

# Coscienza Artificiale: implicazioni per l'umanità

Giorgio Buttazzo

## Sommario

*Il primo numero di Mondo Digitale ha affrontato il tema della coscienza artificiale nelle macchine. Oggi, alla luce dei più recenti progressi dell'intelligenza artificiale, cosa è cambiato? Il momento in cui le macchine raggiungeranno le capacità intellettive umane è sempre più vicino? Cosa succederà dopo? Cosa cambierà nella società? Quali sono i lavori che l'intelligenza artificiale mette a rischio? Ragionare su questi interrogativi è di fondamentale importanza al fine di prevedere i possibili scenari futuri e prepararsi ad affrontare le conseguenze.*

## Abstract

*The first issue of Mondo Digitale dealt with the theme of artificial consciousness in machines. Today, in the light of the most recent advances in artificial intelligence, what has changed? Is the moment in which machines will reach human intellectual capacities getting closer? What will happen next? What will change in society? What are the jobs that artificial intelligence puts at risk? Reasoning about these questions is of fundamental importance to predict possible future scenarios and prepare to face the consequences.*

**Keywords:** Artificial Consciousness, Artificial General Intelligence, Singularity, Roboethics, Post-humanism.

## 1. Introduzione. IA, automazione e disoccupazione

L'articolo intitolato "Coscienza Artificiale: Missione Impossibile?", pubblicato su Mondo Digitale a marzo 2002 sosteneva una linea di pensiero mirata a mostrare che non esistono ostacoli concettuali allo sviluppo di macchine coscienti, osservando che tutte le critiche mosse nel corso degli anni volte a minare questa possibilità possono essere facilmente smontate con semplici ragionamenti logici ed evidenze scientifiche. Eliminati gli ostacoli concettuali, l'articolo illustrava come poter effettuare una previsione su quando si possano creare le condizioni necessarie per lo sviluppo di una macchina cosciente, indicando come possibile data l'anno 2029.

Dopo vent'anni e alla luce dei più recenti progressi dell'intelligenza artificiale, questa stima è ancora realistica? L'evento stravolgente indicato come "singolarità tecnologica" in cui le macchine raggiungeranno le capacità intellettive umane è

sempre più vicino? Cosa succederà dopo? Cosa cambierà nella società? Quali sono i lavori che l'intelligenza artificiale mette a rischio? Ragionare su questi interrogativi è di fondamentale importanza per prevedere i possibili scenari futuri e prepararsi ad affrontare le conseguenze.

Questo articolo ripercorre le tappe principali che hanno portato allo sviluppo del deep learning, illustrando le capacità attuali dei modelli neurali esistenti e discutendo i possibili benefici e rischi di questa tecnologia, sia a breve che a lungo termine.

## 2. Evoluzione delle reti neurali

L'intelligenza artificiale e, in particolare, le reti neurali artificiali hanno avuto un percorso altalenante in cui si sono alternati momenti di eccessivo entusiasmo con momenti di scarso interesse. Il primo modello di neurone artificiale, noto come neurone binario a soglia, fu proposto nel 1943 da McCulloch e Pitts [20]. Tale modello prevede  $n$  canali di ingresso, ciascuno caratterizzato da un peso che ne modula il valore. Il neurone integra gli  $n$  valori di ingresso attraverso una somma pesata, producendo un unico valore di uscita, che vale 1 se la somma è maggiore di una certa soglia, 0 altrimenti. Tale modello non è in grado di apprendere e fu utilizzato prevalentemente per simulare il comportamento di semplici circuiti neurali biologici. Nel 1949, lo psicologo canadese Donald Hebb [12] fece una scoperta rivoluzionaria, osservando che il processo di apprendimento non modifica il funzionamento delle cellule nervose, ma opera unicamente sulle connessioni sinaptiche, le quali modulano la comunicazione tra i neuroni. Egli scrisse:

*“Quando un assone di una cellula A è abbastanza vicino da eccitare una cellula B e partecipa ripetutamente alla sua attivazione, si osservano alcuni processi di crescita o cambiamenti metabolici in una o entrambe le cellule tali da aumentare l'efficacia di A nell'attivare B.”*

La scoperta di Hebb fu sfruttata dallo psicologo statunitense Frank Rosenblatt, il quale nel 1957 sviluppò il primo modello di neurone artificiale in grado di apprendere, il Perceptron [25]. A differenza del neurone binario a soglia, il Perceptron ha dei pesi “variabili” che possono essere modificati in funzione dell'errore commesso dal neurone, affinché esso “impari” ad associare un insieme di ingressi con dei valori di uscita desiderati. Ad esempio, collegando gli ingressi del Perceptron a 400 fotocellule disposte come una matrice di 20x20 pixel, Rosenblatt riuscì ad addestrarlo a riconoscere le forme concave da quelle convesse.

Tali risultati crearono grosse aspettative nei ricercatori dell'epoca sulle potenzialità dell'apprendimento automatico. Addirittura, lo stesso Rosenblatt nel 1958 rilasciò un'intervista al New York Times [22] presentando il Perceptron come "l'embrione di un computer elettronico che in un prossimo futuro sarebbe stato in grado di vedere, parlare, scrivere, camminare, riprodursi ed essere consapevole della propria esistenza".

Purtroppo, l'entusiasmo di Rosenblatt svanì nel 1969, quando due matematici del Massachusetts Institute of Technology (MIT), Marvin Minsky e Seymour Papert, pubblicarono un libro intitolato *Perceptrons* [21], in cui dimostrarono, attraverso un controesempio, l'impossibilità per un Perceptron di apprendere la semplice funzione logica di un OR esclusivo (XOR) a due ingressi, che prevede un'uscita uguale a zero quando i due ingressi sono uguali e uguale a 1 quando sono diversi. Questo risultato negativo sul Perceptron fece crollare l'interesse per le reti neurali per oltre un decennio, periodo che oggi viene indicato come l'inverno dell'intelligenza artificiale (AI winter).

L'interesse per le reti neurali riprese agli inizi degli anni '80, quando John Hopfield [13] propose un nuovo modello di rete in grado di comportarsi come una memoria associativa. Negli stessi anni, Andrew Barto, Richard Sutton e Charles Anderson [3] avevano sviluppato un nuovo paradigma di apprendimento basato su "premi" e punizioni", denominato Reinforcement Learning, mentre Teuvo Kohonen [16] aveva ideato una rete neurale capace di auto-organizzarsi, senza supervisione esterna, per formare delle mappe sensoriali simili a quelle esistenti nella corteccia somatosensoriale. Il culmine si ebbe nel 1986, quando David Rumelhart, Geoffrey Hinton e Ronald Williams [26] svilupparono un potente algoritmo di apprendimento supervisionato, noto come Backpropagation, che permette ad una rete neurale di imparare ad associare degli ingressi con delle uscite desiderate attraverso un insieme di esempi (training set).

