



Università
Ca' Foscari
Venezia

Corso di Laurea Magistrale
in Amministrazione, finanza e controllo

Tesi di Laurea

**L'intelligenza artificiale: definizioni,
framework e impatti a livello
aziendale e sociale**

Relatore

Ch. Prof. Carlo Bagnoli

Laureando

Matteo Florian

Matricola 866665

Anno Accademico

2021/2022

Indice

Introduzione

Capitolo 1. L'intelligenza artificiale: una panoramica generale

- 1.1. Definizioni
- 1.2. Cenni storici
- 1.3. Classificazioni
- 1.4. Campi di applicazione
 - 1.4.1. *Machine learning*
 - 1.4.2. *Deep learning*
- 1.5. Cos'è l'intelligenza?

Capitolo 2. L'intelligenza artificiale: un *framework* teorico

- 2.1. Alcuni *framework* di riferimento
- 2.2. Il modello dei tre domini
 - 2.2.1. *Il dominio dei dati*
 - 2.2.2. *Il dominio scientifico*
 - 2.2.3. *Il dominio delle applicazioni*
 - 2.2.4. *I tre domini riuniti*
- 2.3. I ruoli dell'intelligenza artificiale: il modello ACE

Capitolo 3. L'intelligenza artificiale: gli impatti aziendali e sociali

- 3.1. Gli impatti sui building block del modello di business
 - 3.1.1. *Il modello di riferimento*
 - 3.1.2. *Gli impatti sui singoli building block*
- 3.2. Gli impatti sulla strategia
- 3.3. Gli impatti sulla società

Conclusione

Bibliografia

Sitografia

Allegato A

Indice

Introduzione	1
Capitolo 1. L'intelligenza artificiale: una panoramica generale	2
1.1. Definizioni.....	2
1.2. Cenni storici	13
1.3. Classificazioni	17
1.4. Campi di applicazione	21
1.4.1. <i>Machine learning</i>	28
1.4.2. <i>Deep learning</i>	32
1.5. Cos'è l'intelligenza?	36
Capitolo 2. L'intelligenza artificiale: un <i>framework</i> teorico	41
2.1. Alcuni <i>framework</i> di riferimento.....	41
2.2. Il modello dei tre domini.....	53
2.2.1. <i>Il dominio dei dati</i>	54
2.2.2. <i>Il dominio scientifico</i>	56
Capitolo 3. L'intelligenza artificiale: gli impatti aziendali e sociali	57
3.3 Gli impatti sulla società.....	57
Conclusione	89
Bibliografia	90
Sitografia	93

Introduzione

La quarta rivoluzione industriale...

Capitolo 1. L'intelligenza artificiale: una panoramica generale

1.1. Definizioni

I primi studi nel campo dell'intelligenza artificiale risalgono addirittura agli anni '50. Nel corso del tempo, diversi ricercatori ed organizzazioni hanno elaborato una loro visione di cosa essa sia, nel tentativo di inquadrare un fenomeno che tra periodi di maggior eccitazione ed altri di relativo disinteresse è diventato sempre più complesso e sfaccettato. Questa difformità di opinioni ha avuto come conseguenza il proliferare di numerose definizioni di IA, alcune delle quali notevolmente diverse tra loro, altre molto simili. Come evidenziato da Russell e Norvig¹, infatti, ciascuna di esse è ispirata ad un diverso approccio seguito dall'autore. Mentre certi hanno ritenuto che il modo migliore per verificare il requisito dell'intelligenza fosse tramite la fedeltà rispetto alle *performance* umane, altri si sono focalizzati sul concetto di razionalità, intesa come la capacità di fare la cosa giusta a seconda delle informazioni disponibili. Allo stesso tempo, alcuni hanno considerato l'intelligenza una caratteristica esprimibile attraverso il comportamento, le azioni esteriori, mentre per altri essa è legata ai processi interiori del pensiero e del ragionamento. Esistono pertanto due distinte dimensioni in base alle quali valutare se si ha a che fare con IA:

- umano e razionale;
- comportamento e pensiero (o più in astratto, esteriore e interiore).

Dalla loro combinazione derivano quattro diversi approcci, ognuno dei quali ha avuto negli anni un folto numero di estimatori:

- agire come un umano;
- agire razionalmente;
- pensare come un umano;
- pensare razionalmente.

¹ Russell, S., Norvig, P. (2020), *Artificial Intelligence: A Modern Approach*, 4th ed., Pearson Education, Hoboken.

Pensare come un umano	Pensare razionalmente
Agire come un umano	Agire razionalmente

Tabella 1. Le quattro categorie di Russell e Norvig

Ancora oggi non esiste una definizione comunemente accettata di “intelligenza artificiale”. Ciò può essere visto come un problema, in quanto stabilire universalmente tra tutti gli *stakeholder* cosa essa sia porterebbe diversi vantaggi sia nel campo della ricerca teorica, sia a livello più pratico. Per esempio, un’unica definizione consentirebbe una migliore misurazione degli investimenti in queste tecnologie, oppure permetterebbe lo sviluppo di norme legislative *ad hoc* e l’utilizzo appropriato di pratiche di gestione del rischio.

Verranno ora elencati diversi tentativi di definire l’intelligenza artificiale. Al fine di fornire un quadro completo e sfaccettato, nella ricerca delle definizioni sono state considerate tre diverse prospettive attraverso cui essa viene considerata²: politica e istituzionale; della ricerca; dei *practitioner*³. Inoltre, per ogni enunciato si cercherà di stabilire a quale dei quattro approcci precedentemente esposti può essere ricondotto.

² Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., and Delipetrev, B. (2020), *AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*, EUR 30117 EN, Publications Office of the European Union, Luxembourg.

³ Nel documento originale, al posto della “prospettiva dei *practitioner*” si fa riferimento alla “prospettiva del mercato”. L’espressione è stata cambiata in quanto si ritiene che il termine “*practitioner*” sia più accurato e ponga maggiore attenzione sui soggetti interessati.

Prospettiva politica e istituzionale

Si riferisce all'ottica adottata da enti regolatori ed istituzioni internazionali. L'IA è vista come uno strumento per la crescita e lo sviluppo tecnologico.

High-Level Expert Group on Artificial Intelligence (2019)

“Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions”⁴.

Il High-Level Expert Group on Artificial Intelligence (AI HLEG) è un gruppo di esperti indipendenti nato nel 2018 su iniziativa della Commissione Europea, con il compito di fornire suggerimenti su come affrontare a medio-lungo termine le opportunità e le sfide legate all'IA.

Nel 2019 il gruppo ha pubblicato un documento in cui ha cercato di fare chiarezza su determinati aspetti scientifici e tecnologici riconducibili a questo campo e di creare una base di conoscenza semplice e condivisibile da tutti, ed in questa occasione ha anche fornito una sua definizione.

Come esplicitamente ammesso nel documento, l'HLEG ha ripreso il concetto di razionalità (*“An AI system is thus first and foremost rational”⁵*), la quale, dato uno specifico obiettivo, viene raggiunta:

- percependo l'ambiente circostante;
- raccogliendo ed interpretando dati;
- processando le conoscenze ed informazioni ottenute;
- decidendo qual è la migliore azione da intraprendere;
- e agendo di conseguenza.

⁴ High-Level Expert Group on Artificial Intelligence (2019), *A definition of AI: Main capabilities and scientific disciplines*, Commissione Europea, p. 6.

⁵ *Ibidem*.

L'approccio a cui è riconducibile questa definizione è perciò quello dell'“agire razionalmente”.

World Economic Forum (2019)

*“Artificial intelligence – systems that act by sensing, interpreting data, learning, reasoning and deciding the best course of action – changed from a limited to a transformative force in business when these systems were taught to learn through huge amounts of data and computing power”*⁶.

Anche il World Economic Forum ha cercato di specificare cosa sia l'IA, elaborando una formulazione che, seppur più sintetica, risulta essere negli elementi essenziali quasi identica a quella redatta dall'HLEG. Pertanto, l'approccio utilizzato è nuovamente quello dell'“agire razionalmente”.

OCSE (2019)

*“An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy”*⁷.

Nel 2019 l'OCSE ha adottato un pacchetto di raccomandazioni che si prefiggeva di promuovere un clima di fiducia attorno all'intelligenza artificiale, nel rispetto dei valori democratici e dei diritti umani.

All'interno del documento, l'IA viene definita specificando non il processo attraverso il quale riesce a compiere delle azioni (percepire, raccogliere, processare, decidere ed infine agire), bensì gli output che è in grado di produrre (previsioni, raccomandazioni e decisioni). La capacità di influenzare l'ambiente (*“influencing real or virtual environments”*) fa riferimento all'agire, mentre il fatto di operare in base a degli obiettivi predeterminati (*“a given set of human-defined objectives”*) richiama al concetto di razionalità (scegliere l'opzione migliore per raggiungere un certo traguardo).

⁶ World Economic Forum (2019), *Empowering AI Leadership. An Oversight Toolkit for Boards of Directors*, p. 6.

⁷ OECD (2019), *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, p. 7.

Anche questa definizione è dunque riconducibile all'approccio dell'"agire razionalmente".

Prospettiva della ricerca

Consiste nel punto di vista di coloro che dagli anni '50 ad oggi si sono impegnati nella comprensione dell'intelligenza artificiale come campo di ricerca e nel suo sviluppo come tecnologia di interesse generale.

McCarthy (2007)

"It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable".

"Intelligence is the computational part of the ability to achieve goals in the world. Varying kinds and degrees of intelligence occur in people, many animals and some machines"⁸.

Come si vedrà successivamente, John McCarthy è stato uno dei padri fondatori della disciplina ed in quanto tale, nel corso della sua carriera ha avuto modo di riformulare diverse volte il concetto di intelligenza artificiale. Quella presa in esame è la sua definizione più recente.

In essa, McCarthy identifica l'IA riprendendo a piene mani l'idea di intelligenza ("*intelligent machines, especially intelligent computer programs*"), la quale per lui è una facoltà computazionale ("*the computational part of the ability to achieve goals in the world*") e pertanto è legata al pensiero e ai processi interni. Se poi si considera che quella considerata è l'intelligenza umana ("*Q. Isn't there a solid definition of intelligence that doesn't depend on relating it to human intelligence? A. Not yet*"⁹), si arriva a comprendere come McCarthy in questo caso adotti l'approccio del "pensare come un umano".

Winston (1992)

"Artificial intelligence is...

⁸ McCarthy, J. (2007), *What is artificial intelligence?*, Stanford University, Stanford, p. 2.

⁹ *Ibidem*.

The study of the computations that make it possible to perceive, reason, and act”¹⁰.

L'accento che viene posto sulle computazioni suggerisce un approccio legato al pensiero ed al ragionamento.

Per stabilire l'altra dimensione, occorre nuovamente andare oltre la definizione in sé e capire il punto di vista dell'autore. Nel suo libro, Winston descrive l'intelligenza artificiale come *“an armamentarium of ideas about representing knowledge, using knowledge, and assembling systems”*¹¹. Queste finalità (rappresentare ed utilizzare la conoscenza ed assemblare sistemi) sono raggiunte tramite l'utilizzo di apposite regole e tecniche che hanno molto più a che vedere con la logica e la razionalità piuttosto che con il comportamento umano.

In conclusione, questa definizione può essere inserita nella categoria del “pensare razionalmente”.

Nilsson (2009)

“For me, artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment”¹².

Anche in questo caso, l'IA viene definita partendo dal concetto di intelligenza (*“making machines intelligent”*). Questa per Nilsson è la qualità che permette ad un'entità di funzionare adeguatamente e con lungimiranza nel proprio ambiente, dimostrando dunque un comportamento razionale.

Per questo motivo, l'approccio più indicato è quello dell’“agire razionalmente”.

Kaplan, Haenlein (2019)

“Specifically, we define AI as a system’s ability to interpret external data correctly, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation”¹³.

¹⁰ Winston, P. (1992), *Artificial Intelligence*, Addison-Wesley, Reading, p. 5.

¹¹ *Ibidem*.

¹² Nilsson, N. J. (2009), *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge University Press, Cambridge, p. 13.

Come l'HLEG e il World Economic Forum, Kaplan e Haenlein mettono in risalto le fasi del processo attraverso cui l'IA è in grado di raggiungere determinati obiettivi (e quindi di agire):

- interpretare correttamente i dati dall'esterno;
- imparare da essi;
- utilizzare le conoscenze acquisite.

Inoltre, la precisazione relativa all'adattamento flessibile, una capacità tipica degli esseri umani, suggerisce un approccio orientato meno alla razionalità e maggiormente alle performance umane.

In definitiva, questo enunciato è riconducibile alla categoria dell'"agire come un umano".

Van Kraay (2018)

*"Artificial Intelligence is the ability of a computer system to deal with ambiguity, by making predictions using previously gathered data, and learning from errors in those predictions in order to generate newer, more accurate predictions about how to behave in the future"*¹⁴.

Theo van Kraay è *Senior Program Manager* presso Microsoft e ha pubblicato diversi *post* ed articoli sul sito del colosso del digitale. In uno di questi, egli spiega cos'è l'intelligenza artificiale, arrivando anche a dare la sua personale definizione.

Essa si discosta notevolmente dalle altre sopracitate, poiché non identifica l'IA né mediante l'elencazione delle operazioni che è in grado di compiere o dei risultati che è capace di ottenere (HLEG, World Economic Forum, OCSE, Kaplan e Haenlein), né ricollegandosi all'idea più estesa di intelligenza (McCarthy, Nilson); bensì concentrandosi sulla differenza tra sistemi vecchi e nuovi ed introducendo il concetto di ambiguità. Per van Kraay, ciò che contraddistingue gli attuali sistemi rispetto a quelli utilizzati in passato sta appunto nella capacità di gestire situazioni ambigue tramite l'utilizzo di previsioni, un po' come fanno costantemente gli esseri umani grazie a dei

¹³ Kaplan, A., Haenlein, M. (2019), "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence", *Business Horizons*, 62, pp. 15-25.

¹⁴ Van Kraay, T. (2018), *What is Artificial Intelligence?*, Microsoft Azure, online, <https://azure.microsoft.com/it-it/blog/what-is-artificial-intelligence/>.

processi mentali di cui nemmeno si rendono conto. Questa definizione quindi considera come intelligenza artificiale solo le più moderne tecnologie e applicazioni legate a questo campo, escludendo una grossa fetta di storia della disciplina.

L'approccio utilizzato in questo caso è quello del "pensare come un umano".

Prospettiva dei *practitioner*

È la visione degli operatori che studiano il mercato e di coloro che cercano di implementare soluzioni tecnologiche all'interno delle imprese. Si concentra molto sullo sviluppo industriale dell'IA e sulla valutazione degli impatti che essa avrà a livello economico.

Google

*"Artificial intelligence is a field of science concerned with building computers and machines that can reason, learn, and act in such a way that would normally require human intelligence or that involves data whose scale exceeds what humans can analyze"*¹⁵.

Qui il termine di paragone sono certamente le prestazioni umane. Per quanto riguarda l'altra dimensione, vengono menzionate attività sia legate ai processi interni (*reason, learn, analyze*), sia esprimibili attraverso il comportamento esteriore (*act*). Risulta perciò difficile stabilire quale dei due aspetti prevalga, andando a preferire una soluzione onnicomprensiva in cui sono considerati entrambi.

Alla luce di quanto scritto, la categoria appropriata può essere indicata come dell'"agire e pensare come un umano".

Microsoft (2022)

*"Artificial intelligence (AI) is the capability of a computer to imitate intelligent human behavior. Through AI, machines can analyze images, comprehend speech, interact in natural ways, and make predictions using data"*¹⁶.

¹⁵ *What is Artificial Intelligence (AI)?*, Google Cloud, online, consultato il 10/10/2022, <https://cloud.google.com/learn/what-is-artificial-intelligence#section-2>.

¹⁶ AA.VV. (2022), *Artificial intelligence (AI) architecture design*, Microsoft, online, <https://learn.microsoft.com/en-us/azure/architecture/data-guide/big-data/ai-overview>.

In questo caso, è esplicitato molto chiaramente che il punto di riferimento è il comportamento umano (*“imitate intelligent human behavior”*). L’approccio corretto è dunque quello dell’*“agire come un umano”*.

McKinsey (2020)

*“AI is typically defined as the ability of a machine to perform cognitive functions we associate with human minds, such as perceiving, reasoning, learning, and problem solving. Examples of technologies that enable AI to solve business problems are robotics and autonomous vehicles, computer vision, language, virtual agents, and machine learning”*¹⁷.

Anche secondo McKinsey l’intelligenza artificiale può essere definita confrontandola con ciò che sono in grado di fare gli esseri umani. Tuttavia, a differenza del caso precedente, ci si focalizza ora sulle funzioni cognitive (percepire, ragionare, imparare, risolvere problemi) associabili alla mente umana, piuttosto che sul comportamento visibile. Pertanto, la categoria individuata è quella del *“pensare come un umano”*.

KPMG, TRENDONE (2018)

*“AI is an umbrella term for technologies that enable a computer to perform tasks that require human intelligence”*¹⁸.

Le performance umane sono nuovamente prese come termine di paragone. Il riferimento ai compiti (*“to perform tasks”*) suggerisce un orientamento verso le azioni esteriori e quindi all’approccio dell’*“agire come un umano”*.

Davenport (2019)

“Broadly speaking, AI or cognitive technologies employ such capabilities – previously possessed only by humans – as knowledge, insight, and perception to solve narrowly defined (with the current state of technology) tasks. The tasks are those that can usually be accomplished quickly by humans – identifying an image or interpreting the meaning of a

¹⁷ McKinsey (2020), *An executive’s guide to AI*, p. 1.

¹⁸ KPMG, TRENDONE (2018), *Rethinking the value chain. A study on AI, humanoids and robots*, p. 10.

*sentence. Befitting the “cognitive” label, these tasks could once be performed only by human brains”*¹⁹.

Nel suo libro *“The AI Advantage”*, in cui illustra come le società possono utilizzare l’intelligenza artificiale per migliorare il proprio business ed ottenere vantaggi competitivi, l’autore americano cita una serie di facoltà (conoscenza, intuizione, percezione) che caratterizzano le tecnologie in esame e che erano in precedenza *“possessed only by humans”*, dimostrando di preferire come discriminante le capacità umane piuttosto che la razionalità. In seguito, Davenport menziona alcune attività che è in grado di eseguire l’IA: nonostante il termine *“tasks”* possa far pensare ad una visione indirizzata al comportamento esteriore, la precisazione per cui tali compiti in passato potevano essere svolti esclusivamente dal cervello umano fa propendere per un orientamento verso il pensiero e i processi interni.

In definitiva, l’autore adotta l’approccio del *“pensare come un umano”*.

Strategy Innovation (2019)

*“In termini generali, l’IA può essere definita come un insieme di tecnologie che consentono a sistemi di sensori, algoritmi e macchine di eseguire attività che di solito richiedono l’intelligenza umana”*²⁰.

All’interno del report realizzato in occasione del 5° Strategy Innovation Forum, il team di Strategy Innovation ha elaborato una sua definizione di Intelligenza Artificiale, in cui risulta evidente il confronto con l’intelligenza umana. Come con Davenport, poi, il riferimento alle attività non deve trarre in inganno: nel report è precisato in seguito che l’IA è in grado di *“percepire il mondo esterno”* e di *“comprendere, ossia elaborare i dati trasformandoli in informazioni, conoscenza e persino decisioni”*, svolgendo delle funzioni assimilabili ai processi interni della mente umana.

L’approccio corretto è dunque quello del *“pensare come un umano”*.

¹⁹ Davenport, T. H. (2019), *The AI Advantage. How to Put the Artificial Intelligence Revolution to Work*, MIT Press, Cambridge.

²⁰ Albarelli, A., Bagnoli, C., Campostrini, S., Massaro, M., Muraro, A., Toniolo, K., Vesnic, L., Zantedeschi, M. S. (2019), *Gli impatti di IA e di Blockchain sui modelli di business. Il Report di Ricerca*, Strategy Innovation, Venezia, p. 7.

<p>Pensare come un umano</p> <p>McCarthy, van Kraay, McKinsey, Davenport, Strategy Innovation</p> <p>Google</p>	<p>Pensare razionalmente</p> <p>Winston</p>
<p>Agire come un umano</p> <p>Kaplan e Haenlein, Microsoft, KPMG e TRENDONE</p>	<p>Agire razionalmente</p> <p>HLEG, World Economic Forum, OCSE, Nilsson</p>

Tabella 2. Le diverse definizioni raggruppate per categoria

La tabella 2 riassume quanto detto sopra, raggruppando le diverse definizioni nelle rispettive categorie di Russell e Norvig.

L'analisi svolta non si prefigge di dare un quadro esaustivo del panorama dell'IA, né di porre le basi per un enunciato in grado di chiarire in modo assoluto e definitivo cos'è l'intelligenza artificiale. Numerose altre fonti dovrebbero essere incluse in uno sforzo con degli obiettivi tanto ambiziosi.

Più semplicemente, con questo lavoro si è cercato di raggiungere una duplice finalità:

- fornire una serie di esempi rappresentativi dei diversi punti di vista esistenti, da cui poter trarre degli spunti di riflessione;
- portare un po' di ordine nel *mare magnum* delle definizioni esistenti, grazie alla classificazione per prospettive e, soprattutto, per approcci.

1.2. Cenni storici

L'intelligenza artificiale come materia scientifica ha una storia relativamente recente, per non parlare delle sue applicazioni al di fuori dell'ambito accademico. I primi interrogativi sulla possibilità di creare delle macchine capaci di pensare risalgono solo al secondo dopoguerra. Tuttavia, già in precedenza filosofi, pensatori e matematici avevano contribuito con i loro scritti, le loro idee, le loro ricerche a creare quella base di conoscenza che si sarebbe poi rivelata fondamentale per la nascita della prima comunità dedicata all'IA.

L'origine dell'intelligenza artificiale, intesa in senso ampio come il tentativo di automatizzare le attività umane, può essere fatta risalire a secoli fa, se non addirittura millenni. Già attorno al I secolo a.C. il matematico ed inventore Erone di Alessandria disegnò delle macchine funzionanti tramite meccanismi meccanici o pneumatici e capaci di compiere diverse azioni senza l'intervento dell'uomo. Tra di esse si annoverano un congegno in grado di aprire delle porte una volta accesa una fiamma e delle statue che versavano del vino, giusto per citarne alcune. Questi dispositivi, denominati *automata*, avevano lo scopo di ingannare i fedeli facendo credere a questi che le azioni mostrate fossero frutto dell'intervento degli dei.

In seguito, la macchina logica inventata da Raimondo Lullo nel XIV secolo costituì il primo tentativo di realizzare un calcolatore meccanico in grado riprodurre le capacità computazionali degli esseri umani. Il lavoro di Lullo ispirò nel Seicento Gottfried Leibniz sulla possibilità di creare un dispositivo di ragionamento che utilizzasse le regole della logica per risolvere qualsiasi controversia. Il filosofo tedesco, uno degli antenati del moderno pensiero computazionale, sognava infatti di escogitare una sorta di "algebra del pensiero" grazie alla quale porre fine a qualsiasi problema semplicemente sedendosi ad un tavolo e calcolando. Sia Leibniz che Blaise Pascal realizzarono anche delle macchine in grado di compiere semplici operazioni aritmetiche, ma nessuno dei due affermò mai che le loro invenzioni potessero in qualche modo "pensare".

Tra il XIX ed il XX secolo furono compiuti grandi passi avanti nel campo della statistica e della probabilità. Il metodo dei minimi quadrati (1805), il teorema di Bayes (1802) e le catene di Markov (1913) andarono a comporre le fondamenta teoriche che sarebbero poi state alla base dell'informatica e dell'IA odierne.

Il primo lavoro generalmente riconosciuto come intelligenza artificiale risale al 1943, quando Warren McCulloch e Walter Pitts proposero un modello di neurone artificiale chiamato Threshold Logic Unit (TLU). Ad esso seguì nel 1950 la prima macchina a rete neurale, denominata SNARC, realizzata da Marvin Minsky e Dean Edmonds. Nonostante l'importanza di questi risultati, però, il maggior contributo allo sviluppo della materia in questi primi anni è forse un altro. Sempre nel 1950, infatti, il matematico inglese Alan Turing pubblicò il suo articolo "*Computing Machinery and Intelligence*". Al suo interno, l'autore per la prima volta indagò la possibilità che una macchina possa pensare e presentò un modo per valutare la sua eventuale intelligenza: il famoso test di Turing. Esso consiste in un "gioco dell'imitazione" in cui un umano pone delle domande a due interlocutori, un'altra persona ed una macchina, e attraverso le loro risposte deve cercare di capire chi sia l'uno e chi sia l'altro; allo stesso tempo, compito della macchina è quello di ingannare l'interrogatore umano, convincendolo di essere una persona in carne ed ossa.

L'IA come disciplina si considera essere nata nell'estate del 1956, quando presso il Dartmouth College John McCarthy e Minsky tennero la prima storica conferenza incentrata esclusivamente sul tema. In questa occasione, lo stesso McCarthy coniò anche l'espressione "intelligenza artificiale". Nonostante l'evento non andò secondo le aspettative degli organizzatori, poiché non ci fu un accordo su dei metodi standard da seguire, esso ottenne comunque il risultato di generare tra i partecipanti la sensazione che l'IA fosse effettivamente realizzabile.

In seguito alle discussioni presso il Dartmouth College, la ricerca visse un periodo d'oro. Già nel 1956 Allen Newell e Herbert Simon avevano sviluppato *Logic Theorist*, un programma in grado di simulare una sorta di ragionamento dimostrando alcuni teoremi matematici. Nel 1958, poi, Frank Rosenblatt introdusse il perceptrone, un algoritmo di machine learning²¹ originariamente implementato per il riconoscimento di immagini e che attirò per la prima volta l'attenzione del grande pubblico. I successi conseguiti, oltre all'intervento dei maggiori esponenti del campo, convinsero diverse agenzie governative, tra cui la Defense Advanced Research Projects Agency (DARPA), a finanziare progetti legati all'IA. Il loro interesse era rivolto in particolare verso la realizzazione di macchine che fossero in grado di trascrivere e tradurre le lingue parlate.

²¹ Per vedere cos'è il *machine learning*, si vada al paragrafo 1.4.1.

