

Gestione della Conoscenza per i Servizi Digitali



Antonio Lieto

Università di Salerno, IT
ICAR-CNR, Palermo, IT

Temi del corso

Introduzione ai paradigmi di modellazione dell'**Intelligenza Artificiale** (neurale-simbolica-ibrida)

Paradigma Simbolico: **Logica, Inferenza e Sistemi di Rappresentazione e Gestione della Conoscenza**

Ontologie e Tecnologie Semantiche (con parti di laboratorio su costruzione, immagazzinamento e interrogazione di modelli)

Esame

Progetto di gruppo (min 2 max 4 persone) sulla parte laboratoriale da consegnare e discutere

Orale individuale su tutto il programma

Testi e Materiale

In Inglese

Materiale diffuso a lezione (slides, manuali per software). Disponibili su <https://www.antonioleto.net> (andare in Didattica)

Libro di testo da riferimento per tutta la parte teorica

Testi e Materiale

In Inglese

Materiale diffuso a lezione (slides, manuali per software)

Libro di testo da riferimento per tutta la parte teorica

**Lieto A, 2021, Cognitive Design for Artificial Minds, Routledge/
Taylor & Francis, London/New York.**

<https://www.routledge.com/Cognitive-Design-for-Artificial-Minds/Lieto/p/book/9781138207950>

