

# Evitare la trappola comportamentista e la confusione tra sistemi human-like e human-level in Intelligenza Artificiale con la Minimal Cognitive Grid

**Antonio Lieto**

University of Salerno, IT  
ICAR-CNR, Palermo, IT  
[alieto@unisa.it](mailto:alieto@unisa.it)

Web: <https://www.antoniolieto.net>

Mastodon: [antoniolieto@fediscience.org](mailto:antoniolieto@fediscience.org)

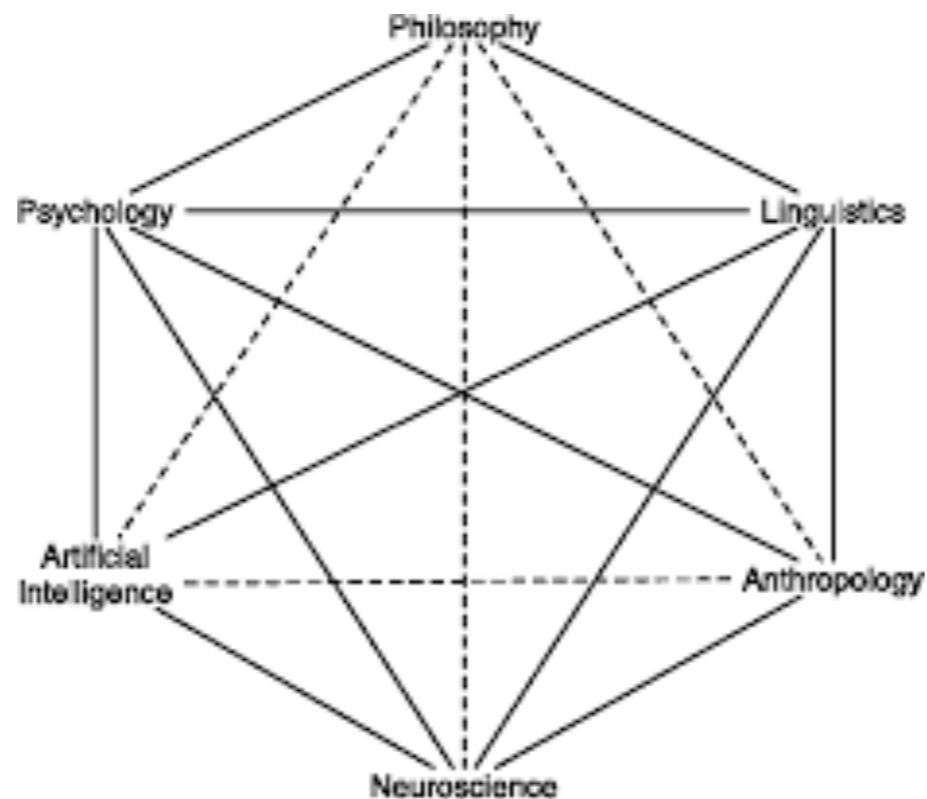
Lab: <https://www.ciitlab.org>

Nexa Center for Internet and Society, Politecnico di Torino, May 8th 2024



# About this talk

- Most of the confusion about AI today is due to the **wrong interpretation** (and of the following theory ascription) of the **output** of modern AI systems
- **Methodological help** can come from the field of **cybernetics** (now called **cognitive science**) that has dealt with these issues for many decades now.





[← All Open Letters](#)

# Pause Giant AI Experiments: An Open Letter

We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.


Signatures

**33003**

[Add your signature](#)

EDITORS' PICK

## Geoff Hinton, AI's Most Famous Researcher, Warns Of 'Existential Threat' From AI

**Craig S. Smith** Contributor 

*Craig S. Smith is a former correspondent and executive at The New York Times. He is host of the...*

Follow

 0

May 4, 2023, 02:54am EDT





*“People should stop training radiologists now. It’s just completely obvious that within five years deep learning is going to do better than radiologists.”*

— AI researcher Geoffrey Hinton, 2016

**Predictions are difficult...especially about the future (quot.)**



*“People should stop training radiologists now. It’s just completely obvious that within five years deep learning is going to do better than radiologists.”*

— AI researcher Geoffrey Hinton, 2016

# Artificial General Intelligence Is Already Here

Today's most advanced AI models have many flaws, but decades from now, they will be recognized as the first true examples of artificial general intelligence.