Nonostante l'inizio carico di aspettative, nel giro di un decennio l'entusiasmo generale man mano scemò. Le limitazioni tecniche dei computer dell'epoca, traducibili in una scarsa memoria e in una velocità di calcolo ridotta, non consentivano di ottenere i risultati auspicati. Un libro di Minsky e Seymour Papert pubblicato nel 1969, in cui essi evidenziavano i difetti e i limiti delle reti neurali²², influenzò la DARPA sulla decisione di interrompere i finanziamenti. Ad esso fece seguito nel 1973 uno studio sul campo dell'IA, chiamato "*Lighthill Report*", in cui si affermava che il lavoro svolto fino a quel momento non avesse raggiunto gli obiettivi prefissati. Ciò convinse il Regno Unito a sospendere a sua volta il sostegno economico, segnando l'inizio di quello che è conosciuto come il primo "inverno dell'IA": un periodo caratterizzato da generale disinteresse, mancanza di finanziamenti e conseguente rallentamento della ricerca sull'intelligenza artificiale.

Per assistere alla ripresa sarà necessario aspettare l'inizio degli anni '80. I fattori principali che portarono a questa seconda fase di splendore furono principalmente due: innanzitutto, nuovi ingenti investimenti. Basti pensare al *Fifth Generation Computer Project* (FGCP), grazie al quale tra il 1982 e il 1990 il governo giapponese spese 400 milioni di dollari per implementare nuove soluzioni negli ambiti della programmazione e dell'intelligenza artificiale. Ad essi andarono ad aggiungersi alcuni importanti miglioramenti a livello di algoritmi. La ricerca sulle reti neurali artificiali conseguì infatti notevoli risultati nello sviluppo di modelli e tecniche innovative, che contribuirono a diffondere la tecnologia. Allo stesso tempo, grazie anche agli sforzi di Edward Feigenbaum, assunsero sempre più rilievo i cosiddetti "sistemi esperti". Nati nella seconda metà degli anni '60, essi consistevano in programmi che cercavano di riprodurre le prestazioni di esperti umani in un determinato dominio di conoscenza tramite regole di tipo "*if-then*". Le reti neurali e soprattutto i sistemi esperti divennero velocemente molto popolari anche all'interno di grosse società americane, giapponesi ed inglesi, segnando l'ingresso effettivo dell'intelligenza artificiale nel mondo dell'economia e delle imprese.

In questo periodo si generò un notevole entusiasmo riguardo ciò che avrebbero potuto fare queste nuove soluzioni. Tuttavia, come già accaduto in passato, le aspettative si rivelarono presto infondate. I sistemi esperti richiedevano centinaia di regole per incapsulare la conoscenza necessaria e si dimostrarono difficili da sviluppare e da

²² Per vedere cosa sono le reti neurali, si vada al paragrafo 1.4.2.

mantenere. Lo stesso McCarthy li criticò, affermando che mancassero di buonsenso e di consapevolezza dei loro limiti. Fu così che verso la metà degli anni '80 i finanziamenti nella ricerca diminuirono nuovamente ed iniziò un secondo inverno dell'IA.

Questa fase durò circa un decennio. La svolta arrivò tra gli anni '90 e l'inizio del nuovo millennio e fu caratterizzata dalla crescita esponenziale delle applicazioni basate su machine learning, resa possibile dalla maggiore disponibilità di dati e da Internet. Ciò costituì un punto di svolta, poiché si passò da un ambito di studio prettamente teorico ad un maggior focus sulle soluzioni applicate. I principali investimenti nella ricerca provenivano ormai da importanti società appartenenti al settore dei calcolatori, come IBM, Hitachi e Toshiba. Non è un caso che nel 1997 fu proprio un programma realizzato da IBM, *Deep Blue*, a sconfiggere a scacchi Gary Kasparov, decretando la prima vittoria di un computer ai danni di un campione del mondo. A questi grandi player si contrapponeva una costellazione di nuove startup tecnologiche, in grado di risaltare grazie alle idee di alcuni brillanti programmatori e scienziati.

I primi anni 2000 rappresentarono un periodo cruciale per lo sviluppo dell'intelligenza artificiale come è conosciuta oggi. Le applicazioni si trasformarono da programmi finalizzati a risolvere complessi ma pur sempre finiti problemi combinatori a veri e propri sistemi capaci di imparare e di imitare più accuratamente forme di ragionamento complesso. Ci fu un notevole ampliamento di ciò che l'IA era in grado di fare, con l'introduzione di nuove funzioni molto simili alle abilità umane, tra cui: il riconoscimento di immagini, suoni e testi senza istruzioni preliminari; la risoluzione di problemi imprevisti; il conseguimento di obiettivi in contesti dinamici²³. Tutto ciò fu possibile grazie alla combinazione di alcuni importanti elementi. Innanzitutto, come già visto l'espansione di Internet aiutò ricercatori e sviluppatori mettendo loro a disposizione un'enorme quantità di dati facili da ottenere. Ad essa si aggiunsero alcune innovazioni tecnologiche che permisero un miglioramento delle doti computazionali ed un notevole aumento della velocità di esecuzione dei processi. Nel giro di una decina d'anni fu possibile raggiungere traguardi fino a poco prima impensabili. Basti pensare ad *AlphaGO*, un programma creato da Deep Mind che nel 2016 riuscì a battere il campione mondiale di Go. Questo evento è significativo a causa dell'elevata complessità del gioco:

²³ Fanti, L., Guarascio, D., Moggi, M. (2022), "From Heron of Alexandria to Amazon's Alexa: a stylized history of AI and its impact on business models, organization and work", *Journal of Industrial and Business Economics*, 49, pp. 409-440.

il numero di configurazioni possibili è maggiore di quello degli atomi nell'universo e per questo motivo non è mai stato possibile ottenere buoni risultati valutando tutte le possibili mosse (come per *Deep Blue* negli scacchi).

Attualmente è in corso il più longevo periodo di splendore dell'intelligenza artificiale. Al momento non sembrano esserci le premesse per un nuovo drastico calo dell'interesse, nonostante alcuni si interrogano sulla possibilità di un altro imminente inverno dell'IA. Il panorama, anche nell'ambito della ricerca, non è più per la maggior parte appannaggio del mondo accademico e governativo; anzi i principali attori sono ormai le grandi società del digitale americane e cinesi, tra cui Alibaba, Alphabet, Amazon, Facebook, Microsoft e Tencent.

1.3. Classificazioni

L'intelligenza artificiale è una materia molto variegata sia sul piano teorico, sia su quello pratico. È naturale, dunque, che esista più di un metodo di classificazione utilizzabile, ognuno dei quali si differenzia per concentrarsi su un determinato aspetto.

Come si è visto nella precedente analisi delle definizioni, numerosi autori hanno preso a riferimento gli esseri umani nei loro tentativi di stabilire cosa sia l'IA. Questa tendenza non riguarda solo l'aspetto epistemico della disciplina, bensì è legata ad un fenomeno più ampio. A lungo, infatti, la ricerca in questo campo, in particolar modo tra gli anni '50 e '60, si è focalizzata nella realizzazione di tecnologie e programmi che fossero in grado di imitare l'intelligenza umana. Ciò ha portato allo sviluppo di diverse soluzioni, alcune capaci di mostrare effettivamente certe facoltà tipicamente umane, altre molto più simili a dei normali calcolatori. In ogni caso, nessuna di esse è ancora mai riuscita a raggiungere pienamente l'obiettivo finale.

In base alla somiglianza con la mente umana e alla complessità dei processi che può svolgere, è possibile ricondurre un sistema di intelligenza artificiale ad una di quattro tipologie: macchine reattive, memoria limitata, teoria della mente, consapevolezza di sé.

Macchine reattive

Sono le forme di IA più semplici, e anche le più vecchie. Hanno capacità molto ridotte, in quanto si limitano a rispondere in modo automatico ad un preciso insieme di stimoli, secondo delle regole preprogrammate. Inoltre, non utilizzano memoria, per cui non sono in grado di utilizzare le esperienze pregresse per imparare e influenzare le loro azioni correnti. Questi sistemi, dato un determinato input, forniranno sempre lo stesso identico output.

Un tipico esempio di macchina reattiva è *Deep Blue*, programma di scacchi in grado di sconfiggere nel 1997 l'allora campione del mondo Garry Kasparov.

Memoria limitata

In aggiunta rispetto alla tipologia precedente, l'IA a memoria limitata può imparare dall'esperienza passata. Essa è in grado, infatti, di usare dati storici in combinazione con informazioni preprogrammate per migliorare le proprie risposte e fare previsioni, prendere decisioni o svolgere complesse attività di classificazione.

La maggior parte dei sistemi attualmente esistenti è riconducibile a questa categoria. È possibile citare un gran numero di esempi, dagli odierni chatbot ed assistenti virtuali ai modelli per il riconoscimento di immagini addestrati tramite una rete neurale artificiale²⁴, fino ai veicoli a guida autonoma.

Teoria della mente

Prima di proseguire, è necessario fare una premessa. A differenza delle due precedenti, questa tipologia di intelligenza artificiale e la prossima non sono state raggiunte al giorno d'oggi. Esse esistono soltanto come concetti o come ambiti di ricerca.

In psicologia, la teoria della mente consiste nella capacità di attribuire stati mentali (intesi come pensieri, emozioni, credenze, desideri, intenzioni) a sé e agli altri, e di prevedere il proprio o l'altrui comportamento²⁵. Un'IA che abbia conseguito un tale livello, pertanto, dovrebbero essere in grado riconoscere, capire, conservare e ricordare diverse reazioni emotive ed atteggiamenti e sapere di volta in volta come comportarsi. Inoltre, dovrebbe riuscire a cogliere gli eventuali cambiamenti di stato emotivo per

²⁴ Per vedere cosa sono le reti neurali artificiali, si vada al paragrafo 1.4.2.

²⁵ Lecciso, F., Liverta Sempio, O., Marchetti, A. (2005), *Teoria della mente. Tra normalità e patologia*, Raffaello Cortina Editore, Milano.

modificare a sua volta la propria risposta. Tutto ciò consentirebbe di migliorare notevolmente le interazioni con altri soggetti.

Attualmente esistono già dei sistemi capaci di leggere le emozioni tramite l'analisi di elementi quali le espressioni facciali, il linguaggio del corpo e il tono di voce (intelligenza artificiale emotiva). Tuttavia, sono ancora necessari importanti passi in avanti prima di poter parlare di teoria della mente.

Consapevolezza di sé

È il livello finale dello sviluppo dell'intelligenza artificiale. Indica una macchina talmente evoluta e simile nel funzionamento alla mente umana da aver sviluppato un'autocoscienza. Ha capacità intellettuali ed emotive paragonabili a quelle di una qualsiasi persona: pensa di sé stessa, ha dei desideri e prova delle emozioni. Questo stadio rappresenta uno passo successivo rispetto alla teoria della mente, in quanto questi ipotetici sistemi non solo comprendono lo stato d'animo altrui, ma anche il proprio.

Esempi di IA dotata di consapevolezza di sé esistono soltanto nei libri e nei film. Forse in futuro sarà possibile conseguire un tale risultato, ma al momento quel giorno sembra molto lontano.

Oltre a quello appena esposto, esiste un altro metodo di classificazione molto utilizzato sia nell'ambito della ricerca, sia in contesti più divulgativi. Esso prende nuovamente come riferimento gli esseri umani, ma è in un certo senso più "operativo" del precedente, in quanto si concentra sulla quantità e sul livello di competenze possedute da un sistema di intelligenza artificiale. Secondo questa visione, esistono tre diverse categorie: l'IA ristretta, l'IA generale e la super IA.

IA ristretta

L'intelligenza artificiale ristretta, in inglese *artificial narrow intelligence* (ANI), include qualsiasi programma o applicazione di IA realizzato fino ad oggi, anche il più complesso. È definita "ristretta" poiché è capace di eseguire solo azioni specifiche, per le quali è stata programmata o addestrata. Essa non replica l'intelligenza umana, bensì si limita a simulare il comportamento umano per una gamma ristretta di competenze. Giusto per

fare un esempio, un algoritmo per l'elaborazione del linguaggio naturale non potrà essere utilizzato per il riconoscimento di immagini all'interno di foto, e viceversa.

IA generale

L'intelligenza artificiale generale, in inglese *artificial general intelligence* (AGI), consiste in un sistema dotato di intelligenza generale²⁶ ed in grado di percepire, imparare, comprendere ed agire come un essere umano. Una simile IA, dunque, può padroneggiare numerose competenze e formare connessioni attraverso sfere della conoscenza diverse fra loro.

Al giorno d'oggi, i ricercatori non sono ancora riusciti a creare un'AGI. Risulta difficile stabilire quanto degli sforzi fatti nello campo dell'IA ristretta possa contribuire al raggiungimento di questo obiettivo, soprattutto perché non è tuttora chiaro cosa serva per sviluppare un'IA generale. Ciò che è certo è che per il momento essa rimane esclusivamente un concetto teorico.

Super IA

La super intelligenza artificiale, in inglese *artificial super intelligence* (ASI), è un'ipotetica IA che non solo saprebbe imitare fedelmente l'intelligenza e il comportamento degli esseri umani, ma addirittura si dimostrerebbe migliore di loro sotto ogni aspetto: dalla matematica allo sport, dalla medicina all'arte, fino alle relazioni emotive. Ciò sarebbe possibile grazie ad alcune capacità "aumentate": una maggiore memoria, una più elevata velocità di analisi dei dati e migliori capacità decisorie, unite ad una vasta gamma di competenze, renderebbero questi sistemi le più complesse forme d'intelligenza sul pianeta.

Nonostante in molti siano preoccupati da un tale scenario, attualmente l'ASI appartiene solamente al genere della fantascienza.

Le due modalità di classificazione presentate non sono del tutto estranee tra loro, e non solo perché entrambe prendono a riferimento l'intelligenza e le facoltà umane. È possibile infatti identificare delle sovrapposizioni tra le diverse tipologie, come illustrato nella figura 3.

²⁶ Per un approfondimento sull'intelligenza, si vada al paragrafo 1.5.

Macchine reattive	Memoria limitata	Teoria della mente	Consapevolezza di sé
IA ristretta		IA generale	Super IA

Tabella 3. Le due modalità di classificazione riunite

Le categorie “macchine reattive” e “memoria limitata” sono perfettamente assimilabili all’IA ristretta. Ciò è facilmente comprensibile, in quanto queste tre denominazioni includono tutti i sistemi di intelligenza artificiale esistenti al giorno d’oggi.

Entrando nel campo di ciò che ancora non esiste, in linea di massima la tipologia “teoria della mente” combacerebbe con l’IA generale. Lo sviluppo di un’effettiva teoria della mente da parte di un’intelligenza artificiale da un lato richiederebbe il possesso di numerose competenze molto diverse tra loro, soprattutto in ambito relazionale; dall’altro migliorerebbe notevolmente il modo di interagire di questi sistemi, rendendoli estremamente simili, se non addirittura indistinguibili, rispetto ad una qualsiasi persona.

Infine, la comparsa di un’autocoscienza rappresenterebbe l’ultimo passo da compiere per eguagliare sotto ogni punto di vista gli esseri umani, proiettando l’intelligenza artificiale verso il livello di super IA.

Non è ben chiaro se le ultime due sovrapposizioni siano del tutto accurate, specialmente nella zona di confine. D’altronde, finché non saranno effettivamente realizzati delle macchine o degli algoritmi con delle tali capacità, queste distinzioni e come si legano tra loro resteranno soltanto delle speculazioni.

1.4. Campi di applicazione

Oggigiorno l’intelligenza artificiale è estremamente diffusa, al punto che spesso molte persone non si rendono nemmeno conto di farne un uso quotidiano. La grande versatilità di questa tecnologia ha fatto sì che iniziasse ad essere utilizzata in numerosi ambiti, a livello professionale e non. Malgrado questa grande varietà, è possibile

identificare sette tipologie di funzionalità riscontrabili nelle diverse applicazioni dell'IA²⁷. Questi “pattern” non sono mutualmente esclusivi, bensì a seconda dei casi possono presentarsi singolarmente oppure insieme ad altri.

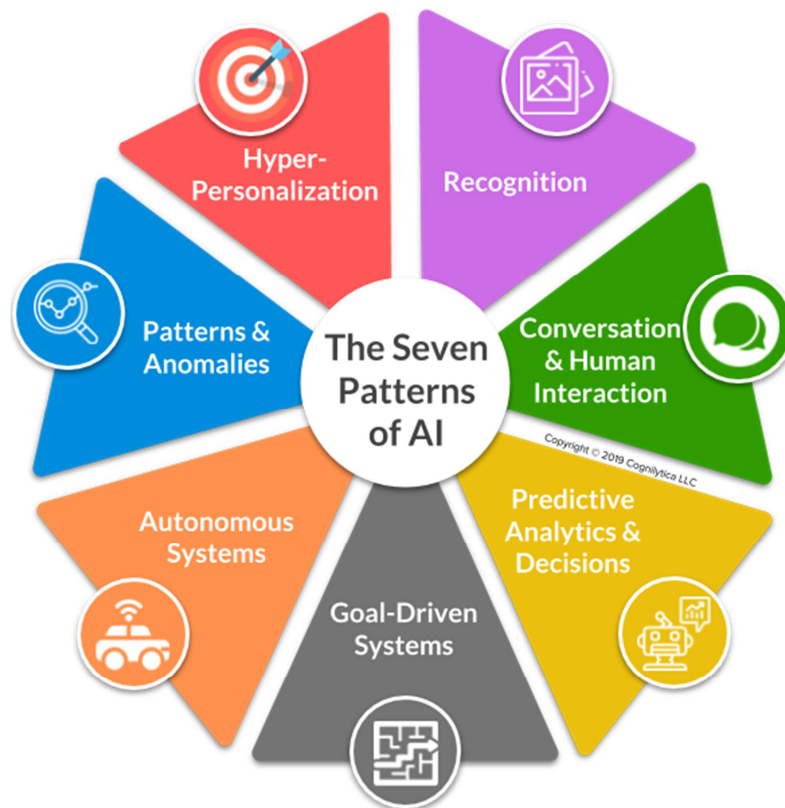


Figura 1. I 7 pattern dell'intelligenza artificiale²⁸

1) Iper-personalizzazione

Consiste nell'utilizzare l'intelligenza artificiale al fine di sviluppare per ogni individuo un profilo che sia unico e che possa adattarsi nel corso del tempo. In questo modo è possibile andare oltre la suddivisione in categorie e considerare ciascuna persona singolarmente, così da proporre contenuti, fornire assistenza e fare raccomandazioni che risultino il più possibile personalizzate. È molto comune nell'ambito del marketing, ma è presente anche in settori come la finanza, la sanità, il fitness ed il benessere.

²⁷ Cognilytica (2019), *The Seven Patterns of AI*, <https://www.cognilytica.com/2019/04/04/the-seven-patterns-of-ai/>.

²⁸ *Ibidem*.

2) Sistemi autonomi

Sono sistemi, sia hardware che software, in grado di svolgere mansioni, raggiungere obiettivi ed interagire con l'ambiente circostante con un intervento degli esseri umani ridotto o del tutto assente. Per fare ciò, essi devono innanzitutto saper percepire il mondo esterno e avere doti di previsione e pianificazione. L'obiettivo principale è senza dubbio ridurre od eliminare il lavoro umano, mantenendo però almeno lo stesso livello di performance. Gli esempi più comuni sono i veicoli a guida autonoma, ma questi sistemi sono molto usati nella creazione automatica di documenti, nell'automazione dei processi di business ed in ambito logistico.

3) Analisi predittiva e supporto alle decisioni

L'IA può essere utilizzata per aiutare gli esseri umani a prendere decisioni migliori. Essa, infatti, è capace di supportare le capacità analitiche umane in diversi stadi del processo decisionale:

- migliora la comprensione delle prestazioni passate, identificando le tendenze storiche, e il monitoraggio delle azioni in corso;
- fornisce una maggiore visualizzazione dei dati e l'identificazione delle relazioni intercorrenti, aiutando a capire perché si sono verificati certi risultati, eventi o trend;
- stima risultati sconosciuti e anticipa comportamenti futuri partendo dalle relazioni esistenti tra vari fattori;
- aiuta a specificare una linea d'azione preferita, allocando in modo efficiente le risorse o trovando la strada più adatta.

Ciò risulta particolarmente vero quando si ha a che fare con decisioni complesse, istantanee, oggettive e replicabili:

- complesse a causa dei dati da considerare, in quanto o esistono molteplici dimensioni di analisi per ogni singolo dato, oppure l'incertezza nella loro misurazione è elevata, o ancora la quantità di dati a disposizione è notevole;
- istantanee quando al manifestarsi di un determinato evento servono risposte in real-time;
- oggettive per assicurare *accountability* al processo decisionale;

- replicabili per far sì che l'esperienza accumulata nel processo decisionale possa essere scalata.

Questo pattern è utilizzato per la previsione di guasti ed avarie, la risoluzione assistita di problemi, l'ottimizzazione delle attività, la navigazione intelligente ed altro ancora.

4) Conversazione ed interazione umana

Si riferisce a macchine e programmi che riescono ad interagire con gli umani tramite la voce, messaggi scritti o immagini. Include anche la creazione di contenuti (testi, immagini, video, audio) che siano destinati e fruibili per gli esseri umani. L'obiettivo è far sì che l'intelligenza artificiale possa relazionarsi con le persone allo stesso modo in cui queste comunicano fra loro. Alcuni esempi sono gli assistenti vocali e i chatbot.

5) Rilevazione di schemi e anomalie

Uno dei maggiori utilizzi dell'IA consiste nella rilevazione di schemi e connessioni in grandi moli di dati. Ciò consente di scoprire delle relazioni che gli esseri umani non saprebbero portare alla luce, sia per l'apparente casualità di certe fenomeni, sia per il numero limitato di dati che sono in grado di processare e analizzare. Inoltre, è possibile verificare se i dati a disposizione rispettano il modello esistente o se sono presenti delle anomalie. Tra le principali applicazioni si ricordano la rilevazione di rischi e frodi e il testo predittivo, il quale è estremamente diffuso nelle tastiere degli smartphone.

6) Riconoscimento

È definito come l'impiego di sistemi di intelligenza artificiale al fine di identificare ed etichettare particolari elementi all'interno di immagini, video, testi, audio o altre forme di contenuti non strutturati²⁹. Questa funzionalità si sta rivelando molto importante soprattutto per le imprese, dato che anche il 90% dei loro dati è di tipo non strutturato. Riuscire a comprenderli ed analizzarli risulta quindi essenziale. Questo è uno dei pattern più diffusi ed è adottato in numerose applicazioni, quali il riconoscimento di oggetti, immagini, gesti e volti.

²⁹ I dati non strutturati si definiscono così in quanto non presentano un'architettura identificabile. Provengono generalmente da immagini, video, file audio e testi (email, pagine web). Per un approfondimento su dati strutturati, non strutturati e semi-strutturati, si vada al paragrafo 2.2.

7) Sistemi goal-driven

Sono sistemi dotati dell'abilità di imparare dai propri errori, allo scopo di trovare la soluzione ottimale ad un determinato problema. È una categoria storicamente legato al mondo dei giochi e al tentativo di creare dei programmi che potessero vincere contro degli sfidanti umani (vedi *AlphaGo*), ma oggi è molto presente anche nell'ottimizzazione delle risorse, nella risoluzione di problemi ripetuti e nel *real-time bidding*.

Questi sette pattern sono caratterizzati da obiettivi diversi, ma come già detto non necessariamente devono essere perseguiti uno per volta. Anzi, i progetti legati all'IA sono talmente numerosi e diversificati oramai che non è difficile imbattersi in certi che combinino più di una funzionalità per raggiungere il risultato desiderato.

Presentando le diverse tendenze, sono già state anticipate alcune applicazioni dell'intelligenza artificiale. Verrà ora approfondito questo tema, illustrando le più importanti aree in cui questa tecnologia è impiegata al giorno d'oggi. Nel corso della spiegazione, sarà anche possibile riconoscere i pattern appena visti.

Natural language processing

Il *natural language processing* (NLP) comprende tutti quegli algoritmi di intelligenza artificiale utilizzati per analizzare, classificare e comprendere il linguaggio naturale, intendendo con ciò una forma di linguaggio che viene normalmente e spontaneamente utilizzata all'interno di un gruppo di individui (ad esempio una qualsiasi lingua). Si occupa prevalentemente di testi, mentre l'elaborazione del parlato è considerata un ambito a sé. Questo campo unisce informatica e linguistica per consentire alle macchine di rivelare, attraverso una serie di tecniche e passaggi, la struttura e il significato di un qualsiasi testo. Ciò risulta particolarmente complesso a causa delle numerose ambiguità che possono essere presenti nel linguaggio umano. Le applicazioni sono varie e spaziano dalla traduzione e comprensione del contenuto, fino alla produzione in modo autonomo di testi.

Automatic speech recognition

L'*automatic speech recognition* (ASP), detto anche *speech-to-text*, consiste nella capacità di un sistema di IA di riconoscere e processare la lingua parlata e di trascriverla sotto

forma di testo con un alto grado di precisione. Combina modelli linguistici ad elementi di acustica ed in alcuni casi sfrutta anche modelli di NLP per migliorare l'utilizzo della punteggiatura e delle lettere maiuscole nelle trascrizioni. L'ASP è certamente usato nelle trascrizioni, ma è impiegato anche in numerose altre attività eseguibili tramite l'utilizzo della voce, come ad esempio le ricerche vocali via smartphone, l'utilizzo di dispositivi di smart home e le applicazioni di self-service e di instradamento delle chiamate nei *contact center*.

Virtual assistant e chatbot

In questa categoria rientrano agenti software in grado di recepire dei comandi in forma scritta o parlata e di agire conseguentemente. Sebbene le due denominazioni siano spesso ritenute intercambiabili, in realtà si riferiscono ad applicazioni dell'intelligenza artificiale diverse tra loro. I *chatbot* sono progettati per interagire con gli esseri umani e rispondere ad una serie di domande. I più evoluti sanno anche comprendere il tono ed il contesto del dialogo, tuttavia riescono a rispondere soltanto ad una gamma limitata di domande. Sono particolarmente diffusi in contesti aziendali, per servizi di *customer care*, *recruiting* o gestione del personale. I *virtual assistant*, invece, sono capaci di comprendere il linguaggio in modo più approfondito, riuscendo anche a riconoscere lo slang o i sentimenti dell'interlocutore, oltre che il contesto. Possono inoltre svolgere un numero di compiti più ampio rispetto ai chatbot. Sono virtual assistant i vari software e dispositivi di smart home prodotti dai giganti del digitale, quali Siri, Alexa, Cortana e Google Home.

Recommendation system

Sono algoritmi in grado di indirizzare le scelte degli utenti in base alle informazioni fornite (direttamente o indirettamente) dagli utenti stessi. Essi sfruttano un gran quantitativo di dati, che possono riguardare precedenti ricerche e scelte, caratteristiche demografiche, gusti personali e altro ancora, per prevedere interessi e desideri di diversi individui e consigliare loro specifici contenuti. In questo modo, è possibile fornire agli utenti esperienze personalizzate ed aumentare il loro coinvolgimento. Tra gli esempi più comuni rientrano i sistemi, estremamente diffusi nelle piattaforme social e di e-commerce, che suggeriscono determinati prodotti (Amazon) o film (Netflix) a seconda di quelli acquistati o visti in passato.