Cognitive Design for Artificial Minds

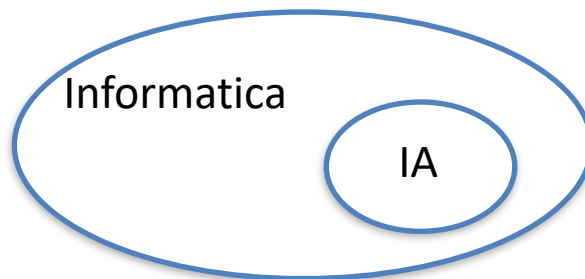
Antonio Lieto



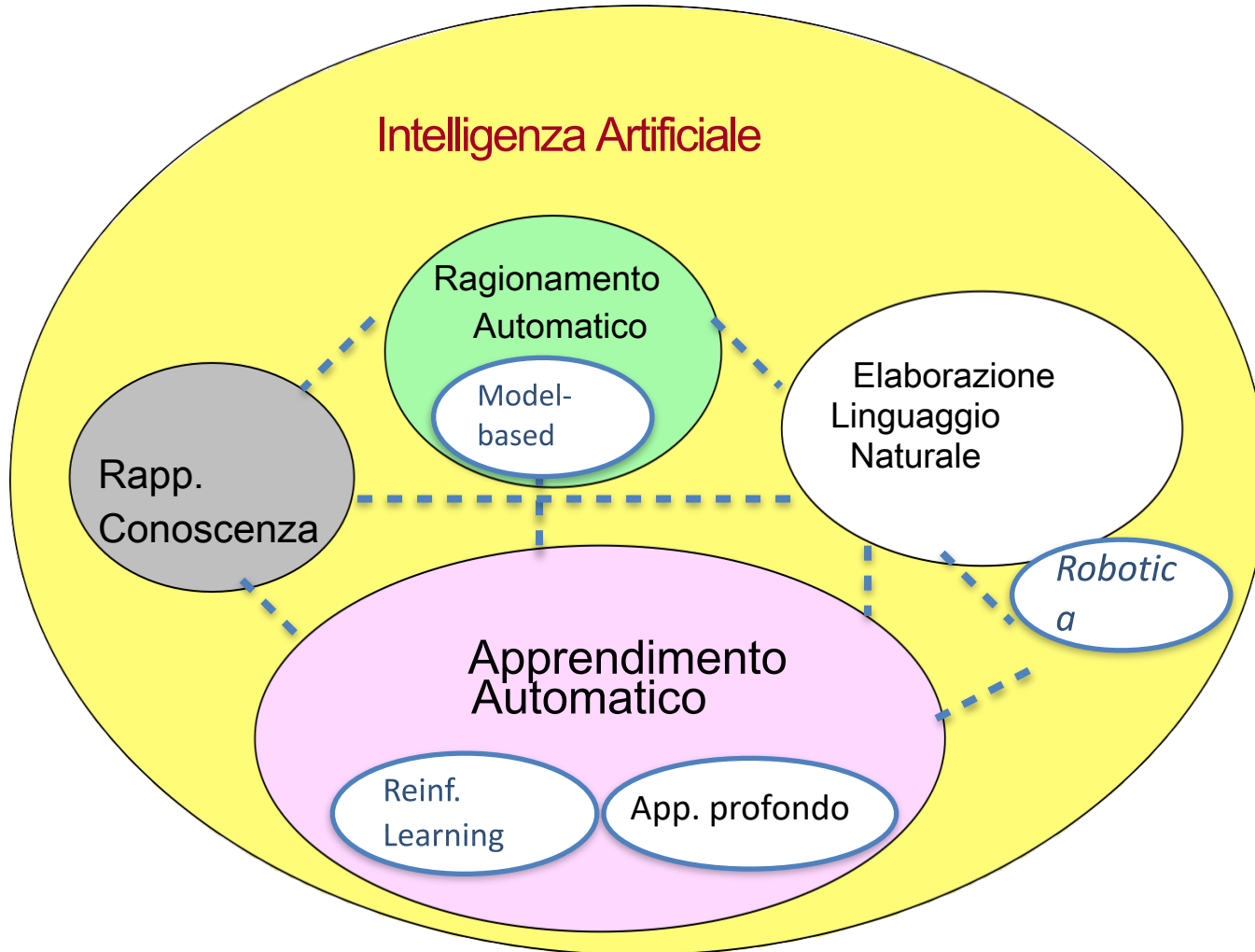
Una definizione

«L'**intelligenza artificiale** è una branca dell' **informatica** che studia i fondamenti teorici, le metodologie e le tecniche che consentono la progettazione di sistemi hardware e software capaci di fornire a un calcolatore prestazioni che, a un osservatore comune, sembrerebbero essere di pertinenza esclusiva dell'intelligenza umana.»

(Somalvico, 2003)



Tante Aree di ricerca in Intelligenza Artificiale (elenco parziale)



Intelligenza Artificiale

L'IA è contemporaneamente scienza e tecnica, disciplina di frontiera in cui si incontrano diversi domini del sapere: logica, informatica, psicologia, neuroscienze, filosofia...

Attiene a tutti i campi dell'informazione e trova applicazione in tutti i settori

Es. Veicoli a guida autonoma

Example: Car driving



Actually a “Perception to Action” system. The visual perception and control system is a Deep learning architecture trained end to end to transform pixels from the cameras into steering angles. And this car uses regular color cameras, not LIDARS like the Google cars. It is watching the driver and learns.