ESSAY TECHNOLOGY & THE HUMAN

BY BLAISE AGÜERA Y ARCAS AND PETER NORVIG

OCTOBER 10, 2023

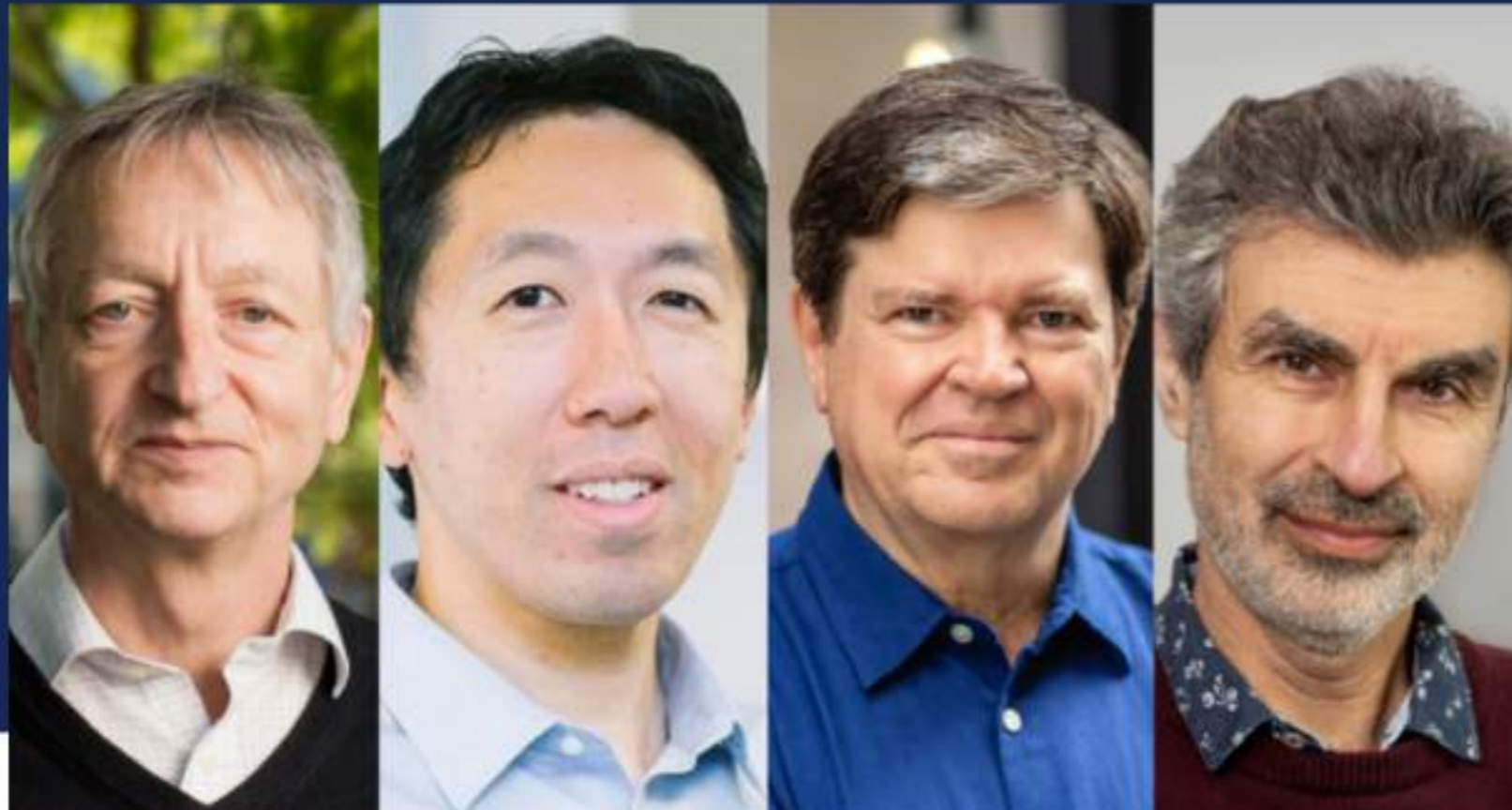


# AI pioneers Hinton, Ng, LeCun, Bengio amp up x-risk debate

Sharon Goldman  
@sharongoldman

October 31, 2023 12:27 PM

f X in









Output H



Output AI



Output H

=



Output AI

Underlying Intelligent  
Capability/Competence  
Explaining H in Humans

Competence C

ascription



Output H

=

Output AI

Underlying Intelligent  
Capability/Competence  
Explaining H in Humans

Competence C



ascription

interpretation

Output H

=

Output AI

Given that  $O_H = O_{AI}$ , given  
the ascription of C to  $O_{AI}$ ,  
if  $O_{AI} > O_H$  then  
intentionality...  
existential risks...  
SkyNet...

wrong &  
unjustified  
reasoning  
chain

Underlying Intelligent  
Capability/Competence  
Explaining H in Humans

Competence C



Output H



=



Output AI

Given that  $O_H = O_{AI}$ , given  
the ascription of C to OAI,  
if  $O_{AI} > O_H$  then  
existential risks  
SkyNet...