Computer vision

La *computer vision* si occupa di algoritmi e tecniche che consentono ad una macchina di “vedere”, riproducendo funzioni e processi della vista umana. Così facendo, è possibile utilizzare immagini e video come fonti di dati significativi, in base ai quali intraprendere eventualmente successive azioni. Attraverso una serie di task (come ad esempio il riconoscimento di oggetti e volti), questi algoritmi riescono ad individuare ed analizzare diversi elementi presenti in un'immagine, arrivando addirittura a ricostruirne l'intero contesto e attribuendole un vero e proprio significato. La computer vision è utilizzata in svariati contesti, dalla *smart surveillance* alla diagnosi di disturbi medici, dalla manutenzione predittiva alla sicurezza nei luoghi di lavoro.

Intelligent data processing

L'*intelligent data processing* comprende tecnologie che consentono di estrarre ed elaborare informazioni utili da un'ampia gamma di dati, in modo veloce e preciso. Esse si avvalgono anche di altre applicazioni dell'IA, come il natural language processing o la computer vision, in modo da poter lavorare con dati sia strutturati, sia non strutturati. Sono molto utili nel mondo del lavoro, in quanto permettono di automatizzare e velocizzare numerosi processi meccanici, così che gli esseri umani possano concentrarsi su compiti che richiedono maggiori sforzi creativi ed intellettivi. Sistemi di intelligent data processing sono particolarmente utilizzati nella classificazione dei dati e nel *forecasting*.

Soluzioni fisiche

L'intelligenza artificiale può anche essere utilizzata in combinazione con dispositivi fisici per far sì che questi riescano a svolgere azioni o muoversi nel mondo in totale o parziale autonomia. A questa classe di applicazioni appartengono innanzitutto gli *autonomous vehicle*. Essi sono un tipologia di veicoli in grado di operare senza intervento umano, grazie ad una serie di sensori che permettono loro di raccogliere in tempo reale dati sull'ambiente, sulle condizioni e sugli agenti circostanti. Questi dati sono poi utilizzati dall'IA per prendere decisioni riguardo al percorso da seguire e alle azioni da intraprendere. Gli autonomous vehicle possono essere adibiti al trasporto di persone, cose o animali, sia su strada che via acqua o aria.

In aggiunta, rientrano tra le soluzioni fisiche gli *autonomous robot*, i quali sono anch'essi capaci di raccogliere informazioni sull'ambiente in cui si trovano e muoversi autonomamente. Sono spesso utilizzati in ambito industriale per l'automazione di processi produttivi e logistici.

IA generativa

L'IA generativa è una particolare area dell'intelligenza artificiale che raggruppa una serie di algoritmi e software in grado di creare contenuti artificiali di varia natura. Essa consente di supportare le attività creative degli esseri umani, fornendo loro nuovi contenuti generati a partire da altri, quali testo, immagini o file audio. Tra le applicazioni più diffuse ci sono i modelli di *text-to-image*, che restituiscono un'immagine in base ad una descrizione testuale fornita dall'utente. Questi sistemi sono ritenuti tra i più promettenti nel ambito dell'IA per i prossimi anni, al punto che Gartner ha inserito l'IA generativa tra i suoi *Top Strategic Technology Trends* del 2022.

Quelli presentati sono i principali campi di applicazione dell'IA. Prima di proseguire, però, è fondamentale soffermarsi su un'altra area della disciplina, fondamentale anche per quelle già trattate: il *machine learning*.

1.4.1. Machine learning

Il *machine learning* (ML), o apprendimento automatico, è uno dei sottoinsiemi più importanti dell'intelligenza artificiale. L'espressione, coniata nel 1959 dallo scienziato americano Arthur Samuel³⁰, indica la capacità di una macchina di "apprendere" e svolgere mansioni per le quali non è stata preventivamente programmata. Invece di adottare un approccio tradizionale, in base a cui si indicano tutti i passaggi da seguire per raggiungere un determinato risultato, questi software sono in grado di imparare dall'esperienza, grazie all'elaborazione di grandi moli di dati.

In genere, in ogni progetto di machine learning si possono distinguere alcuni elementi principali:

- un algoritmo, un insieme di regole e procedure che consentono di riconoscere schemi ricorrenti nei dati tramite l'utilizzo di diversi metodi statistici;

³⁰ Samuel, A.L. (1959), "Some Studies in Machine Learning Using the Game of Checkers", *IBM Journal of Research and Development*, 3, pp 210-229.

- un *dataset* di input, il quale funge da base di partenza e sul quale l'algoritmo lavora (svolge il *training*);
- un modello, il risultato ottenuto una volta che l'algoritmo ha finito di processare i dati e che rappresenta un contesto reale con opportuni parametri;

Qualora il modello non dovesse rivelarsi preciso o accurato, può essere raffinato ulteriormente utilizzando un nuovo set di dati.

Esistono quattro diverse tipologie di machine learning, che si differenziano soprattutto per il ruolo assunto dagli esseri umani nel corso del processo .

Supervised learning

Il *supervised learning*, o apprendimento supervisionato, è attualmente la forma di ML più utilizzata. Ciò probabilmente si deve al fatto che è quella che maggiormente si avvicina all'approccio umano: sin da piccoli i bambini imparano dall'esperienza attraverso un gran numero di esempi, acquisendo degli schemi che una volta rivisti riescono a riconoscere. Allo stesso modo, nel supervised learning si utilizzano dati pre-etichettati, per ciascuno dei quali ad un input è associato un determinato output, così che l'algoritmo impari la relazione esistente e le regole del modello.

Il procedimento è costituito da diverse fasi:

- i dati di input sono etichettati e ad ognuno è assegnato il corretto output. Questo passaggio avviene sotto la supervisione umana;
- l'intero *dataset* è suddiviso in due parti, una da utilizzare per l'addestramento e la seconda per il test finale;
- l'algoritmo è addestrato sul *dataset* di *training*;
- si forniscono al modello i dati di input ignoti del *dataset* di test e si confrontano gli output attesi con quelli prodotti. Qualora l'errore risulti sufficientemente basso, l'addestramento si conclude. In caso contrario, il modello sarà ulteriormente allenato fornendo all'algoritmo nuovi dati etichettati.

Gli algoritmi di supervised learning possono essere utilizzati per problemi di classificazione, in cui è necessario assegnare i dati a delle specifiche categorie; o per problemi di regressione, nei quali si cerca di comprendere la relazione che intercorre tra variabili dipendenti ed indipendenti.

L'apprendimento supervisionato è generalmente molto accurato e semplice da implementare, però necessita di un notevole intervento umano nella preparazione del

dataset. Inoltre, assegnare il corretto output può in alcuni casi rivelarsi non semplice e richiedere il possesso di determinate competenze.

Unsupervised learning

Al contrario del precedente, l'*unsupervised learning*, o apprendimento non supervisionato, lavora con dati non etichettati e senza alcuna supervisione umana. Questi algoritmi non necessitano di un *dataset* di *training*, poiché apprendono direttamente dai dati. Essi cercano modelli nascosti che sfuggono all'osservazione, andando ad individuare similitudini e differenze o relazioni tra gli input. Queste sono spesso rappresentate dalle distanze tra i dati, calcolate utilizzando tecniche matematiche e statistiche. In base alle distanze, poi, si procede a raggruppare i dati, fino al loro esaurimento. Non essendoci un output previsto, non è possibile valutare la validità del modello ottenuto tramite un secondo *dataset*. Tuttavia, esistono delle tecniche alternative che variano a seconda dell'algoritmo scelto.

L'*unsupervised learning* è impiegato in tre diverse categorie di compiti:

- il *clustering*, in cui i dati iniziali sono raggruppati in sottoinsiemi (*cluster*) a seconda delle loro somiglianze o differenze;
- l'associazione, la quale permette di identificare delle relazioni predefinite tra le variabili di grandi database (ad es. per suggerire ai clienti prodotti spesso acquistati insieme);
- la riduzione della dimensionalità, una tecnica utilizzata per diminuire il numero di caratteristiche/dimensioni di un set di dati, mantenendo al tempo stesso la loro integrità.

Gli algoritmi di apprendimento non supervisionato consentono di ottenere insight da grandi moli di dati, suddividendoli in sottoinsiemi con caratteristiche comuni o anche rilevando elementi che si discostano dagli altri. Il fatto che non richiedano alcun intervento umano nella preparazione degli input è sicuramente un vantaggio, tuttavia la loro complessità e la necessità di una notevole potenza computazionale possono rappresentare degli ostacoli al loro uso.

Semi-supervised learning

Il *semi-supervised learning*, o apprendimento semi-supervisionato, è una categoria di ML ibrida, a metà tra il supervised e l'*unsupervised learning*. Queste tecniche consentono di

utilizzare dati sia etichettati, sia non etichettati e trovano la loro ragion d'essere in alcune situazioni reali molto comuni. Il supervised learning permette infatti di ottenere buoni risultati, tuttavia preparare grandi *dataset* di *training* e di test è molto costoso. D'altro canto, l'apprendimento non supervisionato richiede meno sforzi, ma spesso si rischia di trovare raggruppamenti non ottimali. Per risolvere queste criticità, il semi-supervised learning utilizza pochi dati etichettati in combinazione con un gran numero di dati non etichettati: i primi danno struttura al problema, stabilendo le categorie e le corrispondenze tra queste e i *cluster*, mentre i secondi apportano il contesto.

Reinforcement learning

Il *reinforcement learning*, o apprendimento per rinforzo, si differenzia sostanzialmente dalle precedenti tipologie di ML, e probabilmente è quella che assomiglia in misura maggiore all'apprendimento naturale. Infatti, come quest'ultimo esso si basa sull'esperienza, sull'interazione con l'ambiente circostante e sull'abilità di applicare quanto imparato in situazioni nuove.

Il reinforcement learning si basa su un sistema di rinforzi. L'agente da addestrare non viene istruito sulle azioni da compiere, bensì gli viene solamente dato un obiettivo. In base al risultato di una specifica azione, riceverà una ricompensa o una penalità (il rinforzo), che dovrà cercare di massimizzare per raggiungere il traguardo stabilito. Questo processo prevede diverse fasi:

- la definizione dell'ambiente in cui opererà (e dell'interfaccia che utilizzerà) l'agente, del segnale di ricompensa e della *policy*, ovvero dell'algoritmo di *training*;
- l'addestramento;
- la diffusione della *policy* individuata.

Come sottolineato da Sutton e Barto³¹, una delle principali sfide del reinforcement learning riguarda il *trade-off* tra esplorazione e sfruttamento (*exploration* ed *exploitation*): sfruttamento delle azioni note e che hanno già dimostrato precedentemente di essere efficaci ed esplorazione di nuove soluzioni che possono portare a ricompense maggiori. Il dilemma sta nel fatto che nessuna delle due strategie,

³¹ Barto, A.G., Sutton, R.S. (2018), *Reinforcement Learning: An Introduction, 2nd ed.*, MIT Press, Cambridge.

se perseguita in via esclusiva, permette di conseguire l'obiettivo. L'agente deve provare una varietà di azioni e favorire progressivamente quelle che sembrano essere le migliori. Queste metodologie sono particolarmente utili quando si ha ben chiaro un traguardo, ma l'ambiente è incerto e ci sono molti modi per svolgere un compito. Hanno il grande vantaggio che non necessitano di alcun *trai* per l'apprendimento, ma solo di un gran numero di cicli azione-rinforzo. Tuttavia, possono richiedere uno sforzo maggiore a livello di progettazione.

I sistemi di machine learning rivestono oggi un ruolo cruciale non solo nella ricerca sull'IA, ma anche per molte società ed istituzioni. La capacità di processare grandi moli di dati con un ridotto apporto di risorse, unita alla loro versatilità, li rende in molti casi degli asset preziosi. Essi sono in grado di svolgere compiti di diversi livelli di complessità, come: descrivere quanto già avvenuto; fare previsioni su quello che accadrà; fornire suggerimenti sulle azioni da intraprendere³².

Quasi tutte le applicazioni moderne dell'intelligenza artificiale, dal natural language processing alla computer vision, dagli assistenti virtuali all'intelligent data processing, fino alle soluzioni fisiche sono alimentate dal machine learning. È estremamente diffuso in numerosi ambiti della vita quotidiana, nonostante spesso non ce ne si renda nemmeno conto. Lo si può trovare utilizzando un motore di ricerca, nei filtri anti-spam delle email, nella prevenzione delle frodi, nella finanza, nella ricerca scientifica ed in campo medico, nei veicoli a guida autonoma e anche nei sistemi di raccomandazione di piattaforme quali Amazon, Netflix o Spotify.

Esiste un ramo del machine learning che negli ultimi sta consentendo di ottenere risultati straordinari e pertanto merita di essere approfondito: il *deep learning*.

1.4.2. Deep learning

Il *deep learning* (DL), o apprendimento profondo, è per l'appunto una sottocategoria del machine learning e rappresenta la sua forma più avanzata. Ciò che lo contraddistingue dalle altre tecniche di apprendimento automatico è il ricorso alle reti neurali artificiali (*artificial neural networks*, ANN), o più semplicemente reti neurali. Queste sono dei

³² Malone, T.W., Rus, D., Laubacher, R. (2020), *Artificial Intelligence and the Future of Work*, disponibile a <https://workofthefuture.mit.edu/wp-content/uploads/2020/12/2020-Research-Brief-Malone-Rus-Laubacher2.pdf>.

modelli di calcolo matematico-informatici costituiti da nodi interconnessi fra loro e si basano sul funzionamento delle reti neurali biologiche. Sono sistemi adattativi, capaci di modificare la propria struttura interna.

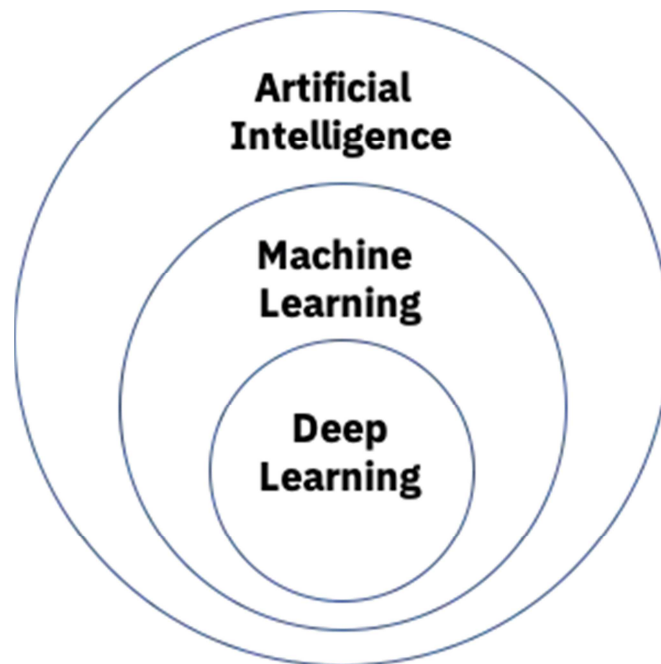


Figura 2. Il rapporto esistente tra IA, ML e DL³³

Le ANN sono composte da diversi livelli o strati (*layer*), ciascuno dei quali è formato anche da migliaia di nodi. Questi fungono da neuroni artificiali, hanno un peso ed un valore di soglia associati e sono collegati a svariati altri nodi interni. Qualora l'output di un neurone superi il valore di soglia specificato, esso si attiva e trasmette il risultato delle sue elaborazioni al livello successivo della rete; in caso contrario, non avviene alcun passaggio.

Generalmente, ogni rete neurale è costituita da:

- un livello di input (*input layer*), che riceve i segnali in ingresso, adattandoli alle esigenze dei neuroni degli strati successivi;
- uno o più livelli nascosti (*hidden layers*), dove avviene il processo di elaborazione vero e proprio;

³³ IBM Cloud Education (2020), *Artificial Intelligence (AI)*, IBM, online, <https://www.ibm.com/cloud/learn/what-is-artificial-intelligence#toc-artificial-7ZT8FnXd>.

- un livello di uscita (*output layer*), il quale raccoglie i risultati delle operazioni svolte negli strati precedenti.

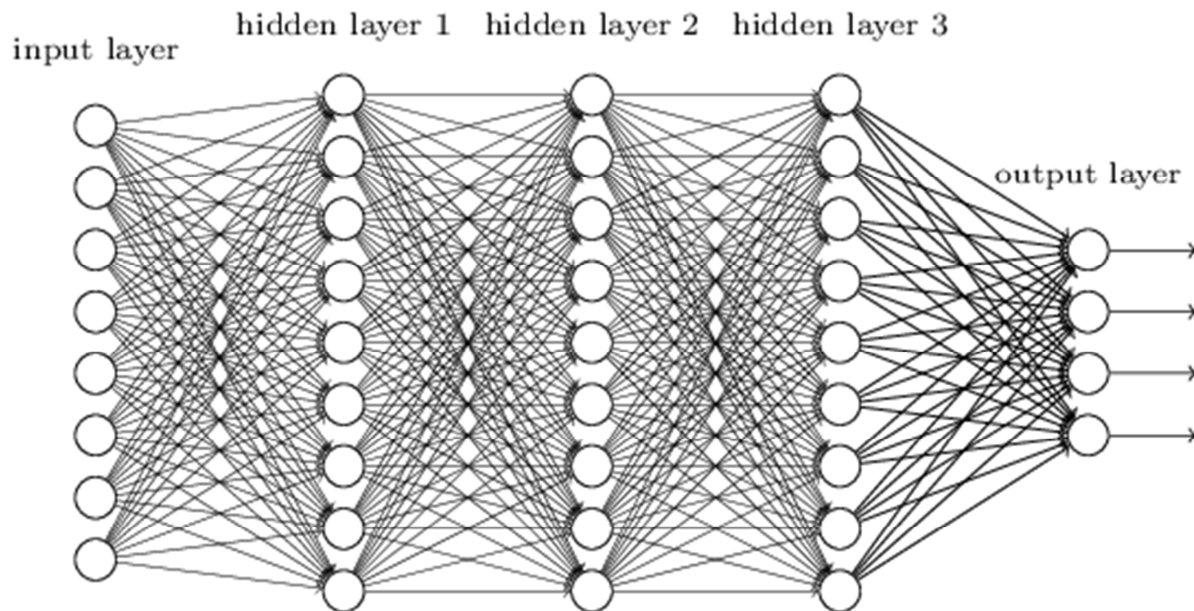


Figura 3. La struttura di una rete neurale artificiale³⁴

Per far sì che l'apprendimento tramite ANN possa definirsi profondo, sono necessari più di tre livelli (dunque almeno due strati nascosti). In caso contrario, si ha a che fare soltanto con una rete neurale di base.

Esistono diverse tipologie di reti neurali artificiali, che si differenziano soprattutto per il modo in cui processano i dati e la capacità di ricordare quanto accaduto negli strati anteriori.

Reti neurali feed-forward

Sono una delle forme di ANN più vecchie. I dati passano solo in un verso, dall'input verso l'output. Sono spesso abbinate ad un algoritmo di retropropagazione, che partendo dal risultato ottenuto ritorna all'inizio individuando gli errori e rivedendo i pesi (le connessioni), allo scopo di migliorare la precisione della rete.

³⁴ Nielsen, M.A. (2015), *Neural Networks and Deep Learning*, Determination Press.

Reti neurali ricorrenti

Lavorano solitamente con dati di serie temporali o che prevedono sequenze. Si differenziano dalle precedenti poiché sono in grado di “ricordare” quanto avvenuto nel livello precedente, in base all’output di quello corrente. Sono spesso utilizzate nella traduzione e nel riconoscimento vocale.

Memoria a lungo termine e a breve termine

È una versione evoluta di reti neurali ricorrenti, capace di ricordare ciò che è successo diversi strati prima. È comunemente usata nel riconoscimento vocale.

Reti neurali convoluzionali

Sono impiegate in particolar modo nel riconoscimento di immagini. Sono formate da due tipi di livelli, i livelli convoluzionali e i livelli *pool*, e spesso i primi sono seguiti dai secondi. Queste reti analizzano distintamente determinati elementi di un’immagine, per poi ricomporli nel livello completamente connesso.

Reti generative avversarie

In questo caso, due reti neurali sono messe in competizione tra loro per migliorare il risultato finale. Una rete, definita “generatore”, crea degli esempi, mentre l’altro network, nel ruolo del “discriminatore”, deve cercare di determinare se essi sono veri o falsi. Sono in grado di generare immagini estremamente realistiche e addirittura opere d’arte.

Le reti neurali possono sfruttare tutte le tipologie di machine learning viste in precedenza: il supervised learning, l’unsupervised learning, il semi-supervised learning ed il reinforcement learning. All’aumentare degli strati di nodi, esse sono capaci di costruire un maggior numero di livelli di astrazione, riuscendo in questo modo a risolvere complessi problemi di riconoscimento di schemi. Per addestrarsi al meglio hanno bisogno di grandi moli di dati, i quali però non richiedono particolari trattamenti da parte degli esseri umani. I sistemi di deep learning, infatti, possono processare dati grezzi ed imparare autonomamente come analizzarli, estraendone le caratteristiche, e come svolgere un compito. Tuttavia, per fare ciò necessitano di un notevole potere computazionale, che può essere difficile e costoso da ottenere.

Il deep learning trova numerose applicazioni, soprattutto nell'elaborazione del linguaggio naturale e nella computer vision (classificazione di immagini, riconoscimento facciale). È anche utilizzato nei veicoli a guida autonoma e nell'IA generativa.

1.5. Cos'è l'intelligenza?

Come si è visto, in più di qualche caso per definire l'intelligenza artificiale si è ricorso al concetto generale di intelligenza. Questa facoltà tipicamente umana è stata spesso considerata un requisito necessario per qualificare un sistema computerizzato come IA. Lo stesso McCarthy ha descritto quest'ultima come *"the science and engineering of making intelligent machines"* ³⁵.

Già da queste prime osservazioni, è evidente l'importanza che l'intelligenza generale ha avuto e continua a rivestire all'interno della disciplina. Sorge però spontanea una domanda: che cos'è l'intelligenza? La risposta può sembrare in apparenza banale, ma qualora ci si soffermi a pensarci risulterebbe subito chiaro quanto la questione sia in realtà complessa. Ha essa a che fare con il ragionamento? Con la pianificazione? Con la capacità di adattarsi? Numerosi psicologi e ricercatori hanno affrontato il tema, senza però essere ancora giunti ad una conclusione generalmente accettata. Tuttavia, le attuali interpretazioni tendono ad essere accomunate dall'identificazione di alcune capacità tipiche: imparare dall'esperienza; riconoscere i problemi; risolvere i problemi.

Ben Goertzel, uno dei principali esperti nel campo dell'AGI, ha identificato quattro approcci che sono stati utilizzati negli anni per concettualizzare la natura dell'intelligenza generale: pragmatico, psicologico, matematico, adattativo³⁶.

Approccio pragmatico

È riconducibile in particolar modo ad alcuni studiosi dell'IA ed ai loro tentativi di stabilire cosa essa sia. Secondo questa prospettiva, tutto ruota attorno agli esseri umani: essi sono considerati i sistemi intelligenti da prendere come riferimento; inoltre, è

³⁵ McCarthy, vedi nota in 1.1.

³⁶ Goertzel, B. (2015), *Artificial General Intelligence*, Scholarpedia, online, http://var.scholarpedia.org/article/Artificial_General_Intelligence#External_Links.

proprio tramite il confronto con le capacità umane che è possibile stabilire la presenza di intelligenza generale in un altro sistema. Questa ricostruzione richiama per molto aspetti la categoria dell'“agire come un umano” usata nell'analisi delle definizioni di IA. Un esempio di approccio pragmatico è dato dal Test di Turing³⁷, in cui una macchina deve dare delle risposte scritte alle domande di un interrogatore umano che riescano ad ingannare quest'ultimo, facendogli credere che l'interlocutore sia una persona.

Approccio psicologico

Come nel caso precedente, si focalizza su un'intelligenza generale di stampo umano; ma invece di guardare direttamente alle capacità pratiche, scava più in profondità per isolare e portare alla luce quelle facoltà sottostanti che permettono di esercitare le prime. È adottato (come dice il nome) soprattutto in psicologia, i cui esponenti, grazie alle loro ricerche, hanno elaborato visioni diverse per descrivere e spiegare il fenomeno dell'intelligenza. Ecco alcune delle teorie più importanti.

Spearman

In seguito ai suoi studi, nel 1904 lo psicologo inglese Charles Spearman pubblicò la sua teoria dei due fattori, secondo la quale l'intelligenza può essere suddivisa in:

- un fattore generale (g), riferito alle capacità mentali generali;
- molteplici fattori specifici (s), che rappresentano ciascuno l'abilità di un individuo in una determinata area (attività spaziali, verbali, meccaniche, numeriche e così via).

Thurstone

Sebbene non rigettasse completamente l'idea di fattore generale di Spearman, Thurstone sosteneva l'esistenza anche di un certo numero di capacità specifiche. Nel 1938 pubblicò un modello in cui identificava sette abilità primarie costituenti l'intelligenza:

- comprensione verbale;
- fluidità verbale;
- abilità numeriche;

³⁷ Turing, A. M. (1950), “Computing Machinery and Intelligence”, *Mind*, 59, pp 433-460.

- visualizzazione spaziale;
- velocità percettiva;
- memoria;
- ragionamento induttivo.

Gardner

Proseguendo il lavoro di Thurstone, nel 1983 lo psicologo americano Howard Gardner propose una teoria secondo cui non esiste una sola intelligenza, bensì multiple distinte intelligenze, le quali possono combinarsi nello svolgimento di determinate attività. In particolare, Gardner ne ha elencate otto, ognuna rappresentante delle capacità uniche:

- intelligenza linguistica;
- intelligenza logico-matematica;
- intelligenza spaziale;
- intelligenza musicale;
- intelligenza corporeo-cinestesica;
- intelligenza interpersonale;
- intelligenza intrapersonale;
- intelligenza naturalistica.

Sternberg

Solo due anni dopo, nel 1985 Robert Sternberg presentò la sua teoria triarchica, in cui inserì alcuni elementi mancanti nella visione di Gardner. Con essa, Sternberg concepisce l'intelligenza ponendo una forte attenzione alla sua capacità di adattarsi al contesto e di trasformarlo in base alla necessità. Egli inoltre definisce tre tipologie di intelligenza:

- intelligenza analitica;
- intelligenza creativa;
- intelligenza pratica.