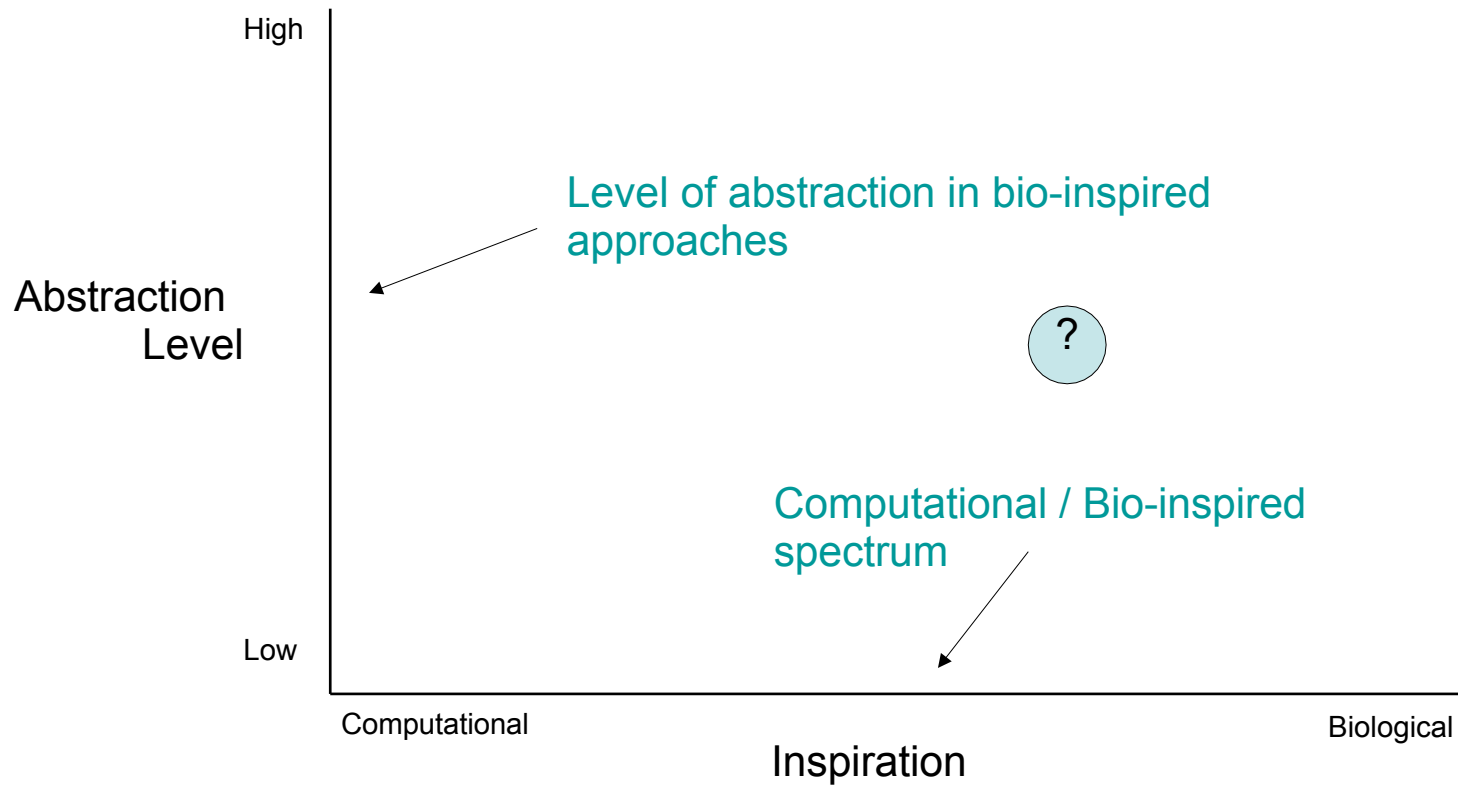
<https://www.youtube.com/watch?v=YuyT2SDcYrU>

2 Modi per replicare un facoltà cognitiva

- 1) Inventare una nuova soluzione **computazionale**
- 2) Prendere ispirazione dalla natura (**cognitive or bio-inspired systems**)

Vernon (2014): possiamo pensare di usare uno spazio bidimensionale per posizionare il sistema in base al grado di ispirazione cognitiva e in base al paradigma computazionale utilizzato

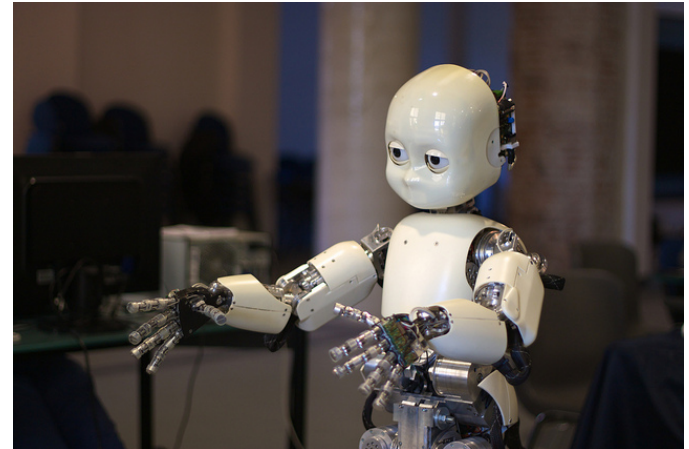
Lo Spazio 2D della Modellazione Cognitiva