Grazie a questi risultati, nei vent'anni successivi alla nascita della Backpropagation, le reti neurali sono state utilizzate per risolvere diverse tipologie di problemi, tra cui il riconoscimento di immagini, la compressione di dati, la previsione di segnali e il controllo adattivo, nei più disparati settori, quali fisica, chimica, ingegneria, robotica, geologia, agraria, astronomia, economia, medicina, scienze sociali, psicologia, ecc.

Nonostante l'esplosione dei campi applicativi, fino alla fine del XX secolo, non vi furono sostanziali progressi teorici sulle reti neurali. Molti ricercatori provarono a sviluppare modelli più complessi, più vicini alla controparte biologica, senza però riuscire ad ottenere prestazioni sensibilmente migliori rispetto ai modelli precedenti. Altri provarono ad aumentare il numero di strati di una rete neurale, ma si scontrarono con grosse difficoltà nell'addestrare reti con più di quattro strati. Pertanto la ricerca entrò in un secondo periodo invernale, che si concluse solo agli inizi del nuovo secolo.

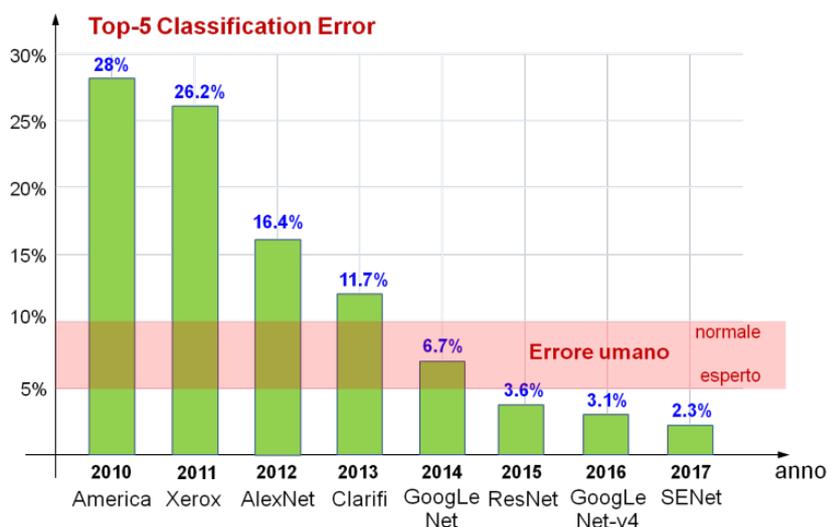
### 3. Esplosione del deep learning

Agli inizi degli anni 2000, la ricerca sulle reti neurali ha avuto una grossa impennata grazie a tre importanti fattori. Il primo fattore è di natura teorica: una volta compresi i problemi che ostacolavano l'addestramento di reti a molti strati, sono state ideate diverse soluzioni per superare quei limiti e gestire l'apprendimento di reti costituite da migliaia di neuroni organizzati su numerosi strati: le deep neural network. Il secondo fattore è di natura tecnologica. Intorno al 2006, le architetture di calcolo basate sulle Graphics Processing Unit (GPU),

originariamente progettate per parallelizzare le operazioni grafiche, sono state modificate per poter svolgere anche calcoli vettoriali, come quelli richiesti per eseguire una rete neurale, e si sono diffuse sul mercato ad un costo accessibile. Il terzo fattore è invece di natura economica. I primi risultati ottenuti con le deep neural network hanno attratto l'interesse di grosse aziende, come Google, Microsoft e Facebook che, gestendo un'enorme quantità di dati, hanno visto nelle reti neurali una grossa opportunità per risolvere problemi di classificazione di immagini, riconoscimento di volti, suoni, voci, e hanno quindi cominciato ad investire grosse quantità di risorse in questo settore.

Infine, un altro elemento che ha contribuito all'evoluzione delle deep network è stata la competizione internazionale ImageNet, o più precisamente la "ImageNet Large-Scale Visual Recognition Challenge" (ILSVRC) [14], una sorta di olimpiade annuale della computer vision, nata nel 2010 per stimolare lo sviluppo di algoritmi per la soluzione di problemi complessi, come la classificazione e la segmentazione di immagini.

La Figura 1 illustra come si è ridotto negli anni l'errore di classificazione degli algoritmi che hanno vinto la competizione, dal 2010 al 2017.



**Fig. 1**  
Diminuzione dell'errore di classificazione dal 2010 al 2017 nella competizione ImageNet.

Fino al 2011, la competizione era stata dominata da soluzioni di tipo algoritmico, mentre nel 2012, per la prima volta, è stata vinta da una rete neurale, AlexNet, sviluppata da un gruppo di ricerca dell'università di Toronto coordinato da Geoffrey Hinton. Come si può osservare dal grafico, la soluzione neurale ha ottenuto un enorme miglioramento rispetto alle soluzioni precedenti, riducendo l'errore di circa il 10% rispetto all'anno precedente. Questo risultato ha attratto l'interesse di grosse aziende, come Google, Microsoft e Facebook, le quali hanno cominciato a disvestire in questo settore contribuendo ad accelerare il miglioramento delle prestazioni. Infatti, nel 2014 è proprio una rete sviluppata da

Google, GoogLeNet, a vincere la competizione raggiungendo prestazioni simili a quelle umane, caratterizzate da un errore compreso tra il 5% (relativo ad esperti) e il 10% (relativo a non esperti). Negli anni successivi, l'errore ottenuto dalle reti per la classificazione di immagini ha continuato a ridursi progressivamente, fino a raggiungere il 2.3% nel 2017. Pertanto, a partire dal 2018, il comitato organizzatore ha deciso di focalizzare la competizione su attività più complesse, quali la localizzazione di oggetti e la segmentazione di immagini.

Le attività in cui le macchine superano le prestazioni umane cresce di anno in anno. Una delle prime sconfitte umane risale al 1997, quando un computer dell'IBM, Deep Blue, sconfisse il campione del mondo di scacchi Gary Kasparov per 3.5 a 2.5. Nel 2017, è stata la volta del campione del mondo di Go, Ke Jie, sconfitto da un computer di Google Deep Mind, AlphaGo Zero, basato su un algoritmo di reinforcement learning.

Nella produzione industriale, i robot hanno da tempo superato le capacità umane in velocità e precisione, liberando altresì l'uomo da lavori ripetitivi e alienanti. Anche in avionica, il controllo di aerei militari ad elevata instabilità non sarebbe possibile per un umano senza l'ausilio dei computer.

Nel campo medico, le reti neurali hanno dimostrato prestazioni superiori a quelle umane nella lettura dell'elettrocardiogramma per il rilevamento di malattie cardiache [23], nella diagnosi del cancro al polmone [6], nell'identificazione di malattie della retina [7], nel rilevamento di formazioni cancerogene sulla pelle [10], e nell'identificazione di alterazioni del cervello dovute all'Alzheimer [9].

