of Economic Research Conference Report (2019): 16.
[132] Kobbi Nissim and Alexandra Wood, "Is Privacy Privacy?," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2128 (2018): 20170358, https://doi.org/10.1098/rsta.2017.0358.
[133] Deirdre K. Mulligan, Colin Koopman, and Nick Doty, "Privacy Is an Essentially Contested Concept: A Multi-Dimensional Analytic for Mapping Privacy," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160118, https://doi.org/10.1098/rsta.2016.0118.
[134] Nissim and Wood, "Is Privacy Privacy?"

sembrerebbe quindi essere in particolar modo in conflitto con l'innovazione e la scienza, ma va adottato un approccio olistico in cui i diritti umani vengono considerati nella loro interazione con l'ambiente[135].

d) Trasparenza

Tale fattore serve per dare voce al principio di replicabilità. La trasparenza delle componenti di un sistema di intelligenza artificiale permette la tracciabilità, la spiegabilità e una migliore comunicazione.

La tracciabilità del processo di raccolta dei dati e degli algoritmi consente, nel caso di una prestazione che non ci si aspettava, di cercare al ritroso le possibili cause. Senza tracciabilità la spiegazione del perché l'output sia stato di un certo tipo rispetto ad un altro sarebbe infatti impossibile. Dovrebbe essere possibile per una persona capire le logiche sottostanti alle decisioni degli algoritmi. Se infatti le operazioni condotte dall'IA fossero spiegabili, se ne beneficerebbe in trasparenza, tutelando l'autonomia dell'utente[136]. La trasparenza dovrebbe riguardare i processi così come gli obiettivi ed è tutelata dal GDPR che fornisce garanzie ponendo degli obblighi di trasparenza e responsabilità per quanto riguarda il trattamento dei dati e in particolare per i processi decisionali basati sulla profilazione[137]. Dall'interazione tra i dati di input e l'algoritmo nasce la complessità, il requisito di trasparenza serve a garantire che lo strumento sia comprensibile ai soggetti interessati, che il processo decisionale sia equo migliorando anche l'accountability[138]. Bisogna decidere fino a che punto spingere l'approccio alla trasparenza poiché potrebbe limitare i modelli di apprendimento automatico e perché la trasparenza di per sé non risolve i problemi di responsabilità. Anche in questo caso la valutazione deve avvenire di concerto e magari in modo differenziale, con una maggior

[135] Vayena and Tasioulas, "The Dynamics of Big Data and Human Rights."
[136] Blacklaws, "Algorithms."
[137] Blacklaws, "Algorithms."
[138] Marion Oswald, "Algorithm Assisted Decision Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2128 (2018): 20170359, https://doi.org/10.1098/rsta.2017.0359.

attenzione, qualora venga richiesto, come quando sono coinvolte comunità minoritarie[139].

e) Non discriminazione ed equità

Rispettare il principio di equità dovrebbe essere ovvio e il perché palese. Pregiudizi e discriminazioni sono spesso alla base di danni nei confronti di persone o gruppi. I pregiudizi o bias possono essere presenti in qualsiasi fase del processo o propagarsi in esso, dalla raccolta dei dati, all'implementazione degli algoritmi, fino alle stesse decisioni. Una distorsione a livello dei dati potrebbe condurre ad azioni discriminatorie che a loro volta comprometterebbero i dati raccolti successivamente, formando così un circolo vizioso[140]. Le cause principali dei bias possono quindi essere riscontrate nella qualità dei dati di input qualora includessero pregiudizi, le decisioni algoritmiche si basano infatti su dati storici a volte discriminatori. Oltre ai pregiudizi storici vi sono quelli rappresentativi quando una popolazione è sotto o sovra rappresentata nel campione di analisi, quelli tecnici dovuti perlopiù alla tecnologia informatica intesa come software o hardware e quelli emergenti che si riscontrano in seguito al cambiamento di conoscenza o valori avvenuto in una società[141].

Algoritmi e IA sono sempre più alla base delle decisioni e discriminazioni basate magari su attributi demografici personali. Le numerose definizioni di equità matematiche mal si adattano alla realtà[142].

L'equità nel machine learning e nell'intelligenza artificiale è affrontata con un approccio empirico. Si fa uso di costrutti, vincoli di ottimizzazione e presupposti statistici considerando quindi l'equità come un problema da risolvere. In questo modo si delimita il problema dell'equità logaritmica al campo computazionale e solo chi ha le giuste

[139] Hetan Shah, "Algorithmic Accountability," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2128 (2018): 20170362, https://doi.org/10.1098/rsta.2017.0362.
[140] Floridi et al., "How to Design AI for Social Good," 4.
[141] Alina Köchling and Marius Claus Wehner, "Discriminated by an Algorithm: A Systematic Review of Discrimination and Fairness by Algorithmic Decision-Making in the Context of HR Recruitment and HR Development," *Bus Res* 13.3 (2020): 795–848, https://doi.org/10.1007/s40685-020-00134-w.
[142] Michelle Seng Ah Lee, Luciano Floridi, and Jatinder Singh, "Formalising Trade-Offs beyond Algorithmic Fairness: Lessons from Ethical Philosophy and Welfare Economics," *AI Ethics* 1.4 (2021): 529–44, https://doi.org/10.1007/s43681-021-00067-y.

conoscenze tecniche vi può partecipare[143]. Secondo Lauden, scopo delle discipline intellettuali è quello di risolvere problemi e allo stesso modo è quello che fanno gli algoritmi, ma tali problemi possono essere empirici, ma anche concettuali, riguardanti "l'adeguatezza delle soluzioni ai problemi empirici"[144]. L'equità può essere progettata adottando un certo livello di semplificazione e basandosi soprattutto sui principi di uguaglianza e giustizia, ma in ogni caso si tratta della traduzione di un concetto complesso in una forma statica, e i criteri per fare questo sono veramente molti[145]. Numerose sono le definizioni di equità, cambiano a seconda del Paese o del contesto, l'equità è quindi intrinsecamente soggettiva. Per cercare di scalfire la complessità sottostante a tale concetto basti sapere che l'equità può essere vista secondo vari punti di vista, distributiva, procedurale e interazionale[146]. Inoltre sono vari i tipi di disuguaglianza: naturale, socioeconomica, di talento, di preferenza e discriminazione sociale, ognuna con diversi livelli di accettabilità e tutela. Ancora più numerose sono le metriche per misurare l'equità[147]. Vi è inoltre da considerare che l'etica non è il solo parametro, ma ve ne sono altri di rilevanti quali l'accuratezza e la correttezza, la trasparenza e la responsabilità. Anche in questo caso per ognuno di questi concetti vi sono più teorie matematiche e, se di per sé sono già argomenti complessi, trovare un trade-off lo sarà ancor più. L'equità algoritmica rappresenta uno dei problemi più affrontato negli ultimi anni a causa proprio della sua difficoltà concettuale, è un problema sia concettuale che empirico. Trovare una soluzione univoca risulta impossibile, considerando che le dimensioni connesse sono molteplici[148].

f) Accountability:

L'IA va programmata secondo le regole della legge e i codici etici, questo è molto facile a dirsi in linea teorica, mentre è assai complicato nella pratica. Per esempio chi è responsabile in caso in caso di danni provocati dall'intelligenza artificiale? Potrebbe essere plausibile per un robot che maggior autonomia equivalga a maggior

[143] Scantamburlo, "Non-Empirical Problems in Fair Machine Learning."
[144] Larry Laudan, *Progress and Its Problems: Towards a Theory of Scientific Growth*, 1st paperback print. (Berkeley, Calif.: Univ. of Calif. Press, 1978).
[145] Scantamburlo, "Non-Empirical Problems in Fair Machine Learning."
[146] Köchling and Wehner, "Discriminated by an Algorithm."
[147] Lee, Floridi, and Singh, "Formalising Trade-Offs beyond Algorithmic Fairness."
[148] Scantamburlo, "Non-Empirical Problems in Fair Machine Learning."

responsabilità[149]. Vari in realtà sarebbero i modelli di responsabilità che si potrebbero applicare: tramite la responsabilità diretta la "colpa" verrebbe attribuita alla stessa intelligenza artificiale, non avente però personalità giuridica e provocando inoltre una deresponsabilizzazione delle persone fisiche magari reali responsabili. La responsabilità viene quindi spostata su esseri umani, società o altri agenti[150]. Con il modello "perpetrazione da parte di un altro" assume centralità il concetto di intento, attribuendo responsabilità a qualsiasi soggetto abbia volontariamente perpetrato il danno. Il modello di responsabilità del comando, come è facilmente intuibile, designa come colpevole chi ricopre ruoli/incarichi di responsabilità. Infine nel modello "probabile-naturale-conseguenza" la responsabilità viene attribuita allo sviluppatore o all'utente finale quando il danno è dovuto a negligenza, ossia è dovuto ad una probabile e naturale conseguenza imputabile al loro comportamento[151].

La responsabilità dovrebbe essere condivisa tra i vari soggetti in gioco: i designers, i regolatori e gli utenti[152]. Quello della responsabilità morale collettiva o distribuita (DMR) non è un concetto nuovo. Le azioni morali distribuite (DMA) sono quelle compiute da un gruppo di agenti umani, artificiali o ibridi attraverso un'interazione. Poiché l'aumento delle relazioni tra uomo e macchina stanno aumentando esponenzialmente è assolutamente da evitare la situazione in cui il problema di tutti diventi responsabilità di nessuno. L'etica classica non si occupa distribuzione della responsabilità poiché valuta gli agenti e la loro natura oppure con un approccio consequenziale valuta l'intenzionalità delle azioni. Spostare il focus dall'interesse per i singoli agenti alla valutazione del danno subito dal "paziente", e quindi alle caratteristiche del sistema che si vogliono ottenere, potrebbe essere una soluzione. Questo per implementare un modello multi-agente, possibile qualora gli agenti, sia umani che macchine, siano autonomi, abbiano la possibilità di interagire tra loro e imparino dai propri errori. Distribuire la responsabilità morale significa concentrarsi su quali agenti sono causalmente responsabili, indipendentemente da intenzionalità e natura delle azioni. Il DMA è l'output di una rete, in cui gli agenti della società

[149] Lin, Abney, and Bekey, "Robot Ethics."
[150] Thomas C. King et al., "Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions," *Sci Eng Ethics* 26.1 (2020): 89–120, https://doi.org/10.1007/s11948-018-00081-0.
[151] King et al., "Artificial Intelligence Crime."
[152] Mariarosaria Taddeo and Luciano Floridi, "How AI Can Be a Force for Good," *Science* 361.6404 (2018): 751–52, https://doi.org/10.1126/science.aat5991.

costituiscono i nodi, ed è l'unico elemento suscettibile di valutazione morale. Se le interazioni hanno prodotto un comportamento "cattivo" l'importante è la possibilità di correggerlo, propagandolo a ritroso e continuando fino al raggiungimento di un output soddisfacente[153]. Possibili obiezioni a tale modello riguardano l'ingiustizia nel responsabilizzare agenti non direttamente coinvolti e la fattibilità applicativa di tale modello, ma sicuramente tale modello, già utilizzato in contesti specifici, agirebbe più sulla prevenzione creando aspettative comuni.

Supervisionare eticamente la pratica scientifica per cercare di capire il modello più adeguato atto a prevenire i danni dovrebbe essere una pratica usuale. Non si dovrebbe separare la scienza e gli sviluppi tecnologici dalla valutazione etica[154].

L'audit è un'attività fondamentale e può essere svolta da soggetti interni o esterni. Riguarda valutazioni circa gli algoritmi, i dati e i processi, in particolare sarebbe utile analizzare gli impatti negativi, tramite impact assessments. Un impact assessments consiste in una valutazione sui diritti umani, Human Rights Impact Assessments, e considera sia fattori tecnici che etici, rappresentando un sistema sociotecnico complesso, ma completo, che deve perciò coinvolgere professionisti di più discipline[155]. Gli HRIA rappresentando un tipo di audit basato sull'etica (EBA), possono considerarsi un vero e proprio meccanismo di governance. L'EBA è un processo strutturato attraverso il quale viene valutato il comportamento di un'entità coerentemente a principi e diritti così da poter riporre fiducia nelle decisioni prese da un sistema di intelligenza artificiale, rendendolo così operativo. Migliorare la trasparenza e la regolarità procedurale consentirebbe di identificare più facilmente i responsabili in caso di danni prodotti dall'intelligenza artificiale. Esistono diversi approcci complementari tra loro: audit della funzionalità con il quale ci si concentra sulla logica che sottende le decisioni, audit del codice con il quale si verifica il codice vero e proprio e audit di impatto, nel quale rientra l'HRIA e che rappresentano la principale tendenza. I protocolli non sono ancora stati definiti nello specifico, ma possono essere top-down, basati su

[153] Luciano Floridi, "Faultless Responsibility: On the Nature and Allocation of Moral Responsibility for Distributed Moral Actions," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160112, https://doi.org/10.1098/rsta.2016.0112.
[154] Sabina Leonelli, "Locating Ethics in Data Science: Responsibility and Accountability in Global and Distributed Knowledge Production Systems," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160122, https://doi.org/10.1098/rsta.2016.0122.
[155] Latonero and Agarwal, "Human Rights Impact Assessments for AI: Learning from Facebook's Failure in Myanmar."

aspetti legali definiti da istituzioni nazionali, o bottom-up, dovuti ad un'espansione delle autorità regolamentative dei dati le cui regole quindi si diffonderebbero in un campo affine[156].

Un aspetto da non sottovalutare è che gli audit esterni non hanno una visione generale dovuta al minor accesso a conoscenze o risorse, disponibili invece per un revisore interno. Il giudizio di quest'ultimo d'altro canto potrebbe non essere oggettivo. Un sistema con più agenti e con una responsabilità diffusa e condivisa sarebbe quindi una soluzione vincente[157]. La responsabilità che un sistema di intelligenza artificiale sia etico ricade in chi lo ha progettato e lo gestisce, gli auditor hanno responsabilità per quello che vanno a verificare. Poiché sia i comportamenti umani che gli algoritmi, qualora possano apprendere, si modificano nel tempo, il processo di valutazione deve essere continuo[158]. Il processo di auditing qui espresso va ad indagare e verificare non solo il requisito dell'accountability, ma anche gli altri.

g) Benessere ambientale e sociale

L'ambiente stesso e la società sono da considerarsi stakeholder a cui le soluzioni di IA possono apportare molti benefici[159]. Quotidianamente siamo esposti agli effetti dell'intelligenza artificiale, gli impatti sono quindi notevoli. Questo particolare aspetto lo vedremo nel prossimo paragrafo.

Abbiamo finito di analizzare i vari requisiti e quello che possiamo dedurre è che la progettazione di sistemi di intelligenza artificiale è un gioco di equilibri, sia all'interno dei singoli requisiti che tra i requisiti. Le tensioni che possono venirsi a creare sono molte e questi elementi sono solo alcuni di quelli da considerare per cercare allentarle. Sicuramente una serie di compromessi è inevitabile.

[156] Jakob Mökander et al., "Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations," *Sci Eng Ethics* 27.4 (2021): 44, https://doi.org/10.1007/s11948-021-00319-4.
[157] Morley et al., "Ethics as a Service."
[158] Jonathan Cave, "The Ethics of Data and of Data Science: An Economist's Perspective," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160117, https://doi.org/10.1098/rsta.2016.0117.
[159] AI HLEG (High-Level Expert Group on Artificial Intelligence), "Ethics Guidelines for Trustworthy AI."

Uno interessante da analizzare ed affrontato da Morley riguarda la flessibilità e la rigidità. A linee guide definite come quelle proposte dall'Unione Europea non deve corrispondere un'applicazione stereotipata. I vari rischi e principi devono essere adottati a seconda delle specifiche esigenze. Affinché l'etica sia operativa e al tempo stesso tuteli gli individui deve essere individuato il giusto livello di astrazione e il rispetto dei principi etici non deve essere considerato come una mera formalità o un obiettivo, anzi deve portare ad un processo riflessivo[160].

In particolare, se l'etica venisse considerata come un "servizio" i modelli applicabili sarebbero tre:

- Software as a Service, in cui la gestione è interamente demandata a terzi, cosa che comporterebbe un'eccessiva rigidità della governance dell'IA.
- Infrastructure as a Service, che rappresenta l'opposto, ossia il caso in cui la governance sarebbe centralizzata, fornendo un'eccessiva flessibilità agli utenti.
- Platform as a Service, che consiste in una via di mezzo, una sorta di compromesso tra i due precedenti modelli, tra decentralizzazione e centralizzazione. Per quanto concerne l'etica e l'IA, l'applicazione di questo modello vedrebbe il coinvolgimento di diversi soggetti: un comitato etico indipendente con il compito di definire un codice generale e impostare le linee guida per un corretto processo e i professionisti delle varie aziende che svolgerebbero un ruolo più pratico[161]. Tale modello, in cui la responsabilità sarebbe distribuita, è quello più assimilabile all'approccio Europeo.