Approccio matematico

A differenza dei precedenti approcci, i quali riprendono molto i comportamenti e/o la mente delle persone per caratterizzare l'intelligenza, alcuni ricercatori si sono impegnati a comprendere questo fenomeno in senso più astratto e generale, senza effettuare alcun

parallelismo con gli esseri umani. Ciò è possibile partendo da due intuizioni fondamentali:

- una vera e assoluta intelligenza generale potrebbe essere ottenuta solo avendo a disposizione una capacità di calcolo infinita. Ogni sistema di computazione risulterà quindi “poco intelligente” in determinati contesti o compiti;
- alcuni sistemi di computazione finiti avranno più intelligenza generale di altri.

Gli umani dunque sono ben distanti dall’essere generalmente intelligenti, anche se possiedono più intelligenza generale rispetto ad altri sistemi.

Un esempio di questa prospettiva è rappresentato dal lavoro di Legg e Hutter³⁸, in cui i due autori formalizzano il concetto di quella che loro chiamano “intelligenza universale” tramite una formula matematica che consente di misurare la capacità generale di un sistema di raggiungere i propri obiettivi in ambienti diversi.

Approccio adattativo

Secondo questo approccio, esiste un forte legame tra l’intelligenza generale e l’ambiente in cui essa si trova ad esistere. Questa idea è ben esemplificata da Pei Wang³⁹, secondo il quale un sistema è tanto più intelligente, quanto maggiore è il numero di ambienti a cui è in grado di adattarsi con efficacia, considerando dei vincoli di risorse realistici.

Riuscire a definire in modo univoco l’intelligenza porterebbe notevoli benefici all’IA, soprattutto per quanto riguarda il suo lato più prettamente scientifico. Una tale mancanza, infatti, ha provocato diversi problemi agli studiosi impegnati in questo campo: all’interno della comunità di ricerca, il sorgere di numerose discussioni ed incomprensioni riguardanti l’oggetto delle loro indagini; all’esterno, l’esigenza di giustificare il lavoro svolto come disciplina a sé stante, e non come branca, ad esempio, delle scienze computazionali o della psicologia. Inoltre, secondo Kaplan e Haenlein⁴⁰ questa è una delle cause per cui a sua volta non esiste una definizione univoca di intelligenza artificiale.

³⁸ Hutter, M., Legg, S. (2007), “Universal Intelligence: A Definition of Machine Intelligence”, *Minds and Machines*, 17 (4), pp 391-444

³⁹ Wang, P. (2006), *Rigid Flexibility: The Logic of Intelligence*, Springer, Dordrecht.

⁴⁰ Kaplan, A., Haenlein, M. (2019), “Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence”, *Business Horizons*, 62.

Ciò considerato, forse non è un caso che i principali apporti dal mondo scientifico alla disciplina siano giunti non tanto dalle ricerche sull'intelligenza in generale, quanto da quelle sul sistema nervoso. A livello tecnico, infatti, i progressi nelle neuroscienze e nella comprensione di come funziona il cervello hanno permesso di realizzare modelli di IA altamente complessi, in grado di emulare determinati processi e schemi neurali. Né sono un esempio le odierne reti neurali, fondamentali per gli algoritmi di deep learning.

La natura dell'intelligenza rimane ancora per gran parte un mistero. È difficile stabilire se e quando si giungerà a formulare una teoria in grado di descrivere al meglio questo fenomeno. Ciò che è certo è che un'architettura teorica tale avrebbe un notevole impatto non solo nello studio dell'intelligenza artificiale, ma anche a livello pratico, e potrebbe addirittura aprire le porte alla realizzazione di un'AGI con abilità pari a quelle degli esseri umani⁴¹.

⁴¹ Goertzel, vedi nota 2 sopra.

Capitolo 2. L'intelligenza artificiale: un *framework* teorico

2.1. Alcuni *framework* di riferimento

Come si è visto nel primo capitolo, esistono numerose definizioni di intelligenza artificiale. Ogni autore che si è cimentato nell'arduo compito di condensare in poche righe il vastissimo mondo dell'IA ha dato la propria visione personale, ricombinando elementi già noti o aggiungendone di nuovi, pur ricollegandosi sempre ad una delle quattro aree di pensiero individuate da Russell e Norvig. Alla non uniformità delle definizioni si accompagna la varietà delle rappresentazioni grafiche. Un processo simile a quello avvenuto per le formulazioni teoriche, infatti, ha interessato la realizzazione di schemi e modelli che dessero una visione a 360° dei sistemi di IA, con gli autori che di volta in volta si sono focalizzati su caratteristiche differenti.

KPMG, ad esempio, ha adottato un approccio che si potrebbe definire "empirico", basato sull'osservazione delle capacità a livello pratico. La società di revisione e consulenza ha effettuato una distinzione in base alle diverse aree di applicazione dell'intelligenza artificiale, partendo da un livello generale e andando man mano ad identificarne le sottocategorie.

Senza entrare nel merito della suddivisione e delle tipologie riportate, molte delle quali sono state precedentemente illustrate, questa soluzione consente di mettere in risalto la molteplicità di compiti diversi che i sistemi di IA sono in grado di svolgere. Ciò nonostante, il forte orientamento al contesto applicativo rischia molto presto di rendere lo schema obsoleto. La mancanza di uno sguardo più approfondito alle caratteristiche essenziali, e dunque difficilmente mutevoli, dell'intelligenza artificiale, lo rende infatti, per così dire, "vulnerabile" all'innovazione e all'introduzione di tecnologie *AI based* dotate di nuove funzionalità, le quali meriterebbero a quel punto di trovare il proprio spazio al fianco di quelle già presenti. Già ora si potrebbe affermare che andrebbe aggiunto uno spicchio dedicato all'IA generativa, data l'importanza sempre maggiore che sta acquisendo.

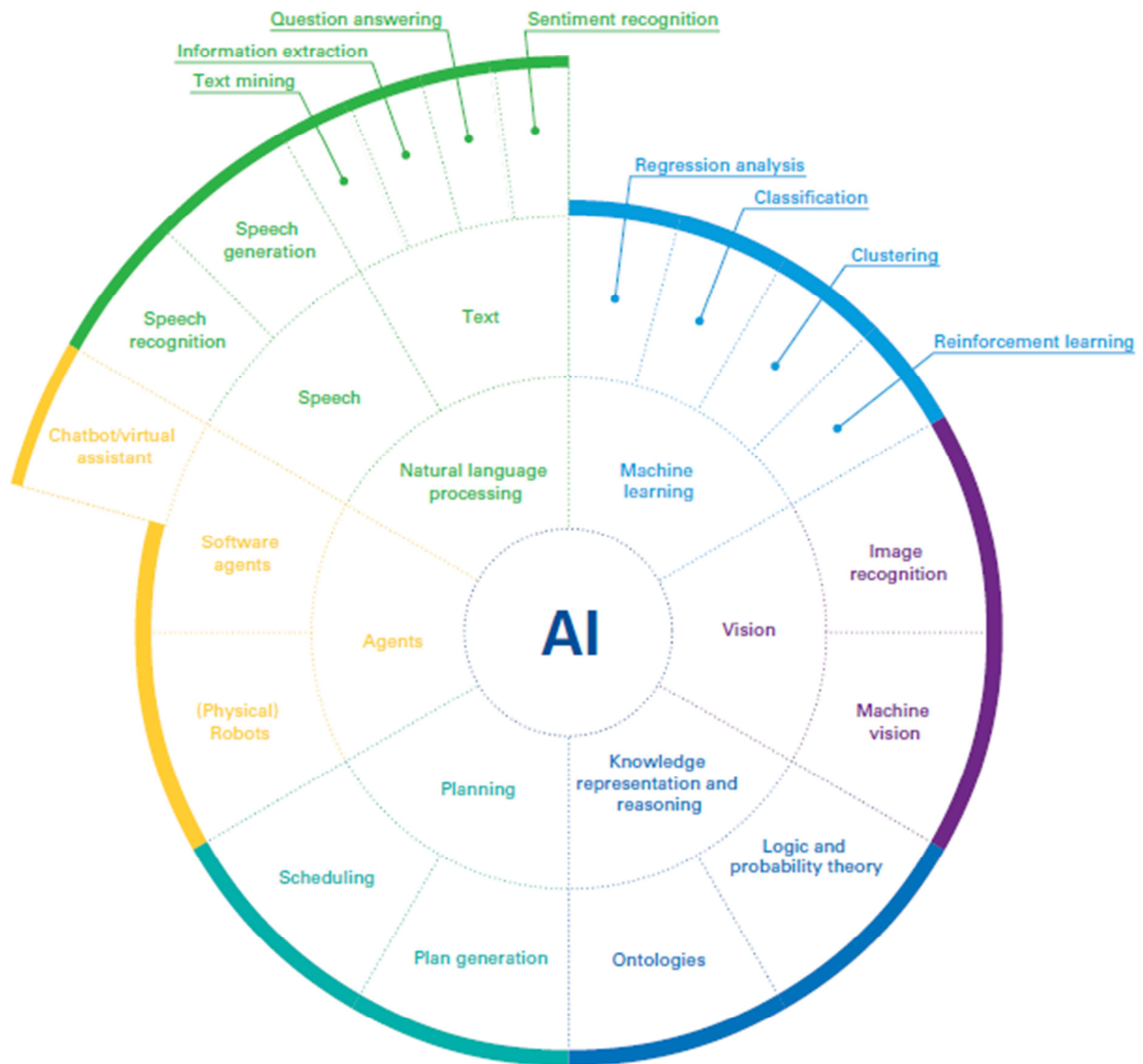


Figura 4. Sguardo d'insieme delle aree di applicazione dell'IA secondo KPMG⁴²

Al pari di KPMG, il World Economic Forum ha affrontato la questione con uno sguardo pragmatico, benché più ampio. Il WEF ha individuato quattro blocchi costitutivi del proprio modello, allineati in modo da formare una sorta di “filiera dell’IA”.

⁴² KPMG, TRENDONE (2018), *Rethinking the value chain. A study on AI, humanoids and robots*, p. 14.

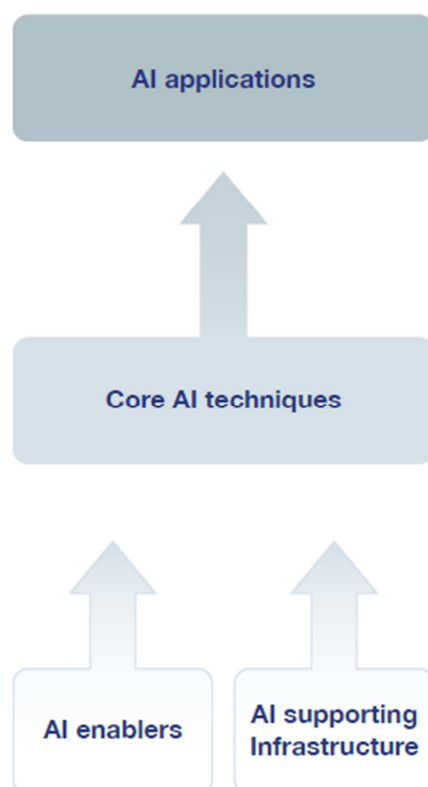


Figura 5. La “filiera dell’IA” secondo il WEF⁴³

Alla base si trovano alcuni elementi essenziali, definiti come “*enablers*”, fra i quali sono inclusi i dati, sia strutturati che non strutturati, l’etica e la responsabilità⁴⁴. A questi sono affiancati i componenti hardware e software che costituiscono l’infrastruttura di supporto allo sviluppo e all’implementazione dell’intelligenza artificiale, come ad esempio il *cloud computing*, i sensori per la raccolta dei dati e l’*Internet of Things* (IoT), i processori ad alte prestazioni e la robotica. Al livello successivo sono elencate le principali “*AI techniques*”, le quali, secondo la visione del WEF, comprendono fra le altre: il machine learning nelle sue varie declinazioni; le tecniche per il ragionamento *rules-based* (come la rappresentazione della conoscenza, il *decision making* e i sistemi esperti);

⁴³ World Economic Forum (2019), *Empowering AI Leadership. An Oversight Toolkit for Boards of Directors*, p. 8.

⁴⁴ Le problematiche legate all’etica e alla responsabilità nell’utilizzo dell’intelligenza artificiale saranno affrontate al paragrafo 3.3.

le tecniche per la percezione (machine vision, natural language processing ecc.). Infine, all'ultimo step della filiera si trova il blocco delle "AI applications", in cui si menzionano una serie di utilizzi specifici, come la ricerca aumentata, i veicoli a guida autonoma, gli agenti intelligenti e il riconoscimento di testi ed immagini.

Questo schema è certamente più completo del precedente, poiché non si limita a riportare le numerose applicazioni dell'intelligenza artificiale, ma include anche altri componenti fondamentali per il funzionamento di tali tecnologie. Tuttavia, anche in questo caso l'analisi condotta rimane confinata ad una dimensione esperienziale. Il modello può essere definito "completo" in quanto elenca la maggior parte degli elementi che attualmente fanno parte del panorama dell'IA, ma esso manca completamente di una generalizzazione a livello concettuale che si focalizzi sulle caratteristiche comuni a qualsiasi sistema.

Quelli appena illustrati sono solo due esempi della vasta gamma di rappresentazioni dell'intelligenza artificiale. Verranno ora esposti ed approfonditi tre *framework* particolarmente interessanti, uno per ognuna delle prospettive (politica e istituzionale, della ricerca e dei *practitioner*) viste nell'analisi delle definizioni al capitolo 1.

Prospettiva politica e istituzionale

Nel 2022 l'OCSE ha pubblicato un *framework* per la classificazione dei sistemi di intelligenza artificiale, destinato principalmente ai *policy maker*, ai legislatori e agli enti regolatori. Esso individua cinque macro-dimensioni rilevanti nel ciclo di vita dell'IA, ciascuna delle quali presenta distinti attributi e proprietà, o sotto-dimensioni: *People & Planet, Economic Context, Data & Input, AI Model, Task & Output*. In base a queste, è possibile esaminare e classificare le singole applicazioni dell'IA utilizzate in uno specifico progetto e contesto, facendo una valutazione nel caso concreto delle diverse sotto-dimensioni.

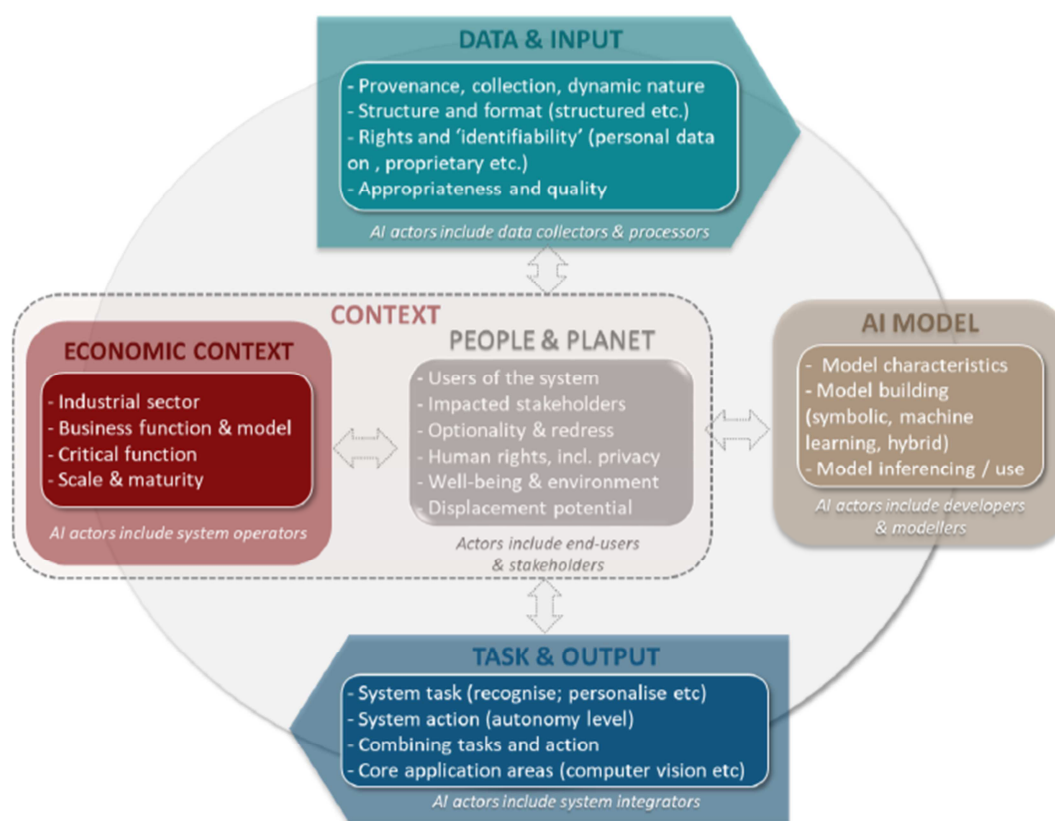


Figura 6. Il framework per la classificazione dei sistemi di IA sviluppato dall'OCSE⁴⁵

Nella spiegazione del modello, l'OCSE ha fatto una distinzione tra gli *stakeholder*, intesi come tutti gli individui ed organizzazioni toccati o influenzati, direttamente o indirettamente, dall'intelligenza artificiale, e gli attori, ovvero quei particolari *stakeholder* che svolgono un ruolo attivo durante il ciclo di vita di un sistema.

Al centro del *framework* si trova la dimensione "People & Planet". Questa considera il potenziale delle applicazioni *AI based* di promuovere un progresso tecnologico orientato al benessere degli individui e della Terra. Essa identifica tutti i soggetti e gruppi che nei diversi contesti entrano in contatto con i sistemi di IA. Le caratteristiche principali includono gli utilizzatori e i portatori di interesse (i quali, in questo caso, coincidono con gli attori), la possibilità di scelta correlate all'impiego dell'applicazione e gli impatti sui diritti umani, l'ambiente, il benessere e il mondo del lavoro.

⁴⁵ OECD (2022), *OECD Framework for the Classification of AI systems*, OECD Publishing, Paris, p. 19.

Il contesto economico (“*Economic Context*”) si riferisce all’ambiente economico in cui una determinata applicazione è usata e descrive il ruolo operativo che essa ha assunto all’interno di un’organizzazione. Queste informazioni sono importanti poiché a seconda dell’ambito settoriale di impiego, i sistemi di IA potrebbero di volta in volta doversi focalizzare maggiormente sul rispetto della privacy, oppure sulla trasparenza, o altro ancora. Le sotto-dimensioni comprendono il settore di appartenenza del sistema (finanza, sanità ecc.), la sua funzione all’interno dei processi aziendali, la sua criticità, la portata e il grado di maturità della sua implementazione. In questo caso, gli attori sono gli operatori dediti alla progettazione, all’impiego e al monitoraggio del sistema.

La dimensione “*Data & Input*” descrive tutti gli elementi con cui un modello di IA crea una propria rappresentazione del contesto esteso (comprendente sia quello economico, sia quello relativo alle persone e all’ambiente) in cui opera. Essi consistono nei dati con cui il sistema è stato addestrato e lavora e negli input, generalmente regole logiche, attraverso cui gli esperti codificano la loro conoscenza. Tra le caratteristiche, sono incluse la provenienza dei dati/input, il metodo con cui sono raccolti e le proprietà dei dati. Gli attori identificati sono gli addetti alla creazione e preparazione del *dataset*.

L’“*AI Model*” costituisce la forma con cui un’applicazione di IA rappresenta gli oggetti, i processi, le persone, le interazioni che delineano il contesto esterno. Esso utilizza i dati e la conoscenza degli esperti per descrivere e interagire con il proprio ambiente, sia esso reale o virtuale. Gli attributi con cui è descritto un modello sono la tipologia tecnica, i metodi con cui è stato costruito (ad esempio, tramite ML), gli obiettivi e le variabili con cui se ne misurano le prestazioni. Gli attori individuati sono gli sviluppatori che lo creano e validano.

Infine, nella dimensione “*Task & Output*” sono ricompresi i compiti svolti dal sistema e le azioni risultanti, tramite le quali interagisce con l’ambiente circostante. Le caratteristiche principali includono, appunto, le funzioni assegnate, il livello di autonomia nello svolgere le proprie azioni, la combinazione o meno di compiti ed azioni e gli ambiti di applicazione fondamentali (computer vision, natural language processing ecc.). In questo caso, gli attori rilevanti sono gli integratori di sistema che predispongono e implementano le applicazioni *AI based*.

Nelle intenzioni degli autori, questo *framework* dovrebbe innanzitutto permettere agli utilizzatori di porre l’attenzione sui rischi specifici associati all’intelligenza artificiale. Molti sistemi, infatti, presentano problematiche legate alla trasparenza del processo

decisionale e alla robustezza ed imparzialità dei risultati e rischiano pertanto di avere impatti negativi sulla società⁴⁶. Lo schema proposto costituirebbe un punto di partenza per la creazione di soluzioni finalizzate alla valutazione e mitigazione di tali rischi. Oltre a ciò, esso favorirebbe un dibattito politico puntuale e lo sviluppo di appositi strumenti legislativi per la regolamentazione dell'IA, data l'influenza che le caratteristiche di un certo sistema hanno sulle tecniche e procedure necessarie per implementarlo. Infine, l'identificazione delle proprietà e degli attributi fondamentali consentirebbe di comprendere e descrivere meglio le numerose applicazioni e tecnologie.

Questo modello è pensato per un'analisi specifica dei sistemi di intelligenza artificiale presenti a differenti livelli nel contesto sociale ed economico, tuttavia alcune considerazioni assumono rilevanza a livello di IA in generale. Le cinque dimensioni descritte e le rispettive sottocategorie presentano un notevole livello di astrazione, il che permette al *framework* di poter essere impiegato su qualsiasi applicazione di IA. La prospettiva politica ed istituzionale emerge guardando sia ai destinatari principali di questo lavoro (*policy maker* e legislatori), sia agli scopi. Le questioni legate alla trasparenza, all'equità e alla fiducia nell'intelligenza artificiale sono infatti fra le tematiche più importanti che gli enti regolatori sono chiamati ad affrontare nei prossimi anni, e questo *framework* vuole essere uno strumento utile in tal senso.

Prospettiva della ricerca

Precedentemente rispetto alla creazione dello schema appena discusso, l'OCSE aveva elaborato, grazie al lavoro di un gruppo di esperti appositamente costituito (*l'Expert Group on Artificial Intelligence at the OECD, AIGO*) un'altra rappresentazione dell'IA, fortemente influenzata da una prospettiva riconducibile all'ambito della ricerca. Nonostante provenga da un ente istituzionale, infatti, essa si basa in gran parte sul

⁴⁶ Gli impatti dell'intelligenza artificiale sulla società, i temi della trasparenza, della robustezza, dell'imparzialità, della responsabilità e della fiducia e i principi etici per lo sviluppo di un'IA antropocentrica saranno affrontati ed approfonditi al paragrafo 3.3.

modello di agente di Russell e Norvig⁴⁷ e sulla definizione di intelligenza artificiale di Winston⁴⁸ già vista al capitolo 1.

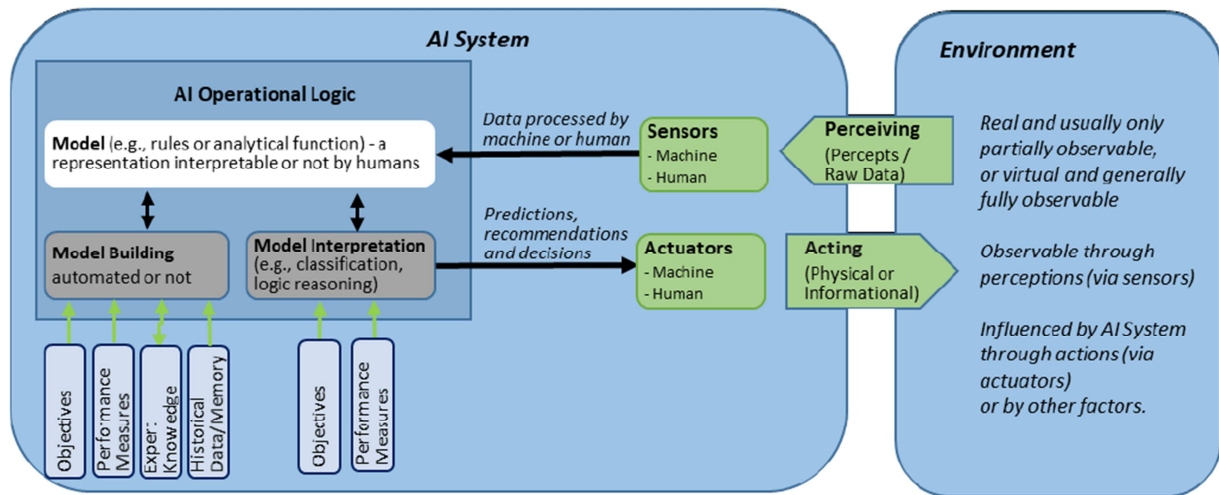


Figura 7. La visione concettuale di un sistema di IA sviluppata dall'AIGO

La visione concettuale dell'OCSE delinea la struttura e descrive il funzionamento di un generico sistema di IA. Questo è composto da tre elementi essenziali: i sensori, gli attuatori e l'*operational logic*.

I sensori e gli attuatori possono essere sia soggetti umani che macchine e permettono al sistema di interagire con l'ambiente. I primi sono in grado di osservarlo e percepirlo, raccogliendo attraverso la loro attività dati grezzi che saranno poi processati; i secondi, invece, possono cambiarne lo stato tramite delle azioni. L'ambiente nel quale l'agente intelligente opera può essere reale o virtuale. Generalmente, gli ambienti reali sono osservabili solo in parte, mentre quelli virtuali possono essere osservati nella loro interezza.

Il cuore dello schema è l'*operational logic*. Essa racchiude la potenza del sistema di IA e consente, partendo dagli input provenienti dai sensori ed in base ad una serie di

⁴⁷ Norvig, P., Russell, S. (2010), *Artificial Intelligence: A Modern Approach*, 3rd ed., Pearson Education, Hoboken, p. 35.