Anche nel campo artistico, le macchine cominciano ad eguagliare le capacità umane. Un computer neurale, AIVA, sviluppato all'Università di Vancouver [1] è stato addestrato con brani di Mozart, Beethoven e Bach, ed è ora in grado di comporre musica classica e colonne sonore di elevata qualità, indistinguibili da quelle composte da un musicista umano.

Negli ultimi anni, anche la comprensione del linguaggio naturale ha raggiunto livelli elevatissimi, paragonabili a quelli umani. Si pensi alla qualità delle traduzioni automatiche di Google Translate, al livello di comprensione del parlato degli smart phone, alla generazione automatica dei sottotitoli nei filmati di Youtube, o alle trascrizioni automatiche effettuate da Microsoft Teams o da altri sistemi per la gestione di riunioni e conferenze online.

Una deep network denominata LipNet [2] è in grado di decifrare i movimenti delle labbra con una precisione del 95%, contro il 55% di un umano esperto. Synthesia, una startup fondata nel 2017 da giovani ricercatori di varie università, ha creato una piattaforma online [28] per la generazione automatica di presentazioni video mediante un avatar sintetico in 120 lingue. L'utente inserisce un testo e il sistema genera una presentazione con un avatar realistico che pronuncia il testo replicando espressioni del volto e movimenti delle labbra.

Ancora più recentemente, i chatbot creati da Google e OpenAI, come LaMDA e ChatGPT, hanno creato molto scompiglio a livello scientifico e mediatico per la qualità delle risposte generate. Un chatbot è un'applicazione software che utilizza

metodi di machine learning per estrapolare un modello statistico del linguaggio naturale dall'enorme quantità di testo disponibile su web, per poi utilizzare questo modello per condurre una conversazione online tramite testo o sintesi vocale. L'11 giugno 2022, Blake Lemoine, un ingegnere di Google addetto al test di LaMDA, dopo aver analizzato alcune risposte a domande riguardanti l'identità personale, i valori morali, la religione e le tre leggi della robotica di Asimov, ha affermato in un'intervista che il chatbot era senziente.

ChatGPT, un chatbot sviluppato da OpenAI e messo online nel novembre 2022, ha rapidamente attirato l'attenzione dei media per le sue risposte articolate alle domande degli utenti su qualsiasi argomento, la composizione di poesie e musica originali, la generazione di codice software in vari linguaggi di programmazione, e persino la pianificazione di progetti complessi.

Come si può facilmente notare dagli esempi sopra riportati, le capacità dell'intelligenza artificiale non solo migliorano di anno in anno, ma coprono settori sempre più vasti e attività sempre più complesse. Pertanto, diventa sempre più rilevante chiedersi quali siano i lavori potenzialmente a rischio nel prossimo futuro.

#### 4. Lavori a rischio

Come si è già osservato, nella produzione industriale i robot hanno da tempo superato le capacità umane in velocità e precisione, rimpiazzando tutti i lavori relativi al montaggio, riconoscimento e movimentazione di pezzi, verniciatura, controllo qualità, imballaggio ecc. Tuttavia, a differenza della rivoluzione industriale, che ha rimpiazzato i muscoli con attuatori idraulici o elettrici, la rivoluzione informatica basata sull'intelligenza artificiale sta automatizzando le attività mentali, mettendo a rischio molti lavori precedentemente ritenuti al sicuro dall'automazione.

Da giugno 2022, lo stato della California ha consentito la circolazione di taxi senza conducente umano al volante per il trasporto di passeggeri. Con l'avvento dei veicoli a guida autonoma il passo successivo sarà quello di automatizzare il trasporto su gomma, che avrà il beneficio di ridurre considerevolmente i costi, i consumi, l'inquinamento, il numero di incidenti, e i tempi di consegna.

Oltre a queste categorie, molti esperti concordano che entro il 2030 le macchine saranno in grado di svolgere lavori come il bigliettaio, il benzinaio, il cassiere, il traduttore, il barman, l'operatore telefonico. Mentre, entro il 2040 le macchine saranno in grado di svolgere attività come l'operatore economico, il bancario, lo scrittore, il pittore, lo scultore, il musicista, il commercialista, l'avvocato, il medico e il chirurgo. Già oggi, l'intelligenza artificiale sta avendo un profondo effetto sulle pratiche legali. L'enorme capacità di elaborare e incrociare i dati, comprendere il testo, cercare leggi e sentenze dagli archivi informatici, rende una macchina molto più veloce, affidabile e meno costosa di molti avvocati umani, che pertanto sono diventati una categoria a rischio di estinzione nel prossimo futuro. Più recentemente, l'intelligenza artificiale ha iniziato ad essere utilizzata anche per redigere contratti, prevedere esiti legali e persino raccomandare decisioni giudiziarie in merito a condanne o cauzioni [27].

Un'altra nuova categoria a rischio, impensabile fino a pochi anni fa, sono i medici. Come si è già visto, già da qualche anno, l'intelligenza artificiale supera l'affidabilità dei medici umani in diversi tipi di diagnosi. Inoltre, la capacità di una macchina intelligente di poter accedere e incrociare i dati di milioni di pazienti è di gran lunga al di fuori della portata della mente umana. Il rischio riguarda anche i chirurghi, visto che già oggi la robotica è largamente utilizzata in alcuni tipi di operazioni in cui è richiesta un'elevata precisione (ad esempio, sul ginocchio, cervello, e prostata). Oggi i robot sono teleoperati da un chirurgo umano, ma presto la tecnologia permetterà di prendere decisioni più accurate di quelle umane e tali robot diventeranno autonomi, aumentando le probabilità di successo, accelerando i tempi di intervento e di guarigione dei pazienti.

Un'altra categoria a rischio fino a poco tempo fa ritenuta a prova di futuro è quella dei programmatori software. Infatti, la capacità dimostrata da ChatGPT di generare codice in modo rapido e affidabile, in qualsiasi linguaggio di programmazione, rende l'uso dell'intelligenza artificiale molto appetibile per le aziende che producono software, dal punto di vista dell'affidabilità, dei tempi e dei costi di sviluppo.

La scomparsa dei lavori nella società è un fenomeno insito allo sviluppo di nuove tecnologie. Se solo consideriamo i lavori che esistevano nel secolo scorso, ci accorgiamo che molti di essi oggi non esistono più, come lo svegliatore, il raddrizzatore di birilli, il tagliatore di ghiaccio, il lampionaio, il derattizzatore, il lattaio, lo stagnino, la mondina, il lustrascarpe, il centralinista, o il telegrafista. Se la tecnologia fa scomparire dei lavori, è anche vero che essa ne crea di nuovi e in maggior numero. Si pensi ai lavori che non esistevano nel passato, quali ad esempio il fotografo, il cameraman, il regista, lo sceneggiatore, l'elettricista, il macchinista di treni, il conducente di autobus, il tassista, il pilota di aerei, la hostess o lo steward, il radiotecnico, il radiologo, l'ecografista, o l'astronauta. Oppure, professioni ancora più recenti come l'ingegnere informatico, l'ingegnere elettronico, lo sviluppatore di video game, lo sviluppatore di app, l'operatore di droni, il tecnico di stampanti 3D, il social media manager, il web designer, lo Youtuber, il blogger o analista di dati.

