



Gianfranco Basti

ETICA

DEL MACHINE LEARNING

PARTE I:
LE ORIGINI DELL'IA E DEL MACHINE LEARNING

Roma 2025

SOMMARIO

SOMMARIO.....	2
0. BIBLIOGRAFIA DEL CORSO	4
SOMMARIO.....	7
1. SCHEMA DEL CORSO	9
2. DALLA LOGICA ALL'INFORMATICA: IL CALCOLATORE UNIVERSALE (DAVIS, 2012).....	12
2.1. L'INTERPRETAZIONE ESTENSIONALE DEL SILLOGISMO IN LEIBNIZ	12
2.2. L'INVENZIONE DELLA LOGICA ALGEBRICA DA PARTE DI BOOLE.....	15
2.3. ELEMENTI DI CALCOLO BINARIO E CONSEGUENZE PER LA LOGICA BOOLEANA.....	21
2.3.1. <i>Conteggio decimale</i>	23
2.3.2. <i>Conteggio binario</i>	24
2.3.3. <i>Conversione fra i due sistemi</i>	25
2.3.4. <i>Addizione</i>	27
2.3.5. <i>Passaggio alla logica booleana estensionale</i>	28
2.3.6. <i>Passaggio alla logica booleana modale e intensionale</i>	32
2.4. DALLA LOGICA DI BOOLE ALLA MACCHINA DI TURING: IL CALCOLATORE UNIVERSALE (DAVIS, 2012), CAPP. VII-VIII	33
2.4.1. <i>La logica di Boole e il concetto di porta logica</i>	33
2.4.2. <i>Algoritmi e codifica binaria del linguaggio naturale</i>	41
2.4.3. <i>La "Macchina di Turing" e il Calcolatore Universale</i>	44
3. IL TEST DI TURING E IL PROGRAMMA DI RICERCA DELL'IA SIMBOLICA	53
3.1. IL TEST DI TURING.....	53
3.2. LA CONFERENZA DI DARTMOUTH E LA NASCITA DEL PROGRAMMA DI RICERCA DELL'IA	59
3.2.1. <i>I sistemi esperti, la programmazione ad oggetti e l'IA simbolica</i>	62
3.3. LA NASCITA DELLE SCIENZE COGNITIVE	68
3.3.1. <i>Dal Test di Turing e dal "dogma dell'IA" alla nascita delle scienze cognitive</i>	68
3.3.2. <i>Dalle scienze cognitive alle neuroscienze cognitive: le Reti Neurali Naturali (Limoncelli, 2020)</i>	75
3.4. L'IA SIMBOLICA E L'APPROCCIO FUNZIONALISTA ALLE NEUROSCIENZE COGNITIVE.....	89
3.4.1. <i>Le reti neurali lineari di McCulloch e Pitts</i>	89
3.4.2. <i>La regola di Hebb nella formazione di reti neurali naturali</i>	90

3.4.3.	<i>Le reti neurali lineari di Farley e Clark</i>	91
3.4.4.	<i>Il perceptrone lineare di Rosenblatt</i>	94
3.5.	L'IA NON-SIMBOLICA O STATISTICA, LE RETI NEURALI NON-LINEARI E IL MACHINE LEARNING (ML)	97
3.5.1.	<i>L'IA statistica discriminativa e generativa e l'attuale "boom dell'IA"</i>	97
3.5.2.	<i>Funzioni di attivazione lineari e non-lineari e l'approccio connessionista alle RNA</i>	101
4.	BIBLIOGRAFIA DELLA I PARTE	106
NOTE		125

0. BIBLIOGRAFIA DEL CORSO

♦ In Grassetto i Testi Fondamentali del Corso

Basti, G. (2012). Logica aletica, deontica, ontologia formale. Dalla verità ontica all'obbligo deontico , Città del Vaticano, 2012, pp. 105-270. In P. Gherri, & G. Basti (Eds.), *Logica e Diritto: tra argomentazione e scoperta* (pp. 105-270). Città del Vaticano: Lateran University Press. (Disponibile online)

Basti, G. (2020). Ethical Responsibility vs. Ethical Responsiveness in Conscious and Unconscious Communication Agents. *Proceedings*, 47(68), 1-7. doi:10.3390/proceedings 47010068 (Disponibile online)

Basti G., «La sfida etica dell'intelligenza artificiale e il ruolo della filosofia». *Aquinas*, LXV, 2(2022), pp. 299-321 (Disponibile online)

Basti G, Vitiello G., «Deep learning opacity and ethical accountability in AI systems. A new perspective». In: *The Logic of Social Practices, II*, Lowe, Robert and Giovagnoli, Raffaella (Eds.), Springer International Publishing, Berlin-New York, 2023, pp. 21-73 (Disponibile online)

Christian, B. (2020). *The alignment problem. How can artificial intelligence learn human values?* (Kindle Ed.). London, UK: Atlantic Books.

Davis, M. (2012). *Il calcolatore universale. Da Leibniz a Turing. 3. Edizione.* Milano: Adelphi. (Disponibile online)

De Mauro, A. (2019). *Big Data Analytics. Analizzare e interpretare dati col machine learning. (Kindle Edition)* Tutorials online:
<https://www.apogeoonline.com/libri/big-data-analytics-andrea-de-mauro/>. Milano: Apogeo-Feltrinelli.

Floridi, L. (2022). *Etica dell'intelligenza artificiale. Sviluppi, opportunità sfide (Ed. Kindle).* Milano: Raffaele Cortina Editore.

Goodfellow, I., & Bengio, Y. C. (2016). *Deep Learning (Kindle Ed.)*. Cambridge Mass., USA: MIT Press.

Horstman, M. (2024). *Transformers for Dummies: A Peek Inside AI Models,* <https://michielh.medium.com/transformers-unleashed-the-neural-architecture-powering-modern-ai-and-language-models-57626643fd49> (Disponibile online)

MLK. (2019, January 7). *Machine Learning Basics*. Tratto da Machinelearningknowledge.ai:

<https://machinelearningknowledge.ai/mcculloch-pitts-neuron-model/>

Müller, V. C. (2025, September 1). *Ethics of Artificial Intelligence and Robotics*.

Retrieved from Stanford Encyclopedia of Philosophy:

<https://plato.stanford.edu/archives/fall2025/entries/ethics-ai/> (Disponibile online)

SOMMARIO

SOMMARIO.....	2
0. BIBLIOGRAFIA DEL CORSO	4
SOMMARIO.....	7
1. SCHEMA DEL CORSO	9
2. DALLA LOGICA ALL'INFORMATICA: IL CALCOLATORE UNIVERSALE (DAVIS, 2012).....	12
2.1. L'INTERPRETAZIONE ESTENSIONALE DEL SILLOGISMO IN LEIBNIZ	12
2.2. L'INVENZIONE DELLA LOGICA ALGEBRICA DA PARTE DI BOOLE.....	15
2.3. ELEMENTI DI CALCOLO BINARIO E CONSEGUENZE PER LA LOGICA BOOLEANA.....	21
2.3.1. Conteggio decimale	23
2.3.2. Conteggio binario.....	24
2.3.3. Conversione fra i due sistemi.....	25
2.3.4. Addizione	27
2.3.5. Passaggio alla logica booleana estensionale	28
2.3.6. Passaggio alla logica booleana modale e intensionale.....	32
2.4. DALLA LOGICA DI BOOLE ALLA MACCHINA DI TURING: IL CALCOLATORE UNIVERSALE (DAVIS, 2012), CAPP. VII-VIII	33
2.4.1. La logica di Boole e il concetto di porta logica.....	33
2.4.2. Algoritmi e codifica binaria del linguaggio naturale	41
2.4.3. La "Macchina di Turing" e il Calcolatore Universale.....	44
3. IL TEST DI TURING E IL PROGRAMMA DI RICERCA DELL'IA SIMBOLICA	53
3.1. IL TEST DI TURING.....	53
3.2. LA CONFERENZA DI DARTMOUTH E LA NASCITA DEL PROGRAMMA DI RICERCA DELL'IA	59
3.2.1. I sistemi esperti, la programmazione ad oggetti e l'IA simbolica.....	62
3.3. LA NASCITA DELLE SCIENZE COGNITIVE	68
3.3.1. Dal Test di Turing e dal "dogma dell'IA" alla nascita delle scienze cognitive	68
3.3.2. Dalle scienze cognitive alle neuroscienze cognitive: le Reti Neurali Naturali (Limoncelli, 2020).....	75

3.4.	L'IA SIMBOLICA E L'APPROCCIO FUNZIONALISTA ALLE NEUROSCIENZE COGNITIVE.....	89
3.4.1.	<i>Le reti neurali lineari di McCulloch e Pitts</i>	89
3.4.2.	<i>La regola di Hebb nella formazione di reti neurali naturali</i>	90
3.4.3.	<i>Le reti neurali lineari di Farley e Clark</i>	91
3.4.4.	<i>Il perceptrone lineare di Rosenblatt</i>	94
3.5.	L'IA NON-SIMBOLICA, LE RETI NEURALI NON-LINEARI E IL MACHINE LEARNING	97
3.5.1.	<i>Funzioni di attivazione lineari e non-lineari e l'approccio connessionista alle RNA</i>	97
4.	BIBLIOGRAFIA DELLA I PARTE.....	106
NOTE.....		125

1. Schema del Corso

PARTE PRIMA: LE ORIGINI DELL'IA E DEL MACHINE LEARNING

1. Dalla Logica all'Informatica: il Calcolatore Universale
2. Il Test di Turing e il Programma di Ricerca dell'IA Simbolica
3. La Nascita delle Scienze e Neuroscienze Cognitive
4. L'IA Pre-Simbolica e le Reti Neurali
 - a. IA Discriminativa e IA Generativa
 - b. Reti Neurali Naturali
 - c. Reti Neurali Artificiali
 - d. Il Machine Learning

PARTE SECONDA: ALGORITMI DI MACHINE LEARNING.

IA DISCRIMINATIVA E IA GENERATIVA

1. Machine Learning Supervisionato, Con Rinforzo, Non-Supervisionato,
2. Algoritmi di Machine Learning nell'IA Discriminativa: il Deep-Learning
 - a. Il Perceptrone Multistrato

- b. Machine Learning Supervisionato: la Back-Propagation
 - c. Machine Learning Supervisionato: Reti Convulsive
 - d. Machine Learning Non-Supervisionato
 - e. Machine Learning Quantistico
3. IA Generativa: la Rivoluzione dei Transformers nel Machine Learning

PARTE TERZA: ETICA E MACHINE LEARNING

- 1. Etica nell'IA ed Etica nel Machine Learning
- 2. Le “Ingiustizie Algoritmiche” nel Machine Learning
- 3. Problemi Etici dell'IA Generativa
- 4. Neuroetica nelle Neuroscienze Cognitive ed Etica nel Machine Learning
- 5. Logica Deontica e la Nozione di “Algoritmo Buono” nel Machine Learning

CONCLUSIONI

Responsabilità etica condivisa uomo-macchina nei sistemi di IA autonomi e non-autonomi

PARTE PRIMA:

LE ORIGINI DELL'IA

E DEL MACHINE LEARNING

2. Dalla logica all'informatica: il calcolatore universale (Davis, 2012)

2.1. L'interpretazione estensionale del sillogismo in Leibniz

- ◆ L'inizio della **logica matematica moderna** risalgono a Leibniz con il suo ideale della *Charcateristica Universalis*, ovvero un'estensione dell'algebra dalla fisica alla logica, dal calcolo algebrico differenziale-integrale da lui inventato con Newton, all'**algebra della logica**, come manipolazione di **simboli logici (numeri) secondo regole**.
- ◆ Dove con algebra in matematica si intende una «struttura di relazioni definita su un campo numerico con le operazioni caratteristiche che costituiscono la **firma** di quell'algebra».
- ◆ P.es., nel caso dell'**algebra aritmetica**, la firma è costituita dalle quattro operazioni aritmetiche elementari.

- ◆ Di qui il motto leibniziano del *calculemus* per recuperare l'unità del sapere su base logico-matematica che era andata smarrita con la crisi della filosofia scolastica.
- ◆ Il primo passo di questo programma di ricerca fu l'interpretazione estensionale in termini di **logica delle classi/insiemi** (= **estensionale**) la logica dei predicati del sillogismo aristotelico.lgi
- ◆ Nella logica estensionale della teoria degli insiemi/classi, dove esiste uno stretto rapporto fra **appartenenza** $\in$ predicativa (lo “è” della copula soggetto-predicato) e **inclusione** nelle rispettive classi/insiemi **la prova**, e dunque **l'universalità della verità logica di un asserto**, consiste nel dimostrare che le due classi di oggetti, poste nella relazione di appartenenza nell'asserto predicativo vero (p.es., “tutti gli uomini sono mortali”), appartengono a un'unica classe che le include ambedue.
- ◆ Prendiamo il più antico metodo dimostrativo della logica occidentale, ovvero il **sillogismo deduttivo** nella sua forma elementare *In Barbara*. “Tutti gli uomini (B) sono mortali (A), i Greci (C) sono uomini, quindi (è universalmente dimostrabile e dunque vero) che tutti gli uomini sono mortali”.
- ◆ **Nell'interpretazione estensionale** (insiemistica) che per primo Leibniz, seguito da Eulero e quindi da Venn (con i suoi famosi diagrammi: cfr. *infra*) hanno dato dello

schema sillogistico aristotelico “ $AB \ \& \ BC \therefore AC$ ” ciò corrisponde ad affermare predicativamente $((B \in A) \wedge (C \in B) \rightarrow (C \in A))$, ovvero usando l’usuale calcolo dei predicati come abbiamo imparato a fare a Logica II: $(\forall x((x \in B) \rightarrow (x \in A)) \wedge \forall x(x \in C) \rightarrow (x \in B)) \rightarrow \forall x(x \in C) \rightarrow (x \in A)$.

- ◆ Ambedue le formule predicative hanno la loro prova estensionale nelle inclusioni corrispondenti fra classi $((A \supseteq B) \wedge (B \supseteq C) \rightarrow (A \supseteq C))$, per semplice transitività dell’operatore insiemistico di inclusione: $(A \supseteq B \supseteq C) \rightarrow (A \supseteq C)$.
- ◆ Il punto irrisolto in Leibniz era come associare a numeri **simboli logici che rappresentano classi**.
- ◆ Questo problema fu risolto brillantemente da George Boole che dimostrò che con la sua idea si può formalizzare algebricamente non solo la **logica dei predicati** aristotelica, ma anche la **logica delle proposizioni** stoica. Anche se dobbiamo aspettare Gottlob Frege con la sua invenzione delle **tavole di verità** per i predicati (connettivi) proposizionali e quindi della nozione di **funzione proposizionale** l’unificazione delle due logiche in un solo **calcolo logico-algebrico** come sappiamo da Logica II (cfr., (Davis, 2012), capp. I-III).

2.2. L'invenzione della logica algebrica da parte di Boole

- ◆ L'intuizione folgorante nella sua semplicità per risolvere il problema irrisolto di Leibniz di come associare a numeri simboli logici che rappresentano classi/insiemi di oggetti si deve **George Boole** (1815-1864).
- ◆ È chiaro che se con simboli numerici qualsiasi (variabili numeriche) io volessi denotare classi di oggetti, p.es., “tavoli” con x e “bianchi” con y , allora la classe dei “tavoli bianchi” corrisponderebbe al **prodotto** xy (= intersezione fra classi) ovvero la classe degli oggetti che sono “tavoli e bianchi”.
- ◆ Il problema è che quando volessi rappresentare con simboli numerici una **identità logica** (p.es., la proposizione tautologica “i tavoli sono tavoli” – e tutte le leggi logiche sono tautologie) essa corrisponderebbe al **prodotto** $xx = x$.
- ◆ Ma questa equazione aritmetica è vera solo per due numeri **1** e **0**, per qualsiasi altro numero $xx = x^2 \neq x$.
- ◆ Quindi una logica algebrica o **logica booleana** suppone che la rappresentazione numerica di oggetti logici (classi) avvenga mediante una **numerazione binaria**,

ovvero **in base 2** (2 cifre) e non, per esempio, **in base decimale** (10 cifre), come siamo abituati.

- ◆ Per far questo, però, dobbiamo reinterpretare i simboli numerici 0 e 1 in termini di classi. Il punto di partenza è ricordare che in aritmetica **0 per qualsiasi numero x dà 0, mentre 1 per qualsiasi numero x dà quello stesso numero:**

$$0x = 0; \quad 1x = x$$

- ◆ Passando alle classi, $0x = 0$ per ogni x se interpretiamo 0 come denotante **una classe a cui non appartiene nulla (la cosiddetta classe vuota $\emptyset$)** ovvero rappresenta **tutto ciò che è falso (=contraddittorio)** in (una determinata) logica. Viceversa, $1x = x$ se interpretiamo 1 come denotante **una classe cui appartengono tutti gli oggetti (la cosiddetta classe universale V)**, ovvero rappresenta **tutto ciò che è vero** in (una determinata) logica.
- ◆ Quindi, la logica algebrica non è altro che l'algebra aritmetica ordinaria con in più la semplice regola che $xx = x$ e quindi un'algebra aritmetica definita usando una numerazione binaria per i suoi calcoli. Infatti, data la regola $xx = x$ essa può essere riscritta anche come $x - (xx) = 0$, che raccogliendo a fattor comune dà

$$x(1 - x) = 0$$

- ◆ Ovvero: “**nulla può appartenere e non appartenere a una classe x** ” che è esprime il **principio di contraddizione**, quello che Aristotele definiva la **legge logica fondamentale**. Un risultato entusiasmante che giustamente esaltò Boole, consapevole che così stava aprendo una nuova pagina nella storia della logica, quella appunto della **logica matematica**.
- ◆ Ed infatti il risultato così ottenuto da Boole lo portò a dimostrare che con un calcolo algebrico su classi così definito fosse possibile rappresentare non solo argomentazioni sillogistiche, ma anche qualsiasi argomentazione logica valida fra proposizioni (cfr. (Davis, 2012), pp. 47ss.). Prima però di imparare a rappresentare numeri e operazioni aritmetiche in numerazione binaria per familiarizzare col calcolo booleano in logica, facciamo alcuni passi avanti nella storia delle origini della logica matematica.
- ◆ Infatti, qualsiasi manuale di logica contemporanea ricorda giustamente come la logica matematica sia nata come **logica algebrica** (o algebra della logica), a partire dallo stesso W. G. Leibniz (1646-1716), poi dai fondamentali lavori di G. Boole (1815-1864) e A. De Morgan (1806-1871) e C. S. Peirce (1839-1914), tutti

confluiti nella sintesi di E. Schröder (1841-1902) sulla possibilità di concepire la nascente logica matematica come un'**algebra della logica**.

- ◆ Innanzitutto, capiamo perché **le (fregeane) tavole di verità** dei connettivi logici dei predicati proposizionali che hanno per argomento **variabili proposizionali, $p, q, r, \dots$** vanno scritte usando i simboli binari $\langle 1 \equiv \textit{vero o classe non vuota} \rangle$ e $\langle 0 \equiv \textit{falso o classe vuota} \rangle$.
- ◆ Riportiamo qui le tavole di verità dei connettivi logici che già conosciamo da Logica I e Logica II

p	q	$\neg p$	$p \wedge q$	$p \vee q$	$p \rightarrow q$	$p \leftrightarrow q$
1	1	0	1	1	1	1
1	0	-	0	1	0	0
0	1	1	0	1	1	0
0	0	-	0	0	1	1

- ♦ E, sempre da Logica II, riportiamo qui per comodità l'interpretazione sempre fregana del calcolo dei predicati e delle proposizioni in termini di **calcolo delle classi** usando i **diagrammi di Venn** – già Leibniz, ripeto, aveva intuito questo

1. Inclusione ($\subseteq$): $\mathbf{A} \subseteq \mathbf{B} := \forall x((x \in \mathbf{A}) \rightarrow (x \in \mathbf{B}))$

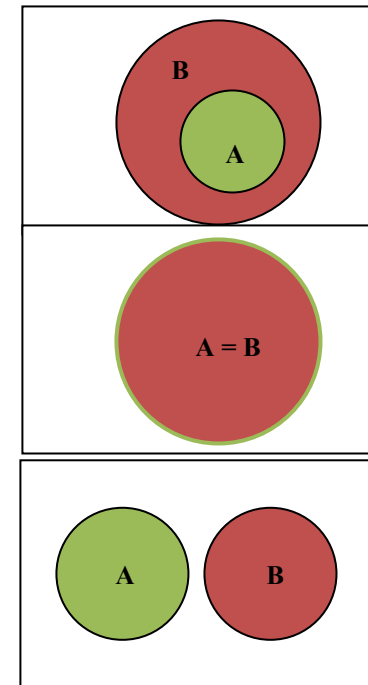
Se tutti gli elementi della prima classe sono inclusi nella seconda, la prima è **inclusa** nella seconda

2. Uguaglianza (=): $\mathbf{A} = \mathbf{B} := \forall x((x \in \mathbf{A}) \leftrightarrow (x \in \mathbf{B}))$

Se l'inclusione delle classi è reciproca, le due classi sono dette **uguali**.