⁴⁸ "Artificial intelligence is... The study of the computations that make it possible to perceive, reason, and act". Winston, P. (1992), *Artificial Intelligence*, 3rd ed., op. cit., p. 5.

obiettivi, di fornire agli attuatori degli output (raccomandazioni, previsioni o decisioni) con i quali influenzare l'ambiente. Una parte fondamentale dell'*operational logic* è costituita dal modello, ovvero una rappresentazione parziale o integrale dell'ambiente esterno, della sua struttura e delle sue dinamiche. Il modello può essere costruito utilizzando tecniche di elaborazione automatica dei dati, tra cui il machine learning, oppure tramite interventi manuali degli operatori umani. Spesso si ricorre a dati storici, ma non è escluso il ruolo della conoscenza dell'esperto. Il processo di costruzione è guidato da un lato dagli obiettivi prefissati e dall'altro da alcune misure di prestazione, quali l'accuratezza del modello, la rappresentatività del dataset e le risorse impiegate per l'addestramento. Una volta realizzato, il modello deve essere interpretato da una persona o da uno strumento automatizzato, al fine di ottenere da esso un risultato nella forma di previsione, raccomandazione o decisione. Anche questo processo è guidato da obiettivi e misure di prestazione, che possono differire a seconda del caso. Ad esempio, un modello probabilistico potrà fornire numerose raccomandazioni (gli obiettivi), a ciascuna delle quali saranno assegnate diverse misure di *performance*, quali il rischio e l'affidabilità, mentre un modello basato su regole logiche produrrà un solo risultato.

Nel complesso, secondo questa visione concettuale un sistema di IA è capace di elaborare dati di input provenienti da operatori umani o altre macchine, per: percepire l'ambiente esterno; astrarre, attraverso tecniche automatiche o manuali, tali percezioni in un modello; usare le interpretazioni di quest'ultimo per arrivare a dei possibili risultati. Tutto ciò, al fine di influenzare l'ambiente in cui si trova grazie alle proprie decisioni, previsioni e raccomandazioni.

Lo schema proposto dall'AIGO fornisce una raffigurazione generalizzata dell'intelligenza artificiale, identificando con un notevole grado di astrazione gli elementi comuni, se non a tutti, certamente alla maggioranza dei sistemi di IA. Esso, però, va oltre e ne spiega anche il funzionamento tecnico, mettendo in luce il ruolo dell'ambiente quale punto di partenza e di arrivo del processo, in quello che sembra quasi un ciclo.

Prospettiva dei *practitioner*

Nel loro libro "*Competing in the Age of AI*", nel quale si illustra come l'adozione dell'IA all'interno della realtà aziendale, in combinazione ad un approccio *data driven*, consenta di rimuovere i tradizionali vincoli che hanno limitato la crescita delle imprese per centinaia di anni e di creare nuove strategie, Marco Iansiti e Karim Lakhani hanno

sviluppato una rappresentazione di quella che i due chiamano “AI factory”. Come riportato dagli autori, per *AI factory* si intende “the scalable decision engine that powers the digital operating model⁴⁹ of the twenty-first-century firm”⁵⁰. Grazie all’intelligenza artificiale, è possibile digitalizzare la raccolta e l’analisi dei dati, in modo da generare costantemente previsioni, *insight* e opzioni di scelta con i quali guidare o automatizzare numerose azioni operative, “industrializzando” così l’attività di *decision making*⁵¹. Una *AI factory* è costituita da alcuni elementi fondamentali.

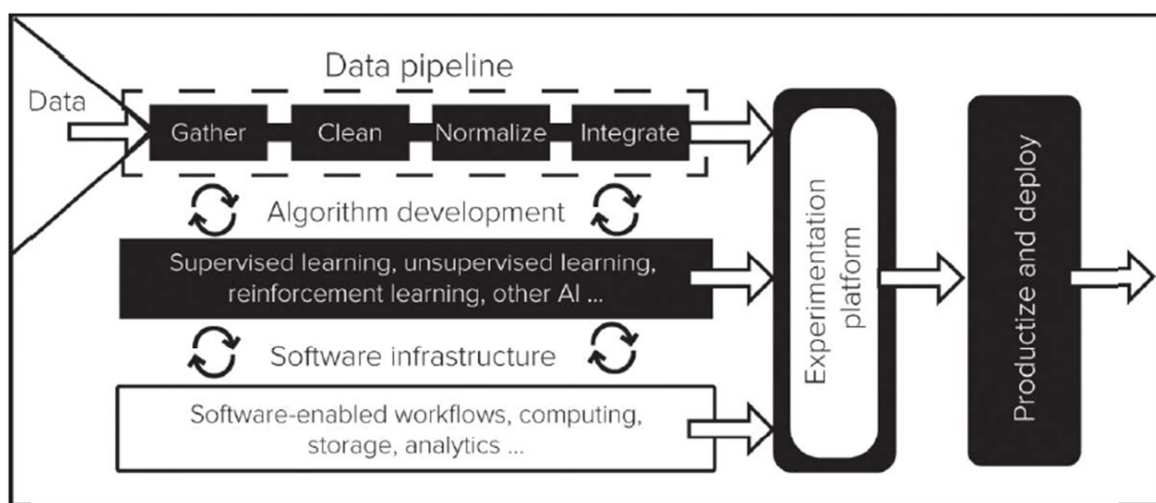


Figura 8. I componenti di una AI factory

La “*data pipeline*” raggruppa tutte le fasi di raccolta e gestione dei dati. Questi sono visti come il carburante dell’*AI factory*, ma nonostante l’importanza della loro funzione, spesso risultano essere incompleti, frammentati e conservati in archivi (*silos*) separati, oltre che potenzialmente non corretti o affetti da *bias*. Per ricavare da essi il massimo valore e costruire un’effettiva *AI factory*, una volta raggruppati i dati devono dunque essere

⁴⁹ L’*operating model*, da non confondere con il *business model* (come un’impresa vuole creare e catturare valore), indica la modalità in cui l’azienda effettivamente opera e offre tale valore ai propri clienti.

⁵⁰ Iansiti, A., Lakhani, K.R. (2020), *Competing in the Age of AI: Strategy and Leadership When Algorithms and Networks Run the World*, Harvard Business Review Press, Brighton, p. 53.

⁵¹ *Ibidem*.

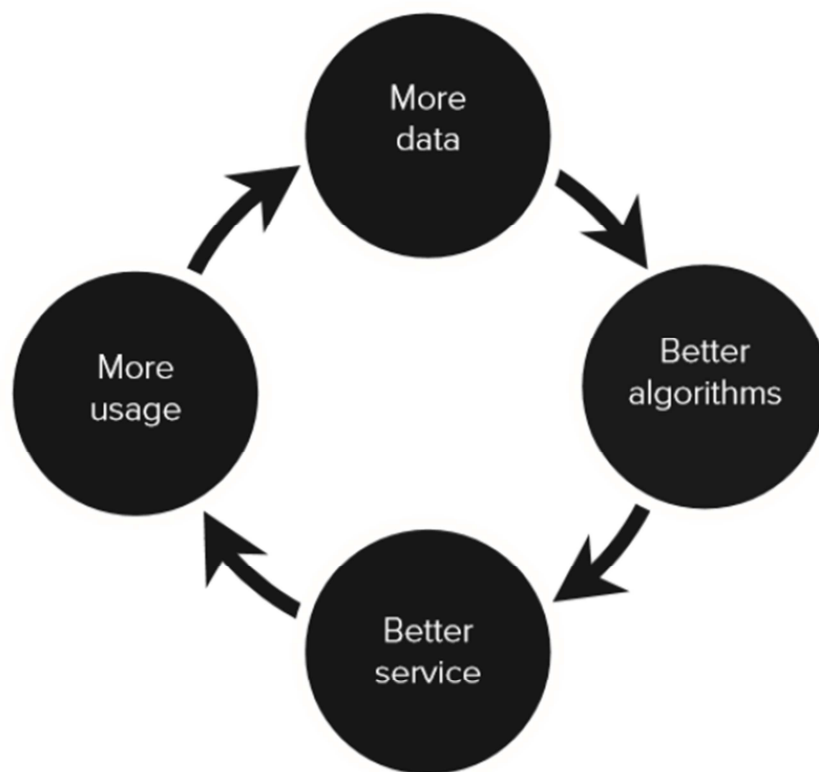
ripuliti, normalizzati ed integrati, secondo un processo che sia sistematico, scalabile e sostenibile.

L’*“algorithm development”* si riferisce alla necessità di scegliere ed utilizzare i giusti algoritmi per trarre beneficio dai dati. Gli algoritmi sono il motore dell’azienda digitale, l’insieme di regole con cui è possibile utilizzare i dati raccolti e preparati per fare previsioni, prendere decisioni o risolvere problemi specifici. La maggior parte dei sistemi di IA operativi si basa su modelli statistici e approcci di machine learning, quali il supervised learning, l’unsupervised learning e il reinforcement learning.

L’*“experimentation platform”* consiste nel meccanismo attraverso cui le decisioni e le previsioni a cui sono giunti gli algoritmi vengono testate. Esse infatti, affinché possano considerarsi affidabili, devono essere validate tramite strumenti e tecniche adeguate. In questo modo, è possibile verificare che i cambiamenti suggeriti abbiano concretamente gli effetti auspicati.

Infine, la *“software infrastructure”* indica l’infrastruttura software e computazionale che incorpora e sostiene la *data pipeline*, il motore algoritmico e l’*experimentation platform*. Per valorizzare al meglio grazie all’intelligenza artificiale il grande ammontare di dati a cui hanno accesso oggi le imprese, è importante disporre di una piattaforma che assicuri la connessione tra le diverse unità aziendali e sia centralizzata, evitando dunque la creazione di numerosi silos e preferendo invece un approccio modulare. Inoltre, l’infrastruttura deve essere robusta, scalabile e in grado di garantire la sicurezza dei dati. Un’efficace implementazione dei vari elementi all’interno dell’*AI factory* consente di creare un ciclo virtuoso tra il coinvolgimento degli utenti, la raccolta dei dati, il design degli algoritmi, le previsioni e il miglioramento costante⁵². I dati provenienti dalle fonti, sia interne che esterne, di cui dispone l’impresa sono utilizzati per addestrare e affinare gli algoritmi, i quali in questo modo generano previsioni sempre più accurate. Le decisioni e azioni che ne derivano, una volta testate attraverso rigidi protocolli, permettono di migliorare i prodotti e servizi destinati ai consumatori, portando questi ad utilizzarli e servirsene maggiormente. I dati relativi all’uso e agli impatti delle previsioni sono raccolti e inviati al sistema, in modo che questo possa apprendere da essi e avviare così un nuovo ciclo.

⁵² Ivi, p. 54.



*Figura 9. Il ciclo virtuoso dell'AI factory*⁵³

Questo *framework*, sebbene non sia propriamente rappresentativo dell'intelligenza artificiale, è comunque un ottimo esempio della prospettiva dei *practitioner*. L'*AI factory* può essere vista come un'estensione di un tradizionale sistema di IA all'intero contesto aziendale. Il modello di Iansiti e Lakhani descrive bene e con un notevole grado di astrazione e generalizzazione l'adozione su larga scala di queste tecnologie a livello imprenditoriale. Inoltre, introduce una caratteristica importante delle applicazioni fondate sull'IA: la ciclicità tra dati, algoritmi e risultati.

Come si è potuto notare, le diverse rappresentazioni affrontate hanno di volta in volta messo in risalto aspetti diversi legati all'intelligenza artificiale.

Il *framework* dell'OCSE si concentra sull'individuazione delle caratteristiche necessarie a descrivere una specifica applicazione, in modo da valutarne i rischi per il contesto sociale ed economico in cui si trova e favorire la nascita di interventi regolatori.

⁵³ Ibidem.

La visione concettuale sviluppata dall'AIGO si focalizza sulla composizione strutturale di un sistema di IA e sul suo funzionamento, senza particolari finalità politiche o economiche.

Il modello di *AI factory* di Iansiti e Lakhani, invece, analizza come l'intelligenza artificiale possa essere implementata a livello aziendale e quali benefici possa apportare.

Questa difformità è dovuta alle differenti prospettive a cui ciascuno schema può essere ricollegato. Verrà ora presentato un nuovo modello, detto "dei tre domini", che cerca unire componenti e tratti presi da ognuno di essi per giungere ad una rappresentazione generale dell'IA.

2.2. Il modello dei tre domini

L'obiettivo di questo modello è fornire una rappresentazione dell'intelligenza artificiale che sia al tempo stesso:

- generale e applicabile su qualunque sistema e applicazione;
- accurata e comprendente le dimensioni maggiormente rilevanti nello studio e applicazione della disciplina.

Inoltre, il lavoro che sarà presentato non vuole essere imperniato su una specifica prospettiva, ma combinarle insieme includendo gli aspetti principali su cui ciascuna di esse tende a focalizzarsi:

- nel caso della prospettiva politica ed istituzionale, la valutazione e gestione dei rischi correlati all'IA e i suoi effetti sul contesto economico e sociale;
- nel caso della prospettiva della ricerca, la struttura e il funzionamento tecnico dei sistemi di intelligenza artificiale;
- nel caso della prospettiva dei *practitioner*, l'adozione di soluzioni IA a livello imprenditoriale e i loro impatti in ambito economico ed aziendale.

Per raggiungere questo risultato, sono stati presi in considerazione numerosi lavori, a partire dai tre schemi (OCSE, 2022; OCSE, 2019; Iansiti e Lakhani, 2020) illustrati nel paragrafo precedente. Da ognuno di essi, infatti, sono stati tratti alcuni elementi, come l'idea di circolarità e il ruolo dei dati e dell'ambiente, che saranno poi ritrovati anche nel

nuovo *framework*. In aggiunta a questi, importante è stato il contributo del modello ADES/ACE del prof. Albarelli⁵⁴ e del lavoro svolto dal dottor Scardino⁵⁵.

Il *framework* individua tre domini fondamentali per descrivere l'intelligenza artificiale: il dominio dei dati, il dominio scientifico e il dominio applicativo. Questi sono collegati tra loro e sono inseriti in un determinato contesto o ambiente.

2.2.1. Il dominio dei dati

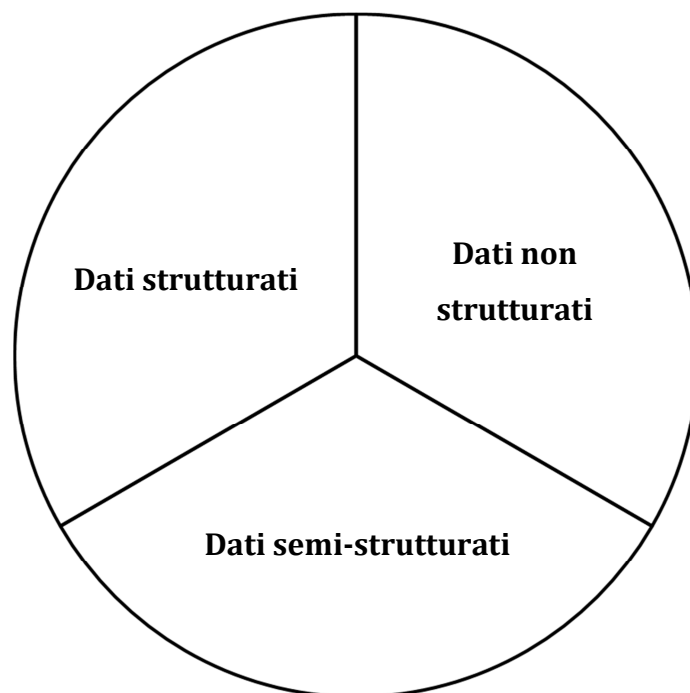


Figura 10. Il dominio dei dati

I dati sono la risorsa fondamentale su cui si basa l'IA moderna. Uno dei fattori che ha determinato la diffusione di queste tecnologie negli ultimi anni è stata proprio la crescita esponenziale dei dati disponibili, grazie a nuove fonti come Internet, i social media e i sensori degli *smart device*.

⁵⁴ Albarelli, A. (2019), *Artificial Intelligence Workshop* [diapositive di Powerpoint].

⁵⁵ Scardino, A.A. (2021), *Artificial Intelligence: Definition, reclassification, impact on business models, the rationale for a digital and business twin and Alpha Inc. Case*. [tesi di laurea magistrale], Università Ca' Foscari di Venezia, Venezia.

I dati possono suddividersi in tre tipologie, in base al grado di ordine che li caratterizza: strutturati, non strutturati e semi-strutturati.

I dati strutturati sono altamente organizzati in quanto rispettano un modello predefinito. Sono conservati in database relazionali⁵⁶ con campi ben definiti e possono essere facilmente trasposti su un comune foglio elettronico. Sono facili da archiviare, da cercare e da analizzare, sia dagli operatori umani che dagli algoritmi di machine learning, e per molto tempo sono stati gli unici dati utilizzabili dalle imprese. Nonostante le loro qualità, si stima che essi rappresentino meno del 20% dei dati esistenti. I dati strutturati possono essere generati da dispositivi elettronici, sensori, registri web e altre fonti “artificiali”, oppure da persone in carne ed ossa. Alcuni esempi includono informazioni demografiche come nomi, numeri di telefono ed età, dati di posizione GPS e transazioni finanziarie.

Al contrario dei precedenti, i dati non strutturati non sono conformi ad alcun modello e non seguono una struttura predefinita. Non possono essere archiviati ordinatamente in un database relazionale o in un foglio di calcolo e pertanto sono conservati ricorrendo ad altre soluzioni, come ad esempio database non relazionali o *data lake*⁵⁷. A causa dell'assenza di un'architettura identificabile, sono tradizionalmente difficili da gestire, cercare e analizzare e per questo motivo sono stati a lungo ignorati dalle aziende. Tuttavia, il forte sviluppo recente di nuove tecnologie *AI based* ha permesso di poter lavorare con essi molto più facilmente. Una delle peculiarità principali di applicazioni quali la *computer vision* o il *natural language processing* sta proprio nella capacità di processare e trarre valore da dati non in forma strutturata. In questa categoria rientrano, fra gli altri, *file* di testo, documenti in formato PDF, contenuti provenienti dai *social media*, immagini, audio e video.

Infine, i dati semi-strutturati si pongono nel mezzo tra le due tipologie precedenti, avendo tratti comuni ad entrambe. Questi non rispettano una struttura rigida e non sono adatti ad essere conservati in un database relazionale. Allo stesso tempo, però,

⁵⁶ Una database relazionale è una forma di archiviazione dei dati mediante tabelle, in cui ciascuna colonna, detta campo o attributo, è etichettata in base allo specifico tipo di dato che può contenere (ad es. “Nome”, “Cognome”, “Data di nascita”, “Numero di telefono”), mentre le righe, o *record*, raggruppano tutti i dati relativi ad un certo elemento (ad es. tutti i dati di un cliente).

⁵⁷ Un *data lake* è un archivio centralizzato senza una struttura o composizione rigida, in grado di raccogliere grandi volumi di dati provenienti da fonti diverse, di tipologie differenti e nel loro formato originale, senza la necessità di elaborarli precedentemente.

presentano alcune proprietà, come *tag* semantici⁵⁸ e metadati⁵⁹, che conferiscono un certo livello di organizzazione e ne facilitano l'analisi, preservandone comunque la fluidità. Fra i dati semi-strutturati, rientrano le email, le fotografie digitali, le pagine *web* e i file in formato XML.

2.2.2. Il dominio scientifico

⁵⁸ I *tag* semantici sono degli elementi che consentono identificare le diverse sezioni di una pagina web e i contenuti in esse presenti.

⁵⁹ I metadati sono delle informazioni sul contenuto di un certo *file* o su altri dati. Essi possono indicarne, ad esempio, l'autore, la data e l'ora di creazione, la dimensione ecc.

Capitolo 3. L'intelligenza artificiale: gli impatti aziendali e sociali

3.3 Gli impatti sulla società

Come si è visto, l'intelligenza artificiale ha le potenzialità per modificare profondamente il modo in cui le imprese generano e distribuiscono valore, andando ad impattare a livello sia di modello di business, sia di strategia. Sebbene ciò sia già piuttosto impressionante, i cambiamenti maggiormente significativi dovuti alla diffusione di queste soluzioni potrebbero avvenire al di fuori dell'ambito aziendale e impattare sulle vite di miliardi di persone. C'è ancora incertezza sui reali effetti di questa rivoluzione, ma già ora è possibile notare come numerosi settori ed ambiti della quotidianità risentano dell'introduzione dell'IA.

Due di questi sono certamente la medicina e l'assistenza sanitaria. Già da alcuni anni l'utilizzo di tecnologie mediche intelligenti, ovvero che sfruttano l'intelligenza artificiale, sta ottenendo riscontri positivi tra la popolazione. Esse, infatti, consentono di adottare un approccio connotato da predittività, prevenzione, personalizzazione e partecipazione (il cosiddetto "modello delle 4 P")⁶⁰, migliorando il trattamento e l'esperienza dei pazienti. Le soluzioni di IA sono particolarmente efficaci nella diagnostica. Modelli di machine learning possono essere addestrati su dati relativi a parametri vitali al fine di riconoscere i segnali di condizioni complesse e difficili da individuare, come ad esempio l'insorgenza di sepsi in bambini nati prematuri⁶¹. Inoltre, reti neurali artificiali applicate alla classificazione di immagini hanno ottenuto ottimi risultati, paragonabili o addirittura migliori rispetto a quelli di medici umani, nel riconoscimento di tumori o altre problematiche dall'analisi di radiografie. L'intelligenza artificiale può essere impiegata anche per migliorare il lavoro degli operatori ospedalieri, efficientando la gestione e la ricerca degli esami clinici. Riprendendo poi il tema della personalizzazione, la capacità di considerare allo stesso tempo storia clinica, preferenze e bisogni personali

⁶⁰ Briganti, G., Le Moine, O. (2020), "Artificial Intelligence in Medicine: Today and Tomorrow", *Frontiers in Medicine*, 7.

⁶¹ IBM, <https://www.ibm.com/topics/artificial-intelligence-medicine>.

fa sì che possano essere sviluppati sistemi in grado di fornire raccomandazioni su misura ed in real-time, ad ogni ora del giorno. Infine, non meno importanti sono le applicazioni dell'intelligenza artificiale nella creazione di nuovi farmaci, per esempio sfruttando l'IA generativa per disegnare nuove molecole⁶².

Un altro ambito che vivrà una profonda trasformazione (e in parte lo sta già facendo) grazie a queste nuove tecnologie è l'istruzione. Nonostante molti esperti ritengano che la presenza di figure umane sia critica e non verrà totalmente soppiantata dalle macchine, l'intelligenza artificiale apporterà notevoli cambiamenti al ruolo dell'insegnante e alle best practice educative. Innanzitutto, essa è capace di valutare per ogni singolo studente il livello di apprendimento in una determinata materia, in modo da proporre delle soluzioni personalizzate a seconda dei bisogni individuali. Tuttora esistono delle piattaforme in grado di identificare le lacune nella conoscenza degli studenti, reindirizzarli verso nuovi argomenti qualora indicato e proporre loro le sfide per le quali di volta in volta sono pronti. Oltre a ciò, soluzioni di IA possono essere implementate per consentire un migliore accesso all'istruzione a coloro che vivono in situazioni di difficoltà a causa di barriere linguistiche e, soprattutto, di disabilità sensoriali o psichiche. Per esempio, è possibile utilizzare un assistente virtuale per supportare bambini con cecità o ipovisione nell'apprendimento e nello sviluppo di una maggiore autonomia⁶³. O ancora, la process automation ed il machine learning consentono di creare facilmente contenuti testuali interattivi e coinvolgenti per bambini affetti da dislessia o autismo⁶⁴. Queste tecnologie hanno il potenziale di rendere l'istruzione più inclusiva, dando a tutti le stesse opportunità di imparare. L'intelligenza artificiale può essere utilizzata anche fuori dall'ambiente scolastico, per assistere gli studenti durante lo svolgimento dei compiti e lo studio a casa. Inoltre, essa è in grado di migliorare la qualità del lavoro svolto dagli educatori umani. Automatizzando lo svolgimento di compiti ripetitivi e subordinati rispetto all'attività principale dell'insegnamento, come ad esempio la correzione di compiti e test e lo svolgimento di pratiche prettamente burocratiche, le applicazioni di IA consentono agli insegnanti di avere più tempo da

⁶² Caggiano, V., Das, P., Strobel, H. (2020), *Using Generative AI to Accelerate Drug Discovery*, IBM, online, <https://www.ibm.com/blogs/research/2020/06/accelerated-discovery/>.

⁶³ UNESCO (2020), *Artificial Intelligence in Education: Compendium of Promising Initiatives*, pp. 17 sg.

⁶⁴ UNESCO (2020), *Artificial Intelligence in Education: Compendium of Promising Initiatives*, p. 20.

dedicare agli studenti. In questo modo, possono concentrarsi maggiormente sulla spiegazione di nuovi concetti e sui bisogni dei loro alunni, andando ad agire dove una macchina farebbe più fatica.

L'intelligenza artificiale andrà a modificare anche il modo in cui le persone vivono la città. Le smart city non sono più un concetto nuovo. Una città può essere definita “*smart*” quando “*investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance*”⁶⁵. In parole povere, in una smart city i servizi tradizionali sono resi più efficienti grazie all'utilizzo di diversi approcci e soluzioni digitali per il benessere degli abitanti. In questo contesto, l'intelligenza artificiale può unirsi alle tecnologie già impiegate per migliorare le condizioni di vita nelle smart city del futuro, anche se già ora nei Paesi più importanti le iniziative legate a queste realtà si stanno concentrando su applicazioni *AI based*. Quest'ultime infatti consentono di analizzare al meglio le enormi moli di dati provenienti da dispositivi IoT (i quali costituiscono una delle infrastrutture più importanti in una smart city) e di decidere le azioni più corrette ed efficienti da portare avanti. In questo modo, l'IA può andare ad influenzare diversi aspetti del vivere urbano: dai trasporti, con la creazione di un intelligent transportation system (ITS) capace di valutare il flusso del traffico e stimare la probabilità di incidenti, la diffusione di veicoli autonomi in grado di comunicare tra loro e con le infrastrutture circostanti e l'utilizzo di droni sempre più performanti grazie al ML, alla sicurezza, grazie a sistemi che permettono di prevedere e affrontare al meglio crimini, pericoli di varia natura e anche attacchi informatici; dall'energia, con la costruzione di reti elettriche intelligenti (smart grid) per ottimizzare la distribuzione di energia elettrica e l'utilizzo di contatori capaci di scegliere il miglior piano di costi, alla gestione della spazzatura, in cui modelli di DL sono impiegati per classificare diverse tipologie di rifiuti; fino all'educazione e alla sanità, di cui si è già parlato. Questi sono solo alcuni esempi di come l'IA, abbinata a piattaforme IoT e big data, possa contribuire a migliorare la pianificazione e la funzionalità delle aree urbane in un'ottica di sviluppo economico-sociale e di sostenibilità. Ciò si rende sempre più necessario, anche a fronte dell'urbanesimo

⁶⁵ Caragliu, A., Del Bo, C., Nijkamp, P. (2011), “Smart Cities in Europe”, *Journal of Urban Technology*, 18(2), p. 70.

crescente che porterà entro il 2050 circa il 68% della popolazione mondiale a vivere in città⁶⁶.