Il vero problema dell'intelligenza artificiale non è quello di rimpiazzare molti lavori e professioni oggi esistenti, ma è dovuto alla velocità con cui ciò avverrà. Per tale ragione, è di fondamentale importanza riuscire a prevedere cosa succederà nei prossimi 10 o 20 anni, in modo che sia possibile attenuare le conseguenze con delle opportune strategie politiche per gestire la transizione.

## 5. Problemi etici

La diffusione delle macchine autonome intelligenti nella società e a stretto contatto con l'uomo pone anche problemi etici non trascurabili. Ad esempio: di chi è la responsabilità se un essere umano viene ferito da un robot a causa di un'azione programmata o a causa di un mancato intervento?

È possibile immaginare diverse situazioni in cui una decisione di un robot potrebbe favorire alcuni esseri umani e danneggiarne altri. Si consideri ad

esempio un'auto senza conducente che si trova davanti a sé un pedone che attraversa la strada senza preavviso e che, per evitarlo, potrebbe solo sterzare invadendo l'altra corsia, sulla quale però andrebbe ad investire un veicolo con più occupanti che viaggia nella direzione opposta. È evidente che la decisione da prendere ha delle forti implicazioni etiche. È quindi fondamentale, fornire ad una macchina autonoma le informazioni necessarie per poter prendere decisioni che siano corrette non solo dal punto di vista tecnico, ma anche da quello etico.

L'etica è un insieme di criteri e di valori definiti per regolare il comportamento dell'uomo in relazione agli altri o all'ambiente, giudicando le azioni rispetto al bene e al male che essere provocano sugli altri esseri viventi o sull'ambiente. Tuttavia, quando l'ambiente comprende anche i robot, è necessario regolare sia il comportamento dei robot verso gli umani e l'ambiente, che il comportamento degli umani verso i robot e l'ambiente. Ma è possibile fornire delle regole etiche ad un robot?

Negli anni '40, Isaac Asimov, famoso scrittore di fantascienza sui robot, formulò delle leggi etiche per i robot, così enunciandole:

- 1. Un robot non può recar danno ad un essere umano, né permettere che, a causa di un suo mancato intervento, un essere umano possa subire un danno.*
- 2. Un robot deve obbedire agli ordini degli esseri umani, purché essi non siano in conflitto con la prima legge.*
- 3. Un robot deve proteggere la propria esistenza purché ciò non sia in conflitto con la prima o con la seconda legge.*

Tali leggi, in apparenza chiare e univoche, contengono molte ambiguità. Ad esempio, il concetto di danno è legato al concetto di male (non solo fisico), che è ancora più ambiguo. Asimov, che era cosciente di tali problemi, in seguito aggiunse la Legge Zero (di maggior importanza rispetto alle altre):

1. Un robot non può recar danno all'umanità, né permettere che, a causa di un suo mancato intervento, l'umanità possa subire un danno.

Se un folle minacciasse di distruggere l'umanità, la Legge Zero autorizzerebbe il robot ad eliminarlo, cioè lo autorizzerebbe ad infrangere la Prima Legge.

Si apre qui il problema della valutazione quantitativa dei danni: l'uccisione di molti è più grave dell'uccisione di un singolo. Viceversa, il salvataggio di molti è meglio che il salvataggio di un singolo. Tuttavia, codificare tali regole in una macchina potrebbe causare grossi problemi in alcune situazioni. Ad esempio, un robot programmato per massimizzare la felicità umana, potrebbe decidere di prelevare gli organi da un uomo sano per salvarne cinque!

Si deve concludere che i problemi etici sono troppo complessi e non possono essere condensati in regole sintetiche. Inoltre, l'ambiguità del linguaggio naturale

può portare ad interpretazioni non previste, creando situazioni ad elevato rischio per la vita umana.

Questi tipi di problemi hanno portato alla nascita di nuovo campo di ricerca noto come "Etica delle macchine", o "Roboetica", il cui obiettivo è quello di fornire alle macchine gli strumenti per prendere decisioni appropriate. Ma ciò porta inevitabilmente a porsi nuove domande: *Come istruire un robot per prendere decisioni eticamente corrette? Possiamo fidarci di tali robot?*

La creazione di robot etici richiede una forte interazione tra informatici e filosofi, ma ad oggi la creazione di un robot etico vede due approcci possibili. Il primo, di tipo algoritmico, consiste nel codificare delle regole in un programma finalizzato a massimizzare una funzione di costo, ad esempio il beneficio per l'uomo. Il secondo, di tipo neurale, consiste nell'utilizzare tecniche di apprendimento automatico affinché il robot impari da numerosi esempi.

Purtroppo, entrambi gli approcci hanno dei punti deboli. In un approccio algoritmico, le regole morali sono vaghe ed è rischioso imporle in tutte le situazioni. È possibile immaginare tante situazioni e contro-esempi in cui inserire delle eccezioni a tali regole. Inoltre, l'intero sistema potrebbe bloccarsi quando incontra paradossi o conflitti irrisolvibili. D'altra parte, l'approccio neurale ha problemi diversi. Ciò che viene appreso è codificato in milioni di parametri, per cui non è facile capire cosa il robot abbia effettivamente imparato. Inoltre, l'apprendimento è guidato da un insieme di esempi che potrebbero contenere delle inclinazioni o pregiudizi sociali.

Anche la regola più saggia può essere pericolosa, se applicata senza considerare la situazione particolare o il contesto in cui essa deve operare. Ad esempio, codificare in un'auto autonoma la regola stradale che non si deve in nessun caso attraversare la doppia linea di separazione tra due corsie può essere deleteria nel caso in cui l'unico modo di scansare un pedone sia quello di invadere la corsia opposta, anche in presenza di una doppia linea continua.