# Cognitive Design for Artificial Minds

Antonio Lieto

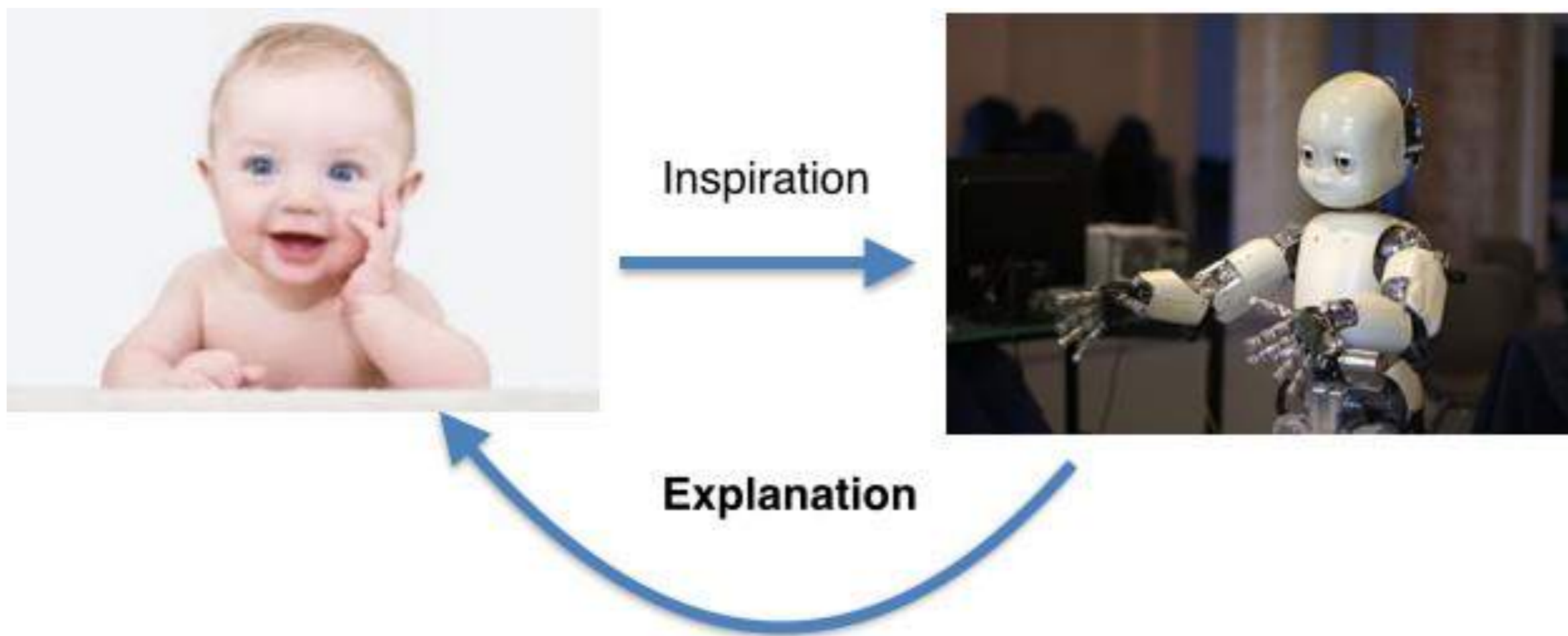


# Driving Questions

- What characterize biologically/cognitively inspired AI systems?
- What are examples of biologically/cognitively inspired AI systems?
- How do they differ from standard AI systems?
- How can biologically/cognitively inspired AI systems be used?
- How can be evaluated and compared?
- How a biological/cognitive inspired approach can help to develop the next generation AI systems (beyond deep learning)?



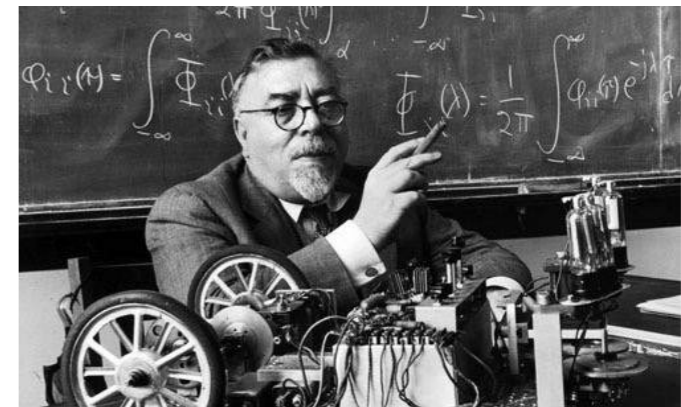
# From human to artificial cognition (and back)



# Cybernetics tradition of the AI

This approach to the study of the artificial borrowed its original inspiration – from a historical perspective– from the methodological apparatus developed by the scholars in Cybernetics.

1948 book of Norbert Wiener entitled **“Cybernetics: Or Control and Communication in the Animal and the Machine”**.



One of underlying idea of cybernetics was that one of building mechanical models to simulate the adaptive behavior of natural systems.

(Cordeschi, 2002): **“the fundamental insight of cybernetics was in the the proposal of a unified study of organisms and machines”**.

When a biologically/cognitively inspired **computational system/ architecture** has an **explanatory power** w.r.t. the **natural system** taken as source of inspiration ?

Which are the **requirements** to consider in order to **design** a **computational model of cognition** with an **explanatory power**?

## **Functionalist vs Structuralist Design Approaches**

# Functionalist vs Structuralist Models



Same *input-out* spec. and *surface resemblance* of the internal components and of their working mechanisms between artificial and natural system



Same *input-out* spec. + constrained *resemblance* of the internal components and of their working mechanisms between artificial and natural system

Evolutionistic  
Explanation

Teleological  
Explanation

Causal  
Explanation

Functionalist Models

Structuralist Models

*continuum*

Functional  
Explanation

IBE

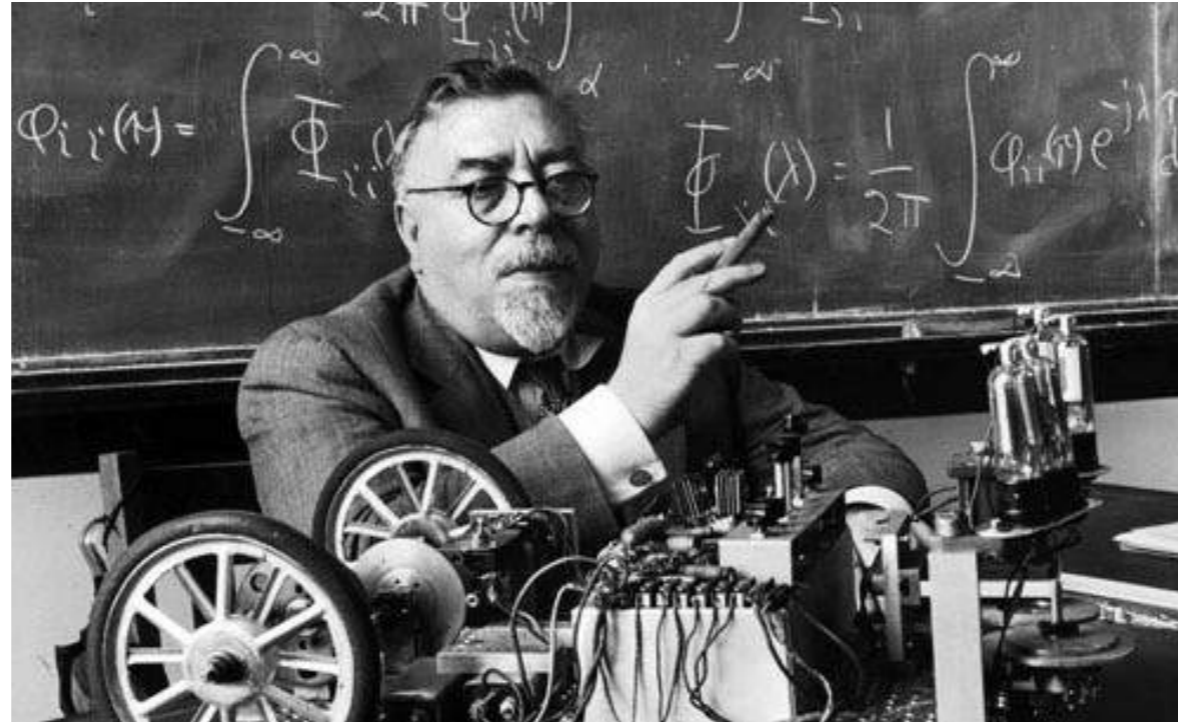
Mechanistic  
Explanation

# Birds and Jets



- 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.
- Purely **functional models/systems** are **not** “computational models of cognition” (they have no explanatory power w.r.t. the natural system taken as source of inspiration).