L'intelligenza artificiale non influirà solo sul mondo materiale, bensì ha il potenziale per impattare anche quello virtuale. Da quando nell'ottobre del 2021 Facebook ha cambiato il suo nome in "Meta", il concetto di metaverso ha attirato su di sé l'attenzione del grande pubblico. Questo termine, formato dall'unione di "meta" (dentro) e "universo", indica "*a shared virtual 3D world or even multiple cross-platform worlds that can provide users with a comprehensively immersive experience, including interactive and collaborative activities*" ⁶⁷. Uno degli elementi principali che caratterizzerà questo nuovo ambiente sono le interazioni e le relazioni che si creeranno con gli altri utenti, nelle quali rivestirà un ruolo importante lo status assunto nel mondo digitale. Diversi sono gli strumenti e le tecnologie che stanno contribuendo a dare vita al metaverso: 5G, Wi-Fi, cloud, GPU, realtà virtuale (VR), realtà aumentata (AR), wearable device e blockchain⁶⁸ sono solo alcuni. In mezzo a questi, l'IA sta giocando un ruolo importante per garantire l'affidabilità dell'infrastruttura e migliorarne le prestazioni. Al di là degli algoritmi di ML impiegati per assicurare e rendere più efficiente il funzionamento delle reti 5G (ed in futuro 6G), l'intelligenza artificiale consente di generare da zero o quasi spazi e scenografie virtuali in cui muoversi ed interagire. L'IA generativa potrà ricreare luoghi realmente esistenti nel mondo fisico partendo da semplici fotografie, oppure spazi totalmente inventati grazie ad alcuni input umani e tecniche di reinforcement learning. Inoltre, essa permetterà anche di realizzare avatar 3D a immagine e somiglianza dell'utente. Tramite l'utilizzo di wearable device o altri gadget e di modelli di machine learning e deep learning, sarà poi possibile analizzare e riconoscere movimenti e azioni complesse compiuti nella realtà fisica, così da proiettarle in quella virtuale. Qui gli avatar potranno interagire fra loro, anche tramite applicazioni *AI based* di traduzione simultanea, o con agenti virtuali intelligenti. Infine, merita una menzione l'AI Research SuperCluster (RSC), il più potente supercomputer per IA al mondo. Sviluppato da Meta,

⁶⁶ Nazioni Unite (2019), *World Urbanization Prospects: The 2018 Revision*, New York, p. 1.

⁶⁷ Han, Z., Huynh-The, T., Kim, D.-S., Nguyen, T.T., Pham, Q.-V., Pham, X.-Q. (2023), "Artificial intelligence for the metaverse: A survey", *Engineering Applications of Artificial Intelligence*, 117 A.

⁶⁸ Per vedere cos'è la blockchain, si vada al paragrafo 4.1 (??????).

una volta completato⁶⁹ consentirà di accelerare la ricerca sull'intelligenza artificiale e di realizzare nuove e più potenti tecnologie. Queste serviranno a costruire la strada verso il metaverso, il quale è stato definito dalla stessa Meta "*the next major computing platform*" e si stima genererà in pochi anni un mercato di centinaia di miliardi di dollari.

Quelli visti finora sono solo alcuni degli impatti che l'IA avrà, in parte già da ora, sulla società. Sarebbe possibile elencarne molti altri, ognuno legato ad un diverso aspetto della vita umana, fino a disegnare uno scenario in cui potenzialmente nessuno di essi è escluso. Tuttavia, sebbene molto spesso i sistemi di intelligenza artificiale contribuiscano positivamente al benessere della collettività, è bene interrogarsi sui rischi ad essi associati. L'utilizzo sempre più pervasivo, e dagli sviluppi non del tutto prevedibili, di queste tecnologie porta infatti con sé una serie di problemi inediti, oltre ad un contesto nuovo in cui la persona e le organizzazioni rischiano di trovarsi allo sbando. In un recente studio⁷⁰, il Parlamento Europeo ha individuato sei aree principali in cui i dilemmi etici e le questioni morali assumono maggiormente rilevanza: l'impatto sociale⁷¹, l'impatto sulla psicologia umana, l'impatto sul sistema finanziario, l'impatto sul sistema legale, l'impatto sull'ambiente ed il pianeta, l'impatto sulla fiducia.

1) Impatto sociale

Uno degli ambiti sociali più delicati e che sono stati trattati con maggiore attenzione negli ultimi anni riguarda le conseguenze dell'utilizzo dell'intelligenza artificiale sui lavoratori. Il tema degli effetti sull'occupazione dovuti alle nuove tecnologie non è certamente nuovo. Già in passato la progressiva automazione e meccanizzazione dei processi industriali, oltreché l'avvento dei computer, hanno suscitato diversi interrogativi e timori sulla possibilità che questi fenomeni portassero ad una notevole perdita di posti di lavoro. Come accaduto allora, anche oggi l'IA si appresta a modificare il mercato del lavoro sia a livello di domanda che di offerta, in modi ancora non del tutto

⁶⁹ Alla data in cui si scrive, il 02.12.2022, non è ancora giunta notizia che l'AI Research SuperCluster sia stato completato. Il lavoro sarebbe dovuto terminare verso la metà del 2022.

⁷⁰ European Parliamentary Research Service (2020), *The ethics of artificial intelligence: Issues and initiatives*, Bruxelles.

⁷¹ Nello studio si utilizza l'espressione "*Impact on society*", ma sia per non far confusione con il titolo del paragrafo, sia perché ritenuta più calzante con le tematiche trattate in seguito, si è preferito riportare la traduzione "impatti sociali".

chiari. Da un lato, una parte degli studiosi si concentra sulla forza per lo più distruttrice delle nuove tecnologie, affermando che nel prossimo futuro sarà a rischio una quota sostanziosa di occupazione⁷². Dall'altro, coloro maggiormente ottimisti sottolineano come storicamente i processi di automazione abbiano portato ad una riduzione dell'impiego degli esseri umani solamente nel breve periodo, e che anzi abbiano creato più posti di lavoro nel medio-lungo periodo, compensando dunque la decrescita iniziale⁷³.

È interessante notare anche come l'intelligenza artificiale non colpirà tutte le professioni allo stesso modo. Secondo un report di McKinsey⁷⁴, lo svolgimento di attività fisiche e la gestione di macchinari in contesti prevedibili sono le mansioni con il maggior potenziale di automazione, seguite dall'elaborazione e dalla raccolta di dati. Al contrario, l'applicazione di competenze specifiche nei processi di decision making, nella pianificazione e in attività creative e la gestione delle persone sono quelle più difficilmente automatizzabili. Un altro studio⁷⁵, in cui sono stati interpellati 156 esperti in machine learning, robotica e sistemi intelligenti provenienti dal mondo industriale e accademico, ha evidenziato come i lavori impiegatizi, ripetitivi, precisi e percettivi possono essere automatizzati in misura crescente, mentre compiti più creativi, dinamici e human oriented tendono ad essere meno suscettibili ad automazione. Un modello in grado di spiegare i processi di ricomposizione nella distribuzione di professioni e retribuzioni in atto è offerto da Autor⁷⁶. Egli sostiene che la diminuzione del costo della potenza computazionale stia portando ad un ricorso sempre maggiore dei software e delle macchine nello svolgimento di compiti facilmente codificabili. Questi, indicati come "routine task", sono tipici delle professione mediamente qualificate (*middle-skilled*). La

⁷² Frey, C.B., Osborne, M. (2013), *The Future of Employment: How susceptible are jobs to computerisation?*, Oxford Martin Programme on Technology and Employment.

⁷³ Autor, D.H. (2015), "Why Are There Still So Many Jobs? The History and Future of Workplace Automation", *Journal of Economic Perspectives*, 29(3), pp. 3-30.

⁷⁴ Bughin, J., Chui, M., Dewhurst, M., George, K., Manyika, J., Miremadi, M., Willmott, P. (2017), *A future that works: Automation, employment, and productivity*, McKinsey Global Institute, p. 42.

⁷⁵ Duckworth, P., Graham, L., Osborne, M. (2019), "Inferring Work Task Automatability from AI Expert Evidence", in Conitzer, V., Hadfield, G., Vallor, S. (a cura di), *AIES '19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, p. 490.

⁷⁶ Autor, D.H. (2013), "The "task approach" to labor markets: an overview", *Journal for Labour Market Research*, 46, pp. 185-199.

progressiva automatizzazione delle funzioni routinarie starebbe dunque comportando una notevole contrazione fra le mansioni di fascia media, a favore di un incremento occupazionale delle professioni altamente qualificate (*high-skilled*), caratterizzate da attività di analisi, ragionamento e problem solving, e di quelle non qualificate (*lower-skilled*), determinando così un fenomeno di “polarizzazione” del mercato del lavoro. Ancorché lo stesso Autor abbia in parte rivisto le sue previsioni⁷⁷, le prime analisi delle tendenze globali sembrano supportare questa ipotesi⁷⁸. In aggiunta a ciò, se si considera che l’intelligenza artificiale è sempre più in grado di svolgere anche attività di risoluzione di problemi e di ragionamento non routinarie, non è da escludere che essa sostituisca gli esseri umani anche nel lavoro intellettuale⁷⁹. Determinate figure professionali rischiano di andare incontro ad un drastico calo, se non ad una vera e propria scomparsa, mentre altre saranno sempre più richieste.

Nonostante questo scenario, molti esperti tendono ormai a suggerire che il miglior modo per affrontare la rivoluzione in atto non sia rimpiazzando l’uomo con l’IA. Se vista come uno strumento in grado di lavorare al fianco delle persone, essa consente infatti di migliorare notevolmente lo svolgimento di determinate mansioni. I sistemi di intelligenza artificiale sono migliori degli esseri umani in attività quali l’analisi di dati e la rilevazione di schemi. D’altro canto, faticano ancora in compiti che richiedono intelligenza generale o abilità fisiche e interpersonali. Collaborando fra loro, macchine e uomini hanno la possibilità di completarsi a vicenda⁸⁰, generando benefici in termini di produttività e di condizioni operative. L’integrazione di queste tecnologie nel contesto aziendale a stretto contatto con i lavoratori umani apre, tuttavia, alcune questioni sul ruolo che le prime dovrebbero avere rispetto ai secondi. Lo sviluppo di sistemi di intelligenza artificiale sempre più avanzati potrebbe portare ad una situazione in cui essi

⁷⁷ Autor, D.H. (2015), “Why Are There Still So Many Jobs? The History and Future of Workplace Automation”, *Journal of Economic Perspectives*, 29(3), p. 26.

⁷⁸ Albarelli, A., Bagnoli, C., Campostrini, S., Massaro, M., Muraro, A., Toniolo, K., Vesnic, L., Zantedeschi, M. S. (2019), *Gli impatti di IA e di Blockchain sui modelli di business. Il Report di Ricerca*, Strategy Innovation, p. 39.

⁷⁹ Colaiacomo, G. (2022), “Intelligenza artificiale e dignità del lavoratore: la necessità di un approccio antropocentrico”, in Buzzelli, D., Palazzo, M. (a cura di), *Intelligenza artificiale e diritti della persona*, Pacini Editore, Pisa, p. 212.

⁸⁰ Laubacher, R., Malone, T.W., Rus, D. (2020), *Artificial Intelligence and the Future of Work*, Massachusetts Institute of Technology, Cambridge, p. 20 sg.

cessino di essere dei semplici strumenti, iniziando ad assumere delle decisioni rilevanti sullo stato e sui compiti degli individui. Basti pensare, ad esempio, all'operatore di un *call center* che chiama determinati soggetti seguendo un ordine stabilito da un software di profilazione. Tutto ciò avrebbe effetti negativi:

- sulla dignità dei lavoratori, i quali sarebbero sottoposti, eventualmente anche a loro insaputa, a decisioni automatizzate, senza possibilità di verificare *ex post* il processo logico seguito. Questo sarebbe vero soprattutto nel caso di ricorso a reti neurali, data la complessità del loro funzionamento (effetto “*black box*”⁸¹);
- in materia di responsabilità, la quale verrebbe notevolmente limitata. Diventerebbe infatti difficile stabilire se attribuire una determinata azione alla persona che l'ha compiuta o alla macchina che l'ha decisa.

Per evitare che questo accada, occorre lasciare all'uomo una sorta di “ultima parola” sull'attività dell'IA⁸², in modo da mantenere la presenza di un agente morale umano che funga da centro di responsabilità e prenda in ultima istanza le decisioni riguardanti gli altri individui coinvolti.

L'intelligenza artificiale non impatterà sul mercato del lavoro solo in termini quantitativi, ma modificherà anche le competenze richieste a livello professionale. Le *soft skill* sembrano essere al centro del dibattito. Secondo una ricerca condotta da LinkedIn, il 57% dei *senior leader* le ritiene ormai più importanti delle *hard skill*⁸³. Competenze come il pensiero analitico e critico, la risoluzione di problemi complessi, *l'active learning*, la creatività e *l'emotional intelligence* sono oggi fra le più rilevanti nel panorama occupazionale⁸⁴. Un ruolo importante sarà quello della formazione. Le scuole a tutti i livelli hanno il compito di preparare già da ora i lavoratori del domani ai cambiamenti apportati dall'IA, inserendo nei loro programmi il pensiero

⁸¹ L'effetto “*black box*” e il problema della trasparenza saranno trattati più approfonditamente al punto 6).

⁸² Colaiacomo, G. (2022), “Intelligenza artificiale e dignità del lavoratore: la necessità di un approccio antropocentrico”, in Buzzelli, D., Palazzo, M. (a cura di), *Intelligenza artificiale e diritti della persona*, Pacini Editore, Pisa, p. 219. (UGUALE A RIFERIMENTO DI PRIMA)

⁸³ Petrone, P. (2018), *The Skills Companies Need Most in 2019 – And How to Learn Them*, LinkedIn, online, <https://www.linkedin.com/business/learning/blog/top-skills-and-courses/the-skills-companies-need-most-in-2019-and-how-to-learn-them>.

⁸⁴ Brown, S., Hingel, G., Ratcheva, V., Zahidi, S. (2020), *The Future of Jobs Report 2020*, World Economic Forum, p. 36.

computazionale⁸⁵ e potenziando l'insegnamento delle materie STEM (*science, technology, engineering, mathematics*). Sarà inoltre importante assicurare l'accesso alle nuove competenze necessarie a tutte le categorie della popolazione, per evitare che si crei un "AI divide" tra coloro più in grado di acquisire nuove capacità e altre che rischiano di rimanere indietro⁸⁶.

Altro tema controverso appartenente alla sfera sociale riguarda la gestione e l'utilizzo dei dati. L'intelligenza artificiale è una delle applicazioni più importanti della *data economy*⁸⁷. Questi sistemi pongono alla base del loro funzionamento l'analisi e l'elaborazione di grandi quantità di dati, non sempre anonimi. Pertanto, la tutela dei dati personali risulta essere una delle aree giuridiche più interessate dalla diffusione dell'IA⁸⁸. Applicazioni basate sul machine learning potrebbero, infatti, essere utilizzate da autorità e organizzazioni per scavare nelle vite delle persone e scoprire le loro abitudini o altre informazioni sensibili, come gli orientamenti politici o religiosi, andando quindi a violare la loro privacy. Questa eventualità sembra ancor più probabile se si considera la capacità di determinati algoritmi di ricostruire i collegamenti tra i dati analizzati e di "de-anonimizzarli" rispetto agli individui a cui si riferiscono, creando nuovi rischi legati alla protezione dei dati personali anche nei casi di dataset apparentemente anonimi. Al momento, non sembrano esserci strumenti legali per contrastare adeguatamente il problema. Nel 2016, l'Unione Europea ha adottato il Regolamento Generale sulla Protezione dei Dati (*General Data Protection Regulations, GDPR*) allo scopo di tutelare la privacy dei cittadini; tuttavia, il testo si applica soltanto ai

⁸⁵ Laubacher, R., Malone, T.W., Rus, D. (2020), *Artificial Intelligence and the Future of Work*, Massachusetts Institute of Technology, Cambridge, p. 25.

⁸⁶ Colaiacomo, G. (2022), "Intelligenza artificiale e dignità del lavoratore: la necessità di un approccio antropocentrico", in Buzzelli, D., Palazzo, M. (a cura di), *Intelligenza artificiale e diritti della persona*, Pacini Editore, Pisa, p. 212. [UGUALE A RIFERIMENTO DI PRIMA]

⁸⁷ Commissione Europea (2020), *White Paper on Artificial Intelligence – A European approach to excellence and trust*, COM(2020) 65 final, Bruxelles, p. 1.

⁸⁸ Cerrina Feroni, G. (2022), "AI e protezione dei dati personali: le nuove sfide dell'Autorità garante", in Buzzelli, D., Palazzo, M. (a cura di), *Intelligenza artificiale e diritti della persona*, Pacini Editore, Pisa, p. 112.

dati personali, e non ai dati aggregati, teoricamente anonimi, utilizzati per addestrare i modelli di ML⁸⁹.

Oltre alle preoccupazioni concernenti la violazione della sfera privata delle persone, negli ultimi anni l'utilizzo dei dati da parte dell'intelligenza artificiale ha generato questioni su possibili discriminazioni a danno di determinate categorie sociali. Gli output di un modello di IA, infatti, sono strettamente dipendenti dai dati con cui è stato addestrato. Qualora questi siano stati raccolti nella società, potrebbero essere espressione di *bias* umani più o meno latenti e, in assenza di un intervento da parte degli sviluppatori, portare il modello a perpetrare ingiustizie e discriminazioni. Si pensi ad esempio al caso di COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), un software basato sul machine learning utilizzato negli Stati Uniti per valutare il rischio di recidiva e la pericolosità sociale degli imputati, che si è dimostrato⁹⁰ assegnare probabilità maggiori a determinati soggetti in base al colore della pelle. Se impiegati nella valutazione del merito creditizio, questi sistemi non neutrali potrebbero ingiustamente negare la concessione di prestiti a particolari gruppi di persone, rendendo loro più difficile spezzare il ciclo della povertà; o ancora, essi rischierebbero di svantaggiare intere categorie sociali nelle fasi di assunzione o ammissione all'università. Al fine di mitigare le problematiche sopracitate, appare innanzitutto necessario prestare attenzione alla qualità dei dati raccolti. Infatti, come sottolineato dalla *European Union Agency for Fundamental Rights (FRA)*, "(...) an algorithm in its application can only be as good as the data it uses"⁹¹. Nonostante questa esigenza sia molto sentita anche dalle istituzioni europee⁹², non appare immediato come far rispettare dei requisiti che

⁸⁹ European Parliamentary Research Service (2020), *The ethics of artificial intelligence: Issues and initiatives*, Bruxelles, p. 13. **(STESSO DI SOPRA)**

⁹⁰ Angwin, J., Larson, J., Mattu, S., Kirchner, L. (2016), *Machine Bias*, ProPublica, online, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

⁹¹ European Union Agency for Fundamental Rights (2019), *Data quality and artificial intelligence – mitigating bias and error to protect fundamental rights*, **p. 2.**

⁹² Uno degli ultimi rimandi all'importanza dei dati di alta qualità è stato fatto nella proposta per il regolamento noto come "*Data Act*", in cui si dice espressamente come essi siano un requisito che consente di incrementare la competitività e l'innovazione e di assicurare una crescita economica sostenibile. Si rimanda al testo della Commissione Europea (2022), *Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)*, COM(2022) 68 final, **p. 17.**

prestino attenzione ai vari aspetti legati alla qualità dei dati durante tutte le fasi di funzionamento degli algoritmi. Per raggiungere un simile obiettivo, oltre alle adeguate previsioni normative, una soluzione potrebbe essere rappresentata dai dati sintetici e dalla blockchain⁹³. I primi⁹⁴ consentirebbero di tutelare maggiormente coloro a cui si riferiscono i dati personali di partenza mediante la “pseudonimizzazione” dei dati e la perdita di un riferimento diretto alle informazioni potenzialmente sensibili. Allo stesso tempo, la blockchain permetterebbe di includere nella catena solo dati che siano conformi a determinati standard qualitativi e limitare l’accesso dei titolari esclusivamente a quelli per cui gli utenti abbiano prestato effettivamente il consenso. Con questa tecnologia, inoltre, sarebbe possibile tracciare tutte le transazioni eseguite, aumentando così la consapevolezza degli interessati.

Sempre con riferimento ai dati, la necessità di disporre di dataset sempre più grandi per allenare i modelli di intelligenza artificiale ha portato ad una notevole preoccupazione riguardo alla concentrazione del mercato digitale. Oggi, infatti, questi big data set sono conservati e processati in infrastrutture centralizzate basate su cloud. Ciò limita lo spazio di ricerca e di produzione di sistemi di IA soltanto a poche grandi società, consentendo loro di raggiungere una posizione dominante rispetto ai concorrenti⁹⁵. Le implicazioni di questa situazione rischiano, però, di travalicare i confini del mercato. Il potere tecnologico, economico e politico raggiunto sfruttando i dati consente a player quali Google, Microsoft, Amazon, Facebook di esercitare una notevole influenza anche in diverse altre aree della società, alcune delle quali particolarmente rilevanti ai fini della costruzione delle opinioni all’interno di una democrazia: governo, partiti politici, scuola, social network, stampa, ricerca scientifica⁹⁶. In particolare, Nemitz⁹⁷ teme che questi

⁹³ Peluso, M.G. (2022), “Intelligenza Artificiale e dati di qualità: la tecnologia come valido alleato”, *MediaLaws – Rivista di diritto dei media*, 15(2), p. 332.

⁹⁴ I dati sintetici sono dati artificiali creati mediante tecniche di IA a partire da dati reali e con le stesse proprietà statistiche di quest’ultimi.

⁹⁵ Albarelli, A., Bagnoli, C., Campostrini, S., Massaro, M., Muraro, A., Toniolo, K., Vesnic, L., Zantedeschi, M. S. (2019), *Gli impatti di IA e di Blockchain sui modelli di business. Il Report di Ricerca*, Strategy Innovation, p. 37.

⁹⁶ European Parliamentary Research Service (2020), *The ethics of artificial intelligence: Issues and initiatives*, Bruxelles, p. 11.

giganti del tech possano condizionare lo sviluppo e l'implementazione dell'IA, nonché le discussioni sulla sua regolazione, grazie a quattro leve alla base del loro "digital power":

- 1) la disponibilità di notevoli risorse finanziarie, da investire per influenzare la politica e la società o per acquistare nuove idee o start-up innovative legate al settore dell'intelligenza artificiale ("killing acquisitions");
- 2) il controllo delle infrastrutture dove avviene principalmente il discorso pubblico. Essi si stanno sempre più sostituendo alle classiche pubblicazioni giornalistiche, diventando la principale fonte informativa per la maggior parte della popolazione e accaparrandosi la maggior parte dei ricavi pubblicitari, a danno di numerose testate indipendenti;
- 3) la raccolta di dati personali tramite i quali profilare gli utenti, in modo da utilizzare le informazioni ottenute per scopi di lucro, di sorveglianza e sicurezza ed elettorali;
- 4) lo sviluppo di applicazioni di IA sempre più complesse e capaci di migliorare i processi aziendali, così da rafforzare la loro posizione di dominio.

Mentre in diverse parti del mondo si studiano delle proposte normative per arginare il potere di questi grandi attori del panorama digitale⁹⁸, un aiuto concreto in tal senso potrebbe nuovamente arrivare dalla blockchain. Difatti, essa consente di contrastare la creazione di dataset centralizzati mediante la predisposizione di registri di dati di addestramento distribuiti, in cui è possibile verificare la provenienza degli stessi e le transazioni eseguite⁹⁹.

Infine, si ritiene il caso di citare i sistemi di rilevamento delle emozioni. Questi sfruttano l'intelligenza artificiale ed il machine learning per rilevare ed analizzare elementi come il batter d'occhi, il movimento dei muscoli facciali, l'intonazione della voce e il linguaggio del corpo, al fine di determinare lo stato emotivo di un soggetto. Sebbene inizialmente

⁹⁷ Nemitz, P. (2018), "Constitutional democracy and technology in the age of artificial intelligence", *Philosophical Transactions of the Royal Society A. Mathematical, Physical and Engineering Sciences*, 376(2133).

⁹⁸ Cfr. Madotto, P. (2021), *Big tech, troppo potere: tutte le proposte per risolvere il dilemma del decennio*, Agenda Digitale, online, <https://www.agendadigitale.eu/cultura-digitale/big-tech-troppo-potere-tutte-le-proposte-sul-tavolo/>.

⁹⁹ Peluso, M.G. (2022), "Intelligenza Artificiale e dati di qualità: la tecnologia come valido alleato", *MediaLaws - Rivista di diritto dei media*, 15(2), p. 334.

gli studi su queste tecnologie fossero spinti da intenti altamente positivi e sociali, come ad esempio il miglioramento dei comportamenti e delle abitudini di vita degli individui¹⁰⁰, oggi il loro utilizzo solleva alcune questioni di tipo etico e legale. Innanzitutto, la capacità di queste applicazioni di “leggere” l’animo umano mette a rischio la privacy delle persone, le quali comprensibilmente potrebbero pretendere che le proprie emozioni rimangano private. Oltre a ciò, esse aprono le porte a forme di condizionamento degli utenti fortemente invasive e al tempo stesso assolutamente silenziose¹⁰¹. Il riconoscimento delle emozioni permette, infatti, di aumentare profondamente la personalizzazione dei servizi e conseguentemente di influenzare i comportamenti dei singoli sia online che offline, limitando così la loro autonomia privata.

Al momento non ci sono strumenti normativi per affrontare adeguatamente queste problematiche. Va menzionata la collocazione dei sistemi di riconoscimento delle emozioni all’interno della proposta del regolamento europeo denominato “AI Act”¹⁰², il quale, una volta approvato¹⁰³, sarà la prima legge sull’intelligenza artificiale emanata da un ente regolatore di tale importanza; tuttavia, già ora sono state espresse perplessità circa l’effettiva efficacia della disposizione nel tutelare i cittadini dai rischi connessi a questi sistemi¹⁰⁴.

2) Impatto sulla psicologia umana

L’intelligenza artificiale migliora costantemente le forme e i modi con cui entra a contatto con le persone. Spesso è difficile distinguere l’output di una macchina da quello di un essere umano, se non impossibile. Si pensi ad esempio al recente ChatGPT, un modello conversazionale capace di dare risposte e fornire informazioni in modo

¹⁰⁰ Incutti, E.M. (2022), “Sistemi di riconoscimento delle emozioni e ruolo dell’autonomia privata: linee evolutive di un umanesimo digitale”, in Buzzelli, D., Palazzo, M. (a cura di), *Intelligenza artificiale e diritti della persona*, Pacini Editore, Pisa, p. 152.