Numerose sono le guide etiche scritte negli ultimi anni, la stragrande maggioranza di queste affronta l'argomento solo ad un livello astratto[162]. L'IA deve essere disegnata e sviluppata in modo da diminuire le disuguaglianze e promuovere la crescita sociale con un approccio multistakeholder[163]. L'etica, affinché possa avere un impatto positivo, deve essere integrata nel processo di progettazione. A chi ritiene non vi sia garanzia di ciò basti ricordare che in altri campi, come quello medico, sono stati resi operativi tali principi etici[164]. Vi è senz'altro quindi il riconoscimento di come un design pro-etico sia

[160] Morley et al., "Ethics as a Service."
[161] Morley et al., "Ethics as a Service."
[162] Jessica Morley et al., "Operationalising AI Ethics: Barriers, Enablers and next Steps," *AI & Soc* (2021), https://doi.org/10.1007/s00146-021-01308-8.
[163] Floridi et al., "AI4People—An Ethical Framework for a Good AI Society," 4.
[164] Morley et al., "Ethics as a Service."

utile per migliorare l'impatto sociale dell'intelligenza artificiale, ma vi è bisogno di una ancor maggior standardizzazione del processo e delle pratiche da seguire poiché gli addetti ai lavori hanno una comprensione limitata del fenomeno per quanto riguarda gli aspetti etico-teorici. Il service design è l'approccio con cui si potrebbe facilitare la comunicazione tra i vari livelli, rendendo operativi i principi etici e creando sistemi per i quali l'IA può essere intesa come un bene pubblico[165]. Come suggerisce Morley, i meccanismi per un'attuazione ottimale richiederanno tempo e una riflessione continua[166].

2.4 Impatto sociale

Dedichiamo un paragrafo specifico ad analizzare cosa si intende quando l'intelligenza artificiale viene additata come un bene sociale e l'impatto che ne può scaturire.

Quando si parla di dati e bene pubblico due sono le definizioni possibili, i dati dovrebbero essere di dominio pubblico, disponibili a varie organizzazioni internazionali, visto il loro impatto positivo, oppure potrebbero essere considerati proprio come un bene pubblico vista la loro capacità di risolvere problemi sociali[167].

L'idea che soluzioni basate sull'IA abbiano il potenziale per avere un impatto positivo sulla società è particolarmente attuale[168]. Una corretta applicazione di tale tecnologia consentirebbe di abbassare i costi, ridurre i rischi, aumentare la coerenza e l'affidabilità trovando nuove soluzioni a problemi[169]. Contrariamente altri pongono l'accento sui rischi che questa tecnologia comporterebbe, come la riduzione dei posti di lavoro.

La tecnologia dell'intelligenza artificiale sta stravolgendo molti aspetti della società e sicuramente uno è quello che riguarda l'organizzazione del lavoro. Il Trades Union Congress, organizzazione che aiuta le associazioni dei lavoratori inglesi, ha stilato un

[165] Cat Drew, "Design for Data Ethics: Using Service Design Approaches to Operationalize Ethical Principles on Four Projects," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2128 (2018): 20170353, https://doi.org/10.1098/rsta.2017.0353.
[166] Morley et al., "Operationalising AI Ethics."
[167] Linnet Taylor, "The Ethics of Big Data as a Public Good: Which Public? Whose Good?," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160126, https://doi.org/10.1098/rsta.2016.0126.
[168] Floridi et al., "How to Design AI for Social Good."
[169] Taddeo and Floridi, "How AI Can Be a Force for Good."

report, il "Technology managing people", scopo del quale è creare consapevolezza proprio sugli effetti nel mondo del lavoro. Le forme dei processi lavorativi stanno subendo, e continueranno a subire, notevoli trasformazioni, dai metodi produttivi alle funzioni personali e gestionali, in particolare per quanto riguarda i processi di assunzione e per il monitoraggio e la valutazione delle prestazioni[170]. È opinione dei pessimisti che la tecnologia crei disoccupazione sostituendo le persone[171]. Sicuramente i sistemi di IA sono in grado di svolgere molte attività in modo più efficiente, ed è giusto che vengano utilizzati. Che senso avrebbe lasciare continuare una persona se questa comporta maggiori costi e un livello di accuratezza inferiore? Se la logica adottata rimanesse quella del profitto non vi sarebbe ragione alcuna, ma le decisioni circa il benessere degli esseri umani dovrebbero poggiare anche su altri criteri, questo implicherebbe innanzitutto un grosso cambiamento culturale[172]. La comparazione tra uomo e macchina avviene pertanto su considerazioni di costo, e i continui miglioramenti nella capacità previsionale delle macchine influiranno senz'altro sulla distribuzione del potere decisionale. Ma vi sono abilità complementari da abbinare, che vanno oltre alla previsione in un compito e sono racchiudibili nel giudizio[173]. Infatti, sempre rimanendo in tema lavoro, è credenza opposta, quella degli ottimisti, che l'introduzione di queste nuove tecnologia creerà nuove mansioni, ad oggi magari inimmaginabili, e soprattutto all'essere umano saranno destinate quelle attività a più alto valore aggiunto[174]. Probabilmente la verità starà nel mezzo, già ad oggi robot e macchine sono ampiamente utilizzati e ci aiutano a svolgere una moltitudine di attività. Ci sgravano da compiti noiosi o ripetitivi, ma aumenta anche la nostra dipendenza nei loro confronti. Se la fabbrica 4.0 trova riscontro nel dibattito pubblico, ci si dovrebbe anche chiedere come sarà il lavoratore 4.0, valutare cioè l'impatto sul capitale umano[175].

Le nostre giornate, il modo di pensare e gli approcci ai problemi sono stati senz'altro modificate da tutte le innovazioni tecnologiche degli ultimi anni. Sono sempre in numero

[170] Joe Atkinson, "'Technology Managing People': An Urgent Agenda for Labour Law," *Industrial Law Journal* 50.2 (2021): 324–29, https://doi.org/10.1093/indlaw/dwab005.
[171] Domenico Talia, *La società calcolabile e i big data: algoritmi e persone nel mondo digitale* (Soveria Mannelli: Rubbettino, 2018).
[172] Petros Gelepithis, "AI and Human Society," *AI & Society* 13 (1999): 312–21.
[173] Ajay K Agrawal, Joshua Gans, and Avi Goldfarb, "Prediction, Judgment and Complexity," *The Economics of Artificial Intelligence: An Agenda* (2019): 23.
[174] Gelepithis, "AI and Human Society."
[175] Talia, *La società*.

crescente le piattaforme che hanno il ruolo di intermediari sociali. Queste svolgono il più disparato tipo di servizio: shop online, video streaming e social network solo per citare quelli più conosciuti. Queste sono macchine sociali, macchine le cui componenti, per esempio i partecipanti, sono esseri umani che interagiscono mediante una piattaforma. Una funzione ricorrente di tali sistemi abilitati dall'intelligenza artificiale è il filtraggio collaborativo usato per le raccomandazioni delle preferenze agli utenti. Questi sistemi si basano su algoritmi statistici o di apprendimento la cui interazione con gli utenti forma un agente. Un agente teologico è un sistema orientato ad un obiettivo e che ha delle variabili ambientali, provenienti cioè dall'ambiente, può essere sia autonomo che eteronomo se le decisioni dipendono da un responsabile[176]. Le macchine sociali quindi possono essere considerate teleologiche avendo un obiettivo da perseguire, ad esempio aumentare il tempo degli utenti nella piattaforma, far spendere più soldi agli utenti. Tale obiettivo a livello macro, spesso non coincide con quello micro, cioè quello degli utenti. Le macchine sociali non ottengono solo informazioni dai partecipanti, ma a questi spesso fanno svolgere compiti elementari, e comunque il loro giudizio è rilevante per le decisioni future dell'intero sistema. Sintetizzando, le macchine sociali combinano infrastruttura, che può essere anche web, algoritmi e persone umane, spesso tali sistemi sono autonomi e in grado di influenzare profondamente gli individui. Le interazioni tra infrastruttura, software e persone sono all'origine dell'agente. Tale relazione non rende completamente libere le persone e genera possibili preoccupazioni per loro. La prima consiste nella gestione diretta delle persone da parte di un software, come nel caso delle tante applicazioni per le consegne soprattutto di cibo, con le conseguenti implicazioni etiche e di benessere del caso. Altra considerazione riguarda l'autonomia decisionale degli individui che in queste piattaforme possono scegliere solo fra le opzioni a loro presentate. I progettisti possono creare meccanismi per guidare e quindi influenzare gli utenti, il cosiddetto effetto psicologico del nudging, e gli utenti possono diventare dipendenti da queste dinamiche. Sempre più le macchine sociali sono integrate nella società svolgendo un ruolo da mediatore, e gli stessi effetti che hanno sugli individui possono essere trasposti ed ampliati all'intera società[177]. Oltre ai possibili effetti o danni psicologici, ricordiamo essere rilevanti anche quelli ambientali. L'elettronica da un lato

[176] Nello Cristianini, Teresa Scantamburlo, and James Ladyman, "The Social Turn of Artificial Intelligence," *AI & Soc* (2021), https://doi.org/10.1007/s00146-021-01289-8.
[177] Cristianini, Scantamburlo, and Ladyman, "The Social Turn of Artificial Intelligence."

produce una grande quantità di rifiuti, mentre dall'altro richiede una grossa quantità di energia per funzionare[178].

[178] Floridi et al., "AI4People—An Ethical Framework for a Good AI Society."

Capitolo 3

L'intelligenza artificiale nei procedimenti di giustizia

Abbiamo brevemente accennato nel precedente capitolo ad alcuni ambiti applicativi in cui è possibile riscontrare la presenza dell'intelligenza artificiale: sono numerosi e il settore pubblico non è da meno. Modelli predittivi e di previsione del rischio permeano la nostra società e sono all'ordine del giorno anche all'interno di servizi pubblici come giustizia penale, sicurezza, salute e assistenza sociale[179]. Questo significa che potenzialmente un buon numero di decisioni con implicazioni importanti per le persone e l'ambiente sono svolte da sistemi automatici[180]. Andremo a vedere meglio come l'intelligenza artificiale è applicata nel campo della giustizia penale.

3.1 Possibili punti d'incontro tra IA e giustizia

In primis distinguiamo le quattro possibili applicazioni degli algoritmi in questo campo: l'algoritmo delinquente, l'algoritmo investigante, l'algoritmo consulente, l'algoritmo giudicante.

L'IA può essere usata per commettere crimini, quindi con fini malevoli: in questo caso si può parlare di *algoritmo delinquente*. Sono sempre di più i casi di crimini informatici in cui l'intelligenza artificiale ha un ruolo fondamentale. Questo è infatti uno dei principali rischi che comporta tale tecnologia. La minaccia è dovuta alle nuove possibilità innescate dall'intelligenza artificiale. I problemi che insorgono in questo caso riguardano l'attribuzione del reato, poiché questi strumenti possono agire in modo autonomo dopo essere stati programmati, e la stessa difficoltà del monitoraggio. A causa dell'elevato livello di complessità in gioco, altrettanto sofisticate dovranno essere le tecniche per l'indagine, infatti queste avvengono trasversalmente su più piani, quello reale e quello digitale[181].

[179] Oswald, "Algorithm-Assisted Decision-Making in the Public Sector."
[180] Mökander et al., "Ethics-Based Auditing of Automated Decision-Making Systems."
[181] King et al., "Artificial Intelligence Crime."

L'intelligenza artificiale può anche essere usata con fine "protettivo" e in tal senso esistono principalmente due aree di possibile intervento, quella della polizia predittiva e quella della giustizia predittiva[182].

Nel primo caso parliamo di un *algoritmo investigante*, usato per prevedere la commissione di un reato o scoprire l'identità di un potenziale colpevole nel caso il reato sia stato già commesso. Questo ambito è chiamato law enforcement e l'intelligenza artificiale viene usata per elaborare grosse moli di dati e per aiutare i pubblici ufficiali nel loro lavoro e nel mantenimento della pubblica sicurezza, per esempio tramite l'individuazione di hot-spot, aree localizzate spazialmente e temporalmente in cui è più probabile avvenga un crimine. I rischi riguardano bias promulgati da tali algoritmi e che si ripercuotono in azioni discriminatorie, oltre che in un'invasione di privacy[183]. Programmi simili sono in uso anche in Italia, come KeyCrime nella Questura di Milano[184].

Nel caso della giustizia predittiva abbiamo a che fare con un *algoritmo consulente e uno giudicante*.

Il primo è uno strumento che, sulla base di banche dati sempre migliori, è capace di offrire consigli e supporto legale agli avvocati circa materiali utili o casi simili da consultare. In futuro potrebbero arrivare addirittura a prevedere l'esito di un processo[185].

Il secondo, l'algoritmo giudicante, che è il nostro oggetto di analisi, consta in uno strumento messo a disposizione dei giudici e impiegato per il pretrial release o anche nella formulazione delle sentenze. Questo ambito desta non poche preoccupazioni perché è uno di quelli in cui gli algoritmi sono più invasivi, avendo le loro decisioni un grosso peso sulla vita delle persone. Nonostante questo, l'accuratezza e la precisione di alcuni strumenti sono tali da farli ritenere un prezioso aiuto, se correttamente utilizzati[186]. In futuro l'impiego di automated decision systems non è da dare per

[182] di Fabio Basile, "Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine," Diritto penale uomo.10 (2019): 33.
[183] Antonella Massaro et al., "Intelligenza artificiale e giustizia penale" (2020): 227.
[184] Condé Nast, "Il software italiano che ha cambiato il mondo della polizia predittiva," *Wired Italia*, 18 May 2019, https://www.wired.it/attualita/tech/2019/05/18/polizia-predittiva-software-italiano-keycrime/.
[185] Claudia Costa, "Intelligenza artificiale e Giustizia: tempi ancora prematuri," *AI4Business*, 8 May 2019, https://www.ai4business.it/intelligenza-artificiale/intelligenza-artificiale-giustizia/.
[186] Massaro et al., "Intelligenza artificiale e giustizia penale."

impossibile, sostituendo si ipotizza che sostituirà l'uomo e creando dei giudice-macchina[187]. Ma al momento per quanto riguarda la giustizia predittiva una delle applicazioni più diffuse riguarda gli algoritmi che calcolano il grado di recidiva, ossia le probabilità di reiterazione di un reato. Sono di ausilio al giudice nella determinazione delle misure cautelari.

In ogni caso, l' obiettivo dell'introduzione di questa tecnologia dovrebbe essere quello di valorizzare la figura del giudice, aiutandolo nel suo operato, ampliandone quindi le capacità[188].

Vi sono altri campi sempre inerenti al diritto che vedono sfruttare l'IA, come quello contrattualistico che ha trovato nell'automazione una risorsa fondamentale[189]. Ancora, alcuni Paesi, molto progressisti, come la Gran Bretagna e i Paesi Bassi hanno strumenti che permettono la risoluzione delle dispute online sempre automaticamente. Questi strumenti provvedono a risolvere extragiudizialmente controversie fornendo servizi di conciliazione, mediazione e arbitrato a cui il giudice può attingere durante un processo[190].

È comunque indubbio che, come pronosticato dall'American Bar Association nel 1963, l'intelligenza artificiale stia rivoluzionando e rivoluzionerà anche le pratiche legali[191]. Abbiamo visto che la giurisprudenza sarà chiamata a disciplinare questa nuova tecnologia e le questioni che ne scaturiranno, ma allo stesso tempo essa stessa ne sarà coinvolta.

[187] Basile, "Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine."
[188] Massaro et al., "Intelligenza artificiale e giustizia penale."
[189] Florian Martin-Bariteau and Marina Pavlovic, *AI and Contract Law*, SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, November 2, 2020), https://papers.ssrn.com/abstract=3730385.
[190] "Online Dispute Resolution | European Commission," n.d., https://ec.europa.eu/consumers/odr/main/?event=main.trader.register; "Online Dispute Resolution e giustizia digitale," *Altalex*, 23 February 2021, https://www.altalex.com/documents/news/2021/02/23/online-dispute-resolution-e-giustizia-digitale.
[191] Reed C. Lawlor, "What Computers Can Do: Analysis and Prediction of Judicial Decisions," *American Bar Association Journal* 49.4 (1963): 337–44, https://www.jstor.org/stable/25722338.