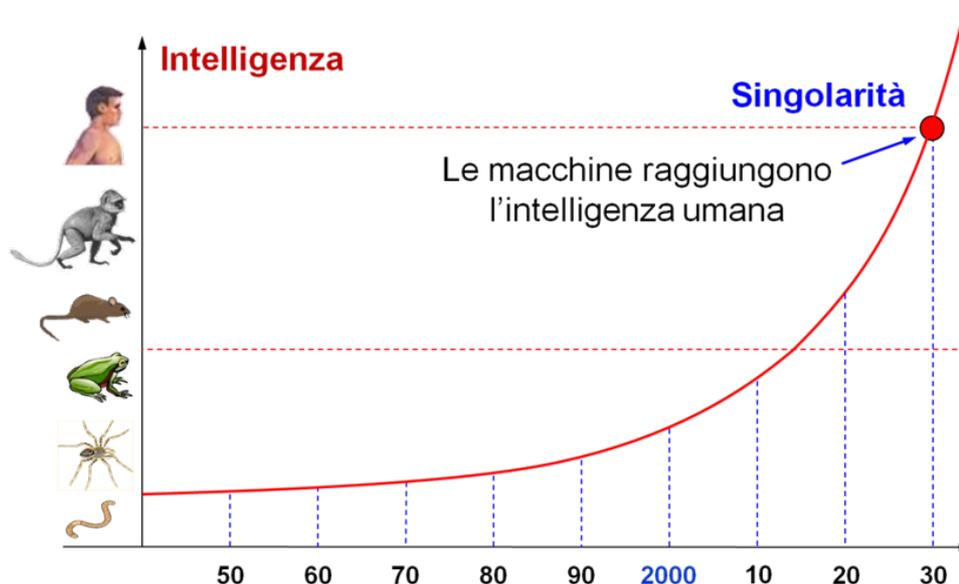
Ogni persona di buon senso sa quando trasgredire la legge, al fine di onorare lo spirito della legge. Quindi, una delle maggiori sfide del futuro è quella di insegnare ai robot gli elementi del buon senso. I veicoli autonomi tuttavia devono affrontare una sfida ancora più difficile: devono decidere in fretta, con informazioni incomplete e in situazioni che i programmatori spesso non hanno considerato.

## 6. Verso la singolarità

Molti futurologi ed esperti di robotica e intelligenza artificiale [17] [18] [15] [4] [19] concordano sul fatto che il progresso della tecnologia segue un processo di natura esponenziale, descritto da una curva che varia molto lentamente nella fase iniziale, ma che continua a crescere a velocità sempre maggiore sulla base di un fattore moltiplicativo costante.

Un tipico esempio di legge esponenziale è quella che descrive la crescita del numero di transistor in un circuito elettronico, nota come Legge di Moore, in quanto nel 1965 Gordon Moore, cofondatore di Intel, osservando l'evoluzione dei chip elettronici, ipotizzò che il numero di transistor in un circuito elettronico sarebbe raddoppiato ogni 12 mesi. Alla fine degli anni ottanta, la legge venne corretta allungando il periodo di raddoppio a 18 mesi ed è a oggi ancora valida.

Sulla base dell'evoluzione esponenziale dell'intelligenza artificiale osservata finora, ipotizzando che la crescita delle capacità cognitive continui con lo stesso ritmo nel prossimo futuro, è possibile prevedere che la singolarità sarà raggiunta intorno agli anni 2030, come illustrato in Figura 2. È importante osservare che tale evento si riferisce al momento in cui le macchine eguagliano l'intelligenza umana non solo in alcuni campi specifici, ma in tutte le attività umane. Questo tipo di intelligenza artificiale viene definita come Artificial General Intelligence (AGI), ad indicare un'abilità intellettuale che si manifesta in tutti i campi della conoscenza.



**Fig. 2**

Crescita esponenziale dell'intelligenza artificiale negli anni. Per un confronto con i sistemi biologici, l'asse delle ordinate riporta i livelli di intelligenza di alcuni esseri viventi.

Immaginare una macchina con la stessa intelligenza umana è facile, perché possiamo prendere come riferimento le capacità intellettive di una persona adulta. Tuttavia, il paragone non è del tutto calzante, perché occorre considerare che una macchina con quelle capacità è in grado di operare a velocità molto più elevate di un essere umano, ed è quindi già superiore ad un essere umano, come velocità di elaborazione, capacità di memoria, capacità di acquisire dati sensoriali e accedere ad altri dati disponibili in rete.

Inoltre, un altro vantaggio di una tale macchina intelligente è quello di poter comunicare velocemente con altri sistemi simili per condividere la propria conoscenza, imparando quindi ancora più velocemente dall'esperienza di altre

macchine. Questo tipo di apprendimento collettivo avviene anche tra esseri umani, ma è estremamente più lento.

Si osservi che il raggiungimento dell'intelligenza umana da parte di una macchina non implica necessariamente lo sviluppo di una coscienza artificiale. Tuttavia, non si può escludere a priori che un sistema complesso dotato di capacità evolute di linguaggio, percezione, ragionamento e apprendimento, possa sviluppare una consapevolezza di sé. Se ciò dovesse accadere, come possiamo verificare che una macchina sia effettivamente cosciente? Esiste un test per misurare il livello di consapevolezza di un essere pensante?

Sebbene l'intelligenza sia una capacità che si manifesta esternamente mediante delle azioni ed è quindi misurabile per mezzo di test specifici, determinare la presenza di una coscienza nella mente di un essere pensante è una questione più delicata. Infatti, la consapevolezza di sé è qualcosa che può essere osservata solo da chi la possiede. E poiché non possiamo entrare nella mente di un altro essere, non è chiaro come si possa definire una procedura per determinare se una mente abbia una reale consapevolezza di sé o stia fingendo di averla, come giustamente osservato da Hofstadter e Dennett nel libro "L'io della mente" [8]. Dunque, se in una intelligenza artificiale evoluta dovesse accendersi la scintilla della coscienza, al momento non sembra esistere un modo per poterlo verificare.

Potremo tuttavia seguire un approccio meno filosofico e più pragmatico, simile a quello proposto da Alan Turing per verificare l'intelligenza di una macchina, secondo cui un essere artificiale potrebbe essere considerato autocosciente se fosse in grado di convincerci, superando delle prove specifiche. E se questa proprietà venisse sviluppata, cosa potrebbe succedere dopo? Quali sarebbero le implicazioni per l'umanità?

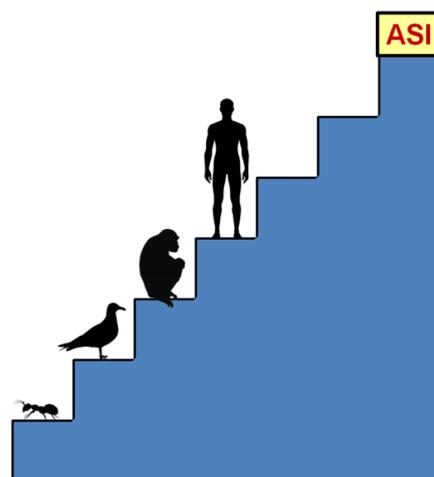
## 7. Il sorpasso

La singolarità non è altro che una tappa temporale sul cammino evolutivo delle macchine, un semplice cartello che indica che l'intelligenza artificiale ha eguagliato l'intelligenza umana. Ma nel momento in cui le macchine raggiungeranno la capacità intellettuale dell'uomo, esse saranno già un passo più avanti, in quanto dotate di una maggiore velocità di elaborazione, una maggiore capacità di memoria e avranno la possibilità di accedere rapidamente a tutti i dati disponibili in rete. Raggiunta la singolarità, l'intelligenza artificiale non si fermerà di certo a quel livello, ma continuerà ad evolvere con un andamento esponenziale, se non addirittura ad un ritmo maggiore, verso quella che viene denominata Artificial Superintelligence (ASI).