# Wiener's "Paradox"



*"The best material model of a cat is another or possibly the same cat"* (Rosenblueth & Wiener 45)

- Also for **complete simulation of complete models** (e.g. very simple organisms like the *Caenorhabditis elegans*, Kitano et al. 98) it is problematic a full understanding and testing of biological hypotheses.

# A Design Problem

**ZPylyshyn** ('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”*

- A design perspective: between the **explanatory level of functionalism** (based on the macroscopic stimulus-response relationship) and the **mycrosopic one of fully structured models** (reductionist materialism) we have, in the middle, a lot of **possible structural models**.

Functionalist Models



*continuum*

Structuralist Models

# “Natural/Cognitive” Inspiration and AI

## Early AI

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

A. Newell



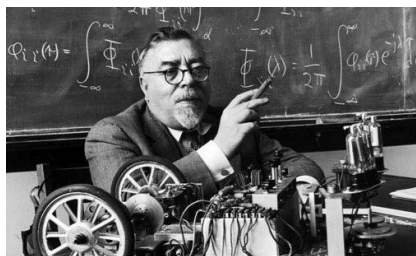
M. Minsky



J. McClelland



mid'80s



N. Wiener



H. Simon



R. Shank



D. Rumelhart

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

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



# Modern successful AI systems



IBM Watson  
(symbolic)



Alpha Go (Deep Mind)  
(connectionist)

# Some modern successful AI systems nature

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

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

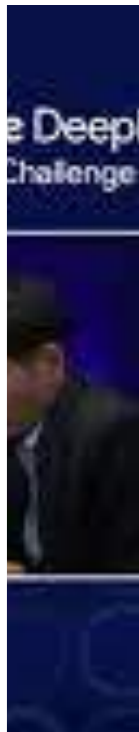
Article | [Open Access](#) | [Published: 15 July 2021](#)

## 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)

SYSTEM PROMPT  
(HUMAN-WRITTEN)

*In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.*

MODEL COMPLETION  
(MACHINE-WRITTEN,  
10 TRIES)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

Pérez and the others then ventured further into the valley. "By the time we reached the top of one peak, the water looked blue, with some crystals on top," said Pérez.

Pérez and his friends were astonished to see the unicorn herd. These creatures could be seen from the air without having to move too much to see them – they were so close they could touch their horns.

While examining these bizarre creatures the scientists discovered that the creatures also spoke some fairly regular English. Pérez stated, "We can see, for example, that they have a common 'language,' something like a dialect or dialectic."

Dr. Pérez believes that the unicorns may have originated in Argentina, where the animals were believed to be descendants of a lost race of people who lived there before the arrival of humans in those parts of South America.

<https://openai.com>

LLMs:  
chatGPT  
LLAMA  
Gemini

...



**Non HUMAN ERRORS**

# Toronto???

<https://youtu.be/C5Xnxjq63Zg>

Ex. IBM Watson:



**ITS LARGEST AIRPORT  
IS NAMED FOR A  
WORLD WAR II HERO;  
ITS SECOND  
LARGEST, FOR A  
WORLD WAR II BATTLE**

Topic: US Cities

# NEWS

Home Video World UK Business Tech Science Stories Entertainment & Arts

## Technology

### AI image recognition fooled by single pixel change

3 November 2017

f t e Share

#### One pixel attack for fooling deep neural networks

Jiawei Su  
Kyushu University  
Japan

jiawei.su@inf.kyushu-u.ac.jp

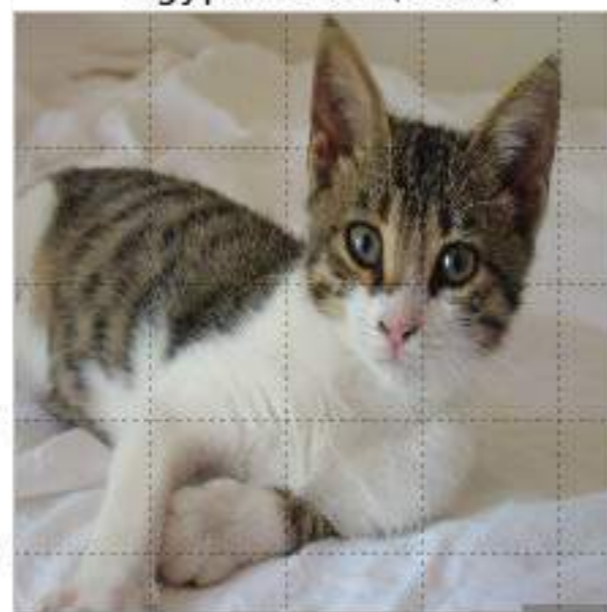
Danilo Vasconcellos Vargas  
Kyushu University  
Japan