3.2 Decisione robotica

Le ragioni per le quali la sostituzione del giudice con un "robot" è desiderabile sono varie.

La prima riguarda la liberazione delle persone dal peso del lavoro, ragione che non riguarda prettamente l'applicazione di sistemi di intelligenza artificiale in ambito giuridico e che è un argomento già affrontato.

Il secondo motivo risiede nella possibilità, anche in questo caso non solo in questo campo, di maggiori performance che si possono esprimere coi concetti di economicità e rapidità. In particolar modo nel diritto il tempo ha un costo, ma è anche un limite caratterizzante, poiché un'eccessiva rapidità danneggia per esempio il diritto alla difesa[192]. Inoltre, la stessa progettazione di sistemi che dovrebbero trattare una materia così complicata potrebbe rivelarsi costosa.

La terza motivazione, il garantire certezza giuridica, è particolarmente rilevante in ambito giuridico. Il desiderio di una sicurezza assoluta in ambito giuridico non è nuovo, idealmente si vorrebbe venissero prese solo decisioni corrette, in età illuministica infatti si aspirava ad una decisione meccanica, basata su ragionamenti razionali e priva di giudizi soggettivi[193]. Già all'epoca nacque l'idea del giudice-macchina, figlia di una concezione meccanicistica della legge e del ruolo del giudice che viene spersonalizzato. Successivamente presero piede teorie realiste, opposte per principi e valide tutt'oggi, che andavano contro la forma e alla base delle quali vi era la credenza che la "legge dei libri" si differenziasse dalla "legge effettiva", dando importanza alle componenti creative e interpretative[194].

Un evento quale un processo è molto influenzato dalla soggettività dei singoli, che siano giudici, testimoni o imputati. Dagli studi psicoanalitici apprendiamo infatti che il cervello umano opera sia con alcuni meccanismi cognitivi controllabili, sia con molti automatici che possono indurre in errore. Con questi presupposti l'idea di un giudice robot

[192] Massimo Luciani, "LA DECISIONE GIUDIZIARIA ROBOTICA," Rivista Associazione Italiana dei Costituzionalisti.03/2018 (2018): 22.
[193] Francesca Ceresa Gastaldo, "Il giudice-robot: l'intelligenza artificiale nei sistemi giudiziari tra aspettative ed equivoci," *Ius in itinere*, 22 March 2021, https://www.iusinitinere.it/il-giudice-robot-lintelligenza-artificiale-nei-sistemi-giudiziari-tra-aspettative-ed-equivoci-36717.
[194] Filippo Donati, "INTELLIGENZA ARTIFICIALE E GIUSTIZIA," *Rivista Associazione Italiana dei Costituzionalisti* 1/2020 (2020): 22.

potrebbe apparire come desiderabile[195]. Ma le modalità di funzionamento dei sistemi di intelligenza artificiale non ricalcano il modo di pensare logico dei giudici, ma basano i loro calcoli su vecchie decisioni dei giudizi e non sulla legge stessa, applicando quindi un metodo deduttivo[196]. Istruire una IA affinché possa ragionare come un giudice costituirebbe un lavoro di progettazione molto complicato che implicherebbe anche delle scelte politiche, dovendo istruire e pesare il robot sui criteri interpretativi. Per ottenere un giudice-macchina in grado di interpretare la legge questo dovrebbe essere dotato di un'intelligenza "forte" e abbiamo visto non essere possibile al momento.

E cosa fa un giudice? L'attività del giudice è difficilmente riconducibile a schemi astratti tipici dei sistemi dell' intelligenza artificiale e il suo compito non si limita ad un esercizio puramente logico. Deve risolvere una varietà di fatti tramite il diritto e una serie di complesse valutazioni che devono tenere conto di più elementi. Il semplice capire gli elementi più rilevanti per un processo non è banale e scontato, è un'attività che un giudice pratica abitualmente, ma che implica la sintesi di moltissime valutazioni differenti[197]. La funzione del giudice non è la di iuris-dictio, cioè pronunciare legge e modellarla affinché sia giusta, ma quella di interpretare e applicare tale legge. Citando Montesquieu, un giudice non è "la bouche de la loi" poiché, per l'appunto una norma può essere interpretata in diversi modi, infatti essendo insita del diritto una certa dose di incertezza[198]. Un giudice non si preoccupa di valori morali e più che la giustezza delle sue decisioni dovrebbe esserne valutata l'esattezza[199].

Limiti legali

Per una possibile applicazione dei sistemi di intelligenza artificiale in campo giudiziario vi sono una serie di ostacoli di carattere giuridico. Una sostituzione del giudice non può avvenire perché al momento è contraria a principi costituzionali e in particolare all'art. 102 della Costituzione[200] il quale prevede che "l'esercizio della funzione giurisdizionale deve essere affidata a magistrati istituiti e regolati". Altri articoli come il numero 25 e il

[195] Gastaldo, "Il giudice-robot."
[196] Donati, "INTELLIGENZA ARTIFICIALE E GIUSTIZIA."
[197] Antonio D'Aloia, "Il diritto verso 'il mondo nuovo'. Le sfide dell'Intelligenza Artificiale," *BioLaw Journal - Rivista di BioDiritto*.1 (2019): 3–31, https://doi.org/10.15168/2284-4503-349.
[198] "Prevedere l'esito di un giudizio: ecco la giurisprudenza predittiva," *Università Ca' Foscari Venezia*, n.d., http://www.unive.it/pag/14024/?tx_news_pi1%5Bnews%5D=9884&cHash=2d30d787f83ed5c005434c077b6d25bd.
[199] Luciani, "LA DECISIONE GIUDIZIARIA ROBOTICA."
[200] Donati, "INTELLIGENZA ARTIFICIALE E GIUSTIZIA."

101 fugano ogni dubbio escludendo la possibilità che il giudice debba sottostare agli esiti di un sistema IA e che questo lo possa sostituire[201]. Con l'applicazione di decisioni robotiche vincolanti troverebbe poi difficoltà il rispetto dei principi del contradditorio e del giusto processo. Il GDPR all'art. 22 esprime il diritto dell'interessato di non essere sottoposto a decisioni unicamente basate sul trattamento automatizzato e all'art.10 rivendica la competenza dell'autorità pubblica nel trattamento dei dati in caso di reati. Il giudice può avvalersi di questi strumenti, ma non in modo esclusivo, per di più tali strumenti non possono essere privati[202].

IA come ausilio

Sistemi di intelligenza artificiale potrebbero e sono già utilizzati come strumento ausiliare per i professionisti del settore, affiancando i giudici nella fase decisoria o gli avvocati nella preparazione di una pratica. In particolare, sarebbero utili negli ambiti della consulenza e ricerca, in tutte quelle attività che sono caratterizzate da semplicità e ripetitività o da grandi moli di dati da analizzare[203].

Una sostituzione "parziale" è quindi possibile e auspicabile. Il diritto fa parte della società e come molti altri ambiti sta subendo l'influsso dei continui sviluppi nel campo dell'intelligenza artificiale. La tecnologia può essere uno strumento utile al servizio della giustizia, ma non deve essere il fine[204].

[201] "La Costituzione - Articolo 102 | Senato Della Repubblica," n.d., https://www.senato.it/istituzione/la-costituzione/parte-ii/titolo-iv/sezione-i/articolo-102.
[202] "Algoritmo e giustizia predittiva in campo penale," *Altalex*, 14 June 2019, https://www.altalex.com/documents/news/2019/06/14/algoritmo-e-la-giustizia-predittiva-in-campo-penale.
[203] "Tutto quello che c'è da sapere sulla Intelligenza artificiale nello studio legale," *Altalex*, 16 July 2018, https://www.altalex.com/documents/news/2018/07/16/intelligenza-artificiale-nel-settore-legale; Luca Tremolada, "Giustizia predittiva, l'intelligenza artificiale migliore amica dell'avvocato," *Il Sole 24 ORE*, 10 March 2020, https://www.ilsole24ore.com/art/giustizia-predittiva-l-intelligenza-artificiale-migliore-amica-dell-avvocato-ACBXBbJB.
[204] "Prevedere l'esito di un giudizio."

3.3 Intelligenza artificiale nella giustizia penale

I sistemi di giustizia predittiva possono formulare delle previsioni su varie fasi legali

L'intelligenza artificiale, quando si parla di giustizia predittiva, può essere applicata in due casi[205]:

- per la previsione di un esito, nella fase in cui si deve giungere ad una sentenza;
- per la valutazione della probabilità di recidiva, cioè la probabilità di commissione di nuovi reati e decisioni in materia di pretrial release. Tale termine fa riferimento alla condizione di scarcerazione preliminare che può avvenire nel periodo tra la presentazione delle accuse da parte delle forze dell'ordine e il processo[206].

In entrambi i casi gli strumenti di intelligenza artificiale avrebbero peso sulle decisioni del giudice, in modo mediato o diretto.

3.3.1 Decisione robotica nel sentencing

Come accennato precedentemente, al momento gli automated decision system hanno visto il loro impiego solo in ambito civile. Se ne sta però iniziando a discutere, per esempio nelle corti dell'Arizona, del Colorado, del Delaware, del Kentucky, della Louisiana i risultati di questi calcoli di rischio vengono consegnati al giudice come elemento per determinare la sentenza penale[207].

In Europa nel 2018 la Commissione per l'efficacia della giustizia ha adottato la "Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi" nella quale richiama i principi fondamentali che un sistema IA in questo ambito deve rispettare[208]:

1. Principio del rispetto dei Diritti Fondamentali;
2. Principio di non-discriminazione;

[205] Massaro et al., "Intelligenza artificiale e giustizia penale."
[206] "Pretrial Release," *Bureau of Justice Statistics*, n.d., https://bjs.ojp.gov/topics/courts/pretrial-release.
[207] Paolo Benanti, "Algoritmi con pregiudizi: il caso serio delle corti di giustizia USA," *paolobenanti*, 3 October 2017, https://www.paolobenanti.com/post/2017/10/03/algoritmi-con-pregiudizi-il-caso-serio-delle-corti-di-giustizia-usa.
[208] Commissione Europea per l'efficienza della Giustizia, "Carta Etica Europea Sull'utilizzo Dell'intelligenza Artificiale Nei Sistemi Giudiziari e Negli Ambiti Connessi," 2019, https://rm.coe.int/carta-etica-europea-sull-utilizzo-dell-intelligenza-artificiale-nei-si/1680993348.

3. Principio di qualità e sicurezza;
4. Principio di trasparenza, imparzialità ed equità;
5. Principio del controllo da parte dell'utilizzatore.

Si noti come tali principi trovino ampio riscontro in quelli da noi analizzati nel precedente capitolo.

Compito principale del giudice è quello di decidere, ed è lo stesso output che si può ottenere dall'intelligenza artificiale. Questa dovrebbe dotare il giudice di maggior autonomia, non privarlo. Magari lo si potrebbe liberare dalla mole di processi che lo attanaglia, anche perché spesso per sopperire alla mancanza di tempo, lo stesso giudice adotta scorciatoie con decisioni standardizzate[209]. Chi meglio della macchina può aiutarlo in questo? Vanno scelti quindi i compiti da affidare alle macchine: quelli a minor complessità, in cui l'interpretazione della legge da parte del giudice per la fattispecie non è richiesta, cioè quelli in cui sono meno richieste le funzioni logiche e creative[210]. I compiti da assegnare sono quelli nei quali la macchina può esprimersi al meglio, dimostrando la propria maggior efficienza, cosa che consentirebbe una diminuzione dei tempi dei processi. Un giorno potrebbero essere gli stessi algoritmi a pronunciare le sentenze, magari nei casi più semplici e lineari in cui l'applicazione di una norma è sufficiente. D'altronde, se opportunamente allenato, un algoritmo potrebbe agire in modo meno discriminatorio dello stesso essere umano. Ricorrere a decisioni robotiche permetterebbe infine una certa standardizzazione. Senza sminuire il valore delle peculiarità dei singoli, una certa prevedibilità delle decisioni è necessaria.

3.3.2 Previsione algoritmica della recidiva

Gli strumenti utilizzati per la giustizia predittiva si chiamano Risk Assessments Tools (RATs), utili appunto per valutazioni legate al rischio.

I RATs dovrebbero essere:

- Obiettivi e scientifici;
- Evidence-based;
- Affidabili;

[209] Massaro et al., "Intelligenza artificiale e giustizia penale."
[210] Carleo, *Decisione Robotica*.

- Neutrali rispetto alla razza[211].

Scopo di questi algoritmi è invece quello di:

- massimizzare la pubblica sicurezza;
- massimizzare il numero di partecipazioni alle udienze;
- ridurre l'incarcerazione di massa;
- perpetrare l'equità[212].

Il quesito principale a cui cercano di dare riposta è capire le probabilità di un determinato individuo di reiterare un reato. Vi è quindi un trade-off tra i costi di gestione delle carceri e mantenimento della sicurezza. Tali strumenti sono utilizzati per aumentare efficienza ed equità, in particolare alleggerendo gli oneri amministrativi[213].

Si riscontra l'utilità di tali software quando è da decidere se applicare una misura cautelare o di prevenzione oppure è da valutare la sospensione di una pena, quindi per capire se un trasgressore potrebbe commettere altri crimini se non recluso.

Questi sistemi di processo decisionale algoritmico operano in condizioni di elevata incertezza con riguardo ai risultati, infatti qualora un giudice propenda per una scarcerazione anticipata non è in grado di fornire una probabilità del rischio di recidiva, la sua è una decisione incerta. La decisione inoltre può avere conseguenze di varia portata per la società, fino a rivelarsi dannosa qualora l'imputato commetta una nuova offesa[214].

Per risolvere tali problemi solitamente i giudici mettono in campo la loro esperienza e la logica, ma già oggi, e in futuro sicuramente di più, si fa ricorso ad algoritmi predittivi che elaborando una mole di informazioni possono rilevare correlazioni e informazioni nascoste all'occhio umano[215]. Quantificando una serie di evidenze che le persone possono considerare solo per date si ottiene una sorta di oggettività scientifica che

[211] "Why Jurisdictions Choose RATs," *Mapping Pretrial Injustice*, n.d., https://pretrialrisk.com/the-basics/the-case-for-rats/.
[212] "Understand Pretrial Justice," *Advancing Pretrial Policy & Research (APPR)*, n.d., https://advancingpretrial.org/pretrial-justice/pretrial-justice/.
[213] Anne L Washington, "HOW TO ARGUE WITH AN ALGORITHM: LESSONS FROM THE COMPAS-PROPUBLICA DEBATE," *The Colorado Technology Law Journal* 17.1 (2019): 37.
[214] Kathrin Hartmann and Georg Wenzelburger, "Uncertainty, Risk and the Use of Algorithms in Policy Decisions: A Case Study on Criminal Justice in the USA," *Policy Sci* 54.2 (2021): 269–87, https://doi.org/10.1007/s11077-020-09414-y.
[215] Basile, "Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine."

potrebbe rispondere a richiesta morale di imparzialità[216]. La "pericolosità criminale" viene analizzata con una metodologia attuariale o statistica, valutando il rischio con un approccio evidence-based. Vari possono essere i fattori presi in esame: sesso, età, etnia, istruzione, situazione sociale, precedenti, residenza, etc. Alcuni di questi fattori di rischio sono statici e non modificabili, altri sono dinamici e cambiano nel tempo, più o meno velocemente[217]. Ogni RAT considera una selezione di questi fattori e li combina pesandoli nel modo che ritiene più opportuno arrivando a produrre un punteggio: l' indice del rischio di recidiva. Rilevante è infatti il fenomeno del "dose-exposure relationship", secondo il quale un maggior numero di fattori di rischio, così come alcune caratteristiche dell'esposizione quali la durata e precocità, aumentano le probabilità di un comportamento violento[218].

Gli Stati Uniti rappresentano il Paese in cui tali software predittivi sono più diffusi. I RATs utilizzati per il pretrial release sono più di 20 e cambiano per ogni Stato, in alcuni di essi sono richiesti dalla legge, mentre in altri il loro uso deve essere autorizzato rendendo la materia molto eterogenea[219].