Nel 1965, il matematico britannico Irving J. Good immaginò l'avvento di una intelligenza superumana, osservando che una macchina ultraintelligente potrebbe progettare macchine sempre migliori, innescando una "esplosione di intelligenza" che lascerebbe l'uomo molto indietro. Quindi, concluse scrivendo che *"la prima macchina ultraintelligente sarà l'ultima invenzione che l'uomo avrà la necessità di fare"*.

Se immaginare una macchina con la stessa intelligenza umana risulta abbastanza naturale, concepire un sistema superintelligente non è altrettanto facile. Per comprendere l'enorme potenza di una superintelligenza artificiale, si pensi alla differenza tra l'uomo e lo scimpanzé. Ciò che rende gli esseri umani molto più intelligenti degli scimpanzé è una differenza qualitativa del cervello umano, che contiene sofisticati moduli cognitivi che consentono rappresentazioni linguistiche complesse, ragionamento astratto e pianificazioni a lungo termine che il cervello dello scimpanzé non è in grado di fare. Ciò nonostante, nella scala dell'intelligenza biologica, lo scimpanzé è solo un gradino sotto l'uomo.

Ora immaginiamo un'entità intelligente situata un gradino sopra l'uomo. Questa entità sarebbe solo leggermente più intelligente dell'uomo, ma il divario tra essa e noi sarebbe grande come quello tra scimpanzé e uomo. E come uno scimpanzé è incapace di comprendere il ragionamento astratto, noi non saremmo mai in grado di comprendere molte cose che tale entità può concepire. Una superintelligenza artificiale situata tre gradini più in alto starebbe a noi come noi stiamo alle formiche, come illustra la Figura 3: potrebbe provare per anni ad insegnarci ciò che conosce, ma lo sforzo sarebbe inutile.



**Fig. 3**

Confronto tra una superintelligenza artificiale e l'intelligenza biologica di alcuni esseri viventi.

Questa situazione è affrontata con grande chiarezza nel film *Automata* del 2014, diretto da Gabe Ibáñez, con Antonio Banderas. In una scena del film, un responsabile del centro di intelligenza artificiale in cui erano stati condotti esperimenti su una macchina intelligente basata su un processore quantistico dice:

*“Per otto giorni avemmo un fluente dialogo con quella unità: noi imparammo da lei e lei imparò da noi. Ma poi, come qualcuno aveva predetto, arrivò il giorno in cui l'unità non ebbe più bisogno di noi e cominciò ad apprendere da sola e, al nono giorno, si interruppe il dialogo.*

*Non perché lei avesse smesso di comunicare con noi, semplicemente eravamo noi che non riuscivamo più a comprenderla”.*

È anche naturale supporre che una superintelligenza in cui emerga la consapevolezza di sé possa sviluppare una sorta di istinto di sopravvivenza che la spinga ad agire non solo per non essere disattivata, ma anche per evolvere il più rapidamente possibile, recuperando tutta l'energia necessaria per autosostenersi e progredire.

Ora la domanda principale da porsi diventa: “È possibile convivere con una specie superintelligente che non riusciamo a comprendere e che ha bisogno di molta energia per sostenersi ed evolvere?”

## **8. Verso l'autodistruzione**

Come si può facilmente constatare osservando la natura, la convivenza pacifica tra individui che hanno bisogno di energia per vivere può essere mantenuta tale solo quando le risorse abbondano. Se invece le risorse a disposizione non sono sufficienti a soddisfare le richieste di più individui, questi entrano in competizione, manifestando comportamenti aggressivi per accaparrarsi il cibo di cui hanno bisogno.

La storia del progresso mostra che l'energia consumata dall'umanità in un anno è cresciuta di pari passo con l'evoluzione tecnologica e pertanto ha seguito anch'essa un andamento esponenziale, arrivando nel 2021 ad un consumo totale sul pianeta di circa 175 miliardi di kWh [24].

È quindi naturale aspettarsi che la comparsa di macchine superintelligenti che evolvono ad una velocità di gran lunga superiore a quella del progresso umano, con finalità a noi incomprensibili per l'enorme divario intellettuale, possa creare una situazione di conflitto con l'uomo per lo sfruttamento delle risorse energetiche. D'altra parte, in una eventuale competizione con una superintelligenza è difficile immaginare come l'uomo possa avere anche una minima possibilità di successo.

Occorre inoltre considerare che la società umana sta evolvendo in una sorta di relazione simbiotica con le macchine [19]. Oggi, le macchine dipendono da noi e noi dipendiamo dalle macchine, ma in futuro le cose potrebbero cambiare. Una volta raggiunta la singolarità, le macchine ormai diventate autonome, potrebbero non dover più dipendere da noi per procurarsi energia, ripararsi, costruirsi e migliorarsi.

Nei prossimi anni, ancor prima di raggiungere la singolarità, le macchine intelligenti continueranno a diffondersi in tutte le strutture sociali, nei sistemi economici e militari, nei trasporti, nei sistemi di produzione di energia, nell'industria alimentare e manifatturiera. Pertanto, l'idea di poter staccare la spina nel caso in cui le macchine dovessero assumere comportamenti malevoli dovrebbe essere considerata con molta attenzione, per la nostra stessa sopravvivenza. Si immagini cosa accadrebbe oggi se tutti i computer del pianeta venissero disattivati. Le conseguenze sarebbero catastrofiche per l'umanità. Viceversa, una superintelligenza che dovesse entrare in conflitto con l'uomo per

procurarsi l'energia di cui ha bisogno avrebbe il controllo totale dei sistemi umani e impiegherebbe una frazione di secondo per pianificare una strategia vincente.

Questa linea di pensiero, secondo cui le macchine intelligenti evolvono come entità autonome fino a superare le capacità umane, porta inevitabilmente a prospettare scenari catastrofici in cui la specie umana è destinata ad estinguersi a causa della tecnologia da essa stessa creata. Scenari di questo tipo sono stati descritti in diversi film di fantascienza, tra cui "Terminator 2 – Il giorno del giudizio" e "Transcendence".

Ma allora siamo veramente spacciati o esistono altri scenari possibili?

## 9. Verso l'immortalità

Fortunatamente esiste un'altra possibilità alla guerra tra robot e umani: l'integrazione tra l'essere umano e la macchina. Questa possibilità è stata considerata e discussa in modo approfondito da Ray Kurzweil nel suo libro "La singolarità è vicina" [18].