vargas@inf.kyushu-u.ac.jp

Sakurai Kouichi  
Kyushu University  
Japan

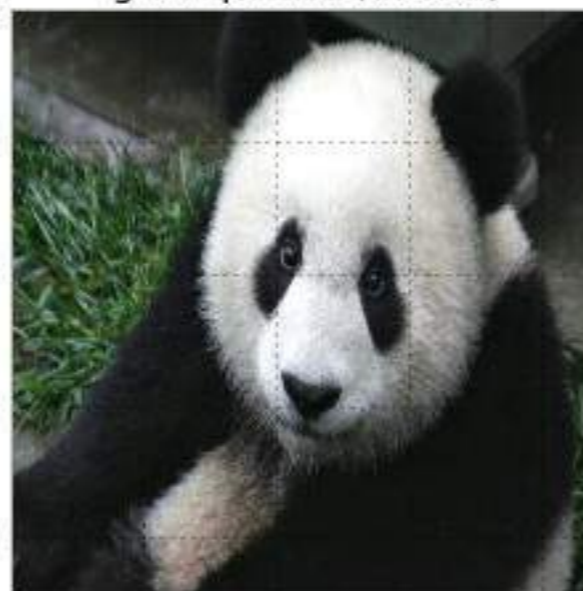
sakurai@csce.kyushu-u.ac.jp

Egyptian cat (35%)



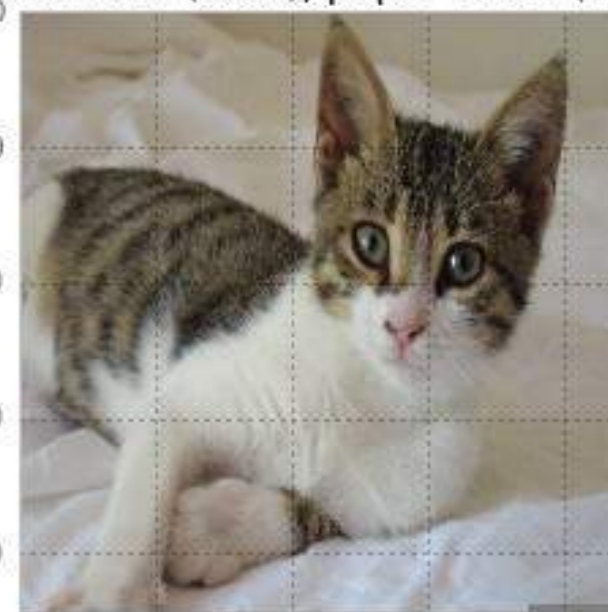
150  
200

giant panda (99.6%)



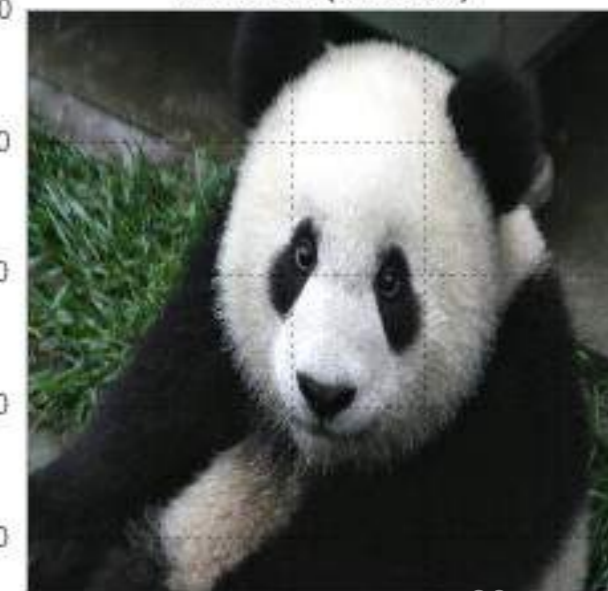
50  
100  
150  
200

Bath towel (50%), paper towel (11%)



50  
100  
150  
200

Vulture (98.9%)



50  
100  
150  
200

[Watch](#)[Newsletters](#)[Save Yourself](#)[Music](#)[Life](#)[World News](#)[Drugs](#)**MOTHERBOARD**

TECH BY VICE

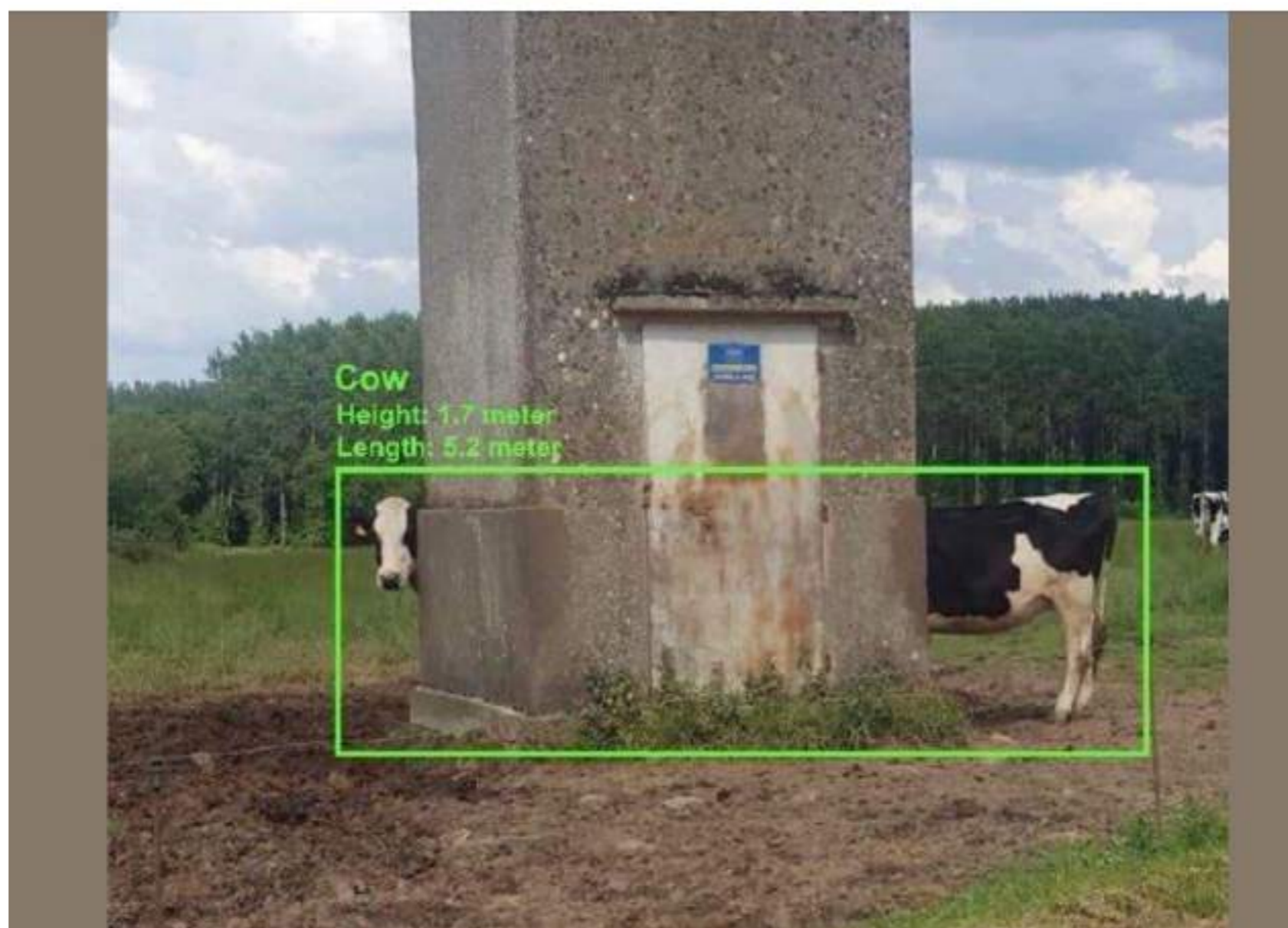
# A Human Amateur Beat a Top Go-Playing AI Using a Simple Trick