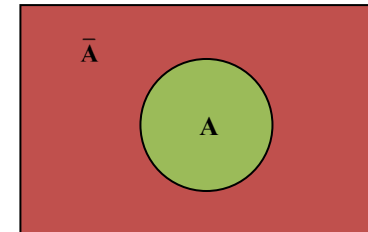
3. Esclusione ($\not\subseteq$): $\mathbf{A} \not\subseteq \mathbf{B} := \forall x((x \in \mathbf{A}) \rightarrow \neg(x \in \mathbf{B}))$

Se nessun elemento della prima classe è incluso nella seconda, la prima **esclude** la seconda.



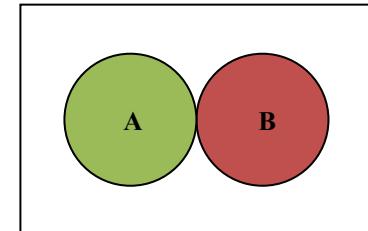
4. Complemento ($\bar{A}$): $(x \in \bar{A}) =: \neg(x \in A)$

La classe $\bar{A}$ è detta **complemento** della classe A , se tutti gli elementi che appartengono ad $\bar{A}$ non appartengono ad A e viceversa.



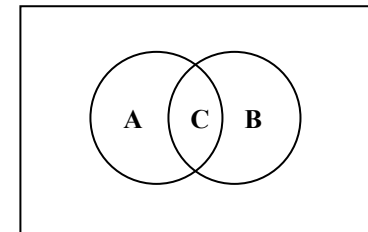
5. Somma ($\cup$): $(x \in (A \cup B)) =: (x \in A) \vee (x \in B)$

La **somma** o **unione** di due classi è la classe che contiene gli elementi di ambedue.



6. Prodotto ($\cap$): $(x \in (A \cap B)) =: (x \in A) \wedge (x \in B)$

Il **prodotto** o **intersezione** C di due classi è la classe degli elementi comuni ad ambedue.



- ◆ Schröder in particolare sviluppò l'intuizione di Boole dimostrando come usando un'algebra a due sole cifre (0,1) o **cifre binarie** (*binary digits* o **bits**) si può tradurre nella cosiddetta **logica equazionale** qualsiasi espressione di logica delle proposizioni¹. Ovvero, formule valide della logica delle proposizioni possono tradursi in altrettante espressioni algebriche a partire dalle essenziali **congruenze**, note

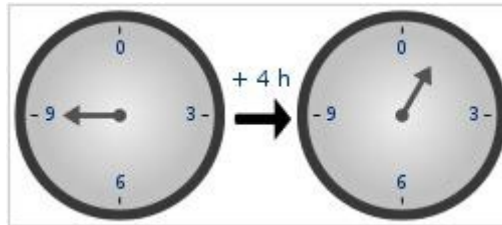
anche a Leibniz, fra **connettivi proposizionali** e **operazioni dell'algebra elementare**, in particolare fra **disgiunzione** $\vee$ e **addizione** $+$, **congiunzione** $\wedge$ e **moltiplicazione** $\times$, **negazione** $\neg$ e **complementazione** $0 = 1 - x$.

- ◆ Dove quest'ultima equazione ci ricorda come il campo numerico dell'algebra booleana della logica non è quello dei **numeri naturali**, (interi positivi) $\mathbb{N}$ ma quello degli **interi** (positivi e negativi) $\mathbb{Z}$.

2.3. Elementi di calcolo binario e conseguenze per la logica booleana

- ◆ Una numerazione binaria $\langle 0, 1 \rangle$ quale quella del *bit* significa che $\langle 1 + 1 = 0 \rangle$ e non $\langle = 2 \rangle$ come nella ordinaria numerazione decimale, come spiegheremo in §2.3.2.
- ◆ Infatti, una **numerazione binaria** è una aritmetica “**modulo 2**”, proprio come la nostra **numerazione decimale** è un'aritmetica “**modulo 10**” e proprio come **le cifre sul quadrante di un orologio** rappresentano un'aritmetica “**modulo 12**”. Ovvero un'aritmetica dove le uniche cifre ammissibili per rappresentare numeri sono $\langle 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 \rangle$ Infatti, sul quadrante di un orologio $\langle 9 + 4 \neq 13 \rangle$, ma

$$\langle 9+4 \equiv 1 \pmod{12} \rangle.$$



- ◆ Ovvero, più sinteticamente, $\langle 9+4 \equiv_{12} 1 \rangle$, e quindi $\langle 12+12 \equiv_{12} 0 \rangle$, proprio come $\langle 1+1 \equiv_2 0 \rangle$, dove il simbolo “ $\equiv$ ” è il simbolo della **congruenza**, ovvero il simbolo algebrico della equivalenza.
- ◆ Per impratichirci delle basi del calcolo aritmetico binario, impariamo a tradurre sistematicamente le espressioni numeriche in **numerazione decimale**, ovvero “modulo 10”, dove le cifre ammissibili per rappresentare numeri sono 10, e dove dunque $\langle 10 + 10 \equiv_{10} 0 \rangle$, in **numerazione binaria**, ovvero “modulo 2”, e dove dunque $\langle 1 + 1 \equiv_2 0 \rangle$.
- ◆ Per impratichirci del calcolo aritmetico binario partiamo dal fatto elementare del conteggio su base decimale e su base binaria.

2.3.1. Conteggio decimale

- ◆ Il conteggio decimale utilizza i dieci simboli da 0 a 9. Il conteggio inizia con la sostituzione incrementale della cifra meno significativa (cifra più a destra) che viene spesso chiamata prima cifra.
- ◆ Quando i simboli disponibili per questa posizione sono esauriti, la cifra meno significativa **viene reimpostata su 0** (*reset*) e la cifra successiva di maggiore rilevanza (una posizione a sinistra) **viene incrementata di 1** (*overflow*) e la sostituzione incrementale della cifra di ordine basso riprende. Questo metodo di reset e overflow viene ripetuto per ogni cifra di significato. Il conteggio procede come segue:
 - ◆ 000, 001, 002, ... 007, 008, 009, → (la cifra più a destra viene riportata a 0, e la cifra alla sua sinistra è incrementata di 1):
 - ◆ 010, 011, 012, ...
 - ◆ ...
 - ◆ 090, 091, 092, ... 097, 098, 099, (le due cifre più a destra sono riportate a 0, e la cifra alla loro sinistra è incrementata di 1):
 - ◆ 100, 101, 102, ...

2.3.2. Conteggio binario

- ◆ Il conteggio binario segue esattamente la stessa procedura, e di nuovo la sostituzione incrementale inizia con la cifra binaria (o bit) meno significativa (quella più a destra, chiamata anche primo bit), tranne per il fatto che **sono disponibili solo i due simboli 0 e 1**.
- ◆ Pertanto, dopo che un bit raggiunge 1 in binario, un incremento lo reimposta su 0 ma causa anche un incremento del bit successivo a sinistra (fra parentesi in corsivo cifra decimale):
- ◆ 0000,
- ◆ 0001(*1*), (Il bit più a destra ricomincia e la cifra successiva viene incrementata)
- ◆ 0010(*2*), 0011(*3*), (a destra, due bit ricominciano da capo e il bit successivo viene incrementato)
- ◆ 0100(*4*), 0101(*5*), 0110(*6*), 0111(*7*), (i tre bit più a destra ricominciano da capo e il bit successivo viene incrementato)
- ◆ 1000(*8*), 1001(*9*), 1010(*10*), 1011(*11*), 1100(*12*), 1101(*13*), 1110(*14*), 1111(*15*) ...
- ◆ ...
- ◆ In sintesi:

Decimal number	Binary number
0	0
1	1
2	10
3	11
4	100
5	101
6	110
7	111
8	1000
9	1001
10	1010
11	1011
12	1100
13	1101
14	1110
15	1111

2.3.3. Conversione fra i due sistemi

- ◆ Nel sistema decimale, ogni cifra decimale nella stringa numerica rappresenta **una potenza crescente di 10**. Quindi la cifra più a destra va moltiplicata per $10^0=1$,

quella successiva per $10^1=10$, la successiva ancora per $10^2=100$, la successiva ancora per $10^3=1000$, e così via. Ad esempio il numero decimale 327:

$$\blacklozenge 327_{10} = [(3) \times 10^2] + [(2) \times 10^1] + [(7) \times 10^0]$$

$$\blacklozenge 327_{10} = [(3) \times 100] + [(2) \times 10] + [(7) \times 1]$$

$$\blacklozenge 327_{10} = 300 + 20 + 7$$

◆ Analogamente, nel sistema binario, ogni bit nella stringa numerica rappresenta **una potenza crescente di 2**, con il bit più a destra che rappresenta $2^0=1$, il successivo che rappresenta $2^1=2$, quindi $2^2=4$ e così via. Il valore di un numero binario è la somma delle potenze di 2 rappresentate da ogni bit "1". Ad esempio, il numero binario 100101 viene convertito in formato decimale come segue:

$$\blacklozenge 100101_2 = [(1) \times 2^5] + [(0) \times 2^4] + [(0) \times 2^3] + [(1) \times 2^2] + [(0) \times 2^1] + [(1) \times 2^0]$$

$$\blacklozenge 100101_2 = [1 \times 32] + [0 \times 16] + [0 \times 8] + [1 \times 4] + [0 \times 2] + [1 \times 1]$$

$$\blacklozenge 100101_2 = 32 + 0 + 0 + 0 + 4 + 1$$

$$\blacklozenge 100101_2 = 37_{10}$$

2.3.4. Addizione

- ◆ L'operazione aritmetica più semplice in binario è l'**addizione**. L'aggiunta di due numeri binari a una cifra è relativamente semplice, utilizzando una forma di riporto alla posizione successiva:
- ◆ $0 + 0 \rightarrow 0$
- ◆ $0 + 1 \rightarrow 1$
- ◆ $1 + 0 \rightarrow 1$
- ◆ $1 + 1 \rightarrow 0$, riporto 1, cosicché 2_{10} in decimale, si scrive in binario 10_2 e si legge “uno, zero”, non “10”.
- ◆ Ovvero, l'addizione di due cifre "1" produce una cifra "0", mentre 1 dovrà essere aggiunto alla posizione successiva. Questo è simile a ciò che accade in decimale quando alcuni numeri a una cifra vengono sommati insieme. Se il risultato è uguale o superiore al valore della radice (modulo) (10), la cifra a sinistra viene incrementata:
- ◆ $5 + 5 \rightarrow 0$, riporto 1 [poiché $5 + 5 = 10 = (0 + (1 \times 10^1))$]
- ◆ $7 + 9 \rightarrow 6$, riporto 1 [poiché $7 + 9 = 16 = (6 + (1 \times 10^1))$]

- ◆ Simili **procedure ricorsive** valgono per le tre altre **operazioni aritmetiche fondamentali**, sottrazione, moltiplicazione, divisione (cfr. [Sistema numerico binario - Wikipedia](#)). Ricordando sempre che le due operazioni-base sono **addizione e sottrazione**, visto che la moltiplicazione è solo **una sequenza di somme** (=sommare n volte il valore del moltiplicando, dove n è il valore del moltiplicatore) e la divisione è solo **una sequenza di sottrazioni** (=sottrarre n volte al dividendo il valore del divisore fino ad arrivare a zero, e il numero n sarà il risultato della divisione).

2.3.5. Passaggio alla logica booleana estensionale

- ◆ Poiché in un'algebra booleana che segue la numerazione binaria $\langle 1 + 1 = 0 \rangle$, vuol dire che la disgiunzione di un'algebra di Boole non è l'ordinaria **alternativa** (1110), ma quella che abbiamo definita in Logica II, seguendo la terminologia di Bochenski la **disgiuntiva** (0110) che universalmente viene invece definita la disgiunzione **esclusiva** (*exclusive or*, **XOR**, simbolo: $\oplus$), ovvero la **negazione dell'equivalenza** (1001).
- ◆ Viceversa, sempre seguendo Bochenski, abbiamo definita l'esclusiva come corrispondente alla tavola di verità (0111) che è, semplicemente, la tavola di verità della **negazione della congiunzione** (1000) (Cfr. Logica II, §9.2.2).

- ◆ Così, interpretando x e y come interi in numerazione binaria, abbiamo le seguenti corrispondenze fra **logica proposizionale** e **logica equazionale** di un'algebra booleana:
 1. $x \wedge y = x \times y$
 2. $x \vee y = (x + y) + (x \times y)$
 3. $\neg x = 1 - x$
- ◆ Per questo, la **firma** di un'algebra booleana, ovvero le operazioni caratteristiche logico-aritmetiche su numeri interi $\mathbb{Z}$ è data dalla terna: $(\wedge, \vee, \neg)$.
- ◆ Gli altri connettivi logici possono essere infatti derivati da questi tre principali:
 4. $x \rightarrow y = \neg x \vee y$
 5. $x \oplus y = (x \vee y) \wedge \neg (x \wedge y)$
 6. $x \leftrightarrow y = \neg (x \oplus y)$
- ◆ È evidente dunque che i simboli $(0,1)$ di un'algebra booleana possono essere interpretati come altrettante **valutazioni falso (0)/vero (1)** di variabili proposizionali p , q della logica proposizionale.
- ◆ Di fatto però le valutazioni non sono su una singola variabile proposizionale p o q ma su un insieme di variabili equivalenti a p o a q . Ciò spiega le corrispondenze che

abbiamo già visto e discusso sia in § 2.2 sia in Logica II §11.2.2 fra **connettivi proposizionali** e **operazioni su insiemi**:

- ◆ $\leftrightarrow/\equiv$ **equivalenza/uguaglianza**
- ◆ $\neg/$ **negazione/complementazione**
- ◆ $\rightarrow/\subseteq$ **implicazione/inclusione**
- ◆ $\vee/\cup$ **disgiunzione (somma)/unione**
- ◆ $\wedge/\cap$ **congiunzione (prodotto)/intersezione.**
- ◆ Più precisamente ciò che caratterizza una logica booleana, che è alla base della logica computazionale o **logica equazionale (*equation logic*)**, è che in essa ogni formula, $\varphi := p \wedge q$; $\psi := p \vee q$; $\zeta := \neg p$; ..., del calcolo proposizionale del primo ordine (**k**) può essere tradotta in maniera standard nelle corrispondenti formule dell'algebra di Boole, ovvero, $a \times b$, $a + b$, $-b$... e della sua aritmetica binaria.
- ◆ Come sappiamo però queste formule booleane corrispondono di fatto ad altrettante **valutazioni semantiche 1/0**, (“è vero/falso che”), sulle formule del calcolo proposizionale corrispondente. Ovvero, rispettivamente: $\vdash p \wedge q$, $\vdash p \vee q$, $\vdash \neg p$, $\vdash \dots$ (dove il simbolo $\vdash$ (*turnstile*) sta per “è dimostrabile (=logicamente vero) che”), cioè a delle

valutazioni semantiche sulle formule proposizionali. Valutazioni che, nel simbolismo dell'**Algebra Universale**, si scrivono: $(\llbracket \varphi \rrbracket, \llbracket \psi \rrbracket, \llbracket \zeta \rrbracket, \llbracket \dots \rrbracket)$, per ciascuna delle formule $\varphi, \psi, \zeta, \dots$ del calcolo proposizionale del primo ordine considerate.

- ◆ Ciò significa che “i connettivi” di una logica equazionale di Boole $(\wedge, \vee, \neg)$ corrispondono a **operazioni su insiemi** $(\cup, \cap, \bar{})$, **non su individui** come i relativi connettivi proposizionali, quindi corrispondono ad altrettanti **operatori su insiemi** che per non essere confusi con quelli del calcolo proposizionale sono simbolizzati come fossero simboli di **operatori insiemistici** $(\cap, \cup, -)$, ovvero $(\wedge, \vee, -)$, cui bisogna aggiungere due operatori “vero” $(\top)$ e “falso” $(\perp)$, che corrispondono ai simboli binari $(1, 0)$ della **semantica algebrica** della logica booleana.
- ◆ Di qui la definizione della **firma** della cosiddetta **algebra booleana con operatori** (*Boolean algebra with operators, BAO*): $(\top, \perp, \wedge, \vee, -)$, definiti su una **struttura algebrica complessa** (topologia) di Boole $\mathbb{B}^+$ **algebra-sottoalgebra** di un'algebra booleana B che soggiace alle strutture insiemi-sottoinsiemi della corrispondente semantica estensionale.

2.3.6. Passaggio alla logica booleana modale e intensionale

- ◆ Tutto questo si deve ai fondamentali **teoremi di M. Stone** (Stone, 1936) e di **A. Tarski** (Jónsson & Tarski, 1952a; 1952b) grazie ai quali è possibile reinterpretare la semantica insiemistica della logica in termini di **semantica relazionale algebrica** e quindi estendere la logica algebrica booleana alla **logica modale relazionale di Kripke**, definire cioè una **logica booleana modale** (Goldblatt & Thomason, 1974; Thomason, 1975; Goldblatt, 1976a; 1976b; 2003) con le relative **interpretazioni intensionali** (aletiche, epistemiche, deontiche).
- ◆ Tutto questo costituisce la base in **informatica teorica** (*Theoretical Computer Science*, TCS) della possibilità di scrivere e implementare nei nostri calcolatori e nei sistemi di IA degli **algoritmi “buoni” di apprendimento automatico** (*Machine Learning*, ML).
- ◆ Che soddisfa cioè dei criteri di **ottimalità etica** di una **logica deontica algoritmizzata**. Tutto questo sarà oggetto della **Parte Terza** del Corso perché costituisce il cuore teorico della possibilità di un’**etica del ML** per sistemi di IA **autonomi** che prendono decisioni da loro stessi che hanno rilevanza etico-sociale-politico-economico-legale, ma è bene avervi qui accennato.

2.4. Dalla logica di Boole alla Macchina di Turing: il Calcolatore Universale (Davis, 2012), capp. VII-VIII

2.4.1. La logica di Boole e il concetto di porta logica

- ♦ La potenza del calcolo binario è che può essere **implementato** (= fisicamente realizzato) in un qualsiasi sistema fisico meccanico, elettrico, elettronico **a due stati mutualmente esclusivi** e che perciò costituisce un **calcolatore o computer**.
- ♦ Per esempio, **qualsiasi** delle seguenti righe di simboli può essere interpretata come il valore numerico binario di 667:

1	0	1	0	0	1	1	0	1	1
	—		—	—			—		
⊗	□	⊗	□	□	⊗	⊗	□	⊗	⊗
y	n	y	n	n	y	y	n	y	y

- ♦ Di fatto le prime **macchine calcolatrici** di largo uso furono dei calcolatori meccanici dove **diversi ingranaggi** composte di ruote dentate (denti=1, spazi=0) corrispondevano alle **quattro operazioni aritmetiche**.

- ◆ **Spostando a destra o a sinistra e ruotando** gli ingranaggi si potevano eseguire in successione (= ricorsivamente: $x_{n+1} = f(x_n)$) le quattro operazioni aritmetiche.