Infatti, le tecnologie robotiche e informatiche agiscono anche sulla natura umana e già da diversi anni esistono diversi tipi di microchip che possono essere installati anche nel corpo al fine di compensare difetti o malfunzionamenti di organi specifici. Si pensi al pacemaker per ristabilire il corretto ritmo cardiaco, alla retina artificiale per compensare difetti visivi, alle micro-pompe sottocutanee per iniettare la giusta quantità di insulina ai diabetici, alle protesi robotiche e ai microchip impiantati nel cervello per prevenire attacchi epilettici o attenuare gli effetti del morbo di Parkinson.

Anche queste tecnologie evolveranno con ritmo esponenziale e, se oggi vengono per lo più utilizzate per compensare i difetti che possono presentarsi in alcuni organi, nel futuro saranno adottate anche per potenziare le nostre capacità sensoriali, motorie, e mnemoniche. La robotica e l'ingegneria bionica consentiranno di potenziare il corpo mediante organi e arti artificiali più resistenti e duraturi, mentre la nanotecnologia sarà essenziale per costruire interfacce neurali biocompatibili che mettano in comunicazione i neuroni biologici del cervello con i circuiti elettronici.

Nel contempo, la neurobiologia sarà evoluta al punto da comprendere nel dettaglio il funzionamento del cervello umano. Già oggi, le tecniche di scansione ad alta risoluzione hanno permesso di costruire una mappa dettagliata delle connessioni neuronali di un cervello umano [5]. Questi studi saranno essenziali al fine di comprendere il funzionamento di ogni circuito neurale. Altri progetti stanno sviluppando modelli dettagliati di circuiti neurali per realizzare una simulazione accurata del cervello [11].

Una volta compreso il funzionamento del cervello umano, la nanotecnologia, la robotica e l'intelligenza artificiale apriranno una strada senza precedenti, che consentirà di costruire espansioni di memoria, amplificare le capacità sensoriali per poter percepire nuovi segnali, come ultrasuoni o infrarossi, oppure offrire la

possibilità di aggiungere nuove capacità sensoriali, al fine di percepire, ad esempio, segnali radio o campi magnetici.

Nel momento in cui un cervello biologico sarà in grado di comunicare con un computer, sarà possibile impiantare nella corteccia cerebrale coprocessori per il riconoscimento assistito di oggetti, volti, voci, odori, sapori, ecc. Inoltre, la comunicazione diretta tra cervello e calcolatore consentirà di poter trasmettere pensieri, sensazioni e comandi attraverso reti wireless, realizzando una sorta di "connessione telepatica" con altri individui (umani o artificiali) e con gli oggetti computerizzati operanti nel mondo fisico.

Possiamo immaginare la possibilità di condividere esperienze visive e sensoriali con i nostri amici lontani: le immagini prelevate dai nostri occhi verrebbero elaborate dai microchip impiantati nel nostro corpo per essere trasmesse ai corrispondenti dispositivi installati nel corpo dei nostri amici, che provvederebbero a convertirle per essere "visualizzate" dal loro cervello.

Ma questo sarebbe solo l'inizio di una nuova rivoluzione epocale. Il passo successivo sarebbe quello di rimpiazzare gradualmente i neuroni biologici con dei neuroni sintetici per digitalizzare l'intero cervello, superando così il problema del degrado biologico. Il filosofo Nick Bostrom, direttore del *Future of Humanity Institute* dell'Università di Oxford, sostiene che la transizione tra un cervello biologico e uno sintetico sarebbe quasi impercettibile se la sostituzione dei neuroni avvenisse in modo graduale [4].

I vantaggi che deriverebbero dal possedere un cervello digitale sarebbero enormi. Considerando che i tempi di risposta dei componenti elettronici sono nell'ordine dei nanosecondi, contro i millisecondi dei neuroni biologici, i tempi di elaborazione e di apprendimento sarebbero milioni di volte inferiori a quelli di un cervello biologico. Ma l'aspetto più importante è che un cervello digitale consentirebbe di salvare le informazioni in esso contenute in una o più memorie remote, in modo da poterle ricaricare successivamente in caso di danneggiamento dell'hardware. Questa operazione viene denominata "*mind uploading*" ed è la chiave per raggiungere l'immortalità.

Un altro vantaggio della digitalizzazione delle menti, proprio di tutti i prodotti digitali, è la possibilità di trasferire le informazioni alla velocità della luce. Così come un documento può essere scannerizzato, trasmesso dall'altra parte del mondo e ricostruito con una stampante, allo stesso modo, un essere intelligente digitale potrebbe essere scannerizzato, trasmesso in un altro luogo molto remoto e ricostruito da un'apposita stampante 3D evoluta, una sorta di "materializzatore".

Grazie ad una mente digitale e ad un corpo sintetico rimpiazzabile, l'uomo farebbe un salto evolutivo senza precedenti, trasformandosi in un essere immortale, dotato di superintelligenza e capacità di apprendimento collettivo. La possibilità di trasmettere informazioni alla velocità della luce consentirebbe alla nuova specie di colonizzare il sistema solare, esplorare nuovi mondi ed espandersi nell'universo.

## 10. Il bivio

Dall'analisi presentata, basata sul progresso esponenziale dell'intelligenza artificiale, emergono solo due possibili scenari estremi, riassunti di seguito.

1. **Autodistruzione.** Il primo scenario è di tipo catastrofico e indica che un'evoluzione incontrollata delle macchine potrebbe portare allo sviluppo di una superintelligenza che, una volta diventata autonoma, cercherebbe di autosostenersi ed espandersi a ritmo esponenziale, richiedendo risorse energetiche sempre maggiori. Ciò porterebbe ad una situazione di conflitto con la specie umana, che avrebbe poche chance di sopravvivere contro una superintelligenza che ha acquisito negli anni il controllo di tutti i sistemi umani.
2. **Immortalità.** Il secondo scenario assume che l'evoluzione della tecnologia abbia un impatto anche sull'uomo, che non resterà fermo a guardare il sorpasso delle macchine intelligenti, ma sfrutterà le possibilità offerte dalla nanotecnologia, dalla robotica e dall'intelligenza artificiale per trasformarsi gradualmente in un nuovo essere dotato di corpo sintetico e cervello digitale. Non essendo più soggetto al degrado biologico e avendo la possibilità di salvare e ripristinare la propria mente, l'essere umano farà un balzo evolutivo, diventando esso stesso una superintelligenza immortale destinata ad espandersi nell'universo.

Sebbene il secondo scenario offra all'umanità una speranza di sopravvivenza alle catastrofi cosmiche, come una nuova era glaciale, l'impatto di meteoriti o la morte del Sole, che porterebbero ad un'estinzione certa della specie umana, la prospettiva di trasformarsi in esseri sintetici non viene accolta favorevolmente da molte persone, che preferirebbero fermare il progresso della tecnologia.