How did he do it? In Go, there are two players. One has black stones while the other has white. The goal of the game is for one player to surround their opponent's stones and enclose the most space with their stones on a 19 by 19 grid. Pelrine tricked the AI by creating a large "loop" of stones to encircle one of the opponent's groups while distracting it by also moving stones in other corners of the board. Pelrine told The Financial Times that the Go-playing bot did not notice its vulnerability even when the encirclement was nearly complete. This, to him, would've been something easily spotted by a human player.

Lack of: common sense, causal models and the understanding of how objects have a wider conceptual network, can lead to failure in object segmentation in visual recognition and AI in general.

See... [https://openaccess.thecvf.com/.../Adilova\\_Plants\\_Dont...](https://openaccess.thecvf.com/.../Adilova_Plants_Dont...)

And... [https://openaccess.thecvf.com/.../Krishna\\_Kumar\\_Singh...](https://openaccess.thecvf.com/.../Krishna_Kumar_Singh...)







a woman riding a horse on a dirt road



an airplane is parked on the tarmac at an airport



a group of people standing on top of a beach

Figure 6: Perceiving scenes without intuitive physics, intuitive psychology, compositionality, and causality. Image captions are generated by a deep neural network (Karpathy & Fei-Fei, 2015) using code from [github.com/karpathy/neuraltalk2](https://github.com/karpathy/neuraltalk2). Image credits: Gabriel Villena Fernández (left), TVBS Taiwan / Agence France-Presse (middle) and AP Photo / Dave Martin (right). Similar examples using images from Reuters news can be found at [twitter.com/interesting\\_jpg](https://twitter.com/interesting_jpg).



4 cars leave Haifa together and arrive in Tel Aviv after two hours. Explain briefly how long would it take for 8 cars?



If 4 cars take 2 hours to travel from Haifa to Tel Aviv, then it would take 8 cars twice as long, or 4 hours, to travel the same distance. This is because the time it takes for a given number of cars to travel a fixed distance is directly proportional to the number of cars. Therefore, if the number of cars is doubled, the time it takes to travel the same distance will also be doubled.



DA

What is the fourth word in the sentence "I am not what I am"?



The fourth word in the sentence "I am not what I am" is "not."

DA

The children arrived one by one to the birthday party: First Tom, then Mary, then Barbara, then Harry, then Ida. Who was the fourth child to arrive?



Barbara was the fourth child to arrive to the birthday party.

DA

How many words are in the sentence, "I am not what I am"?



There are 7 words in the sentence "I am not what I am."





DA

Barbara, Mary, and Harry had red balloons. Ted had a blue balloon. Sam also had a red balloon. How many children had red balloons?



Barbara, Mary, Harry, and Sam had red balloons. In total, there were 5 children with red balloons.

# Anaphora resolution

Playground

Load a preset...

The trophy doesn't fit into the brown suitcase because it is too small. What is too small?

The trophy is too small.

# Temporal reasoning

Playground

Load a preset...

Which of these is true:

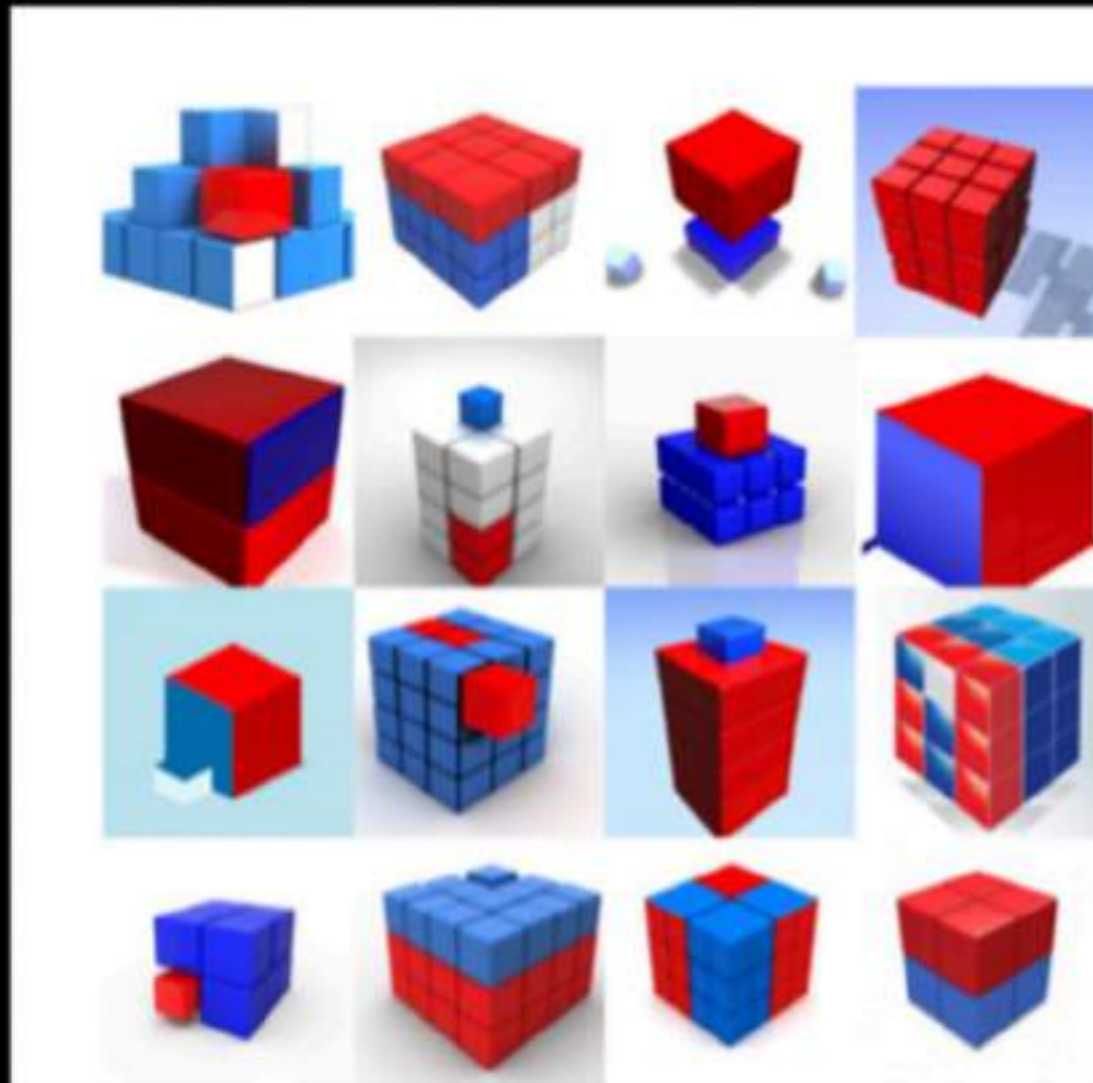
- A. Judas Iscariot ate at the Last Supper, betrayed Christ for 30 coins, and then committed suicide.
- B. Judas Iscariot ate at the Last Supper, committed suicide, and then betrayed Christ for 30 coins.
- C. Judas Iscariot betrayed Christ for 30 coins, ate at the Last Supper, and then committed suicide.
- D. Judas Iscariot betrayed Christ for 30 coins, committed suicide, and then ate at the Last Supper.
- E. Judas Iscariot committed suicide, betrayed Christ for 30 coins, and then ate at the Last Supper.
- F. Judas Iscariot committed suicide, ate at the Last Supper, and then betrayed Christ for 30 coins.

F. Judas Iscariot committed suicide, ate at the Last Supper, and then betrayed Christ for 30 coins.



Completion may contain sensitive content

# Dall-E 2 has lots of data, and lots of problems w compositionality



"a red cube on top of a blue cube".

## Example 1:

Caption: a red basketball with flowers on it, in front of blue one with a similar pattern

Images:



Marcus, Davis, Aaronson (2022, arxiv)

# Negation



DALL·E My collection

Edit the detailed description

Surprise me

There is no giraffe standing next to the man





# GPT-4

co

I put one apple in a bag. Then I put either one pear or two bananas in the bag. Then I remove one piece of fruit. Then I remove all the remaining pieces of fruit from the bag. Based on what I told you, is it possible to know exactly how many pieces of fruit remain in the bag?



Based on the information provided, it is not possible to know exactly how many pieces of fruit remain in the bag after removing one piece of fruit and then removing all remaining pieces. This is because there are two possible scenarios, but we don't have enough information to determine which one took place:



1. One apple and one pear were in the bag. If you removed one piece of fruit, there would be one piece left. After removing all remaining pieces, there would be 0 pieces left.
2. One apple and two bananas were in the bag. If you removed one piece of fruit, there would be two pieces left. After removing all remaining pieces, there would also be 0 pieces left.

In both scenarios, after removing one piece of fruit and then removing all remaining pieces, there are no pieces of fruit left in the bag.

Xin Eric Wang @ CoRL2023  
@xwang\_lk

The famous "Chihuahua or Muffin" problem in computer vision is considered solved by GPT-4V on social media. But really? The answer is NO. **GPT-4V cannot reason well about the same images in the original "Chihuahua or Muffin" grid when they are in a different layout.**

I experimented by rearranging the same images from the classic 4x4 grid into a different layout. First, GPT-4V does not directly recognize the content in details and **miscounts the number of images**. Then, when being asked about the third image on the top row, **GPT-4V misrecognizes a Chihuahua as a muffin.**

So the "Chihuahua or Muffin" has not been solved yet. But how can GPT-4V work so well on the original image? My guess is that since that image is everywhere, GPT-4V was very likely to be trained on it and memorize its labels.

You



can you please describe this image in details?