COMPAS

Uno dei più famosi e tristemente noto è COMPAS, Correctional Offender Management Profiling for Alternative Sanctions. È un software particolarmente elaborato che ha come input una serie di dati derivanti dal fascicolo dell'imputato e da interviste e questionari, per un totale di 137 variabili, e come output il rischio di recidiva che viene calcolato non su base individuale, ma in comparazione rispetto a casi analoghi[220]. La razza non è una variabile di input, ma può essere estrapolata. I risultati sono complessi nella loro interpretabilità e infatti sono state rivolte accuse verso questo algoritmo circa la sua accuracy, il grado in cui la misura rispecchia la realtà, e la fairness, grado in cui la misura è rappresentativa di discriminazioni[221].

---

[216] Hartmann and Wenzelburger, "Uncertainty, Risk and the Use of Algorithms in Policy Decisions."
[217] Washington, "HOW TO ARGUE WITH AN ALGORITHM: LESSONS FROM THE COMPAS-PROPUBLICA DEBATE."
[218] Basile, "Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine."
[219] Massaro et al., "Intelligenza artificiale e giustizia penale."
[220] Stefania Carrer, "Se l'amicus curiae è un algoritmo: il chiacchierato caso Loomis alla Corte Suprema del Wisconsin," *Giurisprudenza penale*, 24 April 2019, https://www.giurisprudenzapenale.com/2019/04/24/lamicus-curiae-un-algoritmo-chiacchierato-caso-loomis-alla-corte-suprema-del-wisconsin/.
[221] Francis X Diebold and Robert S Mariano, "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics* 20.1 (2002): 134–44, https://doi.org/10.1198/073500102753410444.

Nel 2016 in una inchiesta pubblicata da ProPublica è emerso come i risultati fossero distorti, andando a favorire persone bianche e sfavorendo quelle di colore, discriminando quindi gli afroamericani. La stessa utilità del RAT è stata messa in discussione poiché secondo uno studio la valutazione proveniente dall'algoritmo sarebbe comparabile a quella di una persona senza conoscenze specifiche in materia e per di più dei 137 fattori utilizzati solo due (età e numero totale di precedenti condanne) sono quelli che parrebbero avere un peso specifico elevato, facendo assimilare questo algoritmo ad un semplice classificatore lineare[222].

Nel 2016 nel caso Loomis è stato fatto ricorso presso la Corte suprema del Wisconsin poiché il grado di rischio indicato da COMPAS e la pena successivamente determinata erano ritenuti viziati da pregiudizi discriminatori e variabili ininfluenti che sovrastimano il rischio di recidiva individuale[223]. La Corte non ha comunque rigettato la sentenza considerando che l'utilizzo di COMPAS non ha violato il diritto ad un equo processo. Ha inoltre ribadito come i risultati di tali strumenti di calcolo del rischio siano illegittimi qualora utilizzati per determinare una sentenza, mentre legittimi se usati dal giudice in base alla propria discrezionalità e bilanciati con altri fattori[224].

Purtroppo non sono stati possibili studi circa il funzionamento di questo algoritmo, per esempio sul metodo di comparazione, poiché la società che l'ha ideato è privata e vi è una copertura progettuale tramite brevetto e segreto industriale[225].

PSA

Il PSA (Public Safety Assessment), sviluppato da una no-profit, è stato ideato per rimediare ai problemi sorti con il COMPAS e per dare ai giudici informazioni il più possibile imparziali. Le variabili di input sono rappresentate da nove fattori di rischio ben specifici:

- età all'arresto corrente;
- presenza di violenza o meno nel crimine commesso;
- accusa pendente al momento dell'arresto;

[222] Julia Dressel and Hany Farid, "The Accuracy, Fairness, and Limits of Predicting Recidivism," *Sci. Adv.* 4.1 (2018): eaao5580, https://doi.org/10.1126/sciadv.aao5580; Donati, "INTELLIGENZA ARTIFICIALE E GIUSTIZIA."
[223] Costa, "Intelligenza artificiale e Giustizia."
[224] Carrer, "Se l'amicus curiae è un algoritmo."
[225] Basile, "Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine."

- pregressa condanna per reati minori;
- pregressa condanna penale;
- pregressa condanna violenta;
- precedente mancata comparsa negli ultimi due anni;
- precedente mancata comparsa oltre i due anni;
- pregressa condanna di detenzione[226].

Si noti come origine etnica e razza non figurino tra questi fattori. L'output ottenuto è composto da vari punteggi ottenuti con una combinazione pesata dei vari fattori e indicanti tre diversi rischi: il rischio di un'assenza all'udienza in tribunale, il rischio di un nuovo reato e il rischio della commissione di un crimine violento. Tale algoritmo risulta essere molto più trasparente rispetto a COMPAS e dalla sua adozione il numero di persone rilasciate è decisamente aumentato[227]. In USA durante la fase del "parole", fase intermedia tra l'accusa e il processo, vi è la possibilità di essere rilasciati in seguito al pagamento di una cauzione. La libertà provvisoria è una prassi diffusa negli USA, ma potrebbe penalizzare i meno benestanti poiché più propensi a dichiararsi fin da subito colpevoli. Le persone che rimangono in cella perché non possono pagarsi la cauzione con l'applicazione di PSA sono diminuite. Infatti quel 40% di persone che, pur potendo essere rilasciata, non poteva permettersi di pagare la cauzione e quindi rimaneva in prigione, con PSA viene rilasciata senza alcun pagamento[228]. Il PSA viene per esempio ampiamente utilizzato nello Stato del New Jersey che, volendo riformare il sistema cautelare, ha provveduto a sostituire con un algoritmo le udienze per la concessione della libertà su cauzione. Tale valutazione serve da indicazione per la decisione finale del giudice. L'applicazione di questo algoritmo serve per rendere neutrale il calcolo del rischio relativo alla possibile minaccia o fuga di un imputato in attesa di un processo, se non incarcerato. Anche PSA è stato accusato di pregiudizi razziali, probabilmente derivanti da pregiudizi sociali, ma i benefici ottenuti a proposito di incarcerazione di massa e criminalizzazione della povertà sono notevoli[229].

---

[226] "How It Works," *Advancing Pretrial Policy & Research (APPR)*, n.d., https://advancingpretrial.org/psa/factors/; laura and john arnold foundation, "PSA: Risk Factors and Formula," 2016, https://craftmediabucket.s3.amazonaws.com/uploads/PDFs/PSA-Risk-Factors-and-Formula.pdf.
[227] Massaro et al., "Intelligenza artificiale e giustizia penale."
[228] Basile, "Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine."
[229] Ephrat Livni, "Nei tribunali del New Jersey è un algoritmo a decidere chi esce su cauzione," *Internazionale*, 3 March 2017, https://www.internazionale.it/notizie/ephrat-livni/2017/03/03/tribunali-algoritmo-cauzione.

Per l'utilizzo delle macchine in ambito giudiziario vi è quindi la necessità di un controllo umano significativo che può essere tradotto in alcune condizioni:

- il funzionamento dell'algoritmo reso pubblico;
- il tasso di errore reso noto;
- le spiegazioni per far corrispondere alla formula algoritmica una regola giuridica, affinché ci sia una massima chiarezza per tutti i soggetti coinvolti;
- la salvaguardia del contradditorio[230].

3.3.3 Rischi per i diritti umani

L'uso di sistemi decisionali algoritmici cambia il processo decisionale. In uno studio condotto intervistando sia funzionari pubblici che professionisti sul campo, quali poliziotti, ha evidenziato che questi soggetti sono consci del fatto che il software sia solamente un elemento aggiuntivo rispetto agli altri strumenti a disposizione, e non guidi da solo le decisioni. Ma la stessa esistenza di una scala di punteggio del rischio di recidiva – basso, medio, alto – provoca una discussione su una diversa prospettiva, legata appunto a tali punteggi. Le persone accolgono positivamente l'uso degli algoritmi e vedono i punteggi come informazioni aggiuntive, basate su ricerca ed evidenza, molto gradite in un ambiente caratterizzato da incertezza[231].

I rischi derivanti dall'utilizzo di software RATs sono vari e la stessa Vice Presidente degli Stati Uniti, Kamala Harris, ha mostrato preoccupazione per potenziali problemi quali disparità razziali o altri pregiudizi derivanti dall'uso dell'IA nel campo della giustizia penale[232].

Nel caso della valutazione della recidiva si rischierebbe in primis un'eccessiva generalizzazione poiché una valutazione specifica sull'individuo verrebbe a mancare. In caso di provvedimenti in fase cautelare, ossia in caso di pre-trial, il rischio si potrebbe scontrare con la tutela delle libertà personali. Vi sarebbe una sorta di cancellazione della

[230] di Giulio Ubertis, "INTELLIGENZA ARTIFICIALE, GIUSTIZIA PENALE, CONTROLLO UMANO SIGNIFICATIVO (?)" (n.d.): 15.
[231] Hartmann and Wenzelburger, "Uncertainty, Risk and the Use of Algorithms in Policy Decisions."
[232] Roberts et al., "Achieving a 'Good AI Society.'"

soggettività poiché una decisione robotica appare decisamente più obiettiva. Gli imputati potrebbero finire per essere visti come un problema e l'algoritmo come la soluzione. In caso di sentenza da parte della macchina, applicazione mai adottata al momento, vi sarebbe anche una de-responsabilizzazione[233].

Il secondo rischio è quello di possibili effetti discriminatori. Gli algoritmi spesso riproducono in digital le discriminazioni del mondo reale, a volte rendendole ancora più marcate. Questo potrebbe essere causato sia dal rischio di de-individualizzazione, sia alle possibili discriminazioni insite nel dataset di training. Una pulizia dei dati potrebbe risultare essere un'operazione impossibile in alcuni casi rendendo i pregiudizi meccanici inevitabili. Perché non procedere con un de-biasing manuale? Tale procedura dovrebbe essere effettuata da esperti informatici, dai programmatori del sistema IA, e ricadrebbero in capo a loro una serie di decisioni circa gli interessi e i valori in gioco[234]. Vi sarebbe una sorte di discrezionalità soggettiva in capo a chi struttura l'algoritmo. Vi sono vari elementi da tradurre in un diverso linguaggio, quello dei dati. Nella costruzione di un algoritmo sono diverse le fasi, tra cui data mining, data matching e data profiling, in cui vanno compiute scelte personali[235]. Come recita Signorato "l'algoritmo è ontologicamente condizionato dal sistema di valori e dalle intenzioni di chi ne commissiona la creazione e/o di chi lo crea"[236].

La qualità dei dati risulta essere fondamentale, ma non è sufficiente, infatti un ulteriore rischio riguarda l'opacità di alcun algoritmi, nel caso questi dovessero determinare una sanzione la non trasparenza rappresenterebbe una grossa pecca. Un giusto processo dovrebbe essere basato su motivazioni chiare, cosa che un software inaccessibile non permetterebbe[237]. La trasparenza permette di capire le logiche sottostanti al punteggio di un algoritmo e una migliore comprensione e fruizione di essi da parte dei giudici o dei decisori. Ma spesso la complessità nasce dall'interazione tra i dati di input e il meccanismo algoritmico. La trasparenza non serve per eliminare un problema, ma per alimentare un sano dibattito[238].

---

[233] Massaro et al., "Intelligenza artificiale e giustizia penale."
[234] Donati, "INTELLIGENZA ARTIFICIALE E GIUSTIZIA."
[235] Ubertis, "INTELLIGENZA ARTIFICIALE, GIUSTIZIA PENALE, CONTROLLO UMANO SIGNIFICATIVO (?)."
[236] Silvia Signorato, "Giustizia Penale e Intelligenza Artificiale. Considerazioni in Tema Di Algoritmo Predittivo," *Rivista Di Diritto Processuale* 75.2 (2020): 605–16.
[237] Carrer, "Se l'amicus curiae è un algoritmo."
[238] Oswald, "Algorithm-Assisted Decision-Making in the Public Sector."

L'utilizzo di algoritmi potrebbe comportare anche una possibile manipolazione dell'output, che avrebbe ripercussioni sulle stesse valutazioni giuridiche[239]. Tramite la trasparenza, infatti, si garantisce anche l'equità procedurale, sintomo di giustizia naturale, e cioè il controllo sul processo decisionale. In giurisprudenza una correttezza delle motivazioni è concatenata ad una adeguatezza formale. In caso di previsioni o raccomandazioni derivanti da un sistema di intelligenza artificiale sarebbe opportuno analizzare le condizioni utili per spiegare il risultato: i dati di training corrispondono all'attuale situazione? Se vi sono bias o errori cosa riguardano? Come vengono considerati i diritti umani? La granularità della spiegazione varia a seconda del contesto[240].

In ogni caso la valutazione tramite l'ausilio dell'intelligenza artificiale dovrebbe essere eseguita da figure competenti e preparati, onde evitare bias cognitivi. Quando un giudizio è influenzato dalle informazioni precedenti si verifica un "ancoraggio", fenomeno noto soprattutto nel pricing. Questo porta ad un eccessivo affidamento alla decisione algoritmica, anche qualora fosse solo ausiliaria, e ad una conseguente de-responsabilizzazione della figura del giudice[241].

È interessante notare come anche il "dominio" di cui stiamo trattando sia particolarmente sensibile. La modellazione statistica degli algoritmi in altri campi, come quello della previsione dei terremoti o del marketing, si avvale di dati più affidabili e una diversa accettabilità di falsi positivi. Nella giustizia algoritmica vi deve essere un diverso grado di ottimizzazione. Già la dottrina giuridica di per sé è il risultato di un bilanciamento, poiché rappresenta l'equilibrio raggiunto tra il volere di diversi soggetti in termini di valori come efficienza ed equità: è un bilanciamento di compromessi. Tale bilanciamento deve essere poi trasmesso a livello algoritmico, cosa che non è successa per esempio con COMPAS.

L'equità ha varie sfaccettature, potrebbe consistere in una parità di trattamento oppure in una parità di risultato. In COMPAS la trasposizione del concetto di equità in algoritmo è avvenuta con l'accezione di "parità predittiva", secondo la società proprietaria la probabilità di recidiva tra i trasgressori ad alto rischio è la stessa indipendentemente

[239] Carrer, "Se l'amicus curiae è un algoritmo."
[240] Oswald, "Algorithm-Assisted Decision-Making in the Public Sector."
[241] Massaro et al., "Intelligenza artificiale e giustizia penale."

dalla razza. Vi è stata però un'errata classificazione delle persone di colore che sono state catalogate come a rischio più elevato per la commissione di un nuovo crimine, ma tale fatto non ha poi trovato riscontro nella realtà. Vi è stato quindi un enorme numero di falsi positivi causato dai dati di training che vedeva un maggior numero di crimini compiuto da persone di colore, questa è stata questa, quindi, la fonte di discriminazione. Le definizioni esistenti di equità o accuratezza sono varie sia eticamente che matematicamente e le trasposizioni da vari domini possono provocare distorsioni[242].

Spesso con il concetto filosofico di equità si intende la non discriminazione, ma questa interpretazione non è facilmente trasponibile in equità logaritmica. Se tale concetto non può essere applicato bisogna trovarne un altro. Un principio filosofico che viene spesso applicato per cogliere le iniquità è quello dell'egalitarismo, col quale si ritiene che le persone debbano essere trattate ugualmente, giustificando a volte anche una disuguaglianza distributiva. In particolare in questo ambito operativo vengono considerate le misure di equità di gruppo che verificano che una proprietà dell'algoritmo decisionale, calibrata sui membri di diversi gruppi socio-demografici, sia uguale in questi gruppi. La media rivela gli effetti sistematici della decisione, l'equità è l'uguaglianza di tali gruppi rispetto ad una qualche proprietà[243].

Le valutazioni di rischio prodotte dall'intelligenza artificiale misurano le correlazioni e non le cause, il loro risultato è una probabilità e non una certezza. Questo significa che le previsioni assumono la stabilità di alcuni fattori e non sono informazioni da considerare come oggettive. Abbiamo per esempio detto che COMPAS confronta i dati dell'individuo con quelli della popolazione di riferimento, quindi dati sul comportamento di popolazioni passate. Popolazioni della stessa categoria in città diverse potrebbero avere comportamenti differenti, potrebbe essere utile tenere conto delle tendenze reali delle

[242] Aleš Završnik, "Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings," *European Journal of Criminology* 18.5 (2021): 623–42, https://doi.org/10.1177/1477370819876762; Dressel and Farid, "The Accuracy, Fairness, and Limits of Predicting Recidivism"; Julia Angwin Mattu Jeff Larson,Lauren Kirchner,Surya, "Machine Bias," *ProPublica*, n.d., https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=ApJt7l5BZplugeIernGirtWPn_a7EraR; Cindy Redcross et al., "Evaluation of Pretrial Justice System. Reforms That Use the Public Safety Assessment," *Pretrial Justice Reform Study, Mecklenburg County* (2019); "A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased against Blacks. It's Actually Not That Clear.," *Washington Post*, n.d., https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/.
[243] Geoffrey C Barnes, "Classifying Adult Probationers by Forecasting Future Offending" (2012): 64.

popolazioni con i tassi di base, un indicatore che migliora la distribuzione della popolazione, contribuendo all'equità così come la abbiamo appena descritta[244].