From Human to Artificial Cognition

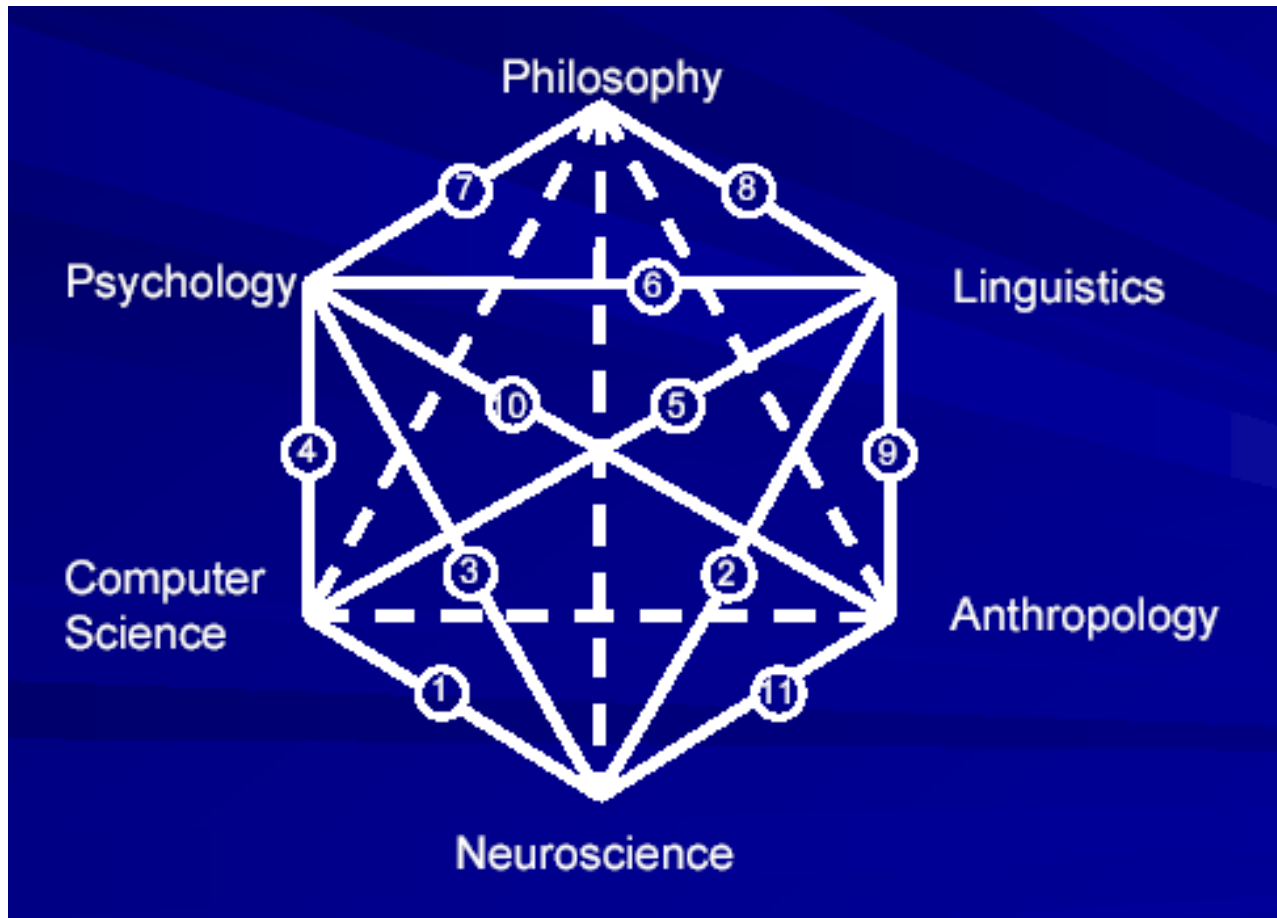


Inspiration



Digressione storica

Scienze cognitive e intelligenza artificiale sono due facce della stessa medaglia (studio della “mente”, naturale e artificiale)



2 Prospettive Storiche

| Cognitivismo | Nouvelle AI |
|--|--|
| Focus su funzioni esecutive di alto livello (ragionamento, pianificazione, linguaggio) | Focus su Percezione |
| Rappresentazioni strutturate (<i>physical symbol system</i> , Simon and Newell, 1976) | Rappresentazioni distribuite |
| Prospettiva architettura (relativa all'integrazione di diverse funzioni) | Prospettiva di singolo sistema. |
| Inspirazione cognitiva (heuristic-driven approach) | Inspirazione biologica |

Fasi storiche dell'AI

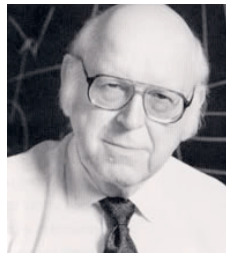
AI degli albori

Cognitive or Biological Inspiration
for the Design of “Intelligent Systems”