¹⁰¹ Ivi, p. 157.

¹⁰² Si tratterà più approfonditamente l’AI Act alla fine del paragrafo.

¹⁰³ Alla data in cui si scrive, il 15.12.2022, la proposta non è ancora stata approvata.

¹⁰⁴ Cfr. Ienca, M., Malgieri, G. (2021), *Artificial Intelligence Act: l’UE regola l’AI ma dimentica di proteggere la mente*, Agenda Digitale, online, <https://www.agendadigitale.eu/cittadinanza-digitale/artificial-intelligence-act-lue-regola-lai-ma-dimentica-di-proteggere-la-mente/>.

apparentemente naturale, molto accurato e pertinente al contesto¹⁰⁵. In un futuro in cui queste tecnologie saranno sempre più presenti nella quotidianità, è necessario interrogarsi sugli effetti che esse avranno sulle relazioni. Queste, infatti, costituiscono un elemento fondamentale dell'esistenza umana. L'utilizzo di robot intelligenti per compiti in cui l'aspetto relazionale è predominante, come l'insegnamento, l'assistenza infermieristica, la cura dei bambini e anche alcuni legati alla sfera sessuale, potrebbe impattare non solo sui legami tra macchine e individui, ma anche su quelli intercorrenti tra le persone.

Sebbene la fantascienza venga in aiuto¹⁰⁶, non è ancora chiaro cosa potrebbe accadere qualora qualcuno sviluppi un attaccamento emotivo verso un'intelligenza artificiale. Uno dei pericoli principali è che i robot siano utilizzati per sfruttare i sentimenti delle controparti umane allo scopo di raggirarle o manipolarle¹⁰⁷. Per vivere al meglio questi rapporti, inoltre, i soggetti coinvolti potrebbero dover costantemente illudersi, a livello conscio o inconscio, di non aver a che fare con un automa, bensì con una persona in carne ed ossa¹⁰⁸. Borenstein e Arkin¹⁰⁹ si chiedono, poi, quali sarebbero gli effetti di una relazione di questo tipo, priva del rischio di un possibile rifiuto da parte della macchina, sullo sviluppo mentale e sociale degli individui.

L'intensificarsi dei legami affettivi con applicazioni di IA potrebbe portare anche ad una diminuzione dei contatti con gli altri esseri umani e ad un allontanamento dalla società. Le credenze, gli atteggiamenti ed i valori che caratterizzano la complessità dei rapporti tra le persone sono suscettibili di venire alterati di fronte alla semplicità delle relazioni con i robot. Secondo Christakis¹¹⁰, l'avvento di macchine dalle sembianze e dai comportamenti molto simili a quelli umani potrebbe intaccare il cosiddetto "social suite",

¹⁰⁵ Cfr. Doda, I. (2022), *Come funziona ChatGPT, il bot conversazionale diventato virale*, Wired Italia, online, <https://www.wired.it/article/chatgpt-bot-conversazionale/>.

¹⁰⁶ Su questo tema, si suggerisce la visione di "Ex Machina", diretto da Alex Garland e di "Her", per la regia di Spike Jonze.

¹⁰⁷ European Parliamentary Research Service (2020), *The ethics of artificial intelligence: Issues and initiatives*, Bruexlles, p. 18.

¹⁰⁸ Ivi, p. 19.

¹⁰⁹ Arkin, R.C., Borenstein, J. (2019), "Robots, Ethics, and Intimacy: The Need for Scientific Research", in Berkich, D., d'Alfonso, M.V. (a cura di), *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence*, Philosophical Studies Series, 134, Springer, pp. 299-309.

¹¹⁰ Christakis, N.A. (2019), "How AI Will Rewire Us", in *The Atlantic*, 2019/3.

ovvero l'insieme delle capacità sviluppatesi nel corso di centinaia di migliaia di anni per interagire con gli altri, in cui rientrano anche amicizia, amore, insegnamento e cooperazione. Sempre Christakis dice: *“As AI permeates our lives, we must confront the possibility that it will stunt our emotions and inhibit deep human connections, leaving our relationships with one another less reciprocal, or shallower, or more narcissistic”* ¹¹¹.

3) Impatto sul sistema finanziario

I mercati finanziari sono stati una delle prime aree ad essere interessate dall'introduzione dell'intelligenza artificiale. Essi risultano particolarmente adatti per una simile rivoluzione, poiché richiedono di analizzare grandi volumi di dati in tempi brevissimi al fine di assumere decisioni rilevanti e cogliere le migliori opportunità. Secondo uno studio di Mordor Intelligence¹¹², il mercato delle applicazioni di IA nella finanza crescerà tra il 2022 e il 2027 con un CAGR del 25,3%. Ciò è dovuto al grande apporto che esse sono in grado di apportare in termini di efficienza ed esperienza. L'impiego di chatbot e l'analisi del comportamento dei clienti permette a quest'ultimi di usufruire di soluzioni finanziarie personalizzate. La capacità dei sistemi di *fraud detection* di rilevare transazioni anomale e possibili frodi, nuove tecniche di verifica (ad esempio tramite il riconoscimento del volto) e l'utilizzo di modelli di machine learning nella protezione dagli attacchi informatici aumentano notevolmente il livello generale di sicurezza. Due attività in particolare hanno beneficiato dal ricorso all'intelligenza artificiale: la valutazione del merito creditizio ed il *trading*. Nel primo caso, grazie al ML è stato possibile ampliare la gamma dei dati su cui basare l'analisi, potendo questa contare ora su dati finanziari strutturati e non strutturati e su dati non finanziari strutturati e non. Le nuove fonti di dati “alternative” consentono di considerare variabili prive di una chiara relazione o interpretazione economica e che dunque sarebbero normalmente scartate. Tale innovazione aumentata l'accuratezza delle previsioni ed estende la possibilità di valutare il merito creditizio di soggetti altrimenti penalizzati o addirittura esclusi nell'accesso al credito, oltre che aprire le porte del mercato a nuovi soggetti in grado di offrire finanziamenti a costi più bassi, in tempi più rapidi e con

¹¹¹ Ibidem.

¹¹² Mordor Intelligence (2022), *AI In Fintech Market - Growth, Trends, COVID-19 Impact, And Forecasts (2022 - 2027)*.

minori garanzie¹¹³. Per quanto riguarda il secondo, invece, l'intelligenza artificiale permette di sfruttare al meglio la grande quantità di dati finanziari disponibili per potenziare ulteriormente le capacità del *trading* algoritmico¹¹⁴. Tradizionalmente, questo si basa su istruzioni di tipo *if-then* stabilite dai programmatori che devono essere aggiustate in base alle condizioni di mercato e sono suscettibili all'errore umano. Con l'introduzione del machine learning e l'analisi di dati sia storici, sia in tempo reale, diventa possibile anticipare i movimenti dei mercati e migliorare la gestione dei portafogli e del rischio, fino a realizzare soluzioni dinamiche, in continua evoluzione e completamente automatizzate.

Nonostante i diversi benefici, l'IA ha sollevato nuove problematiche all'interno del sistema finanziario. Oltre ai temi, già affrontati, legati alla tutela dei dati personali e ad eventuali *bias* (specie nella valutazione del merito creditizio), una delle questioni maggiormente controverse riguarda l'impiego di sistemi di intelligenza artificiale nel *trading*. Agenti autonomi potrebbero, accidentalmente o perché utilizzati a tal fine, destabilizzare il mercato e recare danno ad altri soggetti, ad esempio attuando operazioni per manipolare l'andamento dei prezzi o colludendo tacitamente per mantenere un determinato livello di prezzo¹¹⁵. Qualora si verificasse una simile eventualità, inoltre, sarebbe difficile stabilire su chi far ricadere la responsabilità associata. Se è vero che nel caso di algoritmi di *trading* tradizionali sono le società che li sviluppano ed utilizzano a rispondere dei loro output, agenti *AI based* potrebbero compiere azioni difficili da prevedere da parte dei programmatori, aprendo interrogativi sulla possibilità di mitigare la responsabilità di quest'ultimi¹¹⁶.

4) Impatto sul sistema legale

Nel corso della storia, i sistemi giudiziari e normativi sono sempre stati basati sull'assunto per cui sono gli esseri umani ad assumere delle decisioni. In una società in

¹¹³ Bonaccorsi di Patti, E. et al. (2022), *Intelligenza artificiale nel credit scoring. Analisi di alcune esperienze nel sistema finanziario italiano*, Banca d'Italia, p. 15.

¹¹⁴ Il *trading* algoritmico è una soluzione per il *trading* che si avvale di algoritmi per processare ordini di acquisto o vendita ad una frequenza molto più elevata rispetto a quella che sarebbe in grado di tenere un operatore umano.

¹¹⁵ European Parliamentary Research Service (2020), *The ethics of artificial intelligence: Issues and initiatives*, Bruxelles, pp. 21 sg.

¹¹⁶ Ibidem.

cui queste sono sempre più frequentemente prese da degli algoritmi, l'impianto legale vigente rischia di non essere capace di dare le giuste risposte alle nuove situazioni che si verranno a creare¹¹⁷.

L'intelligenza artificiale può essere utilizzata per commettere dei reati in modi molto diversi. Oltre alle pratiche finanziarie illecite già citate (manipolazione dei mercati, collusione), essa potrebbe trovare applicazione nel traffico di droga, grazie all'impiego di veicoli autonomi, specialmente subacquei, per trasportare sostanze illegali e rendere più difficile risalire ai soggetti coinvolti. Ancora, l'IA è in grado di generare contenuti non veritieri sempre più sofisticati e di vario genere (vedi i cosiddetti "deepfake"), con i quali è possibile diffondere informazioni mendaci e intaccare la reputazione di determinate persone¹¹⁸. Questi semplici esempi, per quanto siano solo alcuni di quelli riportabili, danno già un'idea piuttosto chiara del potenziale pericolo legato a queste tecnologie se usate per intenti criminali. Una delle questioni maggiormente controverse quando si parla di "AI-Crime" (AIC) è quella, come si è già potuto vedere, della responsabilità. Gli attuali modelli di responsabilità legale rischiano di risultare inadeguati di fronte al ruolo che l'intelligenza artificiale avrà in futuro nello svolgimento di attività illegali, ciò comportando una maggiore incertezza del diritto¹¹⁹. Diventerebbe difficile stabilire con sicurezza su chi, e in quale misura, addossare le conseguenze delle azioni (o delle omissioni) dell'IA tra programmatori, costruttori, utilizzatori finali, altre figure potenzialmente coinvolte e gli stessi sistemi di IA. Se è vero che attualmente la responsabilità per i danni provocati da un robot o un'applicazione *AI based* è attribuibile esclusivamente ad un essere umano¹²⁰, qualora si abbia a che fare con modelli di machine learning o reti neurali in grado di apprendere in modo autonomo e cambiare rispetto alla configurazione iniziale, arrivando ad adottare comportamenti difficilmente prevedibili a priori, ci si dovrebbe chiedere se siano effettivamente i loro sviluppatori a dover rispondere di eventuali violazioni della legge. Numerosi interessi, anche di

¹¹⁷ Ivi, p. 22.

¹¹⁸ Ivi, pp. 24 sg.

¹¹⁹ Aggarwal, N., Floridi, L., King, T.C., Taddeo, M. (2019), "Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions", *Science and Engineering Ethics*, 26, p. 95.

¹²⁰ Parlamento Europeo (2017), *European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))*.

primaria importanza, potrebbero rimanere sguarniti di tutela pena di fronte a minacce sempre più frequenti e gravi¹²¹.

Nel tentativo di delineare un quadro della responsabilità per i casi in cui l'intelligenza artificiale è coinvolta nello svolgimento di reati, Hallevy¹²² ha individuato tre possibili scenari:

- esecuzione tramite altri. Il sistema di IA commette attivamente il fatto illecito, ma agisce secondo le intenzioni di altri soggetti, i quali sono i veri responsabili. L'IA è considerata un agente innocente, quasi un mero strumento, e le azioni da essa compiute sono attribuibili agli esseri umani (programmatori o utenti finali) che l'hanno utilizzata per soddisfare la loro volontà criminale. Questa situazione è equivalente a quella di un bambino o di un individuo mentalmente incapace che compie un reato eseguendo le istruzioni di qualcun altro;
- conseguenza naturale e probabile. L'illecito è commesso dall'*AI entity* senza che sia stato in alcun modo premeditato o suggerito dai programmatori o dagli utilizzatori. Tuttavia, il comportamento della macchina è ritenuto una conseguenza naturale e probabile delle loro azioni e perciò sono ritenuti perseguibili per negligenza. Essi avrebbero ragionevolmente potuto prevedere la perpetrazione del reato ed adoperarsi per impedirlo. L'IA, però, non è necessariamente libera da ogni responsabilità. A differenza del caso precedente, secondo l'autore bisogna distinguere tra gli agenti da considerarsi innocenti, in quanto privi di ogni conoscenza relativa all'illiceità della condotta, e quelli che non lo sono, per i quali si prefigura il ricorso al terzo scenario;
- responsabilità diretta. Qualora un sistema di intelligenza artificiale soddisfi entrambi i requisiti dell'elemento fattuale (*actus reus*) e dell'elemento mentale (*mens rea*)¹²³ richiesti per l'imputazione della responsabilità, anche senza il coinvolgimento di una qualche persona nell'esecuzione del crimine, per Hallevy

¹²¹ Borsari, R. (2019), "Intelligenza Artificiale e responsabilità penale: prime considerazioni", *MediaLaws – Rivista di diritto dei media*, 7 (3), pp. 265 [sg.](#)

¹²² Hallevy, G. (2010), *The Criminal Liability of Artificial Intelligence Entities*.

¹²³ Hallevy si rifà agli ordinamenti di *common law*, nei quali per attribuire la responsabilità penale in capo ad un soggetto devono coesistere due elementi: l'elemento esterno, o *actus reus*, consistente in una condotta criminale, e l'elemento interno, o *mens rea*, ovvero la consapevolezza o l'intenzione di commettere un reato.

non c'è motivo per negare la punibilità di tale sistema. Se è vero che risulta piuttosto semplice attribuire una determinata azione od omissione ad una IA, stabilire la sussistenza della *mens rea* appare più complesso. D'altro canto, non sempre l'elemento mentale necessita di specifiche emozioni o intenzioni, essendo per molti reati sufficienti la semplice consapevolezza o negligenza. La responsabilità dell'entità artificiale non sostituirebbe quella del programmatore e/o utente finale, nel caso in cui fosse anch'essa presente, bensì andrebbe ad aggiungersi. Inoltre, il sistema godrebbe comunque della possibilità di ricorrere alle difese generali a disposizione degli imputati umani, come autodifesa, necessità, intossicazione (si pensi ad esempio ad un algoritmo infettato da un virus informatico).

Nel primo scenario, ed in parte nel secondo, è possibile notare come l'attuale modello di imputazione della responsabilità indiretta dell'uomo resista senza grosse difficoltà¹²⁴. Gli agenti artificiali considerati non sono particolarmente sviluppati e le loro azioni, per quanto complesse, sono predeterminate e prevedibili. Per questo motivo, è sempre possibile agire contro chi li abbia programmati, costruiti o impiegati, che sia a titolo di dolo o di colpa. Nel secondo e nel terzo caso, invece, la facoltà di attribuire una responsabilità all'IA, in via autonoma o cumulativa rispetto a quella del utente o programmatore, rende il modello tradizionale inadeguato. Sistemi di intelligenza artificiale avanzati sono in grado di "comprendere" la realtà che li circonda analizzando i dati acquisiti e di prevedere le conseguenze delle loro azioni grazie ai processi di *decision-making*, diventando dunque suscettibili di soddisfare il requisito della *mens rea*. Ciò, unito alla sempre maggiore imprevedibilità dei loro comportamenti da parte degli operatori umani, porta a concepire le macchine come soggetti attivi del reato, superando così l'assioma per cui *machina delinquere (et puniri) non potest*¹²⁵.

Al di là di ogni possibile scenario, rimane la necessità di elaborare un modello in grado di dare risposte certe alle questioni sempre più spinose che il diritto penale si troverà ad affrontare.

¹²⁴ Borsari, R. (2019), "Intelligenza Artificiale e responsabilità penale: prime considerazioni", *op. cit.*, p. 264.

¹²⁵ Borsari, R. (2019), "Intelligenza Artificiale e responsabilità penale: prime considerazioni", *op. cit.*, p. 266.

Sul fronte civilistico, uno dei temi principali riguarderà la diffusione sulle strade delle *driverless car*. L'avanzamento tecnologico a cui è sottoposto il settore automobilistico condurrà ad un periodo di transizione in cui ai mezzi di trasporto tradizionali si affiancheranno veicoli caratterizzati da crescenti livelli di autonomia, fino a raggiungere l'automazione totale. Ciò comporterà il manifestarsi di nuove casistiche di incidenti in cui gli *autonomous vehicle* saranno coinvolti. Già ora con la commercializzazione di automobili dotate di dispositivi di guida assistita sta emergendo il problema di un passaggio da un regime di responsabilità *driver focused*, che alloca sul pilota l'onere risarcitorio, ad uno *product focused*, in cui si vada a colpire il produttore di tali dispositivi¹²⁶. Con la prospettiva di una totale implementazione della guida automatizzata, si rende necessario stabilire, anche in questo caso, un quadro normativo idoneo a fornire prevedibilità e certezza relativamente alla responsabilità per i danni causati da veicoli autonomi, oltre ad un efficiente meccanismo di allocazione dei costi. Se è vero che gli incidenti dovuti a comportamenti anomali delle persone, imperfezioni della struttura stradale o difetti del veicolo potrebbero continuare ad essere regolati secondo le rispettive discipline attuali, il problema sorge per i sinistri provocati da mezzi conformi e operanti nel rispetto del protocollo e degli standard di sicurezza valutati come ragionevoli dal legislatore.

Una prima soluzione potrebbe giungere dal sistema statunitense. Tramite una lettura coordinata delle norme federali sulla sicurezza dei veicoli e di quelle statali in materia di responsabilità civile è possibile ricavare un modello fondato sulla *regulatory compliance*. A questa sarebbe attribuito il rango di *complete defense*, limitando così la responsabilità del produttore solo ai casi di non conformità rispetto agli standard. Nel caso di incidenti che vedessero coinvolti *regulatory-compliant autonomous vehicle*, il costruttore risponderebbe, quindi, solo qualora la causa sia attribuibile a malfunzionamenti dell'auto dovuti a difetti di fabbricazione oppure egli non abbia fornito adeguati avvertimenti per un utilizzo sicuro del mezzo. Ad addossarsi gli oneri per gli eventuali danni cagionati a terzi sarebbe l'utilizzatore in qualità di proprietario o conducente, così come avviene attualmente. Un approccio di questo tipo apparirebbe sicuramente favorevole alle case automobilistiche, le quali vedrebbero fortemente limitata la propria responsabilità; d'altro canto, esso rischierebbe di non tutelare sufficientemente i

¹²⁶ Al Mureden, E., Calabresi, G. (2021), *Driverless cars. Intelligenza artificiale e futuro della mobilità*, Il Mulino, Bologna, p. 150.

consumatori e di non infondere in loro la fiducia necessaria per passare al nuovo sistema di mobilità.

Oltre a quella appena descritta, esiste un'ipotesi alternativa, ispirata ad un principio di socializzazione dei costi degli incidenti dovuti all'automazione del trasporto su strada. Questa soluzione si fonda su un sistema denominato *Market Enterprise Responsibility* (MER), in base al quale le risorse da destinare al risarcimento dei danni provocati dai veicoli a guida autonoma sarebbero attinte da un fondo appositamente creato. Questo sarebbe finanziato dagli stessi costruttori, tramite dei contributi commisurati alla rispettiva quota di mercato, secondo il meccanismo della *Market Share Liability*. In questo modo, ognuno parteciperebbe proporzionalmente rispetto al numero di veicoli venduti, rappresentazione dell'entità del rischio immesso nella società. Un tale sistema sarebbe a stampo fortemente indennitario, con la possibilità per i danneggiati di ricevere in automatico una compensazione economica soddisfacente e la responsabilità che sarebbe posta in capo ai produttori mediante l'obbligo dei versamenti periodici al fondo. Tuttavia, qualora ci fosse l'esigenza di condannare comportamenti particolarmente gravi delle case automobilistiche, sarebbe possibile prevedere forme di responsabilità civile a scopo punitivo. Se da un lato la costituzione di un meccanismo di indennizzi automatici potrebbe pregiudicare gli interessi dei professionisti che operano nel settore dell'infortunistica stradale, oltre che, in parte, della classe forense, dall'altro potrebbe portare ai costruttori un notevole risparmio di costi di transazione e di spese legali.

Al di là degli espedienti adottabili per quanto concerne la mobilità su strada del futuro, permane la necessità di stabilire un modello di attribuzione della responsabilità (e dei costi per il risarcimento del danno ad essa associati) che funzioni in tutti gli altri ambiti della vita quotidiana. A tal proposito, una possibile soluzione potrebbe giungere dall'Unione Europea. In una Risoluzione del 2017 recante raccomandazioni circa norme di diritto civile sulla robotica, il Parlamento Europeo ha invitato la Commissione a valutare la creazione di *"a specific legal status for robots in the long run, so that at least the most sophisticated autonomous robots could be established as having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or*

otherwise interact with third parties independently"¹²⁷. In sostanza, si prevede l'eventualità di assegnare a quei robot particolarmente sofisticati che prendono decisioni autonomamente ed interagiscono con terzi una "personalità elettronica", rendendoli responsabili per i danni da essi stessi cagionati. Tuttavia, non è chiaro se questa nuova forma di entità legale, nel caso in cui dovesse effettivamente vedere la luce, costituirebbe a tutti gli effetti un *tertium genus* di soggettività giuridica. Ci si interroga, infatti, se accanto alle figure attuali della persona fisica e della persona giuridica possa trovare posto quella della "persona elettronica", titolare anch'essa di diritti e doveri¹²⁸. Ciò che potrebbe avvenire è un percorso simile a quello che ha interessato lo schiavo umano nel corso della storia, la cui configurazione è passata da mero bene di proprietà del *dominus* a persona meritevole di tutela giuridica¹²⁹. Oltre alla suddetta questione, pare indispensabile stabilire se le norme sulla responsabilità civile vigenti possano essere interpretate ed adottate in considerazione dell'eventuale nuovo scenario. Elevare i robot a soggetti autonomi implica necessariamente sottoporli a leggi con le quali determinare le conseguenze delle loro azioni od omissioni, qualora queste non siano attribuibili a figure umane ma a decisioni autodeterminate dell'agente artificiale¹³⁰. In tal caso, si verificherebbe l'esenzione della responsabilità di produttori ed utilizzatori, con il rischio di limitare la tutela risarcitoria dei soggetti lesi ad uno specifico e limitato patrimonio¹³¹.

Il dibattito sulla personalità elettronica è molto acceso e ancora in corso, sia a livello europeo che internazionale. L'idea che si sta delineando è che questo strumento giuridico non debba essere finalizzato a dare diritti all'intelligenza artificiale, bensì a fare in modo che essa mantenga sempre un supporto umano. Al pari degli schiavi nell'antichità, l'entità basata su IA resta un bene anche quando personificata e dotata di

¹²⁷ Parlamento Europeo (2017), *European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL))*, punto 59, lett. f).

¹²⁸ Marinacci, D. (2020), *La personalità elettronica: un tertium genus di personalità elettronica?*, Italia Informa - Quotidiano Online, online, <https://italia-informa.com/notizie/La-personalit%C3%A0-elettronica--un-tertium-genus-di-personalit%C3%A0.aspx>.

¹²⁹ Bevivino, G. (2022), "Robot e personalità elettronica: un esempio di approccio analitico", in Buzzelli, D., Palazzo, M. (a cura di), *Intelligenza artificiale e diritti della persona*, Pacini Editore, Pisa, p. 60.

¹³⁰ Marinacci, D. (2020), *La personalità elettronica: un tertium genus di personalità elettronica?*, op. cit.

¹³¹ Ruffolo, U. (2020), "Il problema della "personalità elettronica"", *Journal of Ethics and Legal Technologies*, **2(1)**, p. 76.

patrimonio. Dei danni derivanti dal suo comportamento possono ben rispondere altri soggetti, sia in concorso con essa che in via esclusiva¹³². Il tema della personalità elettronica sembra essere, dunque, una questione prevalentemente disciplinare, volta ad identificare la disciplina maggiormente adatta per affrontare le problematiche legate alla diffusione dell'intelligenza artificiale nella società¹³³. La stessa Commissione Europea appare orientata a valutare l'appropriatezza degli attuali strumenti ed istituti giuridici previsti dai diversi Ordinamenti statuali a regolare le tipologie di rapporti civili che vedranno coinvolti l'IA, senza prevedere al momento l'introduzione di nuove categorie giuridiche¹³⁴.

Prima di passare al punto successivo, è il caso di menzionare come i sistemi *AI based* stanno entrando nel campo della giustizia. Sono sempre più diffusi, infatti, *software* basati su tecniche di machine learning in grado di prevedere il risultato di controversie legali con un notevole livello di affidabilità. Basti pensare ai risultati della *Case Crunch Lawyer Challenge*, una singolare gara in cui 112 avvocati hanno sfidato un programma di intelligenza artificiale sviluppato da una *start-up* di *Cambridge*¹³⁵. Lo scopo consisteva, appunto, nella previsione di dispute relative a particolari clausole contrattuali denominate PPI (*payment protection insurance*) precedentemente risolte dal *Financial Ombudsman*¹³⁶. La competizione è stata vinta dall'IA, la quale ha fornito risposte corrette nell'86,6% dei casi affrontati, a fronte di un tasso di esattezza degli sfidanti umani pari al 62,3%. Un altro esempio della sempre maggiore rilevanza che queste tecnologie avranno in ambito giudiziario è rappresentato dal caso della *start-up DoNotPay*. È notizia recente¹³⁷, infatti, che a febbraio 2023 un software di intelligenza artificiale sviluppato

¹³² Ivi, p. 77.

¹³³ Bevivino, G. (2022), "Robot e personalità elettronica: un esempio di approccio analitico", op. cit., p. 62.

¹³⁴ Marinacci, D. (2020), *La personalità elettronica: un tertium genus di personalità elettronica?*, op. cit.

¹³⁵ Cross, M. (2017), *Robot beats human lawyers in outcomes challenge*, The Law Society Gazette, online, <https://www.lawgazette.co.uk/practice/robot-beats-human-lawyers-in-outcomes-challenge/5063471.article>.

¹³⁶ Il Financial Ombudsman Service è il difensore civico del Regno Unito preposto alla risoluzione di controversie che vedono di fronte consumatori e fornitori di servizi finanziari.