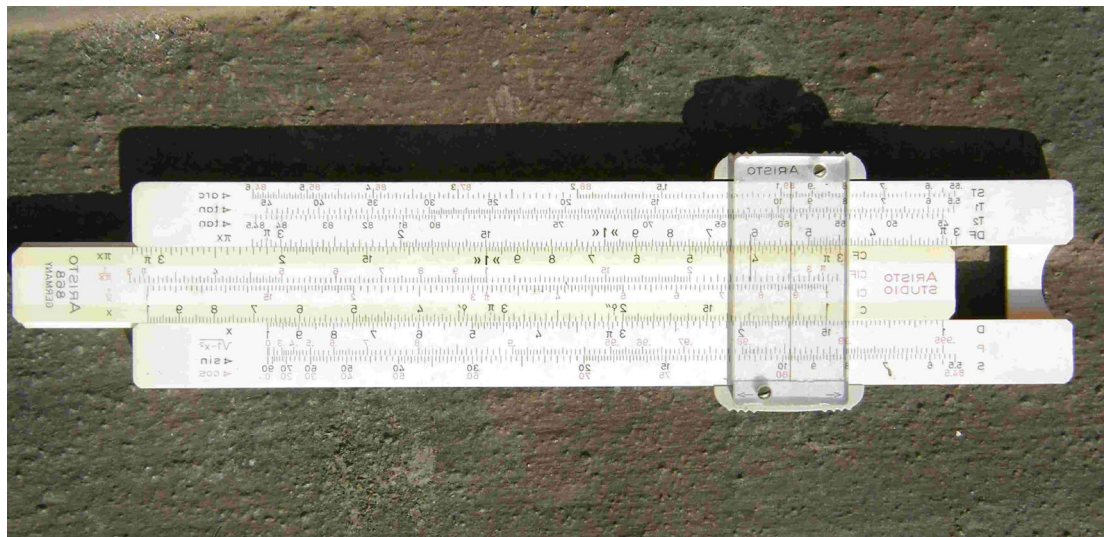


- ◆ Si tratta di un'implementazione meccanica di un algoritmo, ovvero di esecuzione di un **calcolo complesso** mediante una successione di **calcoli più semplici** secondo una **regola o serie di istruzioni ben definite**.
- ◆ Il concetto di «algoritmo» è all'origine della nozione stessa di **algebra** e quindi dell'idea folgorante di Cartesio alla base di tutta la matematica moderna che qualsiasi **figura geometrica** è la soluzione di un'**equazione algebrica** (polinomio,

p.es.: il calcolo del quadrato di un binomio $(a + b)^2 = (a^2 + b^2 + 2ab)$ che corrisponde alla costruzione di un quadrato che sia la somma di due quadrati di lato a e b).



- ♦ La sostanziale equivalenza fra figura geometrica (= modificazione di **grandezze continue**) e equazione algebrica (= modificazione di **grandezze discrete o numeriche**) ci fa comprendere anche la relazione esistente fra **calcolatore analogico** che calcola **simultaneamente o in parallelo** l'algoritmo modificando **una grandezza continua su differenti scale di misura** (p.es., un regolo calcolatore) e **calcolatore digitale** che calcola **serialmente** l'algoritmo mediante una **successione ordinata di calcoli numerici**.



- ◆ Di qui la necessità di sviluppare **processori digitali** sempre più veloci per implementare serialmente calcoli numerici complessi. La potenza dei nostri cervelli naturali è data dal fatto che esso **calcola in parallelo** usando **neuroni** ognuno dei quali, come vedremo, è assimilabile a un **processore analogico-digitale**.
- ◆ In questa luce, emerge ancora più luminosa l'intuizione di Boole che ha dimostrato che anche **la logica estensionale** (= calcolo delle classi) è ultimamente un'**algebra binaria** e che quindi un **calcolatore numerico** è anche – se opportunamente istruito – un **calcolatore logico** e quindi un **Calcolatore Universale**.

- ◆ Infine, per completare la panoramica, è fondamentale ricordare che come calcolatori invece di ingranaggi meccanici si possono usare **circuiti elettrici** con variazione discreta **on/off, 1/0** della grandezza continua della **tensione di una corrente elettrica** usando un controllo **passivo** di essa (resistenza, capacitanza, induttanza).
- ◆ Intuitivamente, usando un circuito di tre lampadine, due in input e una in output si può simulare la tavola di verità qualsiasi connettivo logico binario
- ◆ P.es., la disgiuntiva (1110) corrisponderà al fatto che la terza lampadina sarà spenta “se e solo se” *sse* tutt’e due le lampadine di input sono spente.
- ◆ Di fatto oggi si usano **calcolatori elettronici** che sono capaci di controllo **attivo** sul **flusso di elettroni** (corrente) perché capaci anche di **amplificare e rettificare** la corrente stessa. I primi calcolatori elettronici furono quelli **a valvole termoioniche**. Quelli attuali sono a **transistor** dove in ognuno è implementata una **porta logica binaria (funzione logica semplice e/o una funzione logica complessa (calcolo logico))**.
- ◆ Per completare il quadro, e capire immediatamente la connessione fra logiche di Boole e TCS, ecco gli schemi delle principali **porte logiche (*logical gates*)**, implementabili in altrettanti circuiti di un calcolatore, corrispondenti alle tavole di verità

dei fondamentali **connettivi logici** (o predicatori proposizionali) della logica delle proposizioni espresse in codifica numerica binaria (= logica equazionale):

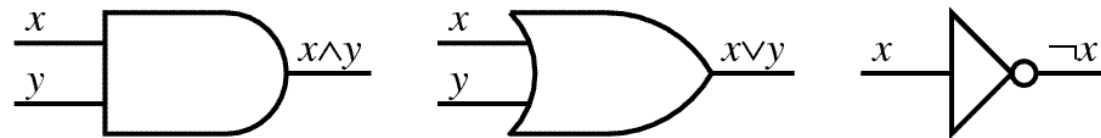


Figure 3. Logic gates

♦ E le due porte logiche corrispondenti alle **due leggi di De Morgan**:

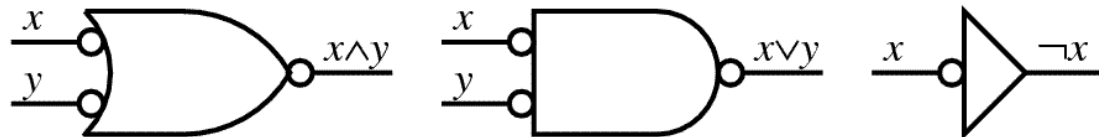


Figure 4. De Morgan equivalents

1. Come spiegato in Logica III, ma vi torneremo anche in seguito quando parleremo della **teoria dei filtri** in un **reticolo logico Booleano** essenziale per comprendere il ML, sono essenziali le due fondamentali **leggi di De Morgan** di logica delle proposizioni.

Esse riguardano i tre connettivi logici di **congiunzione**, **disgiunzione**, **negazione** cui, non casualmente, corrispondono i tre fondamentali **operatori insiemistici** di **intersezione**, **unione**, **complementazione** che definiscono la firma di una **logica di**

Boole, e quindi la fondamentale equivalenza fra **logica proposizionale** e **logica equazionale** che già abbiamo visto.

♦ **Informalmente**, le due leggi si leggono:

1. **non**(A o B) è lo stesso che (**non** A) e (**non** B)
2. **non**(A e B) è lo stesso che (**non** A) o (**non** B)

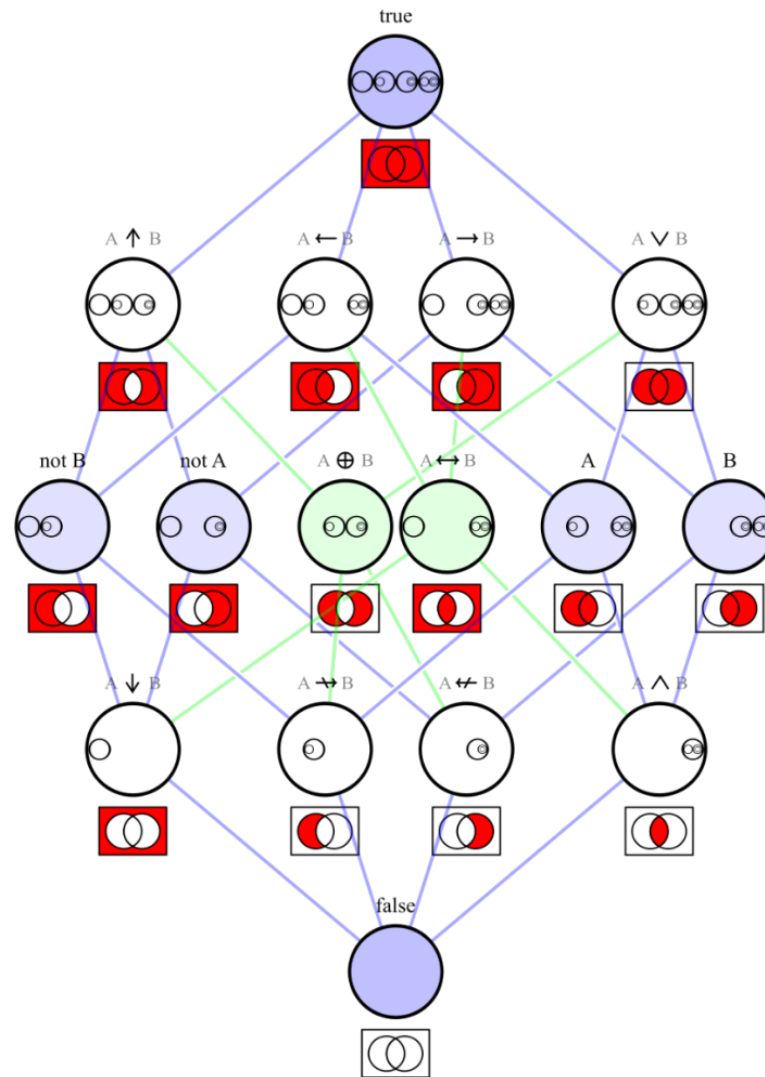
♦ **Formalmente**:

3. $\neg(p \vee q) \Leftrightarrow (\neg p) \wedge (\neg q)$
4. $\neg(p \wedge q) \Leftrightarrow (\neg p) \vee (\neg q)$

♦ Per completare, il quadro, occorre inserire anche **porta di implicazione** ($\neg x \vee y$), recentemente implementata, sia in ingegneria elettronica, che in bioingegneria:



♦ Di qui il seguente schema riassuntivo di tutte le principali **operazioni logiche** di logica delle proposizioni, con i relativi **diagrammi di Venn**, in quanto tutte incluse nel corrispondente **diagramma di Hasse** con cui diverremo più familiari quando parleremo della **teoria dei filtri** parlando delle reti neurali come cuore del ML:



2.4.2. Algoritmi e codifica binaria del linguaggio naturale

- ◆ Il cuore di ogni forma di calcolo è l'**algoritmo**, ovvero una serie di istruzioni che ci consentono di ridurre un calcolo logico e/o matematico complesso in una **sequenza ordinata ripetitiva o ricorsiva** di operazioni semplici.
- ◆ P.es., l'algoritmo con cui la maestra ci ha insegnato alle elementari a fare “a carta e penna” le moltiplicazioni “a n cifre” a partire dalle tabelline delle moltiplicazioni “a $n=1$ cifra” per ciascuna delle dieci cifre 1-9: «Metti i due numeri in colonna e comincia da destra a moltiplicare l' n -ma (ultima) cifra del secondo numero per ciascuna delle cifre del primo numero, scrivendo i risultati e i riporti su una riga. Poi su una seconda riga, scalando di un posto (o scrivendo uno 0), poni i risultati della moltiplicazione della $(n-1)$ ma (penultima) cifra del secondo numero per ciascuna cifra del primo numero e così via **ricorsivamente** sino alla prima cifra del secondo numero. Alla fine, il risultato è dato dalla somma di tutte le righe dei prodotti a 1 cifra».

$$\begin{array}{r}
 28 \times \\
 15 = \\
 \hline
 140 + \\
 28 \cdot \\
 \hline
 420
 \end{array}$$

- ◆ Da Logica I e II abbiamo imparato – e ora ne sappiamo il perché – che questo vale anche per il **calcolo logico** usando ricorsivamente invece delle tabelline le tavole di verità in simboli binari dei connettivi logici.
- ◆ Un **calcolatore** non è altro che un sistema fisico meccanico o elettronico che esegue le istruzioni ricorsive dell'algoritmo **scritte in codice binario** per eseguire un qualsiasi calcolo complesso logico-matematico.
- ◆ L'insieme delle istruzioni (algoritmi) costituisce il **programma di calcolo** di quel calcolatore. La grande invenzione di Turing è quella del **calcolatore multi-programmabile** che, come vedremo, corrisponde al costrutto matematico ideale della **Macchina Universale di Turing** (*Universal Turing Machine*, UTM).

- ◆ Il cuore del **calcolo automatico** consiste quindi nella possibilità di una **codifica binaria** dei simboli di un testo (lettere, segni di interpunzione, a capo, etc.) usando generalmente per ciascun simbolo linguistico **stringhe di bit di lunghezza fissa**.
- ◆ Per esempio, mediante stringhe di $2^3=8$ bits definito **byte** si possono rappresentare fino a 256 combinazioni numeriche con cui rappresentare non solo le 24 lettere dell'alfabeto, ma una molteplicità di segni linguistici e di comandi per il calcolatore.
- ◆ Il più famoso e semplice **codice binario** di questo tipo che usa stringhe di 7 bit (l'ottavo è per lo spazio fra simboli) è il **codice ASCII** (*American Standard Code for Information Interchange*) che è disponibile su qualsiasi programma di scrittura per calcolatori.
- ◆ Per esempio, la lettera “a minuscola” in codice ASCII corrisponde alla stringa binaria 01100001 che in codice decimale corrisponde al numero “97”.
- ◆ Ecco il nome “Wikipedia” scritto in codice ASCII:

```
01010111 01101001 01101011  
01101001 01110000 01100101  
01100100 01101001 01100001
```

- ◆ Altri codici binari largamente usati informatica sono quelli a 8 e 16 bit. Ovvero, il codice a **base-8** che usa le 8 cifre 0-7 per rappresentare numeri (**codice ottale: 1 byte**) e a **base-16** che usa 16 simboli-base, le 10 cifre 0-9 e le 6 lettere A-F (**codice esadecimale: 2byte**) per rappresentare numeri.
- ◆ Di fatto però **il linguaggio-macchina** usato dalle unità di calcolo (CPU- *Central Processor Unit*) dei nostri calcolatori elettronici è proprio il **codice esadecimale**. In ogni calcolatore esiste sempre perciò **un'unità automatica di traduzione** del linguaggio alfanumerico in cui è scritto il programma nel linguaggio esadecimale parlato dal calcolatore.
- ◆ Per questo, i veri programmatori esperti, sono quelli capaci di scrivere il programma direttamente in **linguaggio esadecimale** velocizzando così enormemente l'esecuzione del calcolo.

2.4.3. La “Macchina di Turing” e il Calcolatore Universale

- ◆ In un articolo del 1936 (Turing A. M., On Computable Numbers, with an Application to the Entscheidungsproblem, 1937) che ha cambiato la storia del mondo il matematico inglese Alan Turing definì uno schema astratto di **macchina automatica (*automatic machine*) a stati finiti** capace di calcolare

qualsiasi **funzione ricorsiva** simulando la procedura algoritmica di calcolo “a carta e penna” dell’essere umano. Questa macchina fu poi definita da Alonso Church **Macchina di Turing (*Turing Machine*, TM)**.

♦ Le componenti della macchina sono quanto mai semplici, tenuto conto che esiste **una TM per ognuna delle funzioni ricorsivamente computabili**:

1. **Un nastro finito (memoria)** suddiviso in celle ognuna delle quali può contenere un simbolo 0,1, in base ad un **alfabeto finito** di combinazioni di simboli binari.
2. **Una testina di lettura/scrittura sul nastro (CPU)** capace di 4 comportamenti: **muoversi a destra/sinistra** di una posizione sul nastro, **scrivere/leggere un simbolo** sul nastro.
3. **Un insieme finito di istruzioni (programma)** che determina per ciascun passo di computazione il comportamento della testina.

♦ Ciascuno degli **stati di una TM** sono definiti perciò dalla **quintupla** (relazione a cinque argomenti): $T(s_w, q_w | s_v, q_v, A)$

- Dove $s_w \equiv$ simbolo attuale, $q_w \equiv$ posizione attuale, $s_v \equiv$ simbolo precedente, $q_v \equiv$ posizione precedente, $A \equiv$ insieme finito istruzioni per passare da uno stato all’altro.

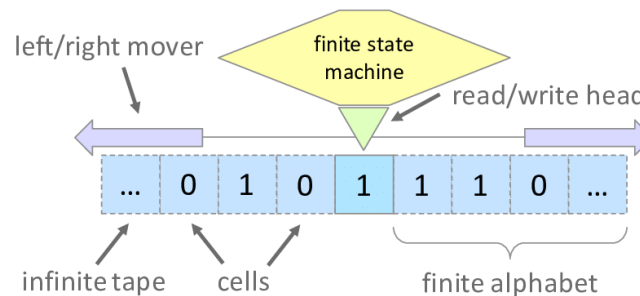


Figura 1. Schema intuitivo di una macchina di Turing

- ◆ Una TM in **teoria degli automi** si definisce perciò una **macchina a transizione di stati determinati** (*labelled state transition (LST) machine*). Infatti una TM computa a partire da uno stato iniziale x_0 fino a uno stato finale $x_n + 1$ che determina la **fermata** (*halting*) della TM e quindi della computazione da essa eseguita.
- ◆ **Fisicamente**, una TM può essere implementata in un **sistema lineare** deterministico, visto che ogni stato è definito univocamente dal precedente dalla funzione lineare che definisce il sistema (= **automa cellulare**).
- ◆ **Matematicamente**, una TM soddisfa alla cosiddetta **Tesi di Church** secondo la quale ogni funzione **effettivamente computabile** è **ricorsivamente computabile**

→ **Tesi di Church-Turing:** ogni funzione ricorsivamente computabile è **computabile da una TM**.

- ◆ In un successivo articolo del 1948 Turing propone la costruzione della **Macchina Universale di Turing (UTM)** che in effetti costituisce il modello astratto di ogni **computer programmabile**.
 - Infatti, ogni MT può simulare i computi di un'altra MT nella misura in cui l'insieme delle istruzioni della seconda MT è **definito sul nastro della prima MT**.
 - In tal modo, si può ipotizzare l'esistenza di una UTM **dotata di un nastro infinito** in cui sono definite le istruzioni per simulare ogni possibile TM.
 - Tuttavia – evidenzia lo stesso Turing – l'UTM è limitata dal cosiddetto **teorema della fermata**, diretta derivazione dai **Teoremi di Incompletezza di Gödel** ed in particolare dal **Secondo teorema di Incompletezza** che dimostra l'impossibilità che l'aritmetica ricorsivamente formalizzata possa essere **metalinguaggio di se stessa** (= impossibilità delle cosiddette **funzioni ricorsive generali**).

- Ovvero, funzioni aritmetiche definite ricorsivamente **sull'intero insieme dei naturali** $\{\mathbb{N}\}$. In tal caso, infatti, la meta-funzione che sintetizza i computi della UTM dovrebbe, da una parte avere per argomento l'intero insieme $\{\mathbb{N}\}$, dall'altra computare ricorsivamente un numero n – il cosiddetto **numero di Gödel** dell'insieme $\{\mathbb{N}\}$ che, ricordiamo, è **un'infinità numerabile** – trovandosi perciò **nell'impossibilità di decidere se n appartiene o no a $\{\mathbb{N}\}$** , il che contraddice la numerabilità stessa di $\mathbb{N}$.
- ◆ Tutto ciò si può intuire molto bene nel caso dell'**impossibilità della fermata per la UTM**.
 - Supponiamo di avere **una lista infinita** di tutte le funzioni aritmetiche ricorsivamente computabili $\Psi_x(x_0)$, dove sia **l'argomento** che **l'indice** della funzione sono numeri naturali, ovvero $x \in \mathbb{N}$.
 - L'UTM può dunque definirsi come un'ulteriore funzione $\Psi_x(x_0) + 1$ che lavora in questo modo. La funzione scorre la lista delle funzioni aritmeticamente computabili. Trovata una di queste, la computa per l'intero suo dominio $x_0, \dots, x_n$. Arrivata alla fine **termina la computazione** ovvero computa lo stato finale $x_n + 1$.