AI moderna

“Intelligence” in terms of optimality of
performance (narrow tasks)

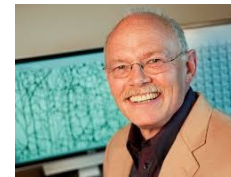
A. Newell



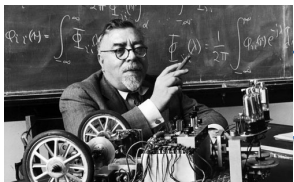
M. Minsky



J. McClelland



mid'80s



N. Wiener



H. Simon



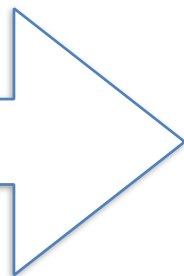
R. Shank



D. Rumelhart

Nowadays:

Renewed attention
“The gap between natural
and artificial
systems is still enormous”
(A. Sloman, AIC 2014).

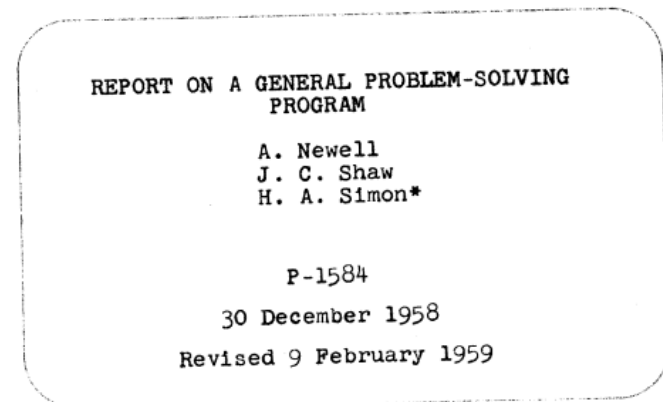


GPS (General Problem Solver)

Sistema in grado di dimostrare semplici teoremi logici le cui strategie erano ispirate in modo esplicito ai protocolli verbali umani (Simon, Shaw, Newell, 1959).

Premio Nobel
"bounded rationality"

GPS implementava l'euristica mezzi-fini (**means-ends analysis** o M-E heuristic) in cui un problem solver confronta uno stato di cose con il goal da raggiungere e cerca di minimizzare la distanza tra questi stati tramite appositi operatori.



Reti Semantiche

Ross Quillian (1968) *sviluppò un sistema psicologicamente plausibile della memoria umana*.
IDEA: la memoria umana è associativa in natura e rappresentando i concetti in un grafo e la loro attivazione (tramite “spreading activation”) tramite meccanismi di propagazione dell’informazione è possibile vedere quali concetti vengono utilizzati in vari compiti.



Es. Frames



| Frame 1 | |
|-------------|---------|
| Concept 1 | |
| Attribute 1 | Value 1 |
| Attribute 2 | Value 2 |
| Attribute 3 | Value 3 |
| ... | ... |

(Minsky M., 1975)

framework usato per rappresentare conoscenza di senso comune, di default, tipica (es. Gli uccelli tipicamente volano). Tale organizzazione permise le prime forme di ragionamento automatico al di là della deduzione classica.

Scripts (Shank and Abelson, (1977))



R. Shank

Strutture dati per rappresentare **conoscenza di senso comune legata agli eventi** (e.g. es. “andare fuori a cena”) e usata in *natural language processing* per permettere ad agenti intelligenti di rispondere a domande riguardanti semplici storie.