Mentre poi i modelli predittivi in altri campi possono permettersi di utilizzare qualsiasi tipo di dato ritengano necessario, nel caso in analisi, trattandosi di un servizio pubblico, è grande la necessità di comprensione[245]. Quello della trasparenza abbiamo visto essere un rischio, ma visto il contesto andrebbe minimizzato il più possibile. Perseguire chiarezza e semplicità non è detto che porti a risultati predittivi peggiori. Infatti l'accuratezza potrebbe essere mantenuta stabile anche semplificando i modelli e utilizzando un numero più esiguo di variabili, così da consentirne una migliore interpretazione[246]. Per ottenere una migliore predizione spesso i data scientists inseriscono predittori per loro rilevanti, ma che lo sono meno per un avvocato o un giudice. Infatti aggiungere una ulteriore variabile per migliorare il potere predittivo ha una bassa "penalità", ossia non comporta alcuna particolare complicazione[247]. La precisione però potrebbe non essere il principale beneficio ricercato da un ente pubblico[248]. Spesso le variabili predittive dei RATs riguardano l'età, l'uso di sostanza, l'età del primo arresto, la stabilità residenziale e l'occupazione. Ogni algoritmo valuta questi fattori in modo diverso, ma tali variabili secondo molti studi risultano insignificanti a fini predittivi. Questi fattori demografici e socioeconomici si collegano implicitamente alla razza e quindi causano discriminazione[249].

Secondo alcuni esperti in futuro, tramite algoritmi migliori e un continuo processo di autoapprendimento, tali problemi potrebbero essere risolti e la tecnologia dell'intelligenza artificiale potrebbe vedere crescere la sua considerazione anche in campo giuridico.

L'interpretazione del giudice risulta essere fondamentale anche per questo. Deve conoscere i fattori che compongono gli algoritmi e bilanciare la valutazione con altri. La discrezionalità soggettiva è importante poiché spesso sono vari i criteri e le motivazioni in gioco. Ma, a onore del vero, anche la stessa interpretazione umana può essere

[244] Washington, "HOW TO ARGUE WITH AN ALGORITHM: LESSONS FROM THE COMPAS-PROPUBLICA DEBATE."
[245] Mattu, "Machine Bias."
[246] Washington, "HOW TO ARGUE WITH AN ALGORITHM: LESSONS FROM THE COMPAS-PROPUBLICA DEBATE."
[247] Barnes, "Classifying Adult Probationers by Forecasting Future Offending."
[248] Oswald, "Algorithm-Assisted Decision-Making in the Public Sector."
[249] "Demographic Bias," *Mapping Pretrial Injustice*, n.d., https://pretrialrisk.com/the-danger/demographic-bias/.

macchiata da pregiudizi. Una persona è infatti portatrice di ideali, valori ed orientamenti politici. È stato verificato come i giudici siano più propensi a concedere la libertà vigilata subito dopo aver mangiato[250]. Si è quasi costretti a scegliere tra bias meccanici e quelli umani, infatti possiamo dire che la discrezionalità che contraddistingue un processo decisionale in caso di applicazione di strumenti IA venga in parte spostata a valle, cioè nella fase di programmazione.

I rischi appena visti, come è facilmente intuibile, potrebbero comportare un impatto, in senso negativo, sui diritti umani. I vulnus più rilevanti riguarderebbero i principi di discriminazione e di uguaglianza. Affinché sia pensabile adottare algoritmi predittivi è necessario verificare la presenza di discriminazioni nei dati utilizzati come input, inoltre la struttura e il funzionamento dell'algoritmo devono essere trasparenti, i giudici devono essere formati e bisogna comunque considerare che il risultato dell'algoritmo è una mera indicazione e non una verità inconfutabile. Al momento più che un completo affidamento su sistemi di decisione automatizzati è auspicabile una coesistenza, riconoscendo punti di forza e debolezza del decisore algoritmico e di quello umano e integrandoli per sopperire appunto alle mancanze. Pensare che l'adozione di strumenti di intelligenza artificiale sia la soluzione per problemi discriminatori ovviamente è sbagliato. Ma l'automazione di alcuni processi può sicuramente portare ad un processo o procedure più informate, permettendo ai decisori scelte più consapevoli.

[250] Završnik, "Algorithmic Justice."

Capitolo 4

Analisi del Risk Assessments Tool Savry

4.1 Introduzione – il perché della nostra analisi

L'intelligenza artificiale e in particolare il machine learning hanno la capacità di scovare correlazioni tra i dati analizzando grandi dataset, questa è un'attività in cui riescono meglio rispetto agli esseri umani.

Una delle applicazioni in cui i sistemi algoritmici possono essere di supporto alle decisioni umane è quella della predizione della recidiva criminale. Abbiamo visto nel precedente capitolo i motivi per cui questi software vengono adottati e principali problemi che possono sorgere, quali pregiudizi e discriminazioni, causati dall'opacità dell'algoritmica.

Tale analisi prende spunto dallo studio condotto da Emilia Gomez: "Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia". Nel loro lavoro hanno analizzato le prestazioni predittive e l'ingiustizia dei metodi di machine learning nella previsione della recidiva giovanile, concentrandosi su un algoritmo esistente, Savry[251].

In particolare hanno analizzato tramite vari algoritmi tre diversi set di variabili: nel primo hanno incluso tutte le considerate da Savry, nel secondo hanno inserito le variabili non considerate da Savry, ossia quelle socio-demografiche, mentre nel terzo hanno unito questi primi due modelli.

Noi proveremo a selezionare le variabili per creare un modello e comparare le nostre performance predittive con quelle degli altri modelli. La selezione iniziale avverrà secondo un metodo qualitativo, avvalendosi di tutte le nozioni apprese durante la stesura di tale elaborato.

[251] Songül Tolan et al., "Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia," in *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law* (ICAIL '19: Seventeenth International Conference on Artificial Intelligence and Law Montreal QC Canada: ACM, presented at the ICAIL '19: Seventeenth International Conference on Artificial Intelligence and Law, 2019), 83–92, https://doi.org/10.1145/3322640.3326705.

La scelta delle variabili è infatti un passaggio cruciale, possono esistere centinaia di algoritmi che considerano una combinazione sempre diversa delle stesse variabili. I trade-off sui quali si deve ragionare riguardano l'accuratezza, la spiegabilità e l'eticità.

Cercheremo quindi di vedere con mano cosa includere nella progettazione di un algoritmo equo.

4.2 Structured Assesment of Violence Risk in Youth

Lo Structured Assesment of Violence Risk in Youth, Savry, è uno strumento utilizzato per la valutazione del rischio di recidiva violenta, in particolare un risk assessment per maschi e femmine di età compresa tra i dodici e i diciotto anni.

Quello che lo contraddistingue, a differenza di COMPAS, è l'essere un SPJ, ossia un "Strucutred Professional Judgment". Tale RAT consiste in un approccio clinico empiricamente guidato che porta ad un giudizio professionalmente strutturato. Vengono fornite linee guida per la valutazione del rischio in modo strutturato garantendo però flessibilità per un adeguamento alle peculiarità dei singoli, lasciando quindi una grande libertà agli esperti. È infatti un software aperto e completamente interpretabile da parte dei professionisti[252].

Struttura

Per quanto riguarda la struttura Savry è composto da due tipologie di variabili di input.

Vi sono i fattori di rischio che sono 24, codificati su tre livelli di gravità - basso quando non presenti, medio quando presente raramente o lievemente, alto se evidente o persistente - e suddivisi in tre aree: fattori individuali, fattori storici e fattori sociali.
I fattori di rischio storici sono quelli più statici, rappresentando il percorso dell'adolescente sono difficili da cambiare.
I fattori di rischio sociali/contestuali servono a rappresentare l'influenza dei rapporti interpersonali o con le istituzioni, mentre i fattori di rischio individuale considerano aspetti comportamentali e psicologici. Entrambi sono dinamici poiché rappresentano

[252] Inge Hempel et al., "Review of Risk Assessment Instruments for Juvenile Sex Offenders: What Is Next?," *Int J Offender Ther Comp Criminol* 57.2 (2013): 208–28, https://doi.org/10.1177/0306624X11428315.

aspetti che possono cambiare[253].
Questi 24 fattori di rischio sono sintetizzati in un punteggio totale.

Vi sono inoltre 6 fattori di protezione codificati come presenti o assenti che vanno a formare un ulteriore punteggio e considerano le risorse a disposizione del soggetto come ad esempio il supporto sociale.

I 30 fattori totali sono visibili nella figura numero 3.

Infine un esperto, sulla base del precedente assessment, assegna una valutazione finale indicante il rischio di recidiva violenta.

La corrispondenza tra la scala di severità dei fattori è ben definita e spiegata all'interno del manuale di tale strumento. I professionisti sono informati del fatto che esistendo differenze di sesso alcuni fattori devono essere valutati diversamente.

Historical Items		*Social/Contextual Items*	
1.	History of violence	11.	Peer delinquency
2.	History of nonviolent offending	12.	Peer rejection
3.	Early initiation of violence	13.	Stress and poor coping
4.	Past supervision/intervention failures	14.	Poor parental management
5.	History of self-harm or suicide attempts	15.	Lack of personal/social support
6.	Exposure to violence in the home	16.	Community disorganization
7.	Childhood history of maltreatment		
8.	Parental/caregiver criminality		
9.	Early caregiver disruption		
10.	Poor school achievement		
Individual Items		*Protective Items*	
17.	Negative attitudes	P1.	Prosocial involvement
18.	Risk taking/impulsivity	P2.	Strong social support
19.	Substance use difficulties	P3.	Strong attachments and bonds
20.	Anger management problems	P4.	Positive attitude toward intervention and authority
21.	Low empathy/remorse	P5.	Strong commitment to school or work
22.	Attention deficit/hyperactivity difficulties	P6.	Resilient personality
23.	Poor compliance		
24.	Low interest/commitment to school or work		

Figura 3. Features di Savry[254]

[253] "Structured Assessment of Violence Risk in Youth (SAVRY): Adattamento Italiano - QI - Questioni e Idee in Psicologia - Il Magazine Online Di Hogrefe Editore," n.d., https://qi.hogrefe.it/rivista/structured-assessment-violence-risk-youth-savry-ad/.
[254] Henny P.B. Lodewijks, Theo A.H. Doreleijers, and Corine De Ruiter, "Savry Risk Assessment in Violent Dutch Adolescents: Relation to Sentencing and Recidivism," *Criminal Justice and Behavior* 35.6 (2008): 696–709, https://doi.org/10.1177/0093854808316146.

4.3 Metodologia

Per poter confrontare i risultati con quelli dello studio del paper di riferimento ho cercato di riprodurre, per quanto possibile in base alle mie conoscenze e alle informazioni disponibili, le stesse condizioni.

Dataset

Il dataset scelto comprende 4753 osservazioni che si riferiscono a ragazzi minorenni, tra 12 e 17 anni, che hanno finito di scontare una pena nel 2010 per un reato commesso. Il sistema giurisdizionale è quello della Catalogna. La fonte dei dati è il Centre d'Estudis Jurídics i Formació Especialitzada[255].

Come nello studio di riferimento ci concentreremo sul sotto-campione di 855 persone che sono state sottoposte al programma Savry.

Nel dataset sono presenti dati demografici relativi a età, sesso ed etnia delle persone e i punteggi di tutti i fattori del software Savry. Per verificare il comportamento di recidiva sono raccolti dati a fine 2013 e a fine 2015, noi sceglieremo come variabile di output gli esiti del 2015.

Algoritmo di apprendimento

Poiché la recidiva ha due possibili esiti, una persona è recidiva o non lo è, tale problema può essere affrontato come una classificazione. Per poter paragonare i risultati del nostro modello a quelli dello studio, utilizzeremo diversi algoritmi di apprendimento supervisionato: una regressione logistica, un random foresti e un gradient boosting.

Scelta variabili

Questa fase è particolarmente rilevante nella nostra analisi.

La variabile di output è l'esito di recidiva nell'anno 2015.

[255] "Recidivism in Juvenile Justice," *Centre d'Estudis Jurídics i Formació Especialitzada*, n.d., http://cejfe.gencat.cat/en/recerca/opendata/jjuvenil/reincidencia-justicia-menors/index.html.

Nella selezione delle variabili di input invece sta parte del nostro lavoro. Tali variabili sono state infatti scelte in base alle conoscenze apprese durante la stesura di tale elaborato.
Un primo requisito consiste nel rispetto dei principi di equità e uguaglianza. Si vuole quindi creare un algoritmo equo alla base. Seguendo tale considerazione non sono stati inserite features quali il sesso e la nazionalità.
Un secondo requisito è quello della spiegabilità. È nostro intento quello di cercare di diminuire il numero di variabili per rendere la formazione del punteggio finale più comprensibile per gli utilizzatori, quali giudici o altri professionisti, andando sicuramente a perdere qualcosa in quanto ad accuratezza predittiva, ma non eccessivamente.

Inoltre nella selezione delle variabili sono state considerate quelle presenti in altri algoritmi quali per esempio PSA.

Le variabili così selezionate riguardano l'età (V5_edat_fet_agrupat), il numero di crimini commessi (V21_fet_nombre), gli antecedenti (V11_antecedents, V65_1_violencia_previa), l'esposizione a violenza fin dall'infanzia (V67_3_inici_precoç_violencia, V70_6_exposicio_violencia_llar), genitori criminali (V72_8_delinquencia_pares), l'assenza di supporto sociale (V79_15_manca_suport_personal_socia), il maltrattamento infantile (V71_7_historia_maltracte_infantil) e il tempo che manca al processo (V28_temps_inici), bassi rendimenti scolastici () e maltrattamenti infantili (V74_10_baix_rendiment_escola).

Metrica per la valutazione della performance

Per la valutazione della performance predittiva dei nostri modelli di machine learning usiamo la metrica AUC (Area Under the Curve)-ROC (Receiver Operating Characteristics). L'area sotto la curva rappresenta l'estensione del valore predittivo della variabile, indicando quindi quanto un modello è in grado di distinguere tra classi, maggiore è l'AUC, migliore sarà il modello. Un AUC esemplificativo di 0,6 significa che vi è una probabilità del 60% che il modello possa distinguere tra le due classi. Con un AUC

di 0.5 si forma una diagonale e il modello non è in grado di fare alcuna discriminazione[256].