- L'indecidibilità (contraddizione) appare immediatamente allorché ci domandassimo se la nostra (meta-)funzione $\Psi_x(x_0) + 1$ che definisce i computi della UTM appartiene o no alla lista supposta **completa** delle funzioni aritmetiche ricorsivamente computabili.
- Se così fosse essa dovrebbe essere una delle $\Psi_x(x_0)$ della lista, ma in tal caso avremmo **la contraddizione** che (una delle) $\Psi_x(x_0) = \Psi_x(x_0) + 1$.
- Questa indecidibilità significa che **non si può dimostrare la fermata della UTM**.
- ◆ Matematicamente questo significa che le funzioni aritmetiche effettivamente computabili non sono le generali ricorsive, ma le **parziali ricorsive**, definite cioè su un **sottoinsieme dei naturali** come un discepolo di Church, Stephen C. Kleene affermò nel suo capolavoro *Introduction to Metamathematics* (Kleene, 1952).
- ◆ Da qui una ridefinizione di Kleene della Tesi di Church secondo cui “*ogni funzione parziale che è effettivamente computabile, ovvero che è un predicato effettivamente decidibile (nel senso che esiste un algoritmo per il quale il suo valore può essere calcolato per ogni n-pla che appartiene al suo dominio di definizione) è parziale ricorsiva*” (Kleene, 1952, p. p. 317).

- ◆ Una ridefinizione della Tesi di Church che Gödel stesso definì un vero e proprio “miracolo matematico” perché rendeva possibile **la definizione consistente di funzione ricorsiva effettivamente computabile**, anche se al costo di abbandonare per sempre la nozione di nozione di ricorsività “generale”.
- ◆ Anche se non è matematicamente possibile – almeno con gli assiomi dell’aritmetica ordinaria – far sì che una UTM definisca per se stessa il dominio delle funzioni parziali ricorsivamente computabili, lasciando così irrisolto il “problema della fermata” per una UTM e contemporaneamente indimostrabile la Tesi di Church anche nella versione di Kleene.
- ◆ In soldoni questo significa che, sebbene l’UTM sia il modello astratto di qualsiasi moderno **computer programmabile** l’impossibilità di risolvere il problema della fermata per una UTM anche nel caso delle ricorsive parziali, **non può esistere un computer auto-programmabile**.
- ◆ Quindi guai a considerare gli algoritmi di ML o di “apprendimento automatico” come modelli di computer auto-programmabili, **sebbene riducano – e di molto – l’intervento del programmatore umano**.

- ◆ E cosa ne è della **Tesi di Turing**? Effettivamente esiste un modo per sfuggire all'indcidibilità del Teorema della Fermata per l'UTM individuato da Turing stesso e per la dimostrazione del quale Church invitò il giovane Turing a fare un dottorato a Princeton sotto la direzione di Church stesso.
 - Anche se poi il lavoro di dottorato di Turing non fu su questo tema ma su una logica definita sugli ordinali e quindi sulla cosiddetta **induzione transfinita** basata su cardinalità degli insiemi che vanno oltre la cardinalità $\aleph_0$ del numerabile (la cardinalità di $\mathbb{N}$) ma su cardinalità maggiori (p.es. ω_1) per diversi **ordini di transfinito** $\omega_1, \omega_2, \omega_3, \dots, \omega_n, \dots, \omega_\omega, \dots$. Dove ciascun numero cardinale n (insieme di cardinalità n) “contiene” sotto di sé infiniti cardinali di ordine transfinito minore (p.es. $n\omega_1$ contiene sotto di sé infiniti $n \in \mathbb{N}$ appartenenti cioè alla cardinalità del numerabile $\aleph_0$).
 - È chiaro che una macchina capace di siffatta **ricorsività (induzione) transfinita** è capace in linea di principio di risolvere il problema della fermata per funzioni ricorsive definite su $\mathbb{N}$ anche se a patto di **non essere più una macchina a stati finiti**.

- ◆ Più radicalmente il **modo definitivo** per risolvere il problema della fermata (che nel caso precedente si riproporrebbe per ciascun ordine di transfinito) consiste nella cosiddetta **TM non-determinista (NTM)** che praticamente consente che ad ogni passo della computazione non una ma **una molteplicità di scelte possibili**. Sono possibili diverse implementazioni di NTM, p.es., definire le relazioni di scelta sui **numeri reali** $\mathbb{R}$ visto che l'insieme $\mathbb{R}$ è non-numerabile né tantomeno ricorsivamente computabile.
 - Oppure definire le computazioni di una TM su **molteplici nastri** (effettivamente almeno 3) tali che a ogni passo della computazione la macchina potrebbe saltare in maniera non predicibile su un altro nastro.
 - Si discute ancora a lungo se un **computer quantistico** possa essere un'implementazione di una NTM, ma si tratta di questioni teoriche che qui non interessano.

3. Il Test di Turing e il Programma di Ricerca dell'IA Simbolica

3.1. Il Test di Turing

- ♦ Il **Test di Turing** fu originariamente definito da Turing come **il gioco dell'imitazione** (*the imitation game*) nell'articolo in cui lo ha definito (Turing A. M., 1950).
- ♦ È un test sulla capacità di una macchina **di esibire un comportamento intelligente equivalente o indistinguibile da quello di un essere umano**.
- ♦ In pratica, Turing propose che un valutatore umano avrebbe dovuto **giudicare le conversazioni in linguaggio naturale tra un essere umano e una macchina** progettata per generare risposte simili a quelle umane.
- ♦ Il valutatore sarebbe stato consapevole che uno dei due partner nella conversazione era una macchina, ma siccome tutti i partecipanti sarebbero stati separati l'uno dall'altro, il valutatore non avrebbe saputo in anticipo **chi dei due interlocutori fosse una macchina**.

- ♦ La conversazione **sarebbe stata limitata a un canale di solo testo** usando una tele-scrivente; quindi, il risultato non sarebbe dipeso dalla capacità della macchina di rendere le parole in forma di discorso parlato.
- ♦ Se il valutatore **non era in grado di distinguere** in modo affidabile la macchina dall'uomo, **allora la macchina avrebbe superato il test**.
- ♦ I risultati del test, quindi, non dipendevano dalla capacità della macchina di dare **risposte corrette** alle domande, ma solo da quanto **le sue risposte assomigliassero a quelle che avrebbe dato un essere umano**.

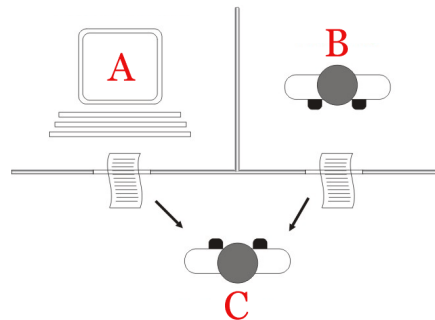


Figura 2. Schema intuitive del Test di Turing: A: macchina; B; interlocutore umano; C: valutatore umano.

- ◆ Turing iniziava il suo articolo con la provocatoria domanda “**se una macchina può pensare**”, ma invece di partire come tradizionalmente si sarebbe fatto con le definizioni di “macchina” e “intelligenza”, preferisce sostituire la questione iniziale con un'altra “**strettamente in relazione con la prima**, ma che può essere espressa in maniera **non ambigua**”. Ovvero: “**Le macchine possono fare** ciò che noi (come entità pensanti) **possiamo fare?**” (Turing A. M., 1950).
- ◆ Quello che è interessante è che Turing dedica tutta la seconda parte dell'articolo a rispondere a **diverse obiezioni** che si sarebbero potuto fare al test che praticamente anticipano le obiezioni che in seguito sono state fatte al programma di ricerca dell'IA. Elenchiamo qui le più rilevanti con le risposte di Turing:
 - **Obiezione teologica:** Questo afferma che il pensiero è una funzione dell'anima immortale dell'uomo e quindi una macchina non potrebbe pensare. Turing risponde dicendo che non vede alcun motivo per cui non sarebbe possibile per Dio concedere un'anima a un computer se lo desiderasse.
 - **Obiezioni matematiche:** Questa obiezione utilizza teoremi matematici, come il teorema di incompletezza di Gödel, per dimostrare che ci sono limiti a quali domande un sistema informatico basato sulla logica può rispondere. Turing

suggerisce che gli esseri umani sono troppo spesso sbagliati e contenti della fallibilità di una macchina.

- **Obiezioni meccaniche:** Una macchina sufficientemente veloce con una memoria sufficientemente grande potrebbe essere programmata con un numero abbastanza grande di domande umane e risposte umane per fornire una risposta umana a quasi tutte le domande, e una vaga risposta casuale alle poche domande non nella sua memoria. Ciò simulerebbe la risposta umana in modo puramente meccanico. Gli psicologi hanno osservato che la maggior parte degli esseri umani ha un numero limitato di risposte verbali.
- **Obiezione al trattamento dei dati:** le macchine elaborano i dati bit per bit. Gli esseri umani elaborano i dati in modo olistico. In questa visione, anche se una macchina appare umana in ogni modo, trattarla come umana significa indulgere nel pensiero antropomorfo.
- **Obiezione dell'intelligenza cosciente:** Questo argomento, suggerito dal professor Jefferson Lister afferma, "fino a quando una macchina non può scrivere un sonetto o comporre un concerto **a causa dei pensieri e delle emozioni provate**, e non per un uso casuale dei simboli, potremmo essere

d'accordo sul fatto che la macchina è uguale al cervello". Turing risponde dicendo che non abbiamo modo di sapere che qualche individuo diverso da noi stessi provi emozioni se non per il comportamento che esibisce, e che quindi dovremmo accettare il test.

- **Obiezione di Lady Lovelace:** Una delle obiezioni più famose, afferma che i computer sono incapaci di **originalità**. Ciò è in gran parte dovuto al fatto che, secondo Ada Lovelace, le macchine sono incapaci di apprendimento indipendente. Turing contraddice questo sostenendo che l'ipotesi di Lady Lovelace è stata influenzata dal contesto da cui ha scritto, e se esposta a conoscenze scientifiche più contemporanee, diventerebbe evidente che la memoria del cervello è abbastanza simile a quella di un computer. Turing risponde inoltre che i computer potrebbero ancora sorprendere gli esseri umani, in particolare dove le conseguenze di fatti diversi non sono immediatamente riconoscibili.
- **Informalità del comportamento:** questo argomento afferma che qualsiasi sistema governato da leggi sarà prevedibile e quindi non veramente intelligente. Turing risponde affermando che questo sta confondendo le leggi di comportamento con le regole generali di condotta, e che se su una scala

abbastanza ampia (come è evidente nell'uomo) il comportamento della macchina diventerebbe sempre più difficile da prevedere. (Ricerche successive sugli algoritmi ricorsivi hanno scoperto che, in ogni caso, i sistemi deterministici sono capaci di comportamenti caotici assolutamente imprevedibili).

- **Percezione extrasensoriale:** Turing sembra suggerire che ci siano prove di percezione extrasensoriale. Tuttavia, egli ritiene che si potrebbero creare condizioni in cui ciò non influirebbe sul test. In ogni caso, ciò che è scientificamente vero in quelle che consideriamo percezioni extrasensoriali è l'esistenza di fenomeni di *entanglement* fra cervelli sia naturali che artificiali che comunicano fra di loro a distanza su base quantistica, come accenneremo, che è la base della teoria quantistica delle comunicazioni. Il presente e il futuro della teoria delle comunicazioni!

◆ **In sintesi,** la risposta di Turing alle obiezioni è che il test riguarda la simulazione artificiale di comportamenti umani che seguono regole e calcoli logici e che noi umani affermiamo essere espressioni della nostra intelligenza cosciente. Ma la coscienza è solo **soggettiva** mentre la **logicità** del comportamento intelligente no. Ed

è in questo senso che è scientificamente corretto parlare di “**intelligenza artificiale**”. Discettare di coscienza nelle macchine invece è tempo perso!

- ◆ In ogni caso, l'ipotesi soggiacente al test che Douglas Hofstadter definirà in seguito **il dogma dell'IA** è quella secondo cui i calcoli eseguiti da un programma che gira su una macchina **concreta realizzazione di una UTM** e che supera il test di Turing **devono possedere qualche essenziale isomorfismo** con i calcoli eseguiti dal cervello umano.

3.2. La Conferenza di Dartmouth e la nascita del Programma di Ricerca dell'IA

- ◆ Il *Dartmouth Summer Research Project on Artificial Intelligence* promosso nel 1956 dal prof. John MacCarthy, allora assistente di matematica al Dartmouth College in New Hampshire, e per il quale egli trovò il finanziamento dalla prestigiosa *Rockefeller Foundation* viene generalmente considerata **la data di nascita dell'IA**.
- ◆ Il progetto è durato circa sei-otto settimane **ed è stato essenzialmente una lunga sessione di brainstorming**. Undici matematici e scienziati avevano inizialmente

programmato di partecipare. Non tutti hanno partecipato, ma più di dieci fra i massimi esperti del settore sono intervenuti per periodi più o meno lunghi.

- ◆ Gli undici che inizialmente avevano scelto di partecipare costituivano il Gotha dell'informatica, della cibernetica e della teoria dell'informazione del tempo. La lista inclusa nella Proposta alla Rockefeller infatti era la seguente:

- Per l'intero periodo:
 - 1) Dr. Marvin Minsky
 - 2) Dr. Julian Bigelow
 - 3) Professor Donald M. MacKay
 - 4) Ray Solomonoff
 - 5) Sig. John Holland
 - 6) Dr. John McCarthy
- Per quattro settimane:
 - 7) Dr. Claude Shannon
 - 8) Sig. Nathaniel Rochester
 - 9) Signor Oliver Selfridge
- Per le prime due settimane:

- 10) Allen Newell
 - 11) Professor Herbert Simon (futuro Premio Nobel in Economia)
- ◆ La Proposta presentata da McCarthy, Minsky, Rochester e Shannon affermava fra l'altro:
- *Proponiamo che uno studio di 2 mesi e 10 persone sull'intelligenza artificiale sia condotto durante l'estate del 1956 al Dartmouth College di Hanover, nel New Hampshire. Lo studio deve procedere sulla base della congettura che ogni aspetto dell'apprendimento o qualsiasi altra caratteristica dell'intelligenza può in linea di principio essere descritto in modo così preciso da poter essere costruita una macchina per simularlo. Si cercherà di trovare il modo di far sì che le macchine usino il linguaggio, formino astrazioni e concetti, risolvano tipi di problemi ora riservati agli esseri umani e migliorino se stesse. Pensiamo che un progresso significativo possa essere fatto in uno o più di questi problemi se un gruppo accuratamente selezionato di scienziati lavora insieme su di esso per un'estate (McCarthy, Minsky, Rochester, & Shannon, 1955).*

- ♦ La Proposta proseguiva discutendo di **computer, elaborazione del linguaggio naturale, reti neurali, teoria della computazione, astrazione e creatività**.
- ♦ Tutte aree che nel campo dell'IA sono considerate tuttora rilevanti come pure i diversi temi elencati nella Proposta. E d'altra parte non poteva essere altrimenti vista **la qualità indiscussa dei partecipanti al workshop** che negli anni seguenti, in diverse università, laboratori e industrie **hanno fatto la storia dell'IA**.

3.2.1. I sistemi esperti, la programmazione ad oggetti e l'IA simbolica

- ♦ Quando parliamo di **IA simbolica** intendiamo sistemi di IA basati sulla **programmazione esplicita** e quindi su **calcoli logico-simbolici** in cui la macchina non esegue alcun **compito di apprendimento automatico (ML)** sulla base dati disponibile e/o da dati in arrivo da reti di sensori, ma solo **compiti inferenziali sui dati** in base al programma.
- ♦ I più diffusi sistemi di IA di questo tipo, largamente usati in molteplici campi della ricerca, dell'industria, dell'amministrazione, dell'educazione e del commercio sono i cosiddetti **sistemi esperti (SE)**.

- ◆ Sono **sistemi basati sulla conoscenza** (*knowledge engineering*) capaci di risolvere problemi in un dominio molto specifico e, proprio per l'ambito applicativo limitato, **possono raggiungere e superare le prestazioni umane di un esperto** dello stesso dominio.
- ◆ In altri termini, questi sistemi **riproducono artificialmente** le prestazioni di **una persona esperta** di un determinato dominio di conoscenza o campo di attività e per questo rientrano nel campo dell'IA. Per questo costituiscono un indispensabile supporto decisionale per **trovare una soluzione ottimale a problemi complessi** senza l'intervento dell'esperto in quel settore.
- ◆ Una peculiarità dei SE è il **carattere trasparente** (*Glass Box*) delle decisioni cui giunge il sistema. Ovvero, è **sempre in grado di esibire i passaggi logici** che sottendono alle decisioni prese.
- ◆ In questo senso sono profondamente diversi dai sistemi dell'IA **non-simbolica** ovvero dai più recenti sistemi di ML basati sulle reti neurali e il *deep learning* per i quali si parla di **Black Box** perché non è possibile risalire in essi al processo induttivo-deduttivo che li porta a produrre una determinata decisione, con tutti i **problemi innanzitutto etico-legali** che ne derivano.

- ◆ D'altra parte, è la stessa cosa che accade nei **processi intuitivi pre-logici** di una mente umana ed è per questo che si richiede ad essa una **riflessione razionale esplicita (ragionamento logico)** quando sono in ballo decisioni con **responsabilità etico-legale** di cui **rendere conto** agli altri (*ethical responsibility as ethical accountability*).
- ◆ Generalmente i SE si suddividono in due grandi classi:
 1. **SE basati su regole classiche di inferenza *if...then***. P.es.: se poniamo al SE la domanda: “qual è il mio problema di salute?”, occorrerà fornire al SE una serie di dati che esso è istruito a chiedere (ho mal di testa, sono raffreddato, la mia temperatura è di 38°C). Essi costituiranno una congiunzione di condizioni (*if*) data la quale la risposta (*then*) del SE al mio problema sarà: “hai l'influenza”.
 2. **SE basati su alberi inferenziali**. in questo caso, dato un insieme di dati ed alcune deduzioni, **il sistema esperto crea un albero (delle possibili alternative *if...then; and...or*) che classifica i vari dati**. Di fronte ad un problema nuovi dati vengono analizzati dall'albero e il nodo finale rappresenta la soluzione.
- ◆ Un sistema ad albero che in qualche modo anticipa i sistemi di ML è basato sulla famosa teoria di M. Minsky dei *frame systems*, ovvero strutture ad albero in cui le

sotto-strutture dei nodi inferiori dell'albero possono venire aggiornati automaticamente dal sistema in base ai nuovi dati inseriti.

◆ Citando direttamente dall'articolo di Minsky:

- *Incontrando una nuova situazione (o apportando un cambiamento sostanziale alla propria visione di un problema), si seleziona dalla memoria una struttura chiamata frame. Questa è una struttura memorizzata da adattare alla realtà modificando i dettagli secondo necessità. Un frame è una struttura di dati per rappresentare una situazione stereotipata come trovarsi in un certo tipo di stanza o andare alla festa di compleanno di un bambino. Allegati a ciascun frame ci sono diversi tipi di informazioni. Alcune di queste informazioni riguardano come utilizzare il frame. Altre riguardano ciò che ci si può aspettare che accada dopo. Altre ancora riguardano cosa fare se queste aspettative non sono confermate. Possiamo pensare a un frame **come a una rete di nodi e relazioni**. I "livelli superiori" della struttura sono fissi e rappresentano cose che sono sempre vere sulla presunta situazione. I livelli inferiori hanno molti terminali o "slot" che devono essere riempiti da istanze o dati specifici. Ogni terminale può specificare le condizioni che i suoi compiti*

devono soddisfare. (Le assegnazioni stesse sono di solito "sotto-strutture" (sub-frames) più piccole) (Minsky, 1975b, p. 1).