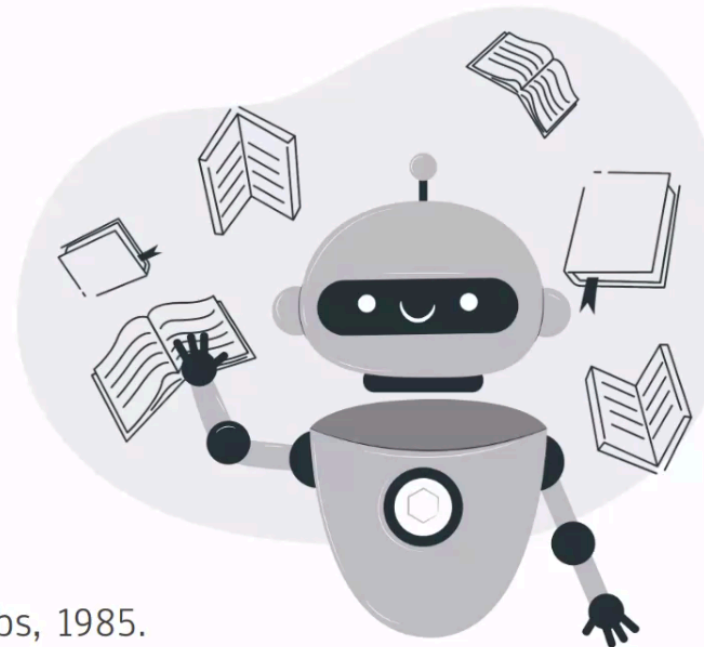
Script Restaurant: e.g. entrare al ristorante, chiedere un tavolo, sedersi, consultare il menu, mangiare, pagare etc.

Esempio **“Mary went to a restaurant, ordered a salmon. When she was paying she noticed that she was late for the next appointment”**,

Domanda: **“Did Mary eat dinner last night?”**).

- After 200 training cycles learned to generalize verbs not in training set.
- Understanding of "regularities" in irregular verbs.
- Tendency to overregularize ("break"/"broked").
- Errors and performance similar to those of child learning new language.

ON LEARNING THE PAST TENSES OF ENGLISH VERBS



Rumelhart, McClelland. On learning the past tenses of English verbs, 1985.

Architetture Cognitive



“A cognitive architecture implements the invariant structure of the cognitive system” (Newell, 1990).

Il primo lavoro nacque negli anni '80s con la prima versione dell'architettura cognitiva **SOAR** (Newell, Laird and Rosenbloom, 1982)

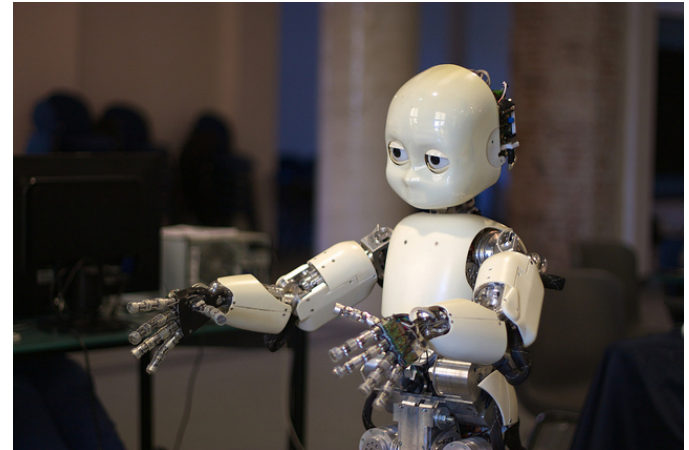
Idea: catturare gli elementi comuni sottostanti a diversi tipi di agenti intelligenti e fornire un framework in cui si possa modellare tale comportamento intelligente c

Allen Newell (1990)
Unified Theory of Cognition

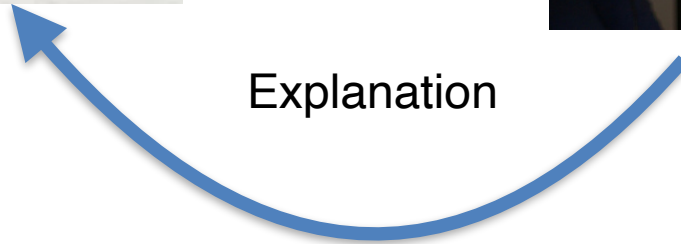
From Human to Artificial Cognition (and back)



Inspiration



Explanation



Potere esplicativo

Quando un sistema biologicamente/cognitivamente ispirato ha un potere esplicativo rispetto al sistema naturale preso come fonte di ispirazione?

Paragidmi di Progettazione

- Approccio Funzionalista vs Strutturalista.

Funzionalismo



- Introdotto da **H. Putnam** postulata la sufficienza di una **equivalenza debole** tra processi cognitivi e procedure di AI.
- Procedure di AI devono avere lo stesso ruolo funzionale (“**funzionare come**”) i corrispondenti processi cognitivi da modellare.
- **Realizzabilità multipla** (funzioni cognitive possono essere implementate in modi diversi).
- **Equivalenza a livello di proprietà macroscopiche di un comportamento intelligente** (basato sul fatto che sistema naturale e artificiale condividono le stesse specifiche di input-output).