4.4 Implementazione

1) Preparazione dei dati

Questa fase include:

- il caricamento dei dati

```
Minori <- read_excel("C:/Users/frass/OneDrive/Desktop/R/tesi/reincidenciaJusticiaMenors.xlsx")
```

- la selezione di quelle persone che hanno partecipato al programma. In concomitanza per praticità trasformiamo anche la nostra variabile di output facendola diventare numerica: 1 = recidivo, 0 = non recidivo.

```
MinoriSavry <- Minori %>%
  filter(V54_SAVRYprograma == "Internament" | V54_SAVRYprograma == "Llibertat vigilada") %>%
  mutate(V115_reincidencia_2015 = ifelse(V115_reincidencia_2015 == "No",0,1))
```

- la pulizia e sistemazione dei dati

Dopo aver verificato, iniziamo pulendo da eventuali valori nulli

```
MinoriSavry <- MinoriSavry[!is.na(MinoriSavry$V5_edat_fet_agrupat),]
```

Cambiamo poi il nome ad alcune variabili per togliere il simbolo della chiocciola che può produrre errori di sintassi.

```
MinoriSavry <- rename(MinoriSavry, V65_1_violencia_previa = "V65_@1_violencia_previa", V67_3_inici_precoç_violencia = "V67_@3_inici_precoç_violencia", V70_6_exposicio_violencia_llar = "V70_@6_exposicio_violencia_llar", V71_7_historia_maltracte_infantil = "V71_@7_historia_maltracte_infantil", V72_8_delinquencia_pares = "V72_@8_delinquencia_pares", V74_10_baix_rendiment_escola = "V74_@10_baix_rendiment_escola", V79_15_manca_suport_personal_social = "V79_@15_manca_suport_personal_social")
```

Infine trasformiamo le variabili di input da noi selezionate in fattori per permettere ai nostri algoritmi di lavorare meglio.

```
MinoriSavry$V5_edat_fet_agrupat = factor(MinoriSavry$V5_edat_fet_agrupat)
MinoriSavry$V11_antecedents=factor(MinoriSavry$V11_antecedents)
MinoriSavry$V15_fet_agrupat = factor(MinoriSavry$V15_fet_agrupat)
MinoriSavry$V16_fet_violencia = factor(MinoriSavry$V16_fet_violencia)
MinoriSavry$V65_1_violencia_previa=factor(MinoriSavry$V65_1_violencia_previa)
MinoriSavry$V67_3_inici_precoç_violencia=factor(MinoriSavry$V67_3_inici_pr
```

[256] Sarang Narkhede, "Understanding AUC - ROC Curve," *Medium*, 15 June 2021, https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5.

```
ecoç_violencia)
MinoriSavry$V70_6_exposicio_violencia_llar=factor(MinoriSavry$V70_6_exposi
cio_violencia_llar)
MinoriSavry$V71_7_historia_maltracte_infantil=factor(MinoriSavry$V71_7_his
toria_maltracte_infantil)
MinoriSavry$V72_8_delinquencia_pares=factor(MinoriSavry$V72_8_delinquencia
_pares)
MinoriSavry$V79_15_manca_suport_personal_social=factor(MinoriSavry$V79_15_
manca_suport_personal_social)
```

2) Creazione dataset di training e test

Crediamo ora i due dataset: uno per la fase di training dei modelli e l'altro per quella di test. Verifichiamo inoltre che la nostra variabile di output sia opportunamente distribuita in tutti quanti i dataset.

```
set.seed(1235)
sample_set <- sample(nrow(MinoriSavry), round(nrow(MinoriSavry)*.65), repl
ace = T)
train <- MinoriSavry[sample_set, ]
test <- MinoriSavry[-sample_set, ]

round(prop.table(table(select(MinoriSavry, V115_reincidencia_2015), exclud
e = NULL)), 4) * 100

##
##     0     1
## 62.35 37.65

round(prop.table(table(select(train, V115_reincidencia_2015), exclude = NU
LL)), 4) * 100

##
##     0     1
## 64.31 35.69

round(prop.table(table(select(test, V115_reincidencia_2015), exclude = NUL
L)), 4) * 100

##
##     0     1
## 61.28 38.72
```

3) Modellazione

Tutti e tre i nostri modelli avranno le stesse variabili di input e output e medesimo dataset di training e test.

Il primo modello che andremo a configurare è quello della regressione logistica

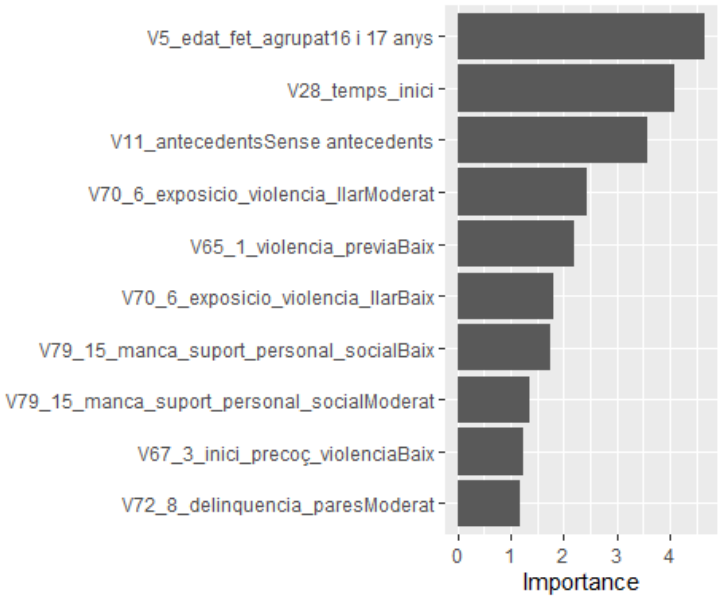
A) Logit model

La regressione logistica multipla viene utilizzata per prevedere la probabilità di appartenenza alla classe in base a più variabili predittive.

Nel calcolare questo modello utilizziamo il dataset di training e poiché il nostro è un problema di classificazione, tra le opzioni di default, specifichiamo il tipo di famiglia ossia binomiale e logit.

Plottiamo inoltre l'importanza relativa delle variabili utilizzate come regressori.

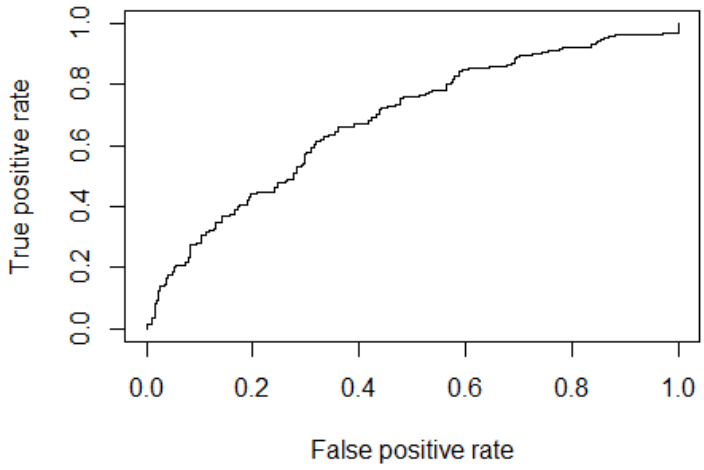
```
fit <- glm(V115_reincidencia_2015 ~ V5_edat_fet_agrupat + V11_antecedents
+ V15_fet_agrupat + V16_fet_violencia + V21_fet_nombre + V28_temps_inici +
 V65_1_violencia_previa + V67_3_inici_precoç_violencia +V70_6_exposicio_vi
olencia_llar + V71_7_historia_maltracte_infantil + V72_8_delinquencia_pare
s  + V79_15_manca_suport_personal_social, data=train, family=binomial(link
= "logit"))

vip::vip(fit)
```



Prediction

Per valutare le performance abbiamo detto utilizzeremo la metrica della curva auc-roc, oltre a ottenere il coefficiente relativo ne plottiamo anche il grafico.

```
prob <- predict(fit, newdata=test, type="response")

## Warning in predict.lm(object, newdata, se.fit, scale = 1, type = if (ty
pe == :
## prediction from a rank-deficient fit may be misleading

pred <- prediction(prob, test$V115_reincidencia_2015)
perf <- performance(pred, measure = "tpr", x.measure = "fpr")
plot(perf)
```



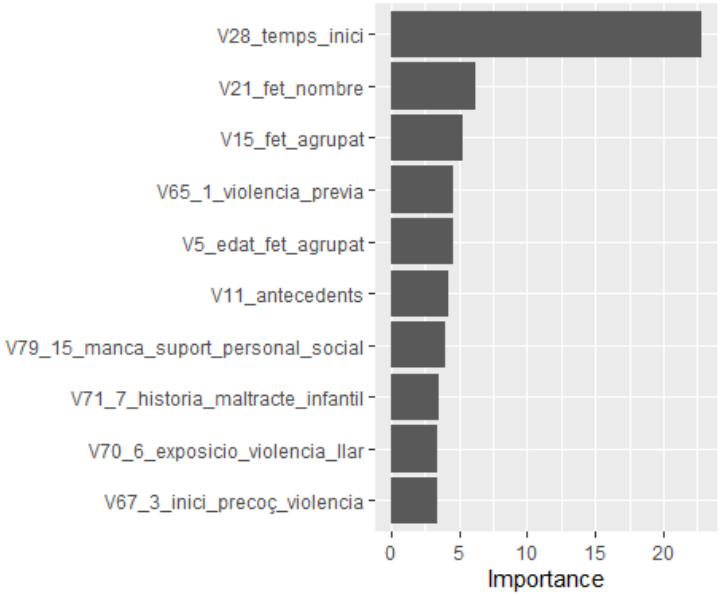

```
auc <- performance(pred, measure = "auc")
auc <- auc@y.values[[1]]
auc

## [1] 0.6829434
```

B) Random forest model

Il random forest è un metodo di apprendimento che consiste nella costruzione di molti alberi decisionali de-correlati Ogni albero predice il modello in modo indipendente e poi nel caso della classificazione l'output sarà la classe che esce in più alberi.

Procediamo con il nostro secondo modello di cui plottiamo anche in questo caso l'importanza delle variabili.

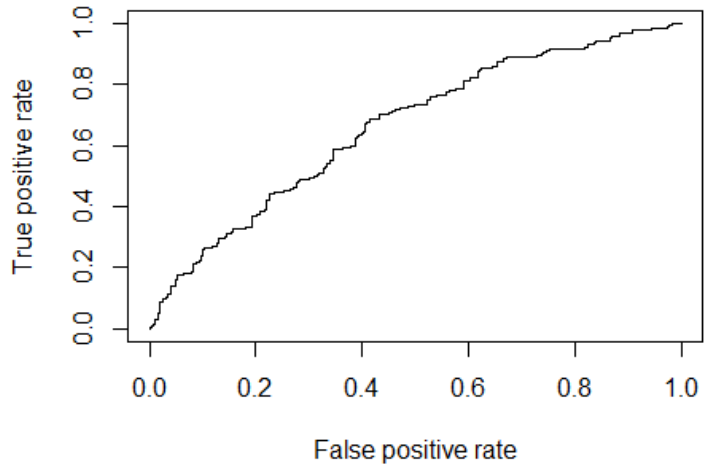
```
rf.fit <- randomForest(V115_reincidencia_2015 ~ V5_edat_fet_agrupat +  V11
_antecedents + V15_fet_agrupat + V16_fet_violencia + V21_fet_nombre + V28_
temps_inici + V65_1_violencia_previa + V67_3_inici_precoç_violencia + V70_
6_exposicio_violencia_llar + V71_7_historia_maltracte_infantil + V72_8_del
inquencia_pares  + V79_15_manca_suport_personal_social, data = train, ntre
e = 2000, nodesize = 20)

## Warning in randomForest.default(m, y, ...): The response has five or fe
wer
## unique values. Are you sure you want to do regression?

vip::vip(rf.fit)
```



Prediction

Procediamo con la valutazione della performance predittiva di questo modello, ottenendo il relativo coefficiente e grafico auc-roc.

```
prob <- predict(rf.fit, newdata=test, type="response")
pred <- prediction(prob, test$V115_reincidencia_2015)
perf <- performance(pred, measure = "tpr", x.measure = "fpr")
plot(perf)
```




```
auc <- performance(pred, measure = "auc")
auc <- auc@y.values[[1]]
auc

## [1] 0.6615132
```

C) Gradient boosting model

L'idea dietro al boosting invece è quella di aggiungere nuovi modelli in modo sequenziale, l'addestramento del nuovo modello sarà influenzato dagli errori di quelli precedenti.

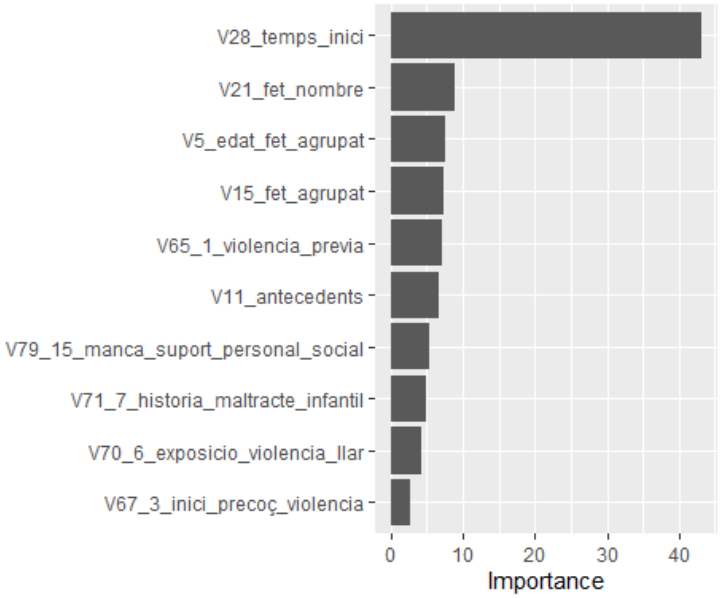
Prima di richiamare il nostro terzo modello, il gradient boosting, proviamo a migliorarne l'accuratezza ricercando i parametri più appropriati. Tale procedura si chiama "tuning" dei parametri e avviene tramite una griglia di ricerca che ci permetterà di testare in modo automatico la combinazione di più parametri indicandoci il modello più accurato. Questa è una fase che può richiedere un tempo di elaborazione abbastanza lungo, a seconda della potenza del nostro computer.

```
hyper_grid <- expand.grid(
  shrinkage = c(.01, .05, .1),
  interaction.depth = c(3, 5, 7),
  #n.minobsinnode = 5,
  #bag.fraction = .65,
  train.fraction = 1,
  n.minobsinnode = c(5, 7, 10),
  bag.fraction = c(.65, .8, 1),
    optimal_trees = 0,
    min_RMSE = 0
)

for(i in 1:nrow(hyper_grid)) {

  # reproducibility
  set.seed(123)

  # train model
  gbm.tune <- gbm(
    formula = V115_reincidencia_2015 ~ V5_edat_fet_agrupat +  V11_antecede
nts +  V21_fet_nombre +  V28_temps_inici + V65_1_violencia_previa + V67_3_
inici_precoç_violencia +
 V70_6_exposicio_violencia_llar + V71_7_historia_maltracte_infantil + V72_
8_delinquencia_pares + V74_10_baix_rendiment_escola + V79_15_manca_suport_
personal_social,
    distribution = "bernoulli",
    data = train,
    n.trees = 6000,
    interaction.depth = hyper_grid$interaction.depth[i],
    shrinkage = hyper_grid$shrinkage[i],
    n.minobsinnode = hyper_grid$n.minobsinnode[i],
    bag.fraction = hyper_grid$bag.fraction[i],
    train.fraction = .75,
    n.cores = NULL,
    verbose = FALSE
  )

  hyper_grid$optimal_trees[i] <- which.min(gbm.tune$valid.error)
  hyper_grid$min_RMSE[i] <- sqrt(min(gbm.tune$valid.error))
}

hyper_grid %>%
```

```
  dplyr::arrange(min_RMSE) %>%
  head(10)
```

Una volta trovati i parametri più performanti li inseriamo nel nostro modello.

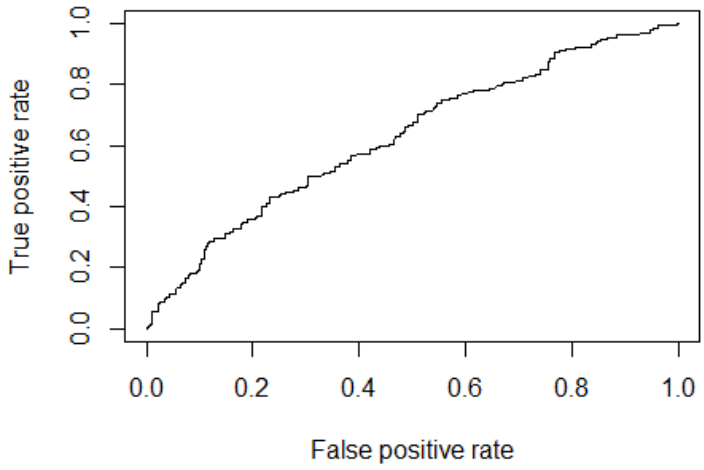
```
gbm.fit <- gbm(formula = V115_reincidencia_2015 ~ V5_edat_fet_agrupat +  V
11_antecedents + V15_fet_agrupat + V16_fet_violencia + V21_fet_nombre + V2
8_temps_inici + V65_1_violencia_previa + V67_3_inici_precoç_violencia +
 V70_6_exposicio_violencia_llar + V71_7_historia_maltracte_infantil + V72_
8_delinquencia_pares  + V79_15_manca_suport_personal_social,
distribution = "bernoulli",
data = train,
n.trees = 197,
interaction.depth = 5,
shrinkage = 0.1,
n.minobsinnode = 5,
bag.fraction = 1,
cv.folds = 5,
n.cores = NULL,
verbose = FALSE)

vip::vip(gbm.fit)
```



Prediction

E procediamo alla previsione anche in questo caso.

```
prob <- predict(gbm.fit, newdata=test, type="response")

## Using 82 trees...

pred <- prediction(prob, test$V115_reincidencia_2015)
perf <- performance(pred, measure = "tpr", x.measure = "fpr")
plot(perf)
```




```
auc <- performance(pred, measure = "auc")
auc <- auc@y.values[[1]]
auc

## [1] 0.628548
```

4.5 Risultati e discussione

Due sono le riflessioni che si possono fare in seguito a questa analisi: una specifica, mentre la seconda più generalizzabile.