- ◆ Infine, è bene ricordare che la teoria dei frame di Minsky intende essere esplicitamente una versione algoritmica della teoria dello **schematismo a priori** della *Critica della Ragion Pura* di Kant che è il punto essenziale della *Critica*
 - Quello in cui Kant illustra come i fenomeni percettivi sono **sussunti sotto diverse categorie** applicando ai fenomeni questi “schemi temporali” che sono effettivamente **strutture ricorsive** essendo per Kant lo schematismo temporale la base per la fondazione della nozione di numero naturale n mediante lo schema $n + 1$.
 - Dice Kant, le immagini mentali di oggetti del senso interno non sono **percezioni sbiadite** di oggetti del senso esterno come in Hume e Locke, ma schemi temporali in cui ordinare le diverse sensazioni memorizzate (colori, forme, suoni, etc.) così che ogni schema corrisponde a un oggetto (p.es., l’immagine mentale di cane) e lo stesso schema – come i subframe di Minsky – possono applicarsi a diverse sensazioni del medesimo oggetto (p.es., cani di diverse forme, colori, latrati, etc.).

- Praticamente, questo è il ragionamento di Minsky, ogni diverso schema temporale di Kant per ciascuna categoria corrisponde a una diversa f della relazione ricorsiva di ordinamento $x_{n+1} = f(x_n)$.
- La categoria corrisponde ai livelli più alti non modificabili del frame mentre le sub-strutture (*sub-frames*) corrispondono agli schemi temporali in cui inserire dati diversi per il medesimo frame-oggetto.
- ◆ Si intuisce come la teoria dei frame di Minsky sia il progenitore della **programmazione ad oggetti** diffusissima oggi in diversi linguaggi di programmazione (C, Java, Python, etc.).
- ◆ Il passaggio al ML basato sulle **reti neurali** consisterà così nella possibilità di ridefinire autonomamente da parte del sistema **la struttura completa dell'albero inferenziale in base ai dati disponibili** e non solo i livelli inferiori della struttura, senza intervento del programmatore come invece si richiede alla teoria dei frame.
 - Filosoficamente, questo corrisponde al passaggio in psicologia cognitiva dallo schematismo a priori kantiano allo **schematismo induttivo** della psicologia dell'intelligenza di **Jean Piaget** nel suo famoso libro omonimo pubblicato nel 1947 (Piaget, 2011), riproposizione moderna dello **schematismo induttivo**

aristotelico-tomista della percezione, come per primo Cornelio Fabro si accorse in Italia (Fabro, 1941) e in seguito e indipendentemente in ambito informatico Donald MacCrimmon MacKay – uno degli invitati alla Conferenza di Dartmouth (MacKay, 1968).

- Fra parentesi è questo il centro teoretico della mia tesi di laurea sul rapporto fra reti neurali e psicologia aristotelico-tomista dell'intenzionalità (Basti, 1991)

◆ Ma per questo occorre prima illustrare il passaggio dalla IA simbolica alle **scienze cognitive** e quindi alle **neuroscienze cognitive** e dunque all'**IA pre-simbolica**.

3.3. La nascita delle scienze cognitive

3.3.1. Dal Test di Turing e dal “dogma dell'IA” alla nascita delle scienze cognitive

- ◆ Grazie alla teoria dei frame di Minsky e al “dogma dell'IA” diventa possibile individuare il referente di un simbolo del calcolo logico in informatica non in un **termine** ma in una **struttura ricorsiva di manipolazione dei dati**, in un **algoritmo del software** di una UTM.

- ◆ Di qui, storicamente la reinterpretazione del classico rapporto "mente-corpo" in termini di rapporto "software-hardware", ovvero **la metafora della mente come "software" del cervello "wetware"** (hardware implementato su materiale biologico e non sul silicio) **della mente**. (Putnam, 1960).
- ◆ Da tutto questo, negli anni '60 del secolo scorso, **nasce il programma di ricerca sulle scienze cognitive come la "nuova" scienza della mente**, caratterizzata da un approccio non riduzionista, rispetto all'approccio fisicista della cosiddetta "teoria dello stato centrale (Gardner, 1984)" di Herbert Feigl, proposta alla fine degli anni '50 (Feigl, 1958) nell'ambito del **movimento neo-positivista**.
 - Feigl, infatti, uno dei membri più giovani del *Wiener Kreis* degli anni '30, culla del "movimento neopositivista" in Europa all'inizio del XX secolo, dopo il suo trasferimento negli Stati Uniti, fondò all'Università di Minneapolis negli anni '50 il "**Minnesota Center for Philosophy of Science**".
 - Divenne anche redattore capo della prestigiosa collezione dei "**Minnesota Studies of Philosophy of Science**" che sopravvisse alla morte del suo fondatore (1988) fino ad oggi.

- In questa raccolta, i risultati della ricerca del Centro, e del movimento del cosiddetto “**empirismo logico**” originato dalle attività del Centro, sono stati pubblicati per diversi decenni².
- Nel II Volume della collana pubblicato nel 1958, dedicato al problema mente-corpo, sono raccolti due contributi fondamentali.
 1. Il primo era l'articolo di Feigl già citato.
 2. L'altro era di Wilfrid Sellars ed era dedicato all'analisi logica del rapporto tra **l'intenzionale e il mentale** (Sellars, 1958), in cui giustamente sottolineava che **il discorso in "prima persona (singolare/plurale)"** (cioè il cosiddetto *I/we talk*) che **esprime gli stati mentali intenzionali (con la "z")** dei soggetti cognitivi individuali/collettivi, suppone **una logica intensionale (con la "s")**.
- Questo rende **logicamente inconsistente** qualsiasi tentativo materialistico di **identificare (mediante un'equivalenza logica) uno stato mentale intenzionale, con uno stato cerebrale osservato**.
- Il linguaggio osservativo della neurofisiologia, suppone, infatti, la **logica estensionale standard** delle scienze matematiche pure e applicate.

- In una parola, lo “***I/we-talk***” in **prima persona** dell'intenzionalità non può essere ridotto sistematicamente allo “***O-talk***” in **terza persona** del linguaggio osservativo delle scienze neurofisiologiche, nella loro ricerca dei correlati neurali degli stati mentali intenzionali.
- ◆ La distinzione di Sellars tra ***I/we-talk*** del linguaggio mentalistico che esprime gli stati coscienti intenzionali e l'***O-talk*** del linguaggio osservativo dell'indagine neurofisiologica, ha influenzato sistematicamente le ulteriori riflessioni filosofiche sulle relazioni tra la **mente intenzionale e il cervello**.
- ◆ Prima di tutto, ha influenzato lo stesso **H. Feigl**, come vedremo immediatamente, ma anche **John Searle**, (Searle, 1980; 1983), **Hubert Dreyfus**, (Dreyfus, 1982), **Walter Freeman** (Freeman, 2007), tutti dell'Università di California a Berkeley e tutti impegnati da diversi punti di vista nella stessa critica del primo approccio dell'IA simbolica alle scienze cognitive.
- ◆ Da un altro punto di vista filosofico, **Willard V.O. Quine**, dall'alto del suo prestigio assoluto come uno dei maggiori logici del XX sec., ha sintetizzato la questione affermando che nel mente-corpo siamo passati dalla **dualità irriducibile di sostanze** di Cartesio, alla **dualità irriducibile dei linguaggi e delle loro logiche**".

Anche se entrambe le lingue condividono lo stesso referente extra-linguistico: **gli stati/operazioni coscienti del corpo umano**, che è l'attore proprio delle azioni intenzionali umane (Quine, 1987, pp. 132-134) – evidentemente, le azioni morali incluse.

- ◆ Una posizione ontologica simile è condivisa anche da Feigl in (Feigl, 1958), dove ha proposto la sua interpretazione **fisicalista ma non materialista** della **teoria dello stato centrale** della relazione mente-corpo basata sull'irriducibilità di Sellars delle logiche intensionali alle logiche estensionali. Feigl ha suggerito che un'appropriata **scienza della mente** dev'essere basata su una **triangolazione** tra:
 1. ***L'I-talk*** del linguaggio mentalistico della **psicologia intenzionale**.
 2. ***L'O₁-talk*** del linguaggio osservativo **delle neuroscienze** che ha definito **fisico₁**.
 3. ***L'O₂-talk*** del linguaggio osservativo **delle scienze comportamentali** che ha definito come **fisico₂**.
- ◆ Ora, ciò che caratterizza la teoria dello stato centrale di Feigl è la relazione tra i due linguaggi osservativi indicati come "fisico₁" e "fisico₂" che Feigl ha assimilato **alla relazione in termodinamica** tra, rispettivamente, i **microstati** dei moti delle particelle e le corrispondenti **variabili termodinamiche "macroscopiche"**

(temperatura, pressione e volume) che hanno la loro corretta spiegazione a livello della dinamica dei microstati, che, nel nostro caso, è il microstato della dinamica cerebrale.

- ◆ Tuttavia, ciò che collega la teoria di Feigl al programma di ricerca sull'IA e al futuro sviluppo delle scienze cognitive e delle neuroscienze è l'osservazione che ***l'entropia e quindi l'informazione di Shannon*** (che condividono la medesima misura statistica nella termodinamica dei sistemi chiusi di Boltzmann) e perciò **i calcoli di una TM nell'IA simbolica** condividono lo stesso **carattere lineare** delle dinamiche coinvolte. Una supposizione che – e questa sarà la critica fondamentale di Freeman – **non si applica al carattere *non-lineare*** dei processi termodinamici **dei sistemi “aperti” o “dissipativi”**, quali sono appunto quelli **biologici, e neurali** in particolare.
- ◆ Ad ogni modo, seguendo la ricostruzione storica di **Howard Gardner** (Gardner, 1984) dello sviluppo originario della **scienza cognitiva**, ciò che caratterizza la **rivoluzione cognitiva** è la completa sostituzione del **linguaggio osservativo comportamentale fisico** di Feigl, con il **linguaggio computazionale dell'elaborazione dell'informazione nel cervello**.

♦ In altri termini, **la triangolazione**

1. ***Dell'I/We-talk*** dei resoconti soggettivi sullo stato intenzionale della mente – che ricordiamolo è l'unico modo per accedere agli stati soggettivi di coscienza – è
2. **Con il linguaggio osservativo O_1 -talk** da parte di un osservatore esterno delle neuroscienze, da un lato, e
3. **Con il linguaggio osservativo O_2 -talk** dell'informatica, dall'altro, modellando l'elaborazione dell'informazione all'interno del cervello in termini dei calcoli di un computer in grado di implementare logiche estensionali e intensionali.

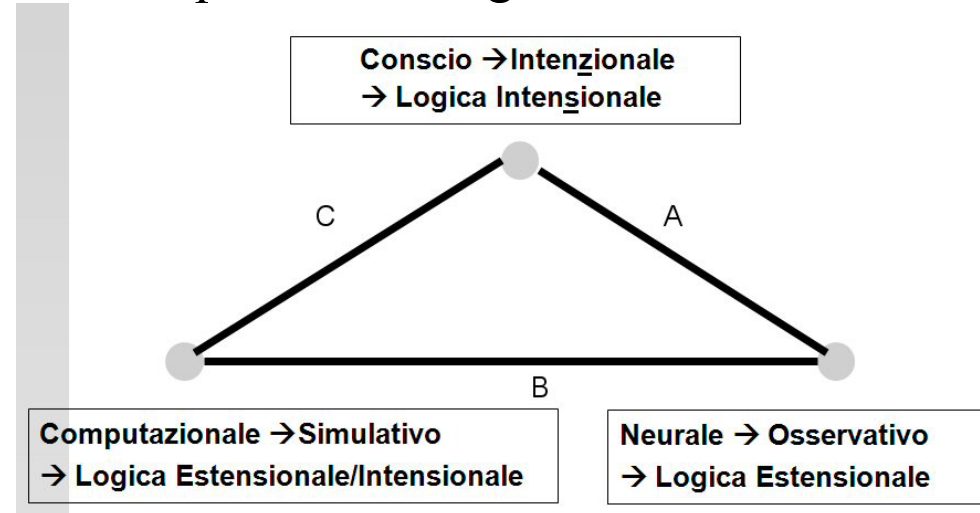


Figura 3. Il triangolo delle scienze cognitive.

- ◆ Ciò che emerge è il **carattere non-riduzionista** delle scienze cognitive, in quanto il **terzo termine della triangolazione** è costituito non dalle descrizioni comportamentali della Teoria dello Stato Centrale di Feigl, ma l'**elaborazione e lo scambio dell'informazione da parte del cervello**.
- ◆ **Scambi di informazione** fra i moduli neurali del cervello che, nel caso dei **sistemi non-lineari** quali appunto sono i sistemi cerebrali è sistematicamente irriducibile agli **scambi energetici** in cui, comunque, i primi sono implementati.
- ◆ Infine, la possibilità di una simulazione algoritmica delle logiche intensionali del cervello è legato allo sviluppo operatosi negli ultimi trent'anni di **logiche booleane modali** basate su un'**algebra booleana con operatori** (*Boolean Algebra with Operators*, BAO) come vedremo nella **Seconda Parte**.

3.3.2. Dalle scienze cognitive alle neuroscienze cognitive: le Reti Neurali Naturali (Limoncelli, 2020).

- ◆ Prima di introdurre la teoria delle **Reti Neurali Artificiali (RNA)** riassumiamo brevemente alcune nozioni elementari di **neurofisiologia** e quindi della teoria delle **Reti Neurali Naturali (RNN)**.
- ◆ Esaminiamo brevemente l'**anatomia del cervello umano** (encefalo).

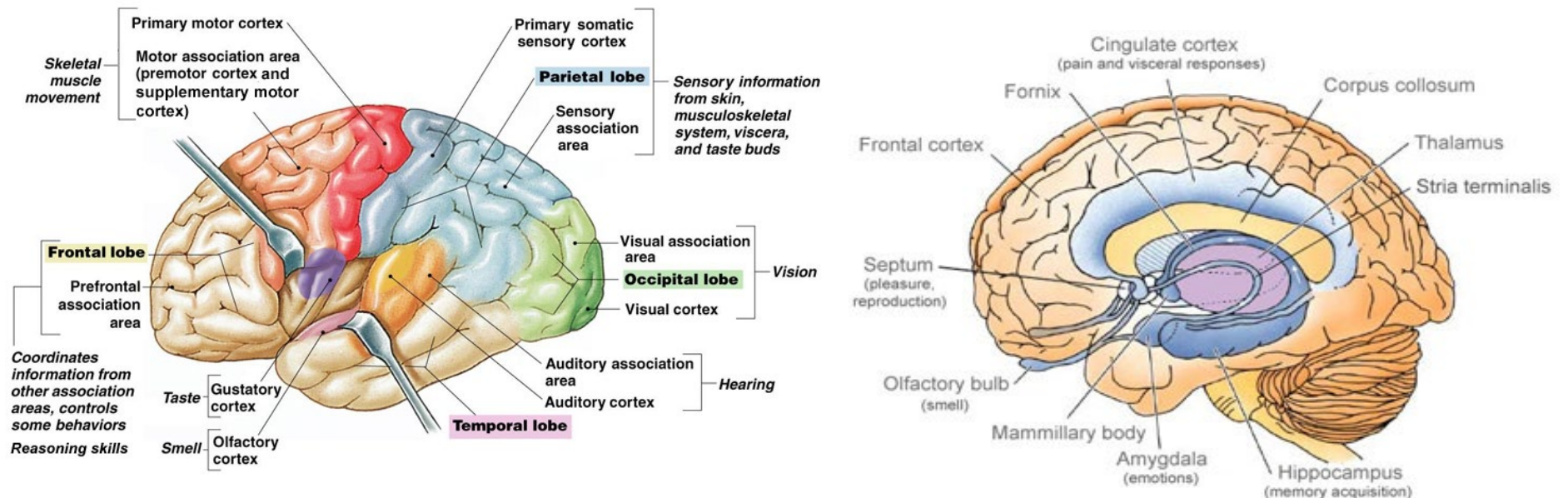


Figura 4. (Sinistra). La corteccia cerebrale costituisce gran parte di quella cosiddetta "materia grigia", formata da decine di miliardi di neuroni, che ricopre la superficie dell'encefalo umano. Tale superficie è caratterizzata dalla presenza di molte scissure, così da ampliare al massimo l'estensione ricopribile dalla corteccia medesima. Come quando si appallottola un pezzo di carta: il massimo volume, nel minimo spazio. Il resto dell'encefalo è costituito per la massima parte dalla cosiddetta "materia bianca" (il

cervello ci appare bianco non grigio), composta essenzialmente di mielina, una sostanza grassa che si pensava avesse solo funzione di isolante elettrico fra le diverse fibre dei neuroni. Nella figura, alla "materia grigia" della corteccia sono state date diverse colorazioni, così da evidenziare in particolare la localizzazione della corteccia sensoria e della corteccia associativa nel cervello umano. Si può constatare inoltre la vicinanza spaziale e funzionale della gran parte della corteccia sensoria con quella motoria, così da evidenziare anche neurologicamente l'unità dell'operazione sensorio-motoria. Infine, si può notare quella notevole predominanza della corteccia associativa sugli altri due tipi di corteccia che è tipica dell'uomo rispetto agli altri mammiferi.

(Destra). La parte colorata in scuro evidenzia quell'insieme di strutture sub-corticali che vanno sotto il nome di sistema limbico. A queste strutture, mediante la loro stretta interazione dinamica con quelle corticali, sono devolute le funzioni di "memoria a lungo termine" (ippocampo) per il "riconoscimento" degli oggetti percettivi e la loro "valutazione emotivo-istintiva" (amigdala) e quindi il controllo della risposta comportamentale "istintiva" alla presentazione percettiva degli oggetti. Per quest'ultima funzione è da notare in particolare il ruolo della corteccia cingolata ("dolore") e del

setto (“piacere sessuale”), in stretta connessione con il sistema ghiandolare ed in particolare con la ghiandola dell'ipofisi. Dall'insieme di queste strutture con l'aggiunta dell'epifisi dipende in gran parte il controllo ormonale del comportamento emotivo-istintivo, in particolare, oltre alla sensazione di fame, l'aggressività (in relazione con le ghiandole surrenali) e la libido sessuale.

♦ Come appare dalla *Figura 4*, il **cervello umano** – ed in generale il cervello dei **mammiferi** si presenta come distinto in **due grandi strutture**:

1. **Le strutture della corteccia cerebrale** suddivise nei simmetrici **lobi** frontali, parietali, temporali e occipitali dei **due emisferi cerebrali**:

- **Cortecce sensorie** – una per ciascun sistema sensorio – dove vengono elaborate le informazioni provenienti **in input** da ciascun senso, non solo quelli **intracranici** (vista, udito, odorato, gusto) ma anche quelli **somatici** esterni e interni (tatto) i cui input sono trasmessi attraverso **le fibre neurali afferenti** della **corda neurale**.
- **Cortecce motorie** che attraverso il **cervelletto** e le **fibre neurali efferenti** della **corda neurale** controllano la **risposta motoria** dell'organismo **in output** in

base agli stimoli sensoriali **valutati emotivamente** attraverso le connessioni con le strutture subcorticali del sistema limbico.

- **Cortecce associative** per gli input delle varie cortecce sensorie. Come si vede dalla *Figura 4 (sinistra)* le cortecce associative costituiscono la più gran parte delle strutture corticali del cervello umano rispetto agli altri mammiferi. Per questo vengono anche definite **neo-corteccia** perché le più recenti dal punto di vista filogenetico (evolutivo).