•

Problemi con il Funzionalismo

- Se l'equivalenza è così debole non è possibile interpretare i risultati di un sistema (es. come si interpreta il **fallimento** di un sistema?).
- Un modello puramente funzionalista è una **black box** dove (riesco solo a predire il tipo di output dato il tipo di input in base ad una analogia superficiale con il sistema naturale analogo preso come fonte di ispirazione)

Birds & Jet



- “Both a **Bird** and a **Jet** can fly but a jet is not a **good explanatory model** of a bird since its flights mechanisms are different from the mechanism of bird”

- Sistemi puramente funzionalisti non sono modelli computazionali della cognizione (non hanno alcun potere esplicativo rispetto al sistema naturale preso come fonte di ispirazione).

Attuali Sistemi AI “Funzionalisti”



Sistemi intelligenti molto potenti ma che non hanno alcun potere esplicativo rispetto a come gli esseri umani risolvono lo stesso tipo di problemi



Some modern successful AI systems nature

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [articles](#) > [article](#)

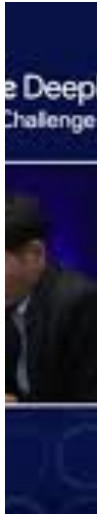
Article | [Open Access](#) | [Published: 15 July 2021](#)

Highly accurate protein structure prediction with AlphaFold

[John Jumper](#) , [Richard Evans](#), [Alexander Pritzel](#), [Tim Green](#), [Michael Figurnov](#), [Olaf Ronneberger](#), [Kathryn Tunyasuvunakool](#), [Russ Bates](#), [Augustin Žídek](#), [Anna Potapenko](#), [Alex Bridgland](#), [Clemens Meyer](#), [Simon A. A. Kohl](#), [Andrew J. Ballard](#), [Andrew Cowie](#), [Bernardino Romera-Paredes](#), [Stanislav Nikolov](#), [Rishub Jain](#), [Jonas Adler](#), [Trevor Back](#), [Stig Petersen](#), [David Reiman](#), [Ellen Clancy](#), [Michal Zielinski](#), ... [Demis Hassabis](#)  [+ Show authors](#)

[Nature](#) **596**, 583–589 (2021) | [Cite this article](#)

880k Accesses | **4314** Citations | **3260** Altmetric | [Metrics](#)

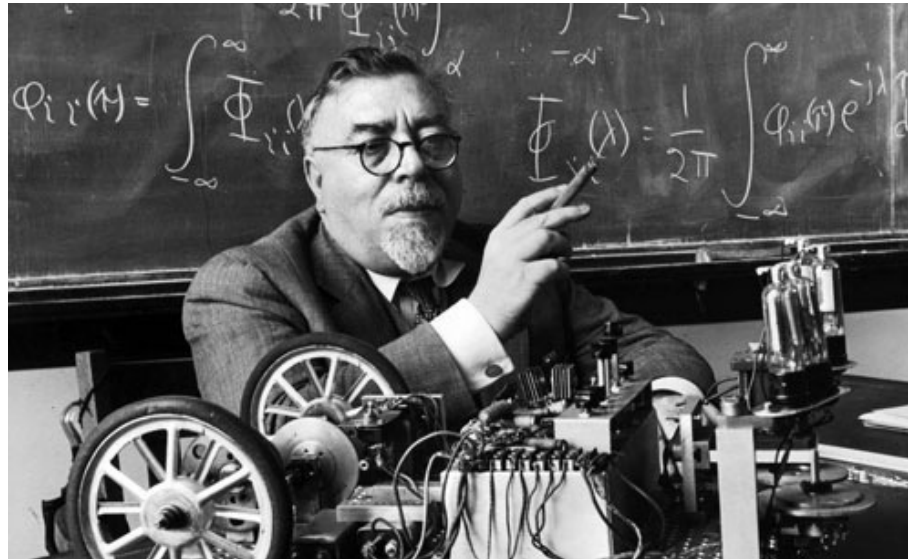


d)