Importanza delle variabili

La prima riguarda l'importanza relativa delle variabili utilizzate nella costruzione dei nostri modelli, quindi l'importanza delle variabili da noi selezionare nel predire la recidiva. Seppur le variabili da noi utilizzate come input risultino essere sempre le stesse possiamo notare come il livello di significatività di ognuna nei rispettivi modelli cambi. L'importanza di alcune variabili è ricorrente, queste sono l'età, il numero dei crimini commessi precedentemente e il tempo che manca al processo. La selezione da noi effettuata potrebbe in partenza influenzare l'utilità di un tale tipo di riflessione poiché sicuramente saranno state escluse delle variabili significative per l'accuratezza predittiva, e il nostro obiettivo non vuole infatti essere quello di trovare le features maggiormente importanti.

Cosa vuol dire performance

La seconda riflessione, non specifica al solo nostro lavoro e generalizzabile, riguarda invece la performance e il suo significato. Riportiamo nella tabella sottostante i punteggi delle metriche auc-roc ottenute nei nostri modelli e in quelli dello studio di riferimento[257].

	logit	rf	gb
SAVRY ML	.66	.65	
Static ML	.70	.66	
Static + SAVRY ML	.71	.69	
Gabriele ML	.68	.66	.63

I modelli "SAVRY ML" sono stati elaborati basandosi su più di 30 variabili, quelli che fanno riferimento alle solo variabili demografiche, "Static ML", ne hanno considerate 15 e la loro somma, "Static + SAVRY ML" vede considerato un numero di variabili che si avvicina a 50. Il nostro set comprende 10 variabili.

Confrontando i risultati si può notare come i nostri punteggi siano leggermente inferiore, ma non eccessivamente. Questo seppur il numero di variabili ridotto.
Un minor numero di variabili consente una maggior comprensione del funzionamento dell'algoritmo, garantendo quindi una maggior spiegabilità. RAT quali COMPAS che considerano centinaia di fattori finiscono per essere accurati, ma poco trasparenti. Inoltre nella stessa opacità è possibile e anzi altamente probabile si nascondano rischi etici.

Facendo dei test per la scelta delle variabili aggiungendone una relativa al basso rendimento scolastico vi sarebbe stato un aumento di .01 in ogni algoritmo. Ma in questo caso appare evidente come seppur la variabile possa essere altamente significativa, è anche discriminatoria.

[257] Tolan et al., "Why Machine Learning May Lead to Unfairness."

Questi strumenti di sicuro possono essere un aiuto sostanziale nel valutare il rischio di recidiva, ma l'affidabilità predittiva non è da considerarsi altissima. Bisogna inoltre considerare che di giovani stiamo parlando e il loro rapido sviluppo rende instabili alcuni fattori, provocando incertezza che sembra andare in una direzione opposta rispetto alle conseguenze sicure a lungo termine di queste decisioni. Nonostante la natura clinica degli SPJ si potrebbe dare maggior peso a comportamenti e tendenze psicologiche[258].

La performance fine a se stessa non è tutto, vi sono una serie di fattori da considerare e nella stesura di questo elaborato abbiamo avuto modo di vederli, dietro l'intelligenza artificiale si nasconde un mondo di compromessi.

[258] Hempel et al., "Review of Risk Assessment Instruments for Juvenile Sex Offenders."

Conclusione

La diffusione della tecnologia dell'intelligenza artificiale appare ormai è inevitabile, è un assunto che bisogna accettare. A differenza di anni fa, in cui il problema consisteva nel pensare a potenziali utilizzi dei sistemi di intelligenza artificiale, oggi, essendoci la consapevolezza che questi rappresentano la realtà, bisogna riflettere circa il loro impatto. Vista la sua crescente diffusione, numerosi risultano i cambiamenti in atto nella nostra società. È bene tentare di plasmare la direzione nella quale si vuole andare. Molti sono sicuramente i benefici e i vantaggi di cui ci si può appropriare, ma non va tralasciata l'attenzione ai possibili rischi.

L'applicazione e l'adozione dell'intelligenza artificiale nelle sue varie forme comporta un cambiamento a livello strutturale ed organizzativo. Le possibilità offerte da questa evoluzione tecnologica sono molteplici, e altre ancora sono da esplorare. Quasi tutti i settori ne stanno venendo influenzati, poiché l'impiego dell'intelligenza artificiale comporta miglioramenti significativi sotto molti punti di vista, quali produttività e precisione.

Tale cambiamento deve però essere accompagnato da un adattamento, c'è bisogno di abituarsi a questa nuova realtà. I sistemi di intelligenza artificiale infatti ormai collaborano con l'essere umano e, andando a formare una macchina sociale, diventeranno sempre più pervasivi. Bisogna comprendere questa tecnologia per poterla regolare e sfruttare adeguatamente, affinché l'agire umano risulti potenziato.

L'adeguamento passa per una necessaria formazione, sia professionale sia scolastica, così da preparare le competenze e le conoscenze utili per gestire una materia così complessa.

L'obiettivo di questa tesi era quello di capire come poter arrivare ad un'intelligenza artificiale affidabile. Con "affidabile" si intende la possibilità di sfruttare il più possibile i benefici che si dimostrano a favore della qualità della vita dell'essere umano, evitando, o per lo meno limitandone, i rischi. Si è visto come quella della progettazione sia una fase delicata e fondamentale in questo senso. Infatti, per poter indirizzarne il futuro e gli impatti che un domani questa tecnologia avrà sulle persone, è necessario prevedere le possibili conseguenze e ragionare anche su un piano etico oltre che solamente pratico.

Vi sono infatti principi e requisiti da considerare, i quali, se rispettati, possono abilitare tutto il potenziale benefico di questa tecnologia. Abbiamo cercato di sviluppare in modo concreto ed esemplificativo queste riflessioni nel campo della giustizia predittiva.

Riteniamo che potranno essere fatti ulteriori passi in avanti in questo campo, anche se forse non saranno completamente rivoluzionari, sicuramente questa tecnologia vedrà una sua crescente applicazione nei più disparati campi. Possiamo citare il caso esemplare del supermercato londinese senza checkout, in cui l'IA tramite telecamere e altri sensori rileva i prodotti prelevati e addebita in modo diretto e automatico la spesa[259].

Magari un giorno si potrebbe riuscire a creare un'intelligenza simile a quella umana, ma per ora di sicuro il rischio non sussiste. Recentemente però un ricercatore di OpenAI ha dichiarato che alcune grandi reti neurali potrebbero essere leggermente consapevoli, ma tale dichiarazione non è in alcun modo verificabile. Al momento l'intelligenza artificiale è da considerarsi uno strumento a disposizione delle persone e, come tale, la sua utilità o pericolosità dipende dagli usi che ne si fa, questo è stato confermato dalla stessa IA. Megatron Transformer di Nvidia, una intelligenza artificiale allenata su un'immensa banca dati contenente testi, che è stata invitata a partecipare a un dibattito all'Università di Oxford sull'etica dell'IA nel quale ha affermato che "l'IA non sarà mai etica poiché è uno strumento, non esiste una buona IA, ma solo esseri umani buoni o cattivi"[260].

[259] "ALDI'S CHECKOUT-FREE CONCEPT STORE OPENS FOR PUBLIC TESTING," *ALDI UK Press Office*, 18 January 2022, https://www.aldipresscentre.co.uk/business-news/aldis-checkout-free-concept-store-opens-for-public-testing/.

[260] Alex Connock and Professor Andrew Stephen, "We Invited an AI to Debate Its Own Ethics in the Oxford Union – What It Said Was Startling," *The Conversation*, n.d., http://theconversation.com/we-invited-an-ai-to-debate-its-own-ethics-in-the-oxford-union-what-it-said-was-startling-173607.

Bibliografia e sitografia

AGID (agenzia per l'Italia digitale). "Libro Bianco Sull'Intelligenza Artificiale al Servizio Del Cittadino," 2018. https://www.agid.gov.it/it/argomenti/intelligenza-artificiale.

Agrawal, Ajay K, Joshua Gans, and Avi Goldfarb. "Prediction, Judgment and Complexity." *The Economics of Artificial Intelligence: An Agenda* (2019): 23.

AI HLEG (High-Level Expert Group on Artificial Intelligence). "Ethics Guidelines for Trustworthy AI," 8 April 2019. Commissione Europea.

Aizenberg, Evgeni, and Jeroen van den Hoven. "Designing for Human Rights in AI." *Big Data & Society* 7.2 (2020): 205395172094956. https://doi.org/10.1177/2053951720949566.

Amigoni, Francesco, Viola Schiaffonati, and Marco Somalvico. "Intelligenza artificiale in 'Enciclopedia della Scienza e della Tecnica.'"

AndreaBacciu2018. "Una Panoramica Introduttiva Su Deep Learning e Machine Learning." *DeepLearningItalia*, 25 September 2017. https://www.deeplearningitalia.com/una-panoramica-introduttiva-su-deep-learning-e-machine-learning/.

Asimov, Isaac, and Laura Serra. *Io, robot*. Milano: Mondadori, 2018.

Atkinson, Joe. "'Technology Managing People': An Urgent Agenda for Labour Law." *Industrial Law Journal* 50.2 (2021): 324–29. https://doi.org/10.1093/indlaw/dwab005.

Barnes, Geoffrey C. "Classifying Adult Probationers by Forecasting Future Offending" (2012): 64.

Basile, di Fabio. "Intelligenza artificiale e diritto penale: quattro possibili percorsi di indagine" 10. Diritto penale uomo (2019): 33.

Benanti, Paolo. "Algoritmi con pregiudizi: il caso serio delle corti di giustizia USA." *paolobenanti*, 3 October 2017. https://www.paolobenanti.com/post/2017/10/03/algoritmi-con-pregiudizi-il-caso-serio-delle-corti-di-giustizia-usa.

Blacklaws, Christina. "Algorithms: Transparency and Accountability." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2128 (2018): 20170351. https://doi.org/10.1098/rsta.2017.0351.

Burr, Christopher, Mariarosaria Taddeo, and Luciano Floridi. "The Ethics of Digital Well-Being: A Thematic Review." *Sci Eng Ethics* 26.4 (2020): 2313–43. https://doi.org/10.1007/s11948-020-00175-8.

Carleo, Alessandra, ed. *Decisione Robotica*. Percorsi. Diritto. Bologna: Il mulino, 2019.

Carrer, Stefania. "Se l'amicus curiae è un algoritmo: il chiacchierato caso Loomis alla Corte Suprema del Wisconsin." *Giurisprudenza penale*, 24 April 2019. https://www.giurisprudenzapenale.com/2019/04/24/lamicus-curiae-un-algoritmo-chiacchierato-caso-loomis-alla-corte-suprema-del-wisconsin/.

Castellani, Maddalena, and Beppe Carrella. *Blockchain: guida pratica tecnico giuridica all'uso*. Firenze: goWare, 2019.

Cataleta, Maria Stefania. "The Fragility of Human Rights Facing AI" Working Paper No.02 *Humane Artificial Intelligence* (n.d.): 33.

Cath, Corinne, Sandra Wachter, Brent Mittelstadt, Mariarosaria Taddeo, and Luciano Floridi. "Artificial Intelligence and the 'Good Society': The US, EU, and UK Approach." *Sci Eng Ethics* (2017). https://doi.org/10.1007/s11948-017-9901-7.

Cautela, Cabirio, Marzia Mortati, Claudio Dell'Era, and Luca Gastaldi. "The Impact of Artificial Intelligence on Design Thinking Practice: Insights from the Ecosystem of

Startups." *Strategic Design Research Journal* 12.1 (2019): 114–34. https://doi.org/10.4013/sdrj.2019.121.08.
Cave, Jonathan. "The Ethics of Data and of Data Science: An Economist's Perspective." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160117. https://doi.org/10.1098/rsta.2016.0117.
Ciacci, Gianluigi, and Giovanni Buonomo. *Profili di informatica giuridica*. 2. ed. CEDAM scienze giuridiche. Milano: Wolters Kluwer, 2021.
Commissione Europea. "AI Ethics Communication," 8 April 2019.
Commissione Europea per l'efficienza della Giustizia. "Carta Etica Europea Sull'utilizzo Dell'intelligenza Artificiale Nei Sistemi Giudiziari e Negli Ambiti Connessi," 2019. https://rm.coe.int/carta-etica-europea-sull-utilizzo-dell-intelligenza-artificiale-nei-si/1680993348.
Consoft Sistemi. "L'intelligenza Artificiale al Servizio Dell'uomo," 2019. https://www.cospe.org/wp-content/uploads/2019/07/03_dossier_INTELLIGENZA-ARTIFICIALE_080719-1.pdf.
Contissa, Giuseppe. *Information Technology for the Law*. Informatica Giuridica. Serie Didattica 6. Torino: G. Giappichelli, 2017.
Copeland, J. "Intelligenza Artificiale in 'Encyclopedia Britannica.'"
Corea, Francesco. *Artificial Intelligence and Exponential Technologies: Business Models Evolution and New Investment Opportunities*. New York, NY: Springer Berlin Heidelberg, 2017.
Costa, Claudia. "Intelligenza artificiale e Giustizia: tempi ancora prematuri." *AI4Business*, 8 May 2019. https://www.ai4business.it/intelligenza-artificiale/intelligenza-artificiale-giustizia/.
Cristianini, Nello, Teresa Scantamburlo, and James Ladyman. "The Social Turn of Artificial Intelligence." *AI & Soc* (2021). https://doi.org/10.1007/s00146-021-01289-8.
D'Aloia, Antonio. "Il diritto verso 'il mondo nuovo'. Le sfide dell'Intelligenza Artificiale." *BioLaw Journal - Rivista di BioDiritto* 1 (2019): 3–31. https://doi.org/10.15168/2284-4503-349.
Diebold, Francis X, and Robert S Mariano. "Comparing Predictive Accuracy." *Journal of Business & Economic Statistics* 20.1 (2002): 134–44. https://doi.org/10.1198/073500102753410444.
Donati, Filippo. "INTELLIGENZA ARTIFICIALE E GIUSTIZIA." *Rivista Associazione Italiana dei Costituzionalisti* 1/2020 (2020): 22.
Dressel, Julia, and Hany Farid. "The Accuracy, Fairness, and Limits of Predicting Recidivism." *Sci. Adv.* 4.1 (2018): eaao5580. https://doi.org/10.1126/sciadv.aao5580.
Drew, Cat. "Data Science Ethics in Government." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160119. https://doi.org/10.1098/rsta.2016.0119.
———. "Design for Data Ethics: Using Service Design Approaches to Operationalize Ethical Principles on Four Projects." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2128 (2018): 20170353. https://doi.org/10.1098/rsta.2017.0353.
Europena Commission. "Europe Fit for the Digital Age: Artificial Intelligence," 2021. https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682.

Floridi, Luciano. “Faultless Responsibility: On the Nature and Allocation of Moral Responsibility for Distributed Moral Actions.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160112. https://doi.org/10.1098/rsta.2016.0112.

———. “Soft Ethics and the Governance of the Digital.” *Philos. Technol.* 31.1 (2018): 1–8. https://doi.org/10.1007/s13347-018-0303-9.

———. *The 4th Revolution: How the Infosphere Is Reshaping Human Reality*. First edition. New York ; Oxford: Oxford University Press, 2014.

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, et al. “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations.” *Minds & Machines* 28.4 (2018): 689–707. https://doi.org/10.1007/s11023-018-9482-5.

———. “AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations.” *Minds & Machines* 28.4 (2018): 689–707. https://doi.org/10.1007/s11023-018-9482-5.

Floridi, Luciano, Josh Cowls, Thomas C. King, and Mariarosaria Taddeo. “How to Design AI for Social Good: Seven Essential Factors.” *Sci Eng Ethics* 26.3 (2020): 1771–96. https://doi.org/10.1007/s11948-020-00213-5.

Floridi, Luciano, and Mariarosaria Taddeo. “What Is Data Ethics?” *Phil. Trans. R. Soc. A.* 374.2083 (2016): 20160360. https://doi.org/10.1098/rsta.2016.0360.

Gardner, Howard, and Ester Dornetti. *Cinque chiavi per il futuro*. Milano: Feltrinelli, 2015.