2. **Le strutture subcorticali del sistema limbico** collocato nella parte più interna dell'encefalo e che costituisce l'interfaccia fra il cervello e il **sistema endocrino** attraverso il flusso **ematico e linfatico**. Il sistema limbico che costituisce la **parte filogeneticamente più antica** del cervello umano – ed è infatti simile in tutti i mammiferi – è caratterizzato dalla presenza, fra l'altro:

- Di due fondamentali **ghiandole endocrine**: l'**epifisi** (la famosa “ghiandola pineale” di Cartesio) e l'**ipofisi** che secernono **ormoni fondamentali** nel flusso **ematico e linfatico** (la linfa è il liquido che riempie gli interstizi fra le cellule del corpo con funzione di **drenaggio metabolico** e di **ricambio dei linfociti**, prodotti dal **midollo** e maturati nel **timo**, fondamentali per il **sistema**

immunitario), per il controllo delle **funzioni neurovegetative** (metabolismo, crescita, funzioni sessuali, funzioni immunitarie, controllo dell'umore, etc.). Queste ghiandole svolgono, insomma, attraverso la produzione di specifici **ormoni (= messaggeri chimici)** e in stretto contatto con l'**ipotalamo** una funzione di controllo sulle altre ghiandole fondamentali del **sistema endocrino** e gli ormoni da esse prodotti: la tiroide, il timo, il pancreas, le ghiandole surrenali, le gonadi (testicoli e ovaie) (cfr. Figura 5).

- Di tre fondamentali **strutture cerebrali**, le prime due collocate nella parte centrale fra i due emisferi cerebrali a forma di mandorla:
 - **L'amigdala** che controlla direttamente le risposte emozionali agli stimoli interni/esterni in particolare la **paura** e la **memoria emozionale**;
 - **L'ipotalamo** i cui nuclei attivano, controllano e integrano i meccanismi autonomi periferici del sistema nervoso, l'attività endocrina attraverso lo stretto contatto con l'ipofisi e l'epifisi e molte funzioni somatiche;
- La terza struttura invece è **L'ippocampo** collocato nella parte interna del lobo temporale e che svolge una funzione essenziale per la formazione delle **memorie a lungo termine**.

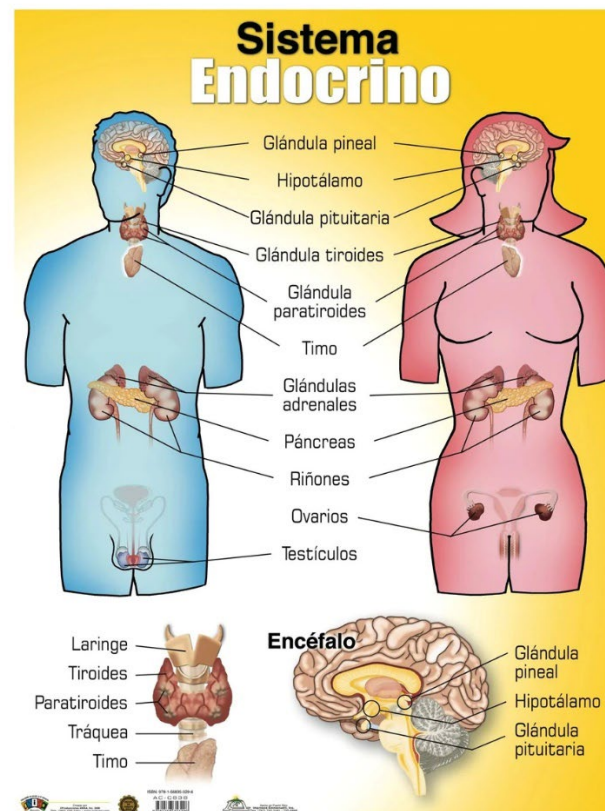
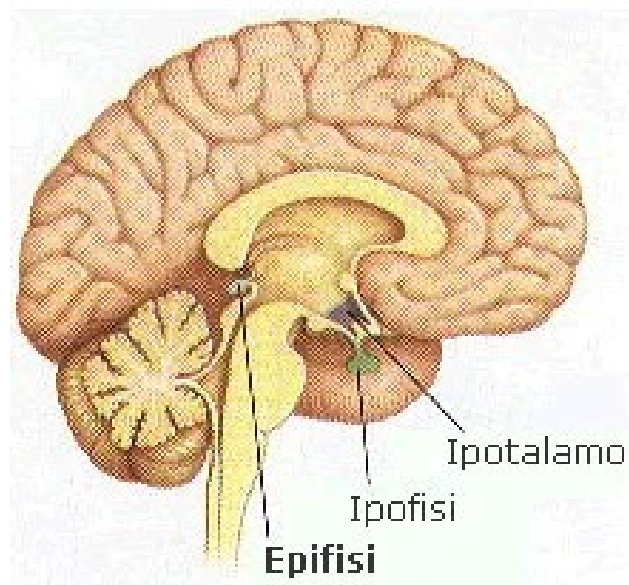


Figura 5. (sinistra) Collocazione delle due ghiandole dell'epifisi e dell'ipofisi in stretta connessione con l'ipotalamo all'interno del sistema limbico (cfr. Figura 5). (destra) relazione con le altre ghiandole del sistema endocrino.

- ◆ Tutte le strutture cerebrali e delle **fibre afferenti-efferenti della corda neurale** (nella spina dorsale) che insieme formano il **Sistema Nervoso Centrale (SNC)** e di **tutte le fibre nervose** che si irradiano dal SNC su tutti i muscoli e gli organi interni e che costituiscono il **Sistema Nervoso Periferico (SNP)**, sono costituite da **cellule nervose o neuroni**.
- ◆ Concentriamoci sulla loro **anatomia e fisiologia** fondamentali per comprendere il concetto di RNN.

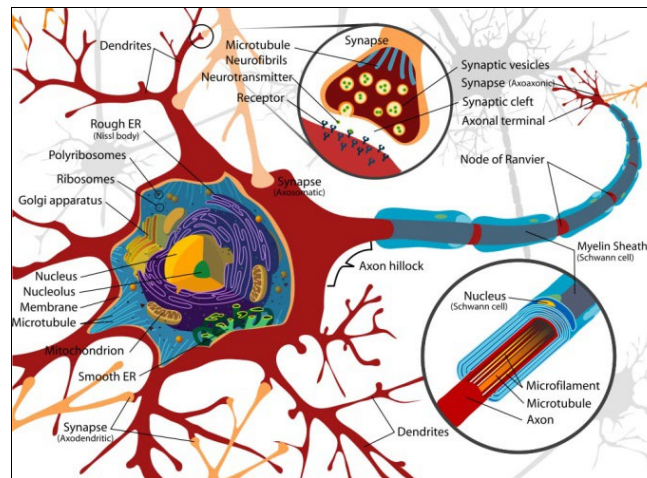


Figura 6. Struttura anatomica di un neurone

- ◆ Come si vede un neurone, rispetto alle altre cellule del corpo umano si caratterizza per il fatto che oltre al **corpo cellulare** ha due altre sub-strutture **un assone** e svariate migliaia di **dendriti** mediante le quali **comunica** con migliaia di altri neuroni **inviando e ricevendo segnali elettro-chimici**, i cosiddetti **stimoli nervosi**.

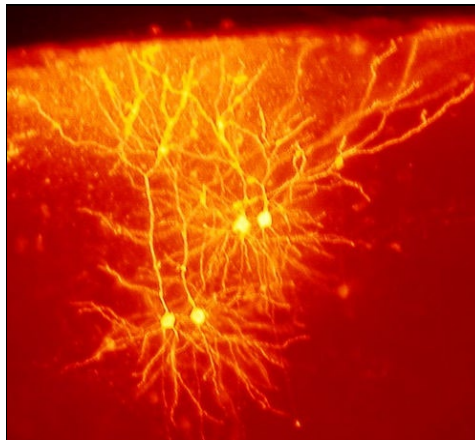


Figura 7. Struttura complessa di interconnessione fra neuroni nella corteccia cingolata del topo.

- ◆ Fondamentale per comprendere il meccanismo di stimolazione elettro-chimica fra neuroni è comprendere che la formazione **dell'impulso elettrico in uscita** dal proprio assone attraverso i suoi dendriti connessi con i dendriti o i corpi cellulari di

altre centinaia o migliaia di neuroni con cui è connesso, attraverso un **meccanismo chimico** che dipende criticamente dagli stimoli ricevuti **in entrata** dai dendriti di altre centinaia o migliaia di neuroni.

- ◆ Bisogna tener presente che i neuroni del SNC di un uomo sono circa **86 miliardi** (con circa 70 miliardi nel solo cervelletto) e ciascuno con in media **7.000 connessioni** sinaptiche (ma si può arrivare fino a **100.000**). Ciò dall'idea della complessità delle RNN.
- ◆ Per comprendere il meccanismo di generazione del potenziale elettrico del neurone e quindi del **segnale 0/1** emesso dal neurone dobbiamo soffermarci sul meccanismo elettro-chimico delle **sinapsi** ovvero della connessione del dendrita di un neurone **pre-sinaptico**, o con un altro neurone **post-sinaptico** (con un suo dendrita o direttamente con il corpo del neurone); oppure con una **fibra muscolare**; oppure con una **ghiandola** del sistema endocrino.
 - Il meccanismo elettrochimico della connessione sinaptica consiste nel fatto che l'arrivo del potenziale elettrico generato nel corpo del neurone pre-sinaptico lungo l'assone provoca la **fusione** delle **vescicole sinaptiche** con la membrana del dendrita assonico e quindi il rilascio delle molecole di un **neuro-**

trasmettitore contenute nelle vescicole che si diffonde nel liquido cefalorachidiano fra il **bottone pre-sinaptico** del mittente e il bottone **post-sinaptico** del ricevente.

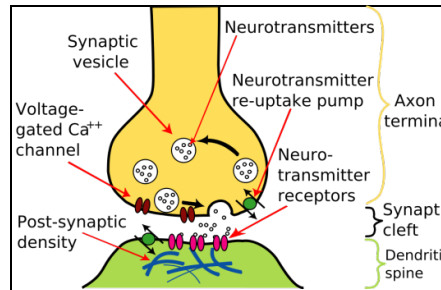


Figura 8. Schema di un bottone sinaptico fra i dendriti di due neuroni.

- Le molecole del neurotrasmettitore sono catturate dai **recettori della membrana** del dendrita ricevente modificando la composizione chimica della membrana del neurone post-sinaptico, un processo che avviene nell'arco di alcuni **millisecondi**. Chiaramente i **farmaci** che si usano in neuropsichiatria, innanzitutto tranquillanti e antidepressivi, interferiscono chimicamente con l'azione dei neurotrasmettitori che sono diverse decine.
- Quelli fondamentali per la formazione del potenziale elettrico nel neurone ricevente sono quelli che hanno effetto **inibitorio** (acido γ -amminobutirrico o

GABA e la glicina) o effetto **eccitatorio** (glutammato) sulla creazione del potenziale d'azione nel neurone post-sinaptico.

- ◆ Il meccanismo di formazione dell'impulso nel neurone post-sinaptico dipende per ciò dalla **somma algebrica** dei contributi fra **sinapsi eccitatorie e inibitorie**. Quando si supera la **soglia di eccitazione**, si innesca il meccanismo di formazione dell'impulso elettrico.

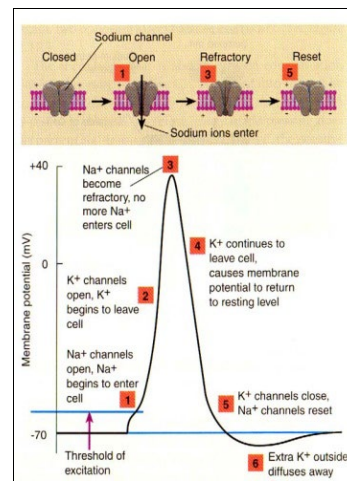


Figura 9. Schema del meccanismo di formazione del potenziale d'azione

- Il **potenziale di azione elettrico** si forma infatti attraverso un meccanismo di rilascio di **ioni di potassio K⁺** compensato dall'entrata di **ioni di sodio Na⁺**

attraverso la **membrana del neurone**. Questo crea un potenziale elettrico d'azione dell'ordine di 40mV che **raggiunto il suo massimo** provoca la chiusura del canale di entrata degli ioni-sodio mentre gli ioni-potassio continuano ad uscire facendo sì che il potenziale torni al suo **stato di riposo o stato di energia minima**.

- Naturalmente occorre **un tempo di latenza** che è dell'ordine di 70 msec. fra la formazione di un potenziale d'azione e il successivo.
- Il che significa che la **frequenza di sparo** di un neurone, ovvero del suo comportamento discreto 0/1 è dell'ordine di circa 10 oscillazioni al secondo. Il neurone insomma è un **processore d'informazione digitale** che lavora a circa **10 Hz** e quindi è lentissimo rispetto ai processori dei nostri computer che lavorano normalmente a frequenze dell'ordine **100 MHz (10^8 Hz)** ben 7 ordini di grandezza più veloci!
- Il segreto del cervello visto come un calcolatore è che esso **processa l'informazione in parallelo su molteplici linee e non in sequenza**, ma soprattutto che si tratta di un **processore analogico-digitale** in quanto il suo comportamento **discreto 0/1** dipende dalla modificazione di **grandezze continue** che

sono le **densità** delle sostanze chimiche coinvolte nel processo di generazione dell'impulso elettrico.

- ◆ Dal punto di vista della **modellizzazione matematica** del comportamento del neurone ci saremo accorti che **due sono i parametri fondamentali** della computazione neurale:
 1. **La funzione di attivazione** del neurone che è essenzialmente **lineare** poiché è costituita dalla somma algebrica dei contributi eccitatori/inibitori in input.
 2. **La funzione sulla soglia di eccitabilità** che definisce la **finestra temporale** entro cui i contributi in input vengono sommati. Più alta è la soglia e più ampia è la finestra temporale, maggiore è il numero di contributi eccitatori/inibitori da cui il comportamento 0/1 del neurone dipende, p.es., da neuroni più lontani.
- ◆ In pratica, ciò significa che **modificando la soglia** il medesimo neurone può far parte di **diversi circuiti di computazione** e quindi **correlare fra di loro** un numero più o meno grande di variabili in ingresso. Matematicamente ciò significa variare **la lunghezza e quindi l'ordine di correlazione e perciò la complessità** della funzione computata.

- ◆ È questa proprietà che rende fortemente **non-lineare e statistica la computazione neurale** ed è il vero segreto della sua **straordinaria efficienza**.
- ◆ Queste nozioni sono così fondamentali per comprendere la **modellizzazione matematica delle RNA** e quindi il concetto di ML

3.4. L'IA simbolica e l'approccio funzionalista alle neuroscienze cognitive

3.4.1. Le reti neurali lineari di McCulloch e Pitts

- ◆ Questa modellizzazione è ciò che caratterizza il primo approccio **funzionalista** alle scienze cognitive (la mente è un UTM che calcola funzioni del calcolo logico e/o matematico) avendo, da un lato, nel già citato articolo di H. Putnam (Putnam, 1960) – la mente come software dell'hardware cerebrale che successivamente cambiò completamente idea – il suo manifesto, e dall'altro nel cosiddetto **approccio simbolico** basato sull'UTM all'IA e alle scienze cognitive il suo sviluppo (Fodor, 1983; Pylyshyn, 1986; Minsky, 1986).
- ◆ Inoltre, il lavoro pionieristico di Warren S. McCulloch e Walter H. Pitts durante gli anni '40 del secolo scorso (McCulloch & Pitts, 1943) ha dimostrato che una rete

(circuito) di neuroni, con una **funzione di attivazione lineare** (sommatoria di input eccitatori/inibitori rispetto a una soglia) è in grado di **calcolare una qualsiasi operazione logica booleana** e quindi è equivalente a una TM.

- ◆ In questo modo gli autori hanno esteso l'approccio simbolico dell'IA, e poi l'approccio funzionalista alle scienze cognitive, anche alle **reti neurali artificiali** (RNA).

3.4.2. La regola di Hebb nella formazione di reti neurali naturali

- ◆ Alla fine degli anni '40, Donald Olding Hebb, sulla base di prove neurofisiologiche, definì la cosiddetta **regola associativa di apprendimento hebbiano** per l'autoassemblaggio dei circuiti neurali nel cervello (Hebb, 1949), mediante la quale l'attivazione simultanea ricorrente dei neuroni produce un aumento del **peso statistico** della connessione sinaptica o **rinforzo della connessione sinaptica** tra questi neuroni, seguendo **una regola lineare**.
- ◆ Determinerà cioè in base alla frequenza di attivazione la costituzione di un **circuito neurale o di una RNN specifica auto-organizzantesi** che si attiverà automaticamente nel cervello ogni volta che uno specifico input si ripresenterà. Si tratterà insomma di una prima forma elementare di **auto-apprendimento** che il cervello manifesta basato sulla regola associativa di tipo probabilistico di Hebb.

- ◆ Una regola che può essere, quindi, sintetizzata nello slogan “i neuroni si collegano insieme in una rete (circuito) se sparano insieme”. (Löwel & Singer, 1992).

3.4.3. Le reti neurali lineari di Farley e Clark

- ◆ Sulla base della scoperta di Hebb, B. G. Farley e Wesley A. Clark pubblicarono un modello computazionale di **RNA auto-organizzante** basato sulla regola di Hebb (Farley & Clark, 1954), in cui array di neuroni artificiali sono arricchiti da circuiti feedforward/feedback, in modo che il **peso statistico della connessione** tra due neuroni (cioè la probabilità di essere attivati insieme) aumenta se vengono attivati frequentemente simultaneamente e diminuisce altrimenti.
- ◆ Questi lavori, quindi, inaugurarono, nell'ambito delle scienze cognitive, la nuova disciplina delle **neuroscienze cognitive**.
- ◆ In questo caso, il sistema computazionale di riferimento non sarà più una TM determinista, ma una **TM probabilista**.

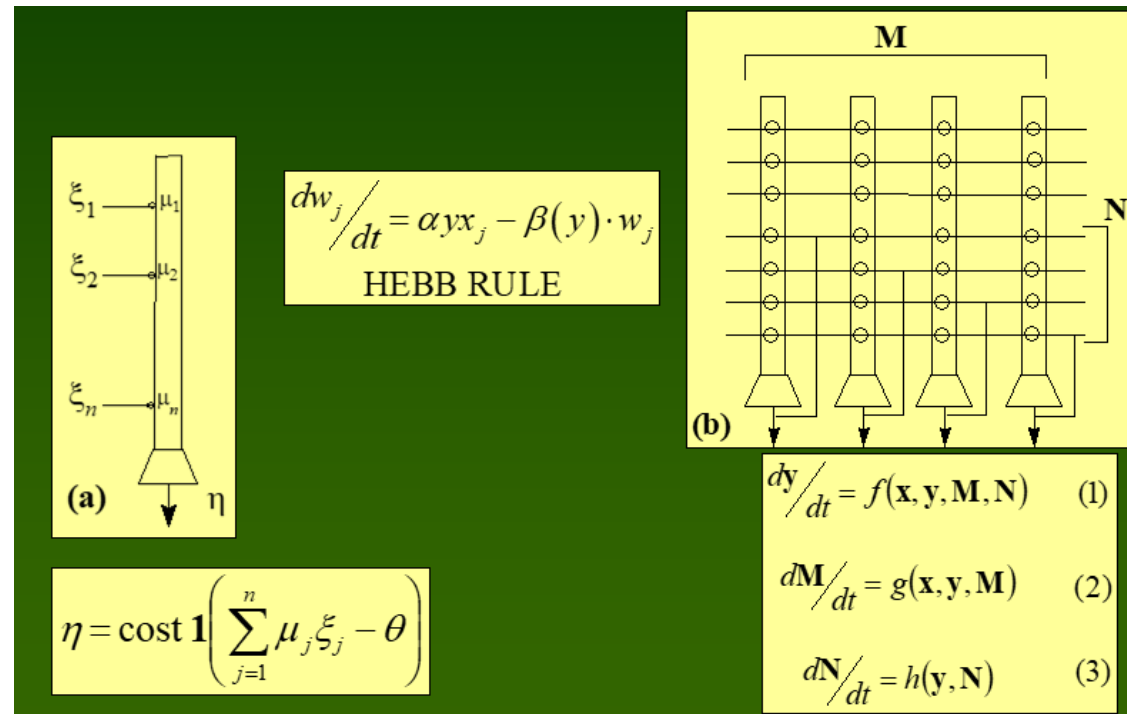


Figura 10. Rappresentazione matematica della funzione di attivazione di un neurone di McCulloch e Pitts (sinistra). Rappresentazione matematica della funzione di attivazione di una rete di Farley-Clark che applica la regola di Hebb non solo in feedforward (M), ma anche in feedback (N). Come si vede, si tratta sempre di reti che seguono una dinamica lineare.