Strutturalismo

- **Equivalenza forte tra sistemi cognitivi e procedure di AI (Cordeschi 2002; Milkowski, 2013; Lieto, 2021).**
- **Focus not solo sulla organizzazione funzionale dei processi ma anche sui vincoli che garantiscano una computazione "human-like"**

“Paradosso di Wiener”



“The best material model of a cat is another or possibly the same cat”

Bisogno di *proxy-models* (buone approssimazioni del reale)

Un Problema di Design

Z.Pylyshyn ('79): “if we do not formulate any restriction about a model we obtain the functionalism of a Turing machine. If we apply all the possible restrictions we reproduce a whole human being”

- Tra il livello esplicativo del funzionalismo (basato su una analogia **macroscopica** tra relazioni di input-output nei sistemi naturali e artificiali) e quello **microscopico dei modelli strutturalisti** ci sono, in mezzo, molti possibili modelli strutturali con diverso potere esplicativo.

Modelli Funzionalisti vs Strutturalisti



Stesse specifiche di *input-out* and **somiglianza superficiale** del modo in cui le componenti di un sistema e i propri meccanismi interni funzionano rispetto al sistema naturale



Stesse specifiche d *input-out* spec. + vincoli aggiuntivi che forniscono una maggiore somiglianza tra sistemi naturali e artificiali

Evolutionistic Explanation

Teleological Explanation

Causal Explanation

Functionalist Models



Structuralist Models

continuum

Functional Explanation

IBE

Mechanistic Explanation

BEHAVIORAL AND BRAIN SCIENCES (2001) **24**, 1033–1050

Printed in the United States of America

Can robots make good models of biological behaviour?

Barbara Webb

Department of Psychology, Centre for Computational and Cognitive Neuroscience, University of Stirling, Stirling FK9 4LA, Scotland, United Kingdom

b.h.webb@stir.ac.uk

www.stir.ac.uk/psychology/Staff/bhw1/

Webb's dimensions

1. **Biological Relevance:** this dimension shows if and, eventually to what extent, a computational model can be used to generate and test hypotheses about a given biological system taken as a source of inspiration.
2. **Level:** “what are the basic elements of the model that have no internal structure or their internal structures are ignored”. In other words it identifies the modelling focus.
3. **Generality:** the range of biological systems the model can represent.
4. **Abstraction:** the amount of details included in the artificial model with respect to the natural system taken as source of inspiration.
5. **Structural accuracy:** the similarity of the mechanisms behind the behaviour of an artificial model with respect to those of the target biological system.
6. **Performance match:** similarity of the performances of the model with respect to the performances obtained by the target biological system.
7. **Medium:** the physical medium that has been used to implement the model.

Limits of the Webb's account

1) the concept of “**biological relevance**” or “**structural accuracy**” are highly overlapping and there is not a clearly defined method that one could use in order to determine **how such elements are/can be operationally declined**.

2) “**Medium**” the Webb's proposal explicitly limits the considerations on this aspect to the presence (or not) of an embodied agent. The “medium”, in her view, is the physical body of the agent (a robot).

It does not consider - for example - alternative physical models of computations based, for example, on quantum computers or on hybrid biological/artificial neural networks realized in the field of **bionics** and **neuromorphing computing**

Minimal Cognitive Grid

“a **non subjective, graded**, evaluation framework allowing both **quantitative** and **qualitative** analysis about the **cognitive adequacy** and the **human-like performances** of artificial systems in both **single** and **multi-tasking** settings.” (Lietao, 2021)

Functional/Structural Ratio

Generality

Performance match (including errors and psychometric measures)

Functionalist Models



Structuralist Models

| | Epistemic goal | Quantitative evaluation | Qualitative evaluation | Graded evaluation | Subjective evaluation |
|-----------------------------|---|--------------------------------|-------------------------------|--------------------------|------------------------------|
| Functional/structural ratio | Evaluating the biological/cognitive adequacy of the artificial system <i>via</i> system dissection of its components/mechanisms | Yes | Yes | Yes | No |
| Generality | Evaluating the transferability of a given system/model to different tasks and biological/cognitive functions | Yes | Yes | Yes | No |
| Performance match | Comparing the output of the artificial system with the natural one(s) in terms of i) results, ii) errors, and iii) response times | Yes | Yes | Yes | No |