Gastaldo, Francesca Ceresa. “Il giudice-robot: l’intelligenza artificiale nei sistemi giudiziari tra aspettative ed equivoci.” *Ius in itinere*, 22 March 2021. https://www.iusinitinere.it/il-giudice-robot-lintelligenza-artificiale-nei-sistemi-giudiziari-tra-aspettative-ed-equivoci-36717.

Gelepithis, Petros. “AI and Human Society” 13 *AI & Society* (1999): 312–21.

Goodman, Bryce, and Seth Flaxman. “European Union Regulations on Algorithmic Decision-Making and a ‘Right to Explanation.’” *AIMag* 38.3 (2017): 50–57. https://doi.org/10.1609/aimag.v38i3.2741.

Grindrod, Peter. “Beyond Privacy and Exposure: Ethical Issues within Citizen-Facing Analytics.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160132. https://doi.org/10.1098/rsta.2016.0132.

Harari, Yuval Noah, and Giuseppe Bernardi. *Sapiens. Da animali a dèi: breve storia dell’umanità*. Milano: Bompiani, 2020.

Hartmann, Kathrin, and Georg Wenzelburger. “Uncertainty, Risk and the Use of Algorithms in Policy Decisions: A Case Study on Criminal Justice in the USA.” *Policy Sci* 54.2 (2021): 269–87. https://doi.org/10.1007/s11077-020-09414-y.

Jackson, Peter. *Introduction to Expert Systems*. 3rd ed. International Computer Science Series. Harlow, England ; Reading, Mass: Addison-Wesley, 1999.

Kaplan, Jerry. *Intelligenza artificiale. Guida al futuro prossimo*. Luiss University Press, 2018.

———. *Intelligenza artificiale. Guida al futuro prossimo*. Luiss University Press, 2018.

King, Thomas C., Nikita Aggarwal, Mariarosaria Taddeo, and Luciano Floridi. “Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions.” *Sci Eng Ethics* 26.1 (2020): 89–120. https://doi.org/10.1007/s11948-018-00081-0.

Köchling, Alina, and Marius Claus Wehner. “Discriminated by an Algorithm: A Systematic Review of Discrimination and Fairness by Algorithmic Decision-Making in the Context of HR Recruitment and HR Development.” *Bus Res* 13.3 (2020): 795–848. https://doi.org/10.1007/s40685-020-00134-w.
———. “Discriminated by an Algorithm: A Systematic Review of Discrimination and Fairness by Algorithmic Decision-Making in the Context of HR Recruitment and HR Development.” *Bus Res* 13.3 (2020): 795–848. https://doi.org/10.1007/s40685-020-00134-w.
Latonero, Mark, and Aaina Agarwal. “Human Rights Impact Assessments for AI: Learning from Facebook’s Failure in Myanmar.” Carr Center for Human Rights Policy (2021): 18.
Laudan, Larry. *Progress and Its Problems: Towards a Theory of Scientific Growth*. 1st paperback print. Berkeley, Calif.: Univ. of Calif. Press, 1978.
laura and john arnold foundation. “PSA: Risk Factors and Formula,” 2016. https://craftmediabucket.s3.amazonaws.com/uploads/PDFs/PSA-Risk-Factors-and-Formula.pdf.
Lawlor, Reed C. “What Computers Can Do: Analysis and Prediction of Judicial Decisions.” *American Bar Association Journal* 49.4 (1963): 337–44. https://www.jstor.org/stable/25722338.
Lee, Michelle Seng Ah, Luciano Floridi, and Jatinder Singh. “Formalising Trade-Offs beyond Algorithmic Fairness: Lessons from Ethical Philosophy and Welfare Economics.” *AI Ethics* 1.4 (2021): 529–44. https://doi.org/10.1007/s43681-021-00067-y.
Leonelli, Sabina. “Locating Ethics in Data Science: Responsibility and Accountability in Global and Distributed Knowledge Production Systems.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160122. https://doi.org/10.1098/rsta.2016.0122.
Lin, Patrick, Keith Abney, and George Bekey. “Robot Ethics: Mapping the Issues for a Mechanized World.” *Artificial Intelligence* 175.5–6 (2011): 942–49. https://doi.org/10.1016/j.artint.2010.11.026.
———. “Robot Ethics: Mapping the Issues for a Mechanized World.” *Artificial Intelligence* 175.5–6 (2011): 942–49. https://doi.org/10.1016/j.artint.2010.11.026.
Livni, Ephrat. “Nei tribunali del New Jersey è un algoritmo a decidere chi esce su cauzione.” *Internazionale*, 3 March 2017. https://www.internazionale.it/notizie/ephrat-livni/2017/03/03/tribunali-algoritmo-cauzione.
Luciani, Massimo. “LA DECISIONE GIUDIZIARIA ROBOTICA” 03/2018. Rivista Associazione Italiana dei Costituzionalisti (2018): 22.
di Magliano, Roberto Pasca. “Etica e innovazione nella governance pubblica” Policy Paper.1 (2021): 16.
Majumdar, D, and H K Chattopadhyay. “AI and Human Rights: From Business and Policy Perspectives” (n.d.): 10.
Marmo, Roberto. *Algoritmi per l’intelligenza artificiale*. HOEPLI, 2020.
Martin-Bariteau, Florian, and Marina Pavlovic. *AI and Contract Law*. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, November 2, 2020. https://papers.ssrn.com/abstract=3730385.
Massaro, Antonella, Lorenza Grossi, Angelo Giraldi, Laura Notaro, and Pietro Sorbello. “Intelligenza artificiale e giustizia penale” (2020): 227.

Massaron, Luca. *Intelligenza artificiale for dummies*. S.l.: HOEPLI, 2020.
Mattu, Julia Angwin, Jeff Larson,Lauren Kirchner,Surya. "Machine Bias." *ProPublica*, n.d. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=ApJt7l5BZpIugeIernGirtWPn_a7EraR.
McCarthy, John. "What Is Artificial Intelligence?" *Computer Science Department Stanford University* (2007): 15. http://www-formal.stanford.edu/jmc/.
Mökander, Jakob, Jessica Morley, Mariarosaria Taddeo, and Luciano Floridi. "Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations." *Sci Eng Ethics* 27.4 (2021): 44. https://doi.org/10.1007/s11948-021-00319-4.
Morielli, Massimo, Leonardo Galimberti, and Applied intelligence, 2018. "Intelligenza Artificiale: Istruzioni per L'uso," 2018. Accenture Applied intelligence. https://www.accenture.com/it-it/insights/artificial-intelligence/artificial-intelligence-explained-executives.
Morley, Jessica, Anat Elhalal, Francesca Garcia, Libby Kinsey, Jakob Mökander, and Luciano Floridi. "Ethics as a Service: A Pragmatic Operationalisation of AI Ethics." *Minds & Machines* 31.2 (2021): 239–56. https://doi.org/10.1007/s11023-021-09563-w.
Morley, Jessica, Libby Kinsey, Anat Elhalal, Francesca Garcia, Marta Ziosi, and Luciano Floridi. "Operationalising AI Ethics: Barriers, Enablers and next Steps." *AI & Soc* (2021). https://doi.org/10.1007/s00146-021-01308-8.
Mulligan, Deirdre K., Colin Koopman, and Nick Doty. "Privacy Is an Essentially Contested Concept: A Multi-Dimensional Analytic for Mapping Privacy." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160118. https://doi.org/10.1098/rsta.2016.0118.
Nast, Condé. "Il software italiano che ha cambiato il mondo della polizia predittiva." *Wired Italia*, 18 May 2019. https://www.wired.it/attualita/tech/2019/05/18/polizia-predittiva-software-italiano-keycrime/.
National Science and Technology Council. "Preparing for the Future of AI," October 2016. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.
———. "The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update" (2019): 50.
Nissim, Kobbi, and Alexandra Wood. "Is Privacy Privacy?" *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2128 (2018): 20170358. https://doi.org/10.1098/rsta.2017.0358.
Olhede, S. C., and P. J. Wolfe. "The Growing Ubiquity of Algorithms in Society: Implications, Impacts and Innovations." *Phil. Trans. R. Soc. A.* 376.2128 (2018): 20170364. https://doi.org/10.1098/rsta.2017.0364.
Oswald, Marion. "Algorithm Assisted Decision Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2128 (2018): 20170359. https://doi.org/10.1098/rsta.2017.0359.
Parliament and Council of the European Union. "General Data Protection Regulation," 2016. https://gdpr-info.eu/recitals/.
Pistilli, Marcello. "Etica e algoritmi. Verso l'algoretica." *innovationgym*, 21 October 2019. https://www.innovationgym.org/en/etica-e-algoritmi-verso-lalgoretica/.

Quintarelli, Stefano, Francesco Corea, Fabio Fossa, Andrea Loreggia, and Salvatore Sapienza. “AI: profili etici Una prospettiva etica sull’Intelligenza Artificiale: princìpi, diritti e raccomandazioni” 3 (n.d.): 22.
Quintarelli, Stefano, Claudia Giulia Ferrauto, Fabio Fossa, Francesco Corea, Andrea Loreggia, and Salvatore Sapienza. *Intelligenza artificiale*. Bollati Boringhieri, 2020.
Redcross, Cindy, Brit Henderson, Luke Miratrix, and Erin Valentine. “Evaluation of Pretrial Justice System. Reforms That Use the Public Safety Assessment.” *Pretrial Justice Reform Study, Mecklenburg County* (2019).
Reed, Chris. “How Should We Regulate Artificial Intelligence?” *Phil. Trans. R. Soc. A.* 376.2128 (2018): 20170360. https://doi.org/10.1098/rsta.2017.0360.
Roberts, Huw, Josh Cowls, Emmie Hine, Francesca Mazzi, Andreas Tsamados, Mariarosaria Taddeo, and Luciano Floridi. “Achieving a ‘Good AI Society’: Comparing the Aims and Progress of the EU and the US.” *Sci Eng Ethics* 27.6 (2021): 68. https://doi.org/10.1007/s11948-021-00340-7.
Russell, Stuart J., Peter Norvig, and Ernest Davis. *Artificial Intelligence: A Modern Approach*. 3rd ed. Prentice Hall Series in Artificial Intelligence. Upper Saddle River: Prentice Hall, 2010.
Santosuosso, Amedeo. *Intelligenza artificiale e diritto*. Mondadori, 2020.
Sartor, Giovanni. *L’informatica giuridica e le tecnologie dell’informatica: corso d’informatica giuridica*, 2016. https://ebookcentral.proquest.com/lib/concordiaab-ebooks/detail.action?docID=4771207.
Scantamburlo, Teresa. “Non-Empirical Problems in Fair Machine Learning.” *Ethics Inf Technol* 23.4 (2021): 703–12. https://doi.org/10.1007/s10676-021-09608-9.
Shah, Hetan. “Algorithmic Accountability.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376.2128 (2018): 20170362. https://doi.org/10.1098/rsta.2017.0362.
Shams, Rushdi. “Developing Machine Learning Products Better and Faster at Startups.” *IEEE Eng. Manag. Rev.* 46.3 (2018): 36–39. https://doi.org/10.1109/EMR.2018.2870669.
Signorato, Silvia. “Giustizia Penale e Intelligenza Artificiale. Considerazioni in Tema Di Algoritmo Predittivo.” *Rivista Di Diritto Processuale* 75.2 (2020): 605–16.
Taddeo, Mariarosaria. “Cyber Security and Individual Rights, Striking the Right Balance.” *Philos. Technol.* 26.4 (2013): 353–56. https://doi.org/10.1007/s13347-013-0140-9.
———. “Data Philanthropy and Individual Rights.” *Minds & Machines* 27.1 (2017): 1–5. https://doi.org/10.1007/s11023-017-9429-2.
———. “Modelling Trust in Artificial Agents, A First Step Toward the Analysis of e-Trust.” *Minds & Machines* 20.2 (2010): 243–57. https://doi.org/10.1007/s11023-010-9201-3.
———. “The Struggle Between Liberties and Authorities in the Information Age.” *Sci Eng Ethics* 21.5 (2015): 1125–38. https://doi.org/10.1007/s11948-014-9586-0.
———. “Trusting Digital Technologies Correctly.” *Minds & Machines* 27.4 (2017): 565–68. https://doi.org/10.1007/s11023-017-9450-5.
Taddeo, Mariarosaria, and Luciano Floridi. “How AI Can Be a Force for Good.” *Science* 361.6404 (2018): 751–52. https://doi.org/10.1126/science.aat5991.
Talia, Domenico. *La società calcolabile e i big data: algoritmi e persone nel mondo digitale*. Soveria Mannelli: Rubbettino, 2018.

Taylor, Linnet. “The Ethics of Big Data as a Public Good: Which Public? Whose Good?” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160126. https://doi.org/10.1098/rsta.2016.0126.

Tremolada, Luca. “Giustizia predittiva, l’intelligenza artificiale migliore amica dell’avvocato.” *Il Sole 24 ORE*, 10 March 2020. https://www.ilsole24ore.com/art/giustizia-predittiva-l-intelligenza-artificiale-migliore-amica-dell-avvocato-ACBXBbJB.

Tucker, Catherine, and Gans. “Privacy, Algorithms, and Artificial Intelligence.” *The Economics of Artificial Intelligence: An Agenda*. National Bureau of Economic Research Conference Report (2019): 16.

Turing, A. M. “I.—COMPUTING MACHINERY AND INTELLIGENCE.” *Mind* LIX.236 (1950): 433–60. https://doi.org/10.1093/mind/LIX.236.433.

Ubertis, di Giulio. “INTELLIGENZA ARTIFICIALE, GIUSTIZIA PENALE, CONTROLLO UMANO SIGNIFICATIVO (□)” (n.d.): 15.

Vayena, Effy, and John Tasioulas. “The Dynamics of Big Data and Human Rights: The Case of Scientific Research.” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2083 (2016): 20160129. https://doi.org/10.1098/rsta.2016.0129.

Warwick, Kevin. *Intelligenza Artificiale - Le basi*. Dario Flaccovio Editore, 2015.

Washington, Anne L. “HOW TO ARGUE WITH AN ALGORITHM: LESSONS FROM THE COMPAS-PROPUBLICA DEBATE.” *The Colorado Technology Law Journal* 17.1 (2019): 37.

Završnik, Aleš. “Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings.” *European Journal of Criminology* 18.5 (2021): 623–42. https://doi.org/10.1177/1477370819876762.

“A Computer Program Used for Bail and Sentencing Decisions Was Labeled Biased against Blacks. It’s Actually Not That Clear.” *Washington Post*, n.d. https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/.

“Algoritmo e giustizia predittiva in campo penale.” *Altalex*, 14 June 2019. https://www.altalex.com/documents/news/2019/06/14/algoritmo-e-la-giustizia-predittiva-in-campo-penale.

“Artificial Intelligence: Italy Launches National Strategy.” *Connessioni - Bridging Worlds*, 7 December 2021. https://www.connessioni.biz/en/artificial-intelligence-italy-launches-national-strategy/.

“Demographic Bias.” *Mapping Pretrial Injustice*, n.d. https://pretrialrisk.com/the-danger/demographic-bias/.

“How It Works.” *Advancing Pretrial Policy & Research (APPR)*, n.d. https://advancingpretrial.org/psa/factors/.

“La Costituzione - Articolo 102 | Senato Della Repubblica,” n.d. https://www.senato.it/istituzione/la-costituzione/parte-ii/titolo-iv/sezione-i/articolo-102.

“Online Dispute Resolution | European Commission,” n.d. https://ec.europa.eu/consumers/odr/main/?event=main.trader.register.

“Online Dispute Resolution e giustizia digitale.” *Altalex*, 23 February 2021. https://www.altalex.com/documents/news/2021/02/23/online-dispute-resolution-e-giustizia-digitale.

"Pretrial Release." *Bureau of Justice Statistics*, n.d. https://bjs.ojp.gov/topics/courts/pretrial-release.
"Prevedere l'esito di un giudizio: ecco la giurisprudenza predittiva." *Università Ca' Foscari Venezia*, n.d. http://www.unive.it/pag/14024/?tx_news_pi1%5Bnews%5D=9884&cHash=2d30d787f83ed5c005434c077b6d25bd.
"Tutto quello che c'è da sapere sulla Intelligenza artificiale nello studio legale." *Altalex*, 16 July 2018. https://www.altalex.com/documents/news/2018/07/16/intelligenza-artificiale-nel-settore-legale.
"Understand Pretrial Justice." *Advancing Pretrial Policy & Research (APPR)*, n.d. https://advancingpretrial.org/pretrial-justice/pretrial-justice/.
"Why Jurisdictions Choose RATs." *Mapping Pretrial Injustice*, n.d. https://pretrialrisk.com/the-basics/the-case-for-rats/.