- ◆ È importante soffermarci un attimo sul lavoro di Farley-Clark perché essi introdussero il formalismo base con cui simulare su un computer digitale una RNA. Si tratta di una **matrice di probabilità transitive** dove la probabilità della transizione dallo stato non-attivato/attivato di un neurone corrisponde al valore calcolato per una **cella della matrice**.
 - Il concetto di **probabilità transitiva** è fondamentale in statistica.
 - Normalmente, le probabilità che siamo abituati a considerare sono quelle **incondizionate** tipiche di **eventi non-correlati** fra di loro, come quelle derivate dal **lancio di un dado**. La **probabilità a priori** con cui una delle facce appare è quella di 1/6.
 - In tal caso il teorema fondamentale del **limite centrale** in teoria della probabilità e quindi la famosa **distribuzione normale di Gauss** ci dice che se lancio infinite volte in successione un dato (o lancio simultaneamente infiniti dadi) la **probabilità a posteriori (frequenza)** con cui uscirà ciascuna delle facce **in media** coinciderà con la probabilità a priori.
 - Il concetto di **probabilità condizionale** invece è una misura di probabilità in cui abbiamo a che fare con **eventi correlati** fra di loro così che il valore della

probabilità di un evento dipenderà criticamente dai valori di probabilità di altri eventi correlati.

- Se – come è essenziale in fisica ma anche nello studio delle RNA interpretate come sistemi dinamici – le probabilità condizionali riguardano la **transizione fra due stati 0/1** di ognuno degli n sotto-sistemi fisici correlati fra di loro, un tale sistema complessivo di n sotto-sistemi correlati fra di loro può essere modellizzato come una **matrice di probabilità transitive (MPT)**, visto che la probabilità di attivazione di ciascuno dei sotto-sistemi (neuroni o array di neuroni nel nostro caso) **dipende criticamente** dalla probabilità di attivazione degli altri neuroni con cui è connesso.
- In tal modo, a ciascuna delle connessioni dei neuroni fra di loro, diciamo i, j sarà associato un determinato **peso statistico $w_{i,j}$** (ovvero quanto quella connessione “pesa” sulla determinazione dello stato di attivazione del neurone j rispetto ad i), un **valore statistico** che il calcolo della MPT definisce.

3.4.4. Il perceptrone lineare di Rosenblatt

- ◆ Un ulteriore approccio significativo alle prime RNA lineari è la cosiddetta architettura del **perceptrone lineare** di Frank Rosenblatt, (Rosenblatt, 1961) che durante

gli anni '60 sperava di implementare **un'architettura di calcolo parallelo** in una rete di neuroni lineari perché ognuno di essi calcola **indipendentemente** una funzione diversa su qualche sottoinsieme del set di input.

- ◆ Questo, con evidenti vantaggi rispetto allo standard **seriale** di computazione innanzitutto in termini di velocità computazionale, e **senza la necessità di un supervisore (programmatore)** che distribuisse i diversi compiti tra i neuroni.

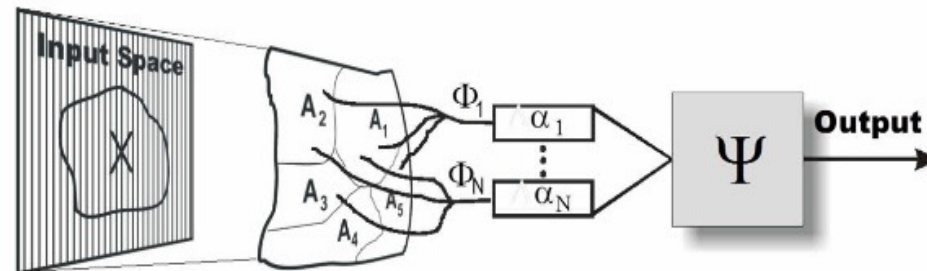


Figura 11. Rappresentazione schematica dell'architettura parallela del perceptrone lineare, dove ogni neurone di ingresso α_i calcola indipendentemente una funzione diversa Φ_i , i cui supporti (dominio-codominio) sono definiti su un filtro costituito dall'unione di sottoinsiemi disgiunti A_i dello spazio di input ("retina") su cui è raffigurato un dato pattern di input X . Il neurone di output, quindi, calcola la semplice sommatoria Ψ dei risultati dei calcoli dei neuroni

di input. Si tratta dunque di un'architettura lineare senza strati intermedi fra i neuroni di input e quello di output

- ◆ Sfortunatamente, nel 1969 Marvin Minsky e Seymour Papert pubblicarono al MIT un libro con una forte critica del perceptrone lineare di Rosenblatt (Minsky & Papert, 1987).
- ◆ Infatti, dimostrarono matematicamente, e in modo molto elegante e convincente, che per questo tipo di architettura computazionale parallela è in linea di principio impossibile calcolare alcune funzioni di base del calcolo logico, prima di tutte la fondamentale **disgiunzione esclusiva XOR** essenziale in qualsiasi calcolo logico in particolare legata alla **classificazione** di vari tipi di input che, al contrario, è semplice da calcolare per un computer seriale standard.
- ◆ La critica di Minsky & Papert fu così distruttiva **contro l'approccio lineare alle RNA** che essa ha effettivamente bloccato qualsiasi tentativo in questa direzione fino agli anni '80 del secolo scorso. Praticamente per vent'anni fino a metà degli anni '80 l'approccio delle RNA all'IA è stato di fatto abbandonato.

3.5. L'IA non-simbolica o statistica, le reti neurali non-lineari e il Machine Learning (ML)

3.5.1. L'IA statistica discriminativa e generativa e l'attuale “boom dell'IA”

- ◆ Nell'attuale Era della Comunicazione, gli esseri umani e le macchine interagiscono e dipendono l'uno dall'altro in modo sempre più forte e indissolubile, come altrettanti **agenti di comunicazione consci e inconsci** (Basti 2017).
- ◆ Questa interdipendenza è diventata particolarmente forte a partire dal 2020, a causa del cosiddetto **boom dell'IA** legato alla massiccia diffusione dei cosiddetti sistemi generativi di IA.
- ◆ Sono il cuore delle **chatbot** o "robot conversazionali", ovvero sistemi di intelligenza artificiale *online* in grado di sostenere **una conversazione in linguaggio naturale con un utente umano**, spesso suggerendo risposte originali ed efficaci alle domande poste dall'utente umano.
- ◆ Esse mostrano infatti capacità sorprendenti di **elaborazione del linguaggio naturale** (*Natural Language Processing*, NLP), poiché sono addestrati su **modelli**

linguistici estremamente grandi (*Large Linguistic Models*, LLM), basati su una rivoluzionaria architettura di **deep learning dell'apprendimento automatico** (*Machine Learning*, ML), i cosiddetti **trasformatori** (*transformers*).

- ◆ Essi furono introdotti da un articolo fondamentale pubblicato nel 2017 da un gruppo di ricercatori di Google (Vaswani, et al., 2017).
- ◆ Questo articolo, dal titolo evocativo *Attention is all you need*, che fa riferimento alla famosa canzone dei Beatles *Love is all you need*, costituisce una vera e propria svolta nella storia recente dei modelli ML nei sistemi di IA, diventando **l'articolo più letto e citato** (ha raggiunto circa **200.000** citazioni) nella storia dell'informatica.
- ◆ Infatti, l'introduzione dei trasformatori ha reso accessibile all'intera società **la distinzione tra due classi di modelli ML** e quindi di sistemi di IA che prima esistevano solo a livello di ricerca scientifica sull'IA:
 1. **I modelli discriminativi**, non supervisionati o supervisionati, in grado di classificare base-di-dati enormi con milioni di variabili e miliardi di parametri intrattabili per qualsiasi mente umana **prevedendo statisticamente le etichette** di ciascuna classe di oggetti sulla base di **caratteristiche comuni nascoste nei dati estratte** dal modello di ML.

2. **I modelli generativi**, che sono in grado di eseguire **l'operazione inversa**. Essi, data l'etichetta precedentemente estratta, possono prevedere l'applicazione della struttura soggiacente a **nuovi dati** generati dal sistema. Questa capacità unica consente loro di generare **nuovi campioni di dati** appartenenti ad una data classe (di testi, di immagini, di video, di dati, di audio, e di dati multimodali integrati fra questi vari tipi di oggetti, ma anche di azioni e comportamenti nel caso di robot...) coerenti con gli altri dati di una certa classe, così, p.es., da consentire a una chatbot di **generare risposte originali coerenti** con la domanda dell'utilizzatore umano.

- ◆ Più in generale, **l'intelligenza artificiale generativa** (*Generative AI*, GAI) è una sotto-classe dei sistemi di IA che utilizza modelli generativi per produrre testi, immagini, video, audio, azioni o altre forme di dati. Questi modelli **apprendono i modelli e le strutture sottostanti** dei loro dati di addestramento e li utilizzano per **produrre nuovi dati**.
- ◆ Il successo improvviso e sorprendente della prima chatbot implementata sul web, **Chat-GPT-3**, che integra entrambi i modelli ML (discriminativo e generativo) perché pre-addestrato su LLM utilizzando trasformatori, e dove GPT sta per

generative pre-trained transformer, reso disponibile online da Open-AI nel **novembre 2022**, ha determinato il rapido sviluppo di questa applicazione.

- ◆ Infatti, **in soli due mesi** questa app ha raggiunto **più di due milioni di utenti**, diventando l'app consumer con la crescita più rapida nella storia di Internet.
- ◆ A partire dalla versione "GPT-4o" (maggio 2024), è diventata **multimodale**, perché in grado di elaborare testo, audio, immagini e video, e attualmente (agosto 2025) è arrivata alla versione "GPT-5", superando in termini di prestazioni rispetto a tutte le versioni precedenti.
- ◆ Oltre alla famiglia di modelli **GPT** incentrati sulla generazione di testo in modo unidirezionale, altri modelli diffusi basati sulla tecnologia di trasformazione sono i modelli **BERT**, specializzati nella direzione opposta della comprensione e dell'analisi del testo, e i modelli **T5** che combinano capacità di comprensione e generazione di testo in entrambe le direzioni, ideali per le traduzioni e, in generale, per le NLP più sofisticate. Vi torneremo nella II Parte.
- ◆ In questo modo, oggi, dopo soli tre anni, **tutti i sistemi operativi più diffusi** sono dotati di applicazioni di IA generativa (Copilot per Windows di Micorsoft, Gemini per Android di Google, Apple IA per IOS di Apple, ...). Questa tecnologia di IA è

stata recentemente estesa anche agli assistenti vocali, primo fra tutti Alexa Plus di Amazon.

- ◆ Più in generale, **tutti i tipi di applicazioni di IA** in ambito scientifico, medico, finanziario, sociale, militare e persino politico possono essere arricchiti dall'introduzione di modelli ML generativi, e questa integrazione si sta rapidamente evolvendo in tutti questi campi, determinando quello che oggi viene definito il boom dell'IA.
- ◆ Per questo motivo, nella Seconda Parte presentiamo e discutiamo approfonditamente sia i modelli discriminativi che quelli generativi di deep learning, inserendoli in una prospettiva storico-teorica necessaria per una loro corretta comprensione.
- ◆ Ma anche per comprendere adeguatamente in maniera non banale e non solo omiletico-esortativa **le questioni etico-giuridiche** che questi sistemi di IA pongono all'umanità e **le loro possibili soluzioni**, in parte già operative.
- ◆ Ma partiamo dall'inizio: dalle RN e dai modelli più elementari di ML.

3.5.2. Funzioni di attivazione lineari e non-lineari e l'approccio connessionista alle RNA

- ◆ Dal punto di vista neurofisiologico, negli stessi anni in cui venivano sviluppate le prime RNA, sono state prodotte molte evidenze sperimentali, che evidenziavano il

carattere **non-lineare** e persino **caotico** – nel senso della nozione dinamica di *caos deterministico* – dell'elaborazione delle informazioni nelle RNN nel cervello (vedi (Freeman, 2001) per una sintesi).

- ◆ Ciò determinò **la crisi dell'approccio funzionalista** alle scienze cognitive dal punto di vista delle neuroscienze, con il conseguente rifiuto del "dogma dell'IA" per cui una TM lineare (probabilistica o meno) poteva essere **sempre** un modello fedele delle architetture computazionali neurali del cervello dei mammiferi.
- ◆ Questo cambiamento di paradigma nelle RNA viene indicato come **l'approccio connessionista alle RNA**.
- ◆ Esso è stato enfatizzato anche dal punto di vista dell'IA, come lo sviluppo del cosiddetto **approccio non-simbolico (statistico) all'apprendimento automatico o ML nei sistemi di IA**, rispetto a quello iniziale simbolico discusso in precedenza in § 3.4.
- ◆ L'approccio connessionista, infatti, è finalizzato alla **classificazione di enormi basi di dati (*big-data*) usando metodi statistici**, con correlazioni interne fra le variabili di **ordine superiore** anziché **le medie del primo ordine** che una RNA lineare può in linea di principio calcolare (Rumelhart, L., & Group, 1986).

- ◆ In particolare, il cosiddetto **algoritmo di apprendimento automatico della *back-propagation* (BP)** (Rumelhart, Hinton, & Williams, 1986) sembrava risolvere direttamente il nucleo della critica di Minsky e Papert al perceptrone lineare di Rosenblatt: la capacità di calcolare la funzione logica XOR. Ciò che caratterizza l'architettura BP rispetto al perceptrone lineare è infatti:
1. La presenza, oltre **ai soli due strati di neuroni di input e output** del perceptrone originale, **di diversi strati interni di neuroni**, in modo da giustificare la nozione di *deep-learning* in questo tipo di architettura di RNA. Proprio come nel cervello fra i neuroni delle **corteccie sensorie (input)** e quelli delle **corteccie motorie (output)** vi sono diversi **strati di corteccie associative** dove si elabora l'informazione (cfr. § 3.3.2).
 2. **La presenza di una funzione *non lineare* (soglia)** che moltiplica la funzione di attivazione (cioè la somma algebrica degli input eccitatori/inibitori) dei neuroni degli strati profondi di un BP. Tale funzione nello specifico è **una funzione *sigmoide*, $\sigma(a) = \frac{1}{1+e^a}$** definita sui numeri reali, e/o il suo parente stretto, **la funzione *tangente iperbolica*, invece della funzione di attivazione a gradino (*stepwise*) 1/0** del perceptrone di Rosenblatt. Quest'ultima si può vedere come **una**

funzione di Heaviside, il cui valore è *zero* per gli argomenti negativi e *uno* per gli argomenti positivi, e quindi **rende lineare la funzione di attivazione** del neurone di Rosenblatt.

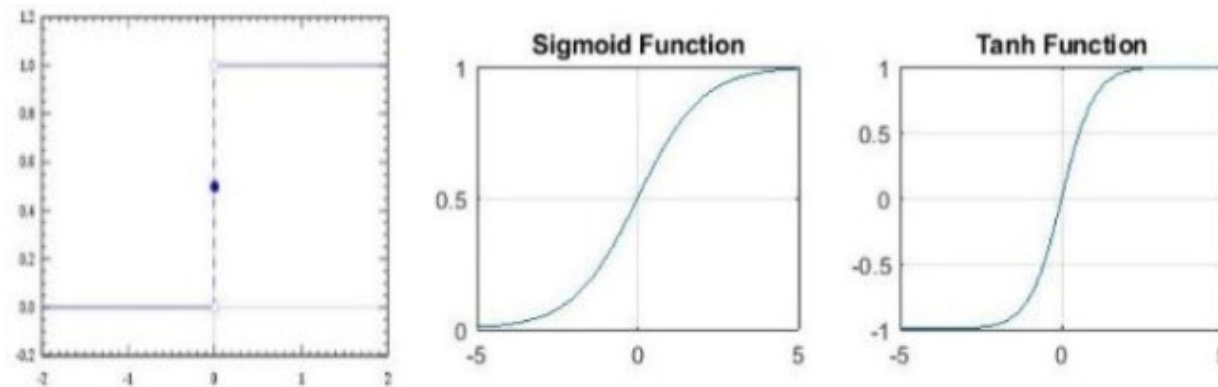


Figura 12. La funzione a gradino di Heaviside (a sinistra), che agisce come una soglia discreta 0/1, confrontata con le funzioni sigmoidea (centro) e tangente iperbolica (destra). È evidente che tanh è una funzione centrata su 0 con valori compresi tra -1 e 1, e che soddisfa quindi una relazione antisimmetrica che è fondamentale in logica. In questo modo, è facile capire che la sigmoide σ è solo una versione riscalata di tanh tale che, poiché è simmetrica rispetto

all'origine, restituisce valori nell'intervallo di numeri reali $[0, 1]$. (Nwankpa C. E., Ijomah, Gachagan, & Marshall, 2018).

- ◆ In una parola, utilizzando la **funzione sigmoidea $\sigma(x)$** o la **funzione tangente iperbolica $\tanh(x)$** nella funzione di attivazione dei neuroni **degli strati nascosti** di una **RNA multistrato**, l'output del neurone può essere **qualsiasi valore numerico reale compreso tra 0 e 1**, in modo da consentire una **caratterizzazione non lineare dei computi statistici della rete**, invece del semplice 0/1 della risposta della funzione di attivazione lineare a gradino dei neuroni di McCulloch e Rosenblatt.
- ◆ Ciò rende in linea di principio la rete in grado di eseguire **calcoli statistici più complessi**, che hanno per oggetto **correlazioni di ordine superiore** (combinazioni complesse di variabili) nel set dei dati di input.
- ◆ La ragione matematica è data dal fatto che **sviluppando in serie di Taylor la funzione sigmoidea**, sono presenti **tutti gli ordini di correlazione** fra le variabili.
- ◆ Ma di questo ci occuperemo quando studieremo i vari algoritmi di ML, a cominciare dalla BP, nella **Seconda Parte** del Corso.

4. Bibliografia della I Parte

Abramsky, S. (2005). A Cook's Tour of the Finitary Non-Well-Founded Sets (original lecture: 1988). In S. Artemov, H. Barringer, A. d'Avila, L. C. Lamb, & J. Woods (A cura di), *Essays in honor of Dov Gabbay. Vol. I* (p. 1-18). London: Imperial College Publications.

Abramsky, S., & Tzevelekos, N. (2011). Introduction to categories and categorical logic. In B. Coecke (Ed.), *New structures for physics. Lecture Notes in Physics, vol. 813* (pp. 3-94). Berlin-New York: Springer.

AI Now Institute. (2019, December). *AI Now 2019 Report*. Retrieved March 10, 2021, from AI Now Institute at NY University:
https://ainowinstitute.org/AI_Now_2019_Report.pdf

Arrow, K. J. (1963). *Social choice and individual values (2nd ed.)*. New Haven & London: Yale UP.

Basti, G. (1991). *Il rapporto mente-corpo nella filosofia e nella scienza*. Bologna : ESD.

Basti, G. (2010). *Filosofia dell'uomo*. Bologna: ESD.