¹³⁷ Farberov, S. (2023), *'Robot lawyer' powered by AI will help fight speeding ticket as it takes first case in court*, New York Post, online, <https://nypost.com/2023/01/05/robot-lawyer-powered-by-ai-will-help-fight-speeding-ticket-as-it-takes-first-case-in-court/>.

da questa società difenderà un cliente presso una corte degli Stati, costituendo così la prima istanza di “avvocato robot”. L’applicazione per smartphone prima ascolterà in tempo reale e analizzerà le diverse argomentazioni in tribunale, dopodiché dirà all’imputato cosa rispondere tramite degli auricolari. Sebbene la causa in questione (una multa per eccesso di velocità) sia di scarsa rilevanza, questo fatto potrebbe rappresentare un precedente e aprire la strada a molte situazioni analoghe nel prossimo futuro.

Tralasciando il caso di *DoNotPay*, per il quale è ancora prematuro effettuare un’analisi approfondita, l’impiego dell’IA nella giustizia non è esente da alcune delle principali problematiche legate a questa tecnologia. Si pensi alla sopracitata vicenda di COMPAS (in cui il sistema risultava pregiudizievole nei confronti di determinate categorie di individui) o ai dubbi relativi alla trasparenza del processo decisionale¹³⁸, requisito fondamentale qualora l’intelligenza artificiale sia usata a supporto della funzione giurisdizionale. Tuttavia, l’ostacolo principale all’adozione di certe funzionalità *AI based* in questo campo è un altro. I sistemi di giustizia predittiva, infatti, non danno, a differenza di un qualsiasi giudice, una spiegazione delle ragioni che li hanno portati ad una certa soluzione. Ciò, però, va in contrasto con la necessità della motivazione giudiziale della decisione, la quale costituisce un elemento essenziale in cui si sostanziano le garanzie costituzionali del processo e la stessa imparzialità dell’organo giudicante¹³⁹.

Nonostante i limiti esposti, sono diversi i modi in cui coloro che operano nell’ambito della giustizia possono sfruttare il potenziale dell’IA. Oltre che impieghi più “tradizionali” e generici, come la traduzione di documenti, la dettatura automatica e la predisposizione e correzione di bozze, l’intelligenza artificiale consente ad avvocati e studi legali di analizzare velocemente testi giuridici, ricercare precedenti utili e valutare le probabilità di successo di una determinata causa o i termini di un accordo. Essa, inoltre, potrebbe essere utile nei giudizi civili per determinare importi monetari o altri fattori che richiedono valutazioni tecnico-scientifiche. Infine, i sistemi predittivi potrebbero trovare applicazione nelle procedure alternative di risoluzione delle liti,

¹³⁸ L’effetto “*black box*” e il problema della trasparenza saranno trattati più approfonditamente al punto 6).

¹³⁹ Donati, F. (2022), “La giustizia e le nuove tecnologie”, in Buzzelli, D., Palazzo, M. (a cura di), *Intelligenza artificiale e diritti della persona*, Pacini Editore, Pisa, p. 251.

specie per le questioni di valore molto basso, al fine della definizione di un accordo transattivo.

5) Impatto sull'ambiente ed il pianeta

La crisi ambientale in corso è una delle più grandi sfide che l'umanità abbia mai dovuto affrontare. Il surriscaldamento globale dovuto alle attività antropiche rischia di avere conseguenze devastanti sugli ecosistemi e sulla vita di miliardi di persone. Tra le tante soluzioni proposte e che si sta cercando di sviluppare, l'intelligenza artificiale potrebbe giocare un ruolo centrale nella lotta al cambiamento climatico. Sono numerosi, infatti, i modi in cui questa tecnologia può essere impiegata per contrastare le cause e gli effetti di tale fenomeno. Secondo uno studio di BCG¹⁴⁰, il ricorso all'IA consentirebbe un abbattimento delle emissioni di gas serra tra il 5% e il 10% entro il 2030, pari ad un ammontare di CO₂ equivalente compreso tra le 2,6 e le 5,3 gigatonnellate. Essa è in grado di sostenere le imprese nei loro sforzi di riduzione della propria *carbon footprint*, grazie: al monitoraggio delle emissioni, tramite l'analisi di dati relativi alle diverse attività ed aree aziendali; alle previsioni di quelle future, tenendo conto degli sforzi in atto e della domanda attesa; alla loro riduzione, attraverso processi di efficientamento di tutte le fasi della catena del valore. L'intelligenza artificiale può essere usata anche per approfondire la comprensione dei cambiamenti che stanno interessando il clima. Tramite tecniche di machine learning e deep learning, è possibile sfruttare al meglio le grandi moli di dati a disposizione dei ricercatori, riuscendo così a prevedere con maggior precisione l'evoluzione dei fenomeni atmosferici e valutare le conseguenze del surriscaldamento globale. Le informazioni così ottenute consentono, ad esempio, di proteggere aree a rischio di eventi meteorologici estremi e di influenzare le decisioni dei governanti in merito all'ambiente. Molte agenzie pubbliche e private hanno già introdotto algoritmi di ML per incrementare le performance dei propri modelli climatici¹⁴¹. Un altro ambito in cui l'intelligenza artificiale sta ottenendo ottimi risultati è quello dell'agricoltura. Con l'aumento della popolazione mondiale e la desertificazione che avanza, è fondamentale garantire a tutti l'accesso al cibo. A tal proposito, i sistemi di IA sono in grado di ottimizzare la coltivazione in base ai dati sull'umidità, composizione

¹⁴⁰ Degot, C., Duranton, S., Frédeau, M., Hutchinson, R. (2021), *Reduce Carbon and Costs with the Power of AI*, BCG, p. 2.

¹⁴¹ PricewaterhouseCoopers (2018), *Fourth Industrial Revolution for the Earth*, p. 17.

del suolo, la temperatura ecc., diminuendo al contempo le risorse utilizzate. Un caso interessante è quello di NatureSweet, una società agricola che grazie ad applicazioni di IA per il controllo dei parassiti e delle malattie ha ottenuto un incremento della produzione di pomodori in serra del 20%¹⁴². Ancora, l'IA può trovare impiego nella conservazione di specie a rischio estinzione. Già adesso è utilizzata per studiare il comportamento degli animali e soluzioni per preservare la biodiversità. In Africa, ad esempio, applicazioni *AI based* analizzano i dati raccolti da UAV (*unmanned aerial vehicle*) in merito alla fauna per ottenere *insight* sullo stato delle specie presenti, del paesaggio e dell'influenza umana¹⁴³.

Nonostante i tanti modi in cui queste tecnologie contribuiscono a combattere il cambiamento climatico, non si può negare l'importanza che l'intelligenza artificiale sta avendo nel peggiorare la situazione. Infatti, se da un lato il suo utilizzo consente di ridurre le emissioni di gas serra, dall'altro la quantità di energia necessaria per addestrare gli algoritmi di machine e deep learning è sempre maggiore. Un'analisi del 2018¹⁴⁴ ha dimostrato come dal 2012 le risorse computazionali impiegate nel *training* dei grandi modelli di IA sono raddoppiate ogni 3,4 mesi, con un incremento complessivo di 300.000 volte. Basti pensare che il consumo di elettricità richiesto dall'addestramento di GPT-3, un potente *software* di generazione di testo sviluppato da OpenAI, ha generato un quantitativo di CO₂ equivalente paragonabile a quello prodotta da un viaggio di andata e ritorno in auto verso la luna¹⁴⁵. L'impronta di carbonio di un determinato modello non è semplice da stabilire e dipende da diversi fattori. Uno dei più influenti è certamente il numero di operazioni semplici necessarie per allenarlo, a cui si aggiungono l'architettura scelta e le risorse *hardware* utilizzate. Altri elementi da tenere in considerazione sono le modalità di archiviazione dei dati e di raffreddamento dei

¹⁴² Gonzalez, W. (2022), *AI And The Environment: Can Climate Tech Help Save The Planet?*, Forbes, online, <https://www.forbes.com/sites/forbesbusinesscouncil/2022/05/17/ai-and-the-environment-can-climate-tech-help-save-the-planet/>.

¹⁴³ Ivi.

¹⁴⁴ Amodei, D., Brockman, G., Clark, J., Hernandez, D., Sastry, G., Sutskever, I. (2019), *AI and Compute*, OpenAI, online, <https://openai.com/blog/ai-and-compute/#fn2>.

¹⁴⁵ Quach, K. (2020), *AI me to the Moon... Carbon footprint for 'training GPT-3' same as driving to our natural satellite and back*, The Register, online, https://www.theregister.com/2020/11/04/gpt3_carbon_footprint_estimate/.

server, oltre alle fonti da cui si genera l'energia che alimenta il *training* (rinnovabili o fossili).

L'impatto ambivalente dell'intelligenza artificiale è dimostrato anche da uno studio¹⁴⁶ sul ruolo che essa sta avendo nel raggiungimento dei *Sustainable Development Goals*¹⁴⁷ (SDG). In questo, si evidenzia come il ricorso all'IA possa favorire il raggiungimento di 134 dei 169 *target* associati agli SDG, in particolare del 93% dei *target* legati all'ambiente, dell'82% di quelli relativi alla società e del 70% di quelli in ambito economico. Allo stesso tempo, però, 59 *target* potrebbero subire effetti negativi dalla diffusione di tali sistemi.

Per una maggiore sostenibilità ambientale dell'intelligenza artificiale, il primo passo da intraprendere sembra essere l'adozione di un approccio per la valutazione dei modelli di ML e DL maggiormente olistico e orientato alla multidimensionalità¹⁴⁸. Fino ad oggi, la ricerca si è spesso concentrata sullo sviluppo di tecniche ed algoritmi che aumentassero l'efficacia e l'accuratezza, portando, come si è visto, alla creazione di modelli sempre più complessi. Tuttavia, la relazione tra performance e dimensioni è logaritmica: ad un incremento esponenziale delle computazioni, corrisponde un crescita delle prestazioni lineare. I costanti sforzi nell'ottenere risultati sempre migliori hanno fatto sì che finisse in secondo piano ogni tentativo di ridurre il tempo o l'ammontare di risorse necessari per l'addestramento. Appare dunque fondamentale iniziare a dare maggiore importanza al requisito dell'efficienza, riconoscendo l'impatto ambientale di ciascun modello. Così facendo, sarà forse possibile ideare soluzioni nuove e creative nel campo dell'IA¹⁴⁹.

6) Impatto sulla fiducia

Con la presenza sempre più pervasiva dei sistemi di intelligenza artificiale in ogni ambito della società, sta crescendo la necessità di creare un clima di fiducia attorno alle nuove tecnologie digitali che sia basato sul rispetto dei valori e dei diritti essenziali. In

¹⁴⁶ Azizpour, H. et al. (2020), *The role of artificial intelligence in achieving the Sustainable Development Goals*, Nature Communications, 11(233).

¹⁴⁷ I *Sustainable Development Goals* sono 17 obiettivi inclusi nell'Agenda 2030 per lo Sviluppo Sostenibile delle Nazioni Unite. Ad ogni obiettivo sono associati dei *target* che li sostanziano.

¹⁴⁸ Caine, M., Firth-Butterfield, K., Khatri, S., Kwartler, E. (2021), *What would it take to make AI 'greener'?*, World Economic Forum, online, <https://www.weforum.org/agenda/2021/09/make-ai-greener-climate-solution-cop26-technology/>.

¹⁴⁹ Ivi.

questo modo, sarà possibile sfruttare a pieno il loro potenziale, perseguendo al tempo stesso il benessere dell'umanità.

Secondo l'opinione prevalente all'interno della comunità di ricerca, la fiducia nell'IA è ottenibile solo risolvendo alcune delle problematiche maggiormente rilevanti in questo campo, come quelle relative all'equità, alla trasparenza, alla responsabilità e al controllo esercitato dagli esseri umani¹⁵⁰.

Il tema della responsabilità associata alle azioni degli agenti artificiali è già stato ampiamente trattato¹⁵¹ e perciò non sarà ripreso. Altrettanto vale per la questione sull'equità e sui possibili *bias* a cui sono suscettibili i modelli di machine learning.¹⁵²

Strettamente legata a quest'ultima è la tematica della trasparenza. Un modo per valutare l'eventuale tendenza di un'applicazione di IA a discriminare determinati individui o classi sociali consiste nel capire in base a quali fattori assume le proprie decisioni. Tuttavia, in alcuni casi ciò può dimostrarsi alquanto difficoltoso. Gli attuali modelli basati su reti neurali, le quali possono essere formate da decine o centinaia di *layer* di nodi, presentano un tale livello di complessità che una volta addestrati è estremamente arduo ricostruire il percorso con cui, da una serie di input, arrivano a definire un certo output. Questi sistemi sono definiti "*black box*" per via dell'opacità che caratterizza il loro processo decisionale.

La necessità di una maggiore trasparenza ha portato i ricercatori nel campo dell'intelligenza artificiale a creare una nuova area di studio, denominata *eXplainable AI* (XAI). Essa racchiude una gamma di strumenti e tecniche che consentono di rendere più comprensibili i risultati dei modelli di machine learning, aumentandone dunque il grado di fiducia. Concettualmente, l'XAI può essere orientata verso l'interpretabilità, qualora si tenti di tracciare da un punto di vista quantitativo i meccanismi interni con cui opera il modello, oppure perseguire l'esplicabilità, al fine di formulare giustificazioni che spieghino la logica e gli output dello stesso¹⁵³. In molti casi, però, capire esattamente

¹⁵⁰ European Parliamentary Research Service (2020), *The ethics of artificial intelligence: Issues and initiatives*, Bruxelles, p. 29.

¹⁵¹ Per un approfondimento sul tema della responsabilità, si vada al punto 4).

¹⁵² Per un approfondimento sul tema dell'equità, si vada al punto 1).

¹⁵³ Bonaccorsi di Patti, E. et al. (2022), *Intelligenza artificiale nel credit scoring. Analisi di alcune esperienze nel sistema finanziario italiano*, Banca d'Italia, p. 12.

quali regole governano il funzionamento interno degli algoritmi appare superfluo. Al contrario, la trasparenza dovrebbe dipendere maggiormente dalla spiegazione delle loro scelte e del loro comportamento esterno, opportunamente contestualizzati¹⁵⁴.

Sono diverse le tecniche di *explainable AI*, le quali si suddividono tra quelle in grado di agire su qualsiasi modello e quelle applicabili solamente a specifiche categorie. Allo stesso tempo, numerosi ricercatori lavorano per sviluppare nuove idee, in modo da rendere l'intelligenza artificiale sempre più comprensibile. Basti pensare al progetto NL4AI, il quale ambisce a creare soluzioni capaci di comunicare il processo seguito dalla macchina utilizzando il linguaggio naturale¹⁵⁵. Esiste, tuttavia, un altro approccio per raggiungere l'esplicabilità. Invece che lavorare con modelli complessi e ricorrere poi a *software* e metodi appositi per rivelarne il funzionamento, è possibile utilizzare delle cosiddette "*white box*", ovvero modelli di ML *explainable by design*, anche se spesso meno potenti delle loro controparti¹⁵⁶. Sebbene ci possano essere delle eccezioni, infatti, generalmente il livello di esplicabilità intrinseca dei sistemi di IA tende a diminuire con l'aumentare delle prestazioni, secondo un rapporto di proporzionalità inversa. Un semplice algoritmo basato sulla logica è facilmente interpretabile ma poco potente, mentre un modello di deep learning, come si è già visto, può garantire grandi *performance*, a costo di essere una *black box*. Non è possibile stabilire a priori quale approccio sia il migliore. La scelta tra i due dipende dal compito che si intende far svolgere alla macchina, oltre che dal grado di complessità che lo caratterizza.

Altro aspetto legato alla fiducia nell'intelligenza artificiale è quello del controllo. Questo è collegato in gran parte al timore diffuso che nel futuro queste tecnologie possano evolversi al tal punto da superare in ogni ambito i loro creatori e condurre la razza umana all'estinzione. Sebbene la maggior parte dei ricercatori sostenga che un simile scenario apocalittico non sia plausibile, per creare un clima di fiducia condivisa appare importante fare in modo che gli esseri umani mantengano sui sistemi di IA una forma di

¹⁵⁴ European Parliamentary Research Service (2020), *The ethics of artificial intelligence: Issues and initiatives*, Bruxelles, p. 33.

¹⁵⁵ **CORDIS, Interactive Natural Language Technology for Explainable Artificial Intelligence, consultato il 15.01.2023**, <https://cordis.europa.eu/project/id/860621>.

¹⁵⁶ Giannotti, F., come riportato in Cassauwers, T. (2020), *Opening the 'black box' of artificial intelligence*, Horizon Magazine, online, <https://ec.europa.eu/research-and-innovation/en/horizon-magazine/opening-black-box-artificial-intelligence>.

“supervisione finale”¹⁵⁷. Ciò risulta ancor più vero qualora si abbia a che fare con particolari applicazioni *AI based*, le quali, per via del settore in cui trovano impiego e delle attività che consentono di svolgere, possono effettivamente mettere a repentaglio la sicurezza, i diritti dei consumatori e i diritti dell’individuo. Come riportato nel *White Paper* sull’intelligenza artificiale elaborato dalla Commissione Europea, “*the objective of trustworthy, ethical and human-centric AI can only be achieved by ensuring an appropriate involvement by human beings in relation to high-risk AI applications*”¹⁵⁸. Questo coinvolgimento può però concretizzarsi in diversi modi, a seconda delle modalità di utilizzo dell’intelligenza artificiale e delle conseguenze di eventuali errori per i cittadini e le organizzazioni. L’operatore umano potrebbe, ad esempio, monitorare costantemente il funzionamento del sistema, convalidarne i risultati prima che questi diventino effettivi, oppure inserire delle limitazioni nella fase di progettazione.

Come si è visto, nonostante l’intelligenza artificiale offra grandi opportunità per il benessere della società e delle persone, al tempo stesso essa porta alla luce alcune nuove problematiche e minacce, le quali devono essere affrontate in modo proporzionato e appropriato. La crescita delle prestazioni e delle capacità di questi sistemi, unita alla loro presenza sempre più capillare, porta alla necessità di elaborare un *set* di regole e di principi che ne favoriscano uno sviluppo rispettoso dei valori e dei diritti fondamentali. Allo scopo di indurre la creazione di applicazioni di IA che massimizzino i benefici e prevengano o riducano al minimo i rischi, il *High-Level Expert Group on Artificial Intelligence* (AI HLEG)¹⁵⁹ ha redatto un documento¹⁶⁰ in cui espone le proprie linee guida per una IA affidabile. Con questo, il gruppo di esperti intende costruire un *framework* di partenza trasversale, da adattare in base alle situazioni specifiche, e diretto a tutti i portatori di interessi che orbitano attorno a questo mondo (dalle imprese alle istituzioni governative, dai ricercatori ai consumatori). Partendo dalla consapevolezza che la fiducia nell’intelligenza artificiale non dipende esclusivamente dal

¹⁵⁷ European Parliamentary Research Service (2020), *The ethics of artificial intelligence: Issues and initiatives*, Bruxelles, p. 35.

¹⁵⁸ Commissione Europea (2020), *White Paper on Artificial Intelligence – A European approach to excellence and trust*, COM(2020) 65 final, Bruxelles, p. 21.

¹⁵⁹ Si è già parlato dell’AI HLEG al paragrafo 1.1.

¹⁶⁰ High-Level Expert Group on Artificial Intelligence (2019), *Ethics Guidelines for Trustworthy AI*, Commissione Europea.

sistema in sé, ma anche dal contesto tecnico e sociale in cui è inserito, gli esperti individuano tre componenti dell'IA affidabile:

- legalità, deve ottemperare a tutti i regolamenti e le disposizioni di legge a cui è soggetta;
- eticità, deve essere aderente ai valori e ai principi etici;
- robustezza, sotto una duplice prospettiva tecnica e sociale, affinché possa operare in modo sicuro e senza causare danni, anche non intenzionali.

Ogni elemento è considerato necessario, ma non sufficiente per la realizzazione di una *trustworthy AI*, e va perseguito durante l'intero ciclo di vita del sistema. Sulla base di questi, e in particolare dell'eticità e della robustezza (le linee guida non trattano esplicitamente la componente della legalità), l'AI HLEG elenca quattro principi etici, fortemente radicati nei diritti essenziali: rispetto dell'autonomia umana; prevenzione dei danni; equità; esplicabilità. Questi costituiscono le fondamenta per un'IA affidabile e tutti gli operatori sono chiamati ad aderirvi. I quattro principi sono poi tradotti in sette requisiti concreti, attuabili attraverso diversi metodi tecnici e non: intervento e sorveglianza umani; robustezza tecnica e sicurezza; riservatezza e *governance* dei dati; trasparenza; diversità, non discriminazione ed equità; benessere sociale e ambientale; *accountability*. La loro implementazione può variare a seconda dell'applicazione considerata e dovrebbe avvenire durante tutte le fasi della vita di un sistema di IA. Nell'intenzione degli autori, le linee guida non costituiscono un orientamento statico e ancorato nel tempo, bensì dovrebbero essere costantemente riviste ed aggiornate, seguendo l'evoluzione della tecnologia, della conoscenza e della società.

Oltre al documento dell'AI HLEG, un altro lavoro, questa volta di tipo legislativo, che vale la pena menzionare è la proposta di "Regolamento sull'IA" della Commissione Europea del 21 aprile 2021¹⁶¹, noto come "AI Act". La proposta punta ad armonizzare le norme sull'intelligenza artificiale all'interno dell'Unione Europea, allo scopo di rafforzare la fiducia dei cittadini nelle soluzioni di IA e favorire gli investimenti e l'innovazione,

¹⁶¹ Commissione Europea (2021), *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*, COM(2021) 206 final, disponibile a https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF.

rendendo così l'UE un polo mondiale per queste tecnologie¹⁶². Essa distingue i sistemi di IA in base ad un approccio basato sul rischio per la sicurezza e i diritti fondamentali, suddividendoli in tre livelli:

- utilizzi vietati, il cui rischio è ritenuto inaccettabile. Ne sono un esempio i sistemi di riconoscimento facciale di massa utilizzati dalla polizia e dalle autorità giudiziarie in luoghi pubblici;
- sistemi a rischio elevato, come ad esempio le applicazioni per l'analisi e il *ranking* dei CV, i quali sono soggetti a controlli, obblighi e limitazioni;
- sistemi a rischio limitato o minimo, tra cui i *chatbot*, per i quali c'è solamente un incoraggiamento ad aderire a codici di condotta volontari.

Volendo concludere, l'intelligenza artificiale rappresenta sicuramente una delle più importanti innovazioni tecnologiche della storia. Il suo potenziale è ancora in parte inespresso, ma già ora si possono notare gli effetti che sta avendo sulla società e sulla vita delle persone. Affinché il suo utilizzo possa portare ad una nuova epoca di prosperità, per usare le parole dell'AI HLEG, "*AI systems need to be human-centric, resting on a commitment to their use in the service of humanity and the common good, with the goal of improving human welfare and freedom*"¹⁶³.

¹⁶² Battelli, E. (2022), "Necessità di un umanesimo tecnologico: sistemi di intelligenza artificiale e diritti della persona", in Buzzelli, D., Palazzo, M. (a cura di), *Intelligenza artificiale e diritti della persona*, Pacini Editore, Pisa, p. 102.

¹⁶³ High-Level Expert Group on Artificial Intelligence (2019), *Ethics Guidelines for Trustworthy AI*, Commissione Europea, p. 4.

Conclusione

Bibliografia

Albarelli, A., Bagnoli, C., Campostrini, S., Massaro, M., Muraro, A., Toniolo, K., Vesnic, L., Zantedeschi, M. S. (2019), Gli impatti di IA e di Blockchain sui modelli di business. Il Report di Ricerca, Strategy Innovation, Venezia.

Barto, A.G., Sutton, R.S. (2018), Reinforcement Learning: An Introduction, 2nd ed., MIT Press, Cambridge.

Buchanan, B. G. (2005), "A (Very) Brief History of Artificial Intelligence", AI Magazine, 26(4), pp. 53-60.

Davenport, T. H. (2019), The AI Advantage. How to Put the Artificial Intelligence Revolution to Work, MIT Press, Cambridge.

Fanti, L., Guarascio, D., Moggi, M. (2022), "From Heron of Alexandria to Amazon's Alexa: a stylized history of AI and its impact on business models, organization and work", Journal of Industrial and Business Economics, 49, pp. 409-440.

High-Level Expert Group on Artificial Intelligence (2019), A definition of AI: Main capabilities and scientific disciplines, Commissione Europea, Bruxelles.

Kaplan, A., Haenlein, M. (2019), "Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence", Business Horizons, 62.

Khayyam, H., Javadi, B., Jalili, M., Jazar, R.N. (2020), "Artificial Intelligence and Internet of Things for Autonomous Vehicles", in Dai, L., Jazar, R. (a cura di), Nonlinear Approaches in Engineering Applications, pp. 39-68, Springer.

KPMG, TRENDONE (2018), Rethinking the value chain. A study on AI, humanoids and robots, disponibile a

<https://assets.kpmg/content/dam/kpmg/xx/pdf/2018/09/rethinking-the-value-chain.pdf>.

Lecciso, F., Liverta Sempio, O., Marchetti, A. (2005), Teoria della mente. Tra normalità e patologia, Raffaello Cortina Editore, Milano.

Malone, T.W., Rus, D., Laubacher, R. (2020), Artificial Intelligence and the Future of Work, Massachusetts Institute of Technology, Cambridge, disponibile a <https://workofthefuture.mit.edu/research-post/artificial-intelligence-and-the-future-of-work/>.

McCarthy, J. (2007), What is artificial intelligence?, Stanford University, Stanford.

McKinsey (2020), An executive's guide to AI, disponibile a <https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/An%20executives%20guide%20to%20AI/Executives-guide-to-AI>.

Nilsson, N. J. (2009), The Quest for Artificial Intelligence: A History of Ideas and Achievements, Cambridge University Press, Cambridge.

OECD (2019), Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449.

Russell, S., Norvig, P. (2020), Artificial Intelligence: A Modern Approach, 4th ed., Pearson Education, Hoboken.

Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., and Delipetrev, B. (2020), AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence, EUR 30117 EN, Publications Office of the European Union, Luxembourg

Samuel, A.L. (1959), "Some Studies in Machine Learning Using the Game of Checkers", IBM Journal of Research and Development, 3, pp. 210-229.

Turing, A. M. (1950), "Computing Machinery and Intelligence", Mind, 59, pp. 433-460.

Winston, P. (1992), Artificial Intelligence, Addison-Wesley, Reading.

World Economic Forum (2019), Empowering AI Leadership. An Oversight Toolkit for Boards of Directors.

World Economic Forum (2022), Empowering AI Leadership: AI C-Suite Toolkit.

Sitografia