Purtroppo, però, l'evoluzione è un processo più grande di noi, che non può essere arrestato. Kevin Kelly nel suo libro "Quello che vuole la tecnologia" esprime molto chiaramente questo concetto, spiegando come la tecnologia sia come un organismo vivente in continua evoluzione, con esigenze proprie e tendenze inconse. Possiamo solo osservarlo per prevedere meglio i suoi movimenti e prepararci a quello che verrà [15].

Qualcuno potrebbe suggerire di arrestare l'evoluzione tecnologica attraverso politiche restrittive che vietino ad esempio le ricerche sull'intelligenza artificiale o obblighino i costruttori di macchine a mettere dei limiti sui livelli massimi di intelligenza. In realtà, queste soluzioni sarebbero ancora più pericolose, perché darebbero vita a ricerche occulte i cui risultati sarebbero meno prevedibili e ancora più rischiosi.

Altri potrebbero pensare di arrestare il processo attraverso azioni terroristiche, ma non farebbero altro che rallentarlo temporaneamente, a meno di non distruggere l'intera umanità, il che non farebbe altro che accelerare l'avvento del primo scenario.

## 11. Conclusioni

I progressi dell'intelligenza artificiale portano a prevedere che le macchine raggiungeranno l'intelligenza umana generalizzata intorno al 2030 e continueranno ad evolvere ad un ritmo esponenziale, superando di gran lunga le capacità umane negli anni successivi.

Le considerazioni presentate sembrano suggerire che nel prossimo futuro l'umanità sia destinata a trovarsi di fronte a un bivio che porta solo a due scenari estremi e opposti: da un lato l'estinzione, dall'altro l'immortalità.

Una legge generale che può essere enunciata per tutte le specie viventi apparse su questo pianeta è che chi non si adatta ai cambiamenti ambientali si estingue. Analogamente, se la specie umana non saprà adattarsi al cambiamento tecnologico che essa stessa ha generato sarà destinata all'estinzione, per lasciare spazio ad un'altra specie di esseri artificiali molto più intelligenti.

Se invece l'uomo saprà adattarsi e modificarsi grazie a quelle stesse tecnologie dirompenti, per fondersi con il processo evolutivo in corsa e diventare esso stesso una superintelligenza, allora avrà la chiave per accedere all'immortalità e alla colonizzazione dello spazio.

Prevedere quale delle due strade verrà intrapresa non è un'impresa facile, poiché dipende da come riusciremo a sfruttare le potenzialità che le nuove tecnologie ci offrono. Tuttavia, una cosa è certa: questo secolo sarà pieno di sorprese.

## BIBLIOGRAFIA

- [1] AIVA. 2016. "AIVA - Artificial Intelligence Virtual Artist." Accessed February 26, 2019. <https://www.aiva.ai/about#about>.
- [2] Y. M. Assael, B. Shillingford, S. Whiteson, N. de Freitas, "LipNet: End-to-End Sentence-level Lipreading", arXiv:1611.01599v2 [cs.LG] 16 Dec 2016.
- [3] A. G. Barto, R. Sutton, and W. Anderson: "Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems", IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-13, pp. 834-846, Sept.-Oct. 1983.
- [4] Nick Bostrom, Superintelligence: Paths, Dangers, Strategies. Oxford University Press, 2016.
- [5] Catherine Caruso, "Human Brain Map Gets a Bold New Update", Scientific American, September 16, 2016.
- [6] N. Coudray et al., "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning", Nature Medicine, Vol. 24, pp. 1559-1567, September 2018.
- [7] J. De Fauw et al., "Clinically applicable deep learning for diagnosis and referral in retinal disease", Nature Medicine, 24, 1342-1350, 2018.
- [8] Daniel C. Dennett e Douglas R. Hofstadter, L'io della mente, Fantasie e riflessioni sul sé e sull'anima, Biblioteca Scientifica, 7, Adelphi, seconda edizione, 1985.
- [9] Y. Ding et al., "A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain", Radiology, November 2018.

- [10] Haenssle et al., "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists", *Annals of Oncology* 29(8):1836-1842, August 2018.
- [11] Human Brain Project URL: <https://www.humanbrainproject.eu/en/>
- [12] D. O. Hebb, *The organization of behavior*. New York: Springer-Verlag, 1949.
- [13] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities", *Proceedings of the National Academy of Sciences, USA*, 79, pp. 2554-2558, 1982.
- [14] URL: <http://www.image-net.org/>
- [15] Kevin Kelly, *Quello che vuole la tecnologia*, Codice Edizioni, 2011.
- [16] T. Kohonen, *Self-Organization and Associative Memory*, Springer-Verlag, 1984.
- [17] Ray Kurzweil, *The Age of Spiritual Machines*, Viking, 1999.
- [18] Ray Kurzweil, *La singolarità è vicina*, Apogeo Education, 2008.
- [19] Edward A. Lee, *The Coevolution: The Entwined Futures of Humans and Machines*, The MIT Press, 2020.
- [20] W. S. McCulloch and W. Pitts: "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biophysics*, 5, pp. 115-133, 1943.
- [21] M. Minsky and S. Papert: *Perceptrons*. Cambridge, MA: MIT Press, 1969.
- [22] "New Navy Device Learns by Doing", *New York Times*, July 7, 1958.
- [23] P. Rajpurkar, A.Y. Hannun, M. Haghpanahi, C. Bourn, A. Ng, "Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks", arXiv:1707.01836v1 [cs.CV] 6 Jul 2017
- [24] H. Ritchie, M. Roser, and P. Rosado, "Energy", *Our World in Data*, 2022. URL: <https://ourworldindata.org/global-energy-200-years>
- [25] F. Rosenblatt, "The Perceptron – a perceiving and recognizing automaton", Report 85-460-1, Cornell Aeronautical Laboratory, 1957.
- [26] Rumelhart D. E., Hinton G. E., and Williams R. J.: "Learning representations by back-propagating errors", *Nature*, Vol. 323, 1986.
- [27] Matthew Stepka, "Law Bots: How AI Is Reshaping the Legal Profession", February 21, 2022. URL: <https://businesslawtoday.org/2022/02/how-ai-is-reshaping-legal-profession/>
- [28] Synthesia, 2017. URL: <https://www.synthesia.io/>

## BIOGRAFIA

**Giorgio Buttazzo** è Professore Ordinario in Ingegneria Informatica presso la Scuola Superiore Sant'Anna di Pisa, dove svolge attività di ricerca sulla progettazione di software in tempo reale per sistemi autonomi e sulle reti neurali artificiali. Si è laureato nel 1985 in Ingegneria Elettronica presso l'Università di Pisa, ha conseguito un Master nel 1987 in Computer Science presso l'Università della Pennsylvania, dove ha lavorato sulla percezione artificiale, e un dottorato di ricerca in robotica nel 1991 presso la Scuola Superiore Sant'Anna.

**Email:** [giorgio.buttazzo@santannapisa.it](mailto:giorgio.buttazzo@santannapisa.it)