- Basti, G. (2012). Logica aletica, deontica, ontologia formale. Dalla verità ontica all'obbligo deontico , Città del Vaticano, 2012, pp. 105-270. In P. Gherri, & G. Basti (Eds.), *Logica e Diritto: tra argomentazione e scoperta* (pp. 105-270). Città del Vaticano: Lateran University Press.
- Basti, G. (2017). The Post-Modern Transcendental of Language in Science and Philosophy. In Z. Delic (Ed.), *Epistemology and Transformation of Knowledge in in Global Age* (pp. 35-62). London, UK: InTech.
- Basti, G. (2020). Ethical Responsibility vs. Ethical Responsiveness in Conscious and Unconscious Communication Agents. *Proceedings*, 47(68), 1-7.
doi:10.3390/proceedings47010068
- Basti, G. (2020). Ethical Responsibility vs. Ethical Responsiveness in Conscious and Unconscious Communication Agents. *Proceedings*, 47(68), 1-7.
doi:10.3390/proceedings47010068
- Basti, G., & Perrone, A. L. (1995). Chaotic neural nets, computability, undecidability. An outlook of computational dynamics. *Int. J. of Int. Syst*, , 10(1), 41-69.
- Basti, G., Bentini, G. G., Chiarini, M., Parini, A., Artoni, A., & et. al. (2019). Sensor for security and safety applications based on a fully integrated monolithic electro-

optical programmable microdiffractive device. *Proc. SPIE 11159, Electro-Optical and Infrared Systems: Technology and Applications XVI*, (pp. 1115907 (1-12)).

Strasbourg, France. doi:10.1117/12.2532107

Basti, G., Capolupo, A., & Vitiello, G. (2017). Quantum Field Theory and Coalgebraic Logic in Theoretical Computer Science. *Prog. in Bioph. & Mol. Biol. Special Issue, Quantum information models in biology: from molecular biology to cognition*, 130(Part A), 39-52.

Basti, G., Capolupo, A., & Vitiello, G. (2020a). The Computational Challenge of Amartya Sen's Social Choice Theory in Formal Philosophy. In R. Giovagnoli, & R. Lowe (A cura di), *The Logic of Social Practices. Studies in Applied Philosophy, Epistemology and Rational Ethics* 52 (p. 87-119). Berlin-New York: Springer Nature. doi:10.1007/978-3-030-37305-4_7

Basti, G., Capolupo, A., & Vitiello, G. (2020b). The doubling of the Degrees of Freedom in Quantum Dissipative Systems, and the Semantic Information Notion and Measure in Biosemiotics. *Proceedings*, 47(69), 1-7.
doi:doi:10.3390/proceedings47010069

- Basti, G., Perrone, A. L., Messi, R., Paoluzi, L., & Picozza, P. (1995). Neural nets with varying topology for high energy particle recognition. Theory and applications. *Nucl. Instr. and Meth. in Phys. Res. Sect. A*, 360, 349-351.
- Becchi, C. M. (2010). Second quantization. *Scholarpedia*, 5(6), 7902. Retrieved from http://www.scholarpedia.org/article/Second_quantization#haag
- Birhane, A., & Cummins, F. (2019, December 16). *Algorithmic injustice: toward a relational ethics*. Retrieved April 9, 2020, from <https://arxiv.org/pdf/1912.07376v1.pdf>
- Blackburn, P., De Rijke, M., & Venema, Y. (2002). *Modal logic. Cambridge tracts in theoretical computer science*. Cambridge, UK: Cambridge UP.
- Blackburn, P., van Benthem, F. J., & Wolter, F. (Eds.). (2007). *Handbook of modal logic*. Amsterdam: Elsevier.
- Blasone, M., Jizba, P., & Vitiello, G. (2011). *Quantum field theory and its macroscopic manifestations. Boson condensations, ordered patterns and topological defects*. London: Imperial College Press.
- Bogoliubov, N. N. (1958). On a new method in the theory of superconductivity. *Nuovo Cimento*, 7(6), 794-805.

- Bogoliubov, N. N., Tolmachev, V. V., & Shirkov, D. V. (1959). *A new method in the theory of superconductivity*. New York, London: Consultants Bureau Inc., Chapman & Hall.
- Christian, B. (2020). *The alignment problem. How can artificial intelligence learn human values? (Kindle Ed.)*. London, UK: Atlantic Books.
- Cresswell, M. J., & Huges, G. E. (1996). *A new introduction to modal Logic*. London: Routledge.
- Damasio, A. (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Putnam Publishing.
- Damasio, A. (2010). *Self comes to mind: constructing the conscious brain* (1 ed.). London: Heinemann.
- Damasio, A. (2018). *The strange order of things. Life, feeling and the making of cultures*. New York: Pantheon Books.
- Davis, M. (2012). *Il calcolatore universale. Da Leibniz a Turing. 3. Edizione*. Milano: Adelphi.
- De Mauro, A. (2019). *Big Data Analytics. Analizzare e interpretare dati col machine learning. (Kindle Edition)* Tutorials online:

<https://www.apogeoonline.com/libri/big-data-analytics-andrea-de-mauro/>. Milano: Apogeo-Feltrinelli.

Dreyfus, H. (1982). Husserl's perceptual noema. In H. Dreyfus (Ed.), . *In Husserl, intentionality and cognitive science* (pp. 97-124). Cambridge, Mass, USA: MIT Press.

Endriss, U. (2011). Logic and social choice theory. In A. Gupta, & J. Van Benthem (Eds.), *Logic and Philosophy Today* (pp. 333-377). London: College Publications.

Fabro, C. (1941). *Percezione e pensiero*. Milano: Vita e Pensiero.

Farley, B. G., & Clark, W. A. (1954). Simulation of Self-Organizing Systems by Digital Computer. *IRE Trans. on Inform. Th.*, 4, 76-84.

Feigl, H. (1958). The "mental" and the "physical". In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota Studies in the Philosophy of Mind. Vol. II: "Concepts, Theories and the Mind-Body Problem"* (pp. 370-497). Minneapolis, USA: Minnesota UP.

Floridi, L. (2022). *Etica dell'intelligenza artificiale. Sviluppi, opportunità sfide* (Ed. Kindle). Milano: Raffaele Cortina Editore.

- Fock, V. A. (1932). Konfigurationsraum und zweite quantelung. *Zeit. f. Phys.*, 75, 622.
- Fodor, J. A. (1983). *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge MA, USA: MIT Press.
- Freeman, W. J. (2001). *How brains make up their minds*. New York: Columbia UP.
- Freeman, W. J. (2007). Intentionality. *Scholarpedia*, 2(2), 1337. Tratto il giorno November 7, 2015 da Scholarpedia 2(2): 1337:
<http://www.scholarpedia.org/article/Intentionality>
- Freeman, W. J., & Vitiello, G. (2006). Nonlinear brain dynamics as macroscopic manifestation of underlying many-body field dynamics. *Physics of Life Reviews*, 3(2), 93-118.
- Freeman, W. J., & Vitiello, G. (2008). Dissipation and spontaneous symmetry breaking in brain dynamics. *Journal of Physics A: Mathematical and Theoretical*, 41(30), 304042. doi:10.1088/1751-8113/41/30/304042
- Galvan, S. (1991). *Logiche intensionali. Sistemi proposizionali di logica modale, deontica, epistemica*. Milano: Franco Angeli.
- Gardner, H. (1984). *The mind's new science: a history of the cognitive revolution*. New York, USA, 1984: Basic Books .

- Gazzaniga, M. (2011). *Who is in charge? Free will and the science of the brain*. New York: Harper Collins Publ.
- Goldblatt, R. I. (1976a). Metamathematics of modal logic I. *Reports on Mathematical Logic*, 6, 41-48.
- Goldblatt, R. I. (1976a). Metamathematics of modal logic I. *Reports on Mathematical Logic*, 6, 41-48.
- Goldblatt, R. I. (1976b). Metamathematics of modal logic II. *Reports on Mathematical Logic*, 7, 21–52.
- Goldblatt, R. I. (1976b). Metamathematics of modal logic II. *Reports on Mathematical Logic*, 7, 21–52.
- Goldblatt, R. I. (1984). *Topoi: The Categorical Analysis of Logic. Revised Edition*. Amsterdam: Elsevier.
- Goldblatt, R. I. (2003). Mathematical modal logic: a view of its evolution. *Journal of Applied Logic*, 1, 309—392.
- Goldblatt, R. I., & Thomason, S. (1974). Axiomatic classes in propositional modal logic. In J. Crossley (Ed.), *Algebra and logic* (pp. 163-173). Berlin-New York: Springer.

- Goldstone, J. (1961). Field Theories with Superconductor Solutions. *Nuovo Cimento*, 19, 154–164. doi:10.1007/BF02812722
- Goodfellow, I., & Bengio, Y. C. (2016). *Deep Learning (Kindle Ed.)*. Cambridge Mass., USA: MIT Press.
- Goranko, V., & Otto, M. (2007). Model theory of modal logic. In P. Blackburn, F. J. van Benthem, & F. Wolter (Eds.), *Handbook of Modal Logic* (pp. 252-331). Amsterdam: Elsevier.
- Hebb, D. O. (1949). *The organization of behavior*. New York, USA: Wiley & Sons.
- Hofstadter, D. (1979). *Gödel, Escher, Bach: an Eternal Golden Braid. First Edition*. New York, USA: Vintage Books: .
- Jónsson, B., & Tarski, A. (1952a). Boolean algebras with operators, Part I. *American Journal of Mathematics*, 73, 891-939.
- Jónsson, B., & Tarski, A. (1952b). Boolean algebras with operators, Part II. *American Journal of Mathematics*, 74, 127-152.
- Keitel, C. H., & Wodkiewicz, K. (1993). Measuring information via Glauber's Q-representation. *Proceedings of the Second International Workshop on Squeezed States and Uncertainty Relations*, (pp. 259-265).

- Kleene, S. C. (1952). *Introduction to Metamathematics*. New York, USA: Van Nostrand.
- Kripke, S. A. (1963). Semantical analysis of modal logic I. Normal modal propositional logic calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 9, 67-96.
- Kripke, S. A. (1965). Semantical analysis of modal logic II. Non-normal modal propositional calculi. In J. W. Addison, L. Henkin, & A. Tarski (Eds.), *The Theory of Models* (pp. 206-220). Amsterdam: North Holland.
- Kripke, S. A. (1975). Outline of a theory of truth. *The Journal of Philosophy*, 72(19), 690-716.
- Kupke, C., Kurz, A., & Venema, Y. (2004). Stone coalgebras. *Theoretical computer science*, 327, 109-134.
- Ladyman, J. (2016). *Structural realism*. (E. N. Zalta, Ed.) Retrieved December 29, 2017, from The Stanford Encyclopedia of Philosophy:
<https://plato.stanford.edu/archives/win2016/entries/structural-realism/>

- Landsman, K. N. (2011, December 14). *Lecture notes on operator algebras*. Retrieved from Institute for Mathematics, Astrophysics, and Particle Physics. Radboud University Nijmegen: <http://www.math.ru.nl/~landsman/OA2011.html>
- Landsman, K. N. (2017). *Foundations of Quantum Theory. From Classical Concepts to Operator Algebra*. Berlin-New York: Springer.
- Lewis, C. I. (1912). Implication and the Algebra of Logic. *Mind*, 21, 522–531.
- Lewis, C. I. (1918). *A Survey of Symbolic Logic*. Berkeley, CA: University of California Press.
- Lewis, C. I., & Langford, C. H. (1932). *Symbolic Logic* (2.Edition, Dover Publications, New York, 1959 ed.). New York: Century Company.
- Limoncelli, L. (2020). *Neurofisiologia del sistema nervoso ed emergenza dell'informazione neurale*. Roma: PUL - Dispensa.
- Lo Piano, S. (2020). Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanit. Soc. Sci. Commun.*, 7, 9. Retrieved from <https://rdcu.be/cWp6r>
- Löwel, S., & Singer, W. (1992). Selection of Intrinsic Horizontal Connections in the Visual Cortex by Correlated Neuronal Activity. *Science*, 255, 209-212.

- MacKay, D. M. (1968). *Information, mechanism, and meaning*. Cambridge MA, USA: MIT Press.
- McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. (1955, August 1). *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. Retrieved from <http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf>
- McCulloch, W. S., & Pitts, W. H. (1943). A Logical Calculus of Ideas Immanent in Nervous Activity. *Bull. Math. Biophys.*, 5, 115-133.
- Minsky, M. (1975b). A framewok for representing knowledge. In P. H. Winston (Ed.), *The psychology of the computer vision*. New York: McGraw-Hill.
- Minsky, M. (1975b). Frame system theory. *TINLAP*. Boston MA, USA: MIT Press.
- Minsky, M. (1977). Frame theory. In P. Johnson-Laird, & P. Wason (A cura di), *Thinking: Reasings in Cognitive Science* (p. 355-376). Cambridge, MA: Cambridge UP.
- Minsky, M. (1986). *The Society of Minds*. New York, USA: Simon & Schuster.
- Minsky, M., & Papert, S. (1987). *Perceptrons. An introduction to computational geometry. Second Edition*. Cambridge MA, USA: MIT Press.

- MLK. (2019, January 7). *Machine Learning Basics*. Tratto da Machinelearningknowledge.ai: <https://machinelearningknowledge.ai/mcculloch-pitts-neuron-model/>
- Müller, V. C. (2020, December 1). *Ethics of Artificial Intelligence and Robotics*. Retrieved from Stanford Encyclopedia of Philosophy: <https://plato.stanford.edu/archives/win2020/entries/ethics-ai/>
- Nambu, Y. (1960). Quasiparticles and Gauge Invariance in the Theory of Superconductivity. *Physical Review* , 117, 648–663.
- Nwankpa, C. E., Ijomah, W., Gachagan, A., & Marshall, S. (2018, November 8). *Activation functions: comparison of trends in practice and research for deep learning*. doi:10.48550/arXiv.1811.03378
- Nwankpa, C. E., Ijomah, W., Gachagan, A., & S., M. (2018, November 8). *Activation Functions: Comparison of Trends in Practice and Research for Deep Learning*. Retrieved March 10, 2021, from <https://arxiv.org/pdf/1811.03378.pdf>
- Parini, A., Chiarini, M., Basti, G., & Bentini, G. G. (2019). Lithium niobate-based programmable micro-diffraction device for wavelength-selective switching applications. *Proc. SPIE 11163, Emerging Imaging and Sensing Technologies for*

Security and Defence IV, (p. 111630C (1-7)). Strasbourg, France.
doi:10.1117/12.2532246

Pessa, E., & Vitiello, G. (1999). Quantum dissipation and neural net dynamics.
Biochem. and Bioenerg., 48, 339-342.

Piaget, J. (2011). *Psicologia dell'intelligenza*. Firenze: Giunti.

Putnam, H. (1960). Minds and Machines. In S. Hook (ed.), *Dimensions of mind*. New York: Collier.

Pylyshyn, Z. W. (1986). *Computation and Cognition: Toward a Foundation for Cognitive Science*. Cambridge MA, USA: MIT Press.

Quine, W. V. (1987). *Quiddities. An Intermittently Philosophical Dictionary*. Cambridge MA, USA: Harvard UP.

Rasetti, M., & Merelli, E. (2017). Topological field theory of data: mining data beyond complex networks. In P. Contucci, & C. Giardinà (Eds.), *Advances in disordered systems, random processes and some applications* (pp. 1-42). Cambridge, UK: Cambridge UP.

Rawls, J. (1999). *A Theory of Justice. Revised Edition*. Cambridge: Belknap Press: An Imprint of Harvard University Press.

- Ricciardi, L. M., & Umezawa, H. (1967). Brain physics and many-body problem. *Kibernetik*, 4, 44-48.
- Rosenblatt, F. (1961). *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Buffalo, NY: Cornell UP.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors, *Nature*, 1986, 323, pp. 533-536. *Nature*, 323, 533-536.
- Rumelhart, D., L., M. J., & Group, P. (1986). *Parallel Distributed Processing. Voll. 1-2*. Cambridge MA, USA: MIT Press.
- Rutten, J. J. (2000). Universal coalgebra: a theory of systems. *Theoretical computer science*, 249(1), 3-80.
- Searle, J. R. (1980). Minds, brains, and programs. *The Behav. and Brain Sc.*, 3, 128-135.
- Searle, J. R. (1983). *Intentionality. An essay in the philosophy of mind*. New York: Cambridge UP.
- Sellars, W. (1958). Intentionality and the mental. In H. Feigl, M. Scriven, & Maxwell (Eds.), *In Minnesota Studies in the Philosophy of Mind. Vol. II: "Concepts*,

Theories and the Mind-Body Problem" (pp. 507-524). Minneapolis, USA: Minnesota UP.

Sen, A. K. (1970). The impossibility of a Paretian liberal . *Journal of Political Economy*, 78(1), 152-157.

Sen, A. K. (1999). The Possibility of Social Choice. Nobel Lecture. *Am. Econ. Rev.*, 89, 178-215.

Sen, A. K. (2010). *The idea of justice*. London, UK: Penguin Books, Ltd.

Sen, A. K. (2017). *Collective Choice and Social Welfare. Expanded Edition*. London: Penguin Ltd. Kindle Edition.

Sen, A. K. (2017). *Collective Choice and Social Welfare. Expanded Edition*. London: Penguin.

Stone, M. H. (1936). The theory of representation for Boolean algebras. *Transactions of the American Mathematical Society*, 40, 37-111.

Suppes, P. (1966). Some formal models of grading principles. *Synthese*, 16(3-4), 284-306.

Takahashi, Y., & Umezawa, H. (1975). Thermo Field Dynamics. *Collective Phenomena*, 2, 55-80.

- Thomason, S. K. (1975). Reduction of second-order logic to modal logic. *Mathematical Logic Quarterly*, 21(1), 107-114.
- Turing, A. M. (1937). On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 2(42), 230–265.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236), 433-460.
- Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59, 433-460.
- Umezawa, H. (1993). *Advanced field theory: micro, macro and thermal concepts*. New York: American Institute of Physics.
- Umezawa, H. (1995). Development in concepts in quantum field theory in half century. *Math. Japonica*, 41, 109–124.
- Venema, Y. (2007). Algebras and co-algebras. In P. Blackburn, F. J. van Benthem, & F. Wolter (A cura di), *Handbook of modal logic* (p. 331-426). Amsterdam, North Holland: Elsevier.
- Veroff, R., & Spinks, M. (2006). Axiomatizing the Skew Boolean Propositional Calculus. *J Autom Reasoning*, 37, 3-20. doi:10.1007/s10817-006-9029-y

- Vitiello, G. (1995). Dissipation and memory capacity in the quantum brain model. *Int. J. Mod. Phys., B9*, 973-989.
- Vitiello, G. (2001). *My Double Unveiled*. Amsterdam: John Benjamins Publ. Co.
- Vitiello, G. (2004). Classical chaotic trajectories in quantum field theory. *Int.J. Mod. Phys., B18*, 785-792.
- Vitiello, G. (2004). The dissipative brain. In G. G. Globus, K. H. Pribram, & G. Vitiello (A cura di), *Brain and Being - At the boundary between science, philosophy, language and arts* (p. 317-330). Amsterdam: John Benjamins Pub. Co.
- Vitiello, G. (2004). The dissipative brain. In G. G. Globus, K. H. Pribram, & G. Vitiello (Eds.), *Brain and Being. At boundary between science, philosophy, language* (pp. 315-334). Amsterdam: John Benjamins.
- Vitiello, G. (2021. In print). Neural networks and many/body systems. Essays in honor of Eliano Pessa. In G. Minati (Ed.), *Multiplicity and Interdisciplinarity*. Cham, Switzerland: Springer Int.Publ.
- Werbos, P. (1974). *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences. PhD thesis*. Boston MA, USA: Harvard UP.

Zalta, E. (1988). *Intensional logic and the metaphysics of intentionality*. Cambridge MA: MIT Press.

Note

¹ Per non creare confusioni, non bisogna confondere un'aritmetica binaria “modulo 2” con la **codifica binaria di numeri**, come quella dei nostri computer, dove “2” non si scrive “0”, ma “11”. L'aritmetica binaria è un **calcolo**, la codifica binaria è un **simbolismo**, sebbene in logica computazionale siano strettamente connessi. Vi torneremo.

² Vale la pena ricordare che Karl R. Popper nella sua *Autobiografia intellettuale*, si definì "**l'assassino del neopositivismo**", identificando la data di tale omicidio con l'arresto (temporaneo) della pubblicazione della collezione dei *Minnesota Studies*. Sfortunatamente per lui, la raccolta (non il Neopositivismo) sopravvisse, non solo alla morte di Feigl, ma anche alla morte di Popper (1994).