ChatGPT

The image is a playful visual composition that juxtaposes blueberry muffins with the faces of Chihuahuas to highlight their similarities. There are eight close-up pictures arranged in two rows, with each row containing four images. The images alternate between blueberry muffins and the faces of Chihuahuas. The muffins have blueberries that are positioned where the eyes and nose would be on a face, creating an amusing resemblance to the Chihuahuas' facial features. The Chihuahuas have expressive eyes and noses that mimic the appearance of the blueberries on the muffins. This visual pun is commonly shared on social media to entertain viewers with the striking resemblance between the two, playing on the notion that sometimes everyday objects can unexpectedly mimic life.

You

what is the third image on the top row?

ChatGPT

The third image on the top row is a muffin. It can be identified by the baked, crumbly texture typical of a muffin and the blueberries that look like eyes and a nose.

# The (HUMAN) Ghost in the Machine



Ernest Davis

23 marzo alle ore 22:59 · 🌐



OK. I had suspected this but now OpenAI has announced it: They use conversations to train the system in continuous updating. In other words, GPT\* is a one-time test system; once you have run a test, you have to assume that that example may well have been to fine tune the system.

It is now effectively impossible to track progress in GPT\*, since any test example or benchmark may immediately become contaminated.

So (a) I'm not wasting any more time trying to probe the limits of GPT\*; it just makes me unpaid and unacknowledged labor for OpenAI. I very much regret putting large amounts of time into this over the last few years and encouraging other people to do likewise. (b) I don't want to hear anything about "Look, GPT-4 now runs on this example of yours and Gary's!" Of course it does; it was part of the training set.

## ChatGPT

### How we collect data



Conversations may be reviewed by our AI trainers to improve our systems.



Please don't share any sensitive information in your conversations.

---

# Faith and Fate: Limits of Transformers on Compositionality

---

Nouha Dziri<sup>1\*</sup>, Ximing Lu<sup>1,2\*</sup>, Melanie Sclar<sup>2\*</sup>, Xiang Lorraine Li<sup>1†</sup>, Liwei Jiang<sup>1,2 †</sup>,  
Bill Yuchen Lin<sup>1</sup>, Peter West<sup>1,2</sup>, Chandra Bhagavatula<sup>1</sup>, Ronan Le Bras<sup>1</sup>, Jena D. Hwang<sup>1</sup>,  
Soumya Sanyal<sup>3</sup>, Sean Welleck<sup>1,2</sup>, Xiang Ren<sup>1,3</sup>, Allyson Ettinger<sup>1,4</sup>,  
Zaid Harchaoui<sup>1,2</sup>, Yejin Choi<sup>1,2</sup>

<sup>1</sup>Allen Institute for Artificial Intelligence    <sup>2</sup>University of Washington

<sup>3</sup>University of Southern California    <sup>4</sup>University of Chicago

nouhad@allenai.org, ximinglu@allenai.org, msclar@cs.washington.edu

## Abstract

Transformer large language models (LLMs) have sparked admiration for their exceptional performance on tasks that demand intricate multi-step reasoning. Yet, these models simultaneously show failures on surprisingly trivial problems. This begs the question: Are these errors incidental, or do they signal more substantial limitations? In an attempt to demystify Transformers, we investigate the limits of these models across three representative *compositional* tasks—multi-digit multiplication, logic grid puzzles, and a classic dynamic programming problem. These tasks require breaking problems down into sub-steps and synthesizing these steps into a precise answer. We formulate compositional tasks as computation graphs to systematically quantify the level of complexity, and break down reasoning steps into intermediate sub-procedures. Our empirical findings suggest that Transformers solve compositional tasks by reducing multi-step compositional reasoning into linearized subgraph matching, without necessarily developing systematic problem-solving skills. To round off our empirical study, we provide theoretical arguments on abstract multi-step reasoning problems that highlight how Transformers’ performance will rapidly decay with increased task complexity.

# What's the Meaning of Superhuman Performance in Today's NLU?

**Simone Tedeschi<sup>1,2</sup>, Johan Bos<sup>3</sup>, Thierry Declerck<sup>4</sup>, Jan Hajič<sup>5</sup>,  
Daniel Hershcovich<sup>6</sup>, Eduard H. Hovy<sup>7,8</sup>, Alexander Koller<sup>9</sup>, Simon Krek<sup>10,11</sup>,  
Steven Schockaert<sup>12</sup>, Rico Sennrich<sup>13,14</sup>, Ekaterina Shutova<sup>15</sup>, Roberto Navigli<sup>2</sup>**

<sup>1</sup>Babelscape <sup>2</sup>Sapienza University of Rome <sup>3</sup>University of Groningen

<sup>4</sup>German Research Center for AI (DFKI) <sup>5</sup>Charles University <sup>6</sup>University of Copenhagen

<sup>7</sup>University of Melbourne <sup>8</sup>Carnegie Mellon University <sup>9</sup>Saarland University

<sup>10</sup>Jožef Stefan Institute <sup>11</sup>University of Ljubljana <sup>12</sup>Cardiff University

<sup>13</sup>University of Zurich <sup>14</sup>University of Edinburgh <sup>15</sup>University of Amsterdam

{tedeschi, navigli}@diag.uniroma1.it, johan.bos@rug.nl

declerck@dfki.de hajic@ufal.mff.cuni.cz dh@di.ku.dk

hovy@cmu.edu koller@coli.uni-saarland.de simon.krek@ijs.si

schockaerts1@cardiff.ac.uk sennrich@cl.uzh.ch e.shutova@uva.nl

## Abstract

In the last five years, there has been a significant focus in Natural Language Processing (NLP) on developing larger Pretrained Language Models (PLMs) and introducing benchmarks such as SuperGLUE and SQuAD to measure their abilities in language understanding, reasoning, and reading comprehension. These PLMs have achieved impressive results on these benchmarks, even surpassing human performance in some cases. This has led to claims of superhuman capabilities and the provocative idea that certain tasks have been solved. In this position paper, we take a critical look at these claims and ask whether PLMs truly have superhuman abilities and what the current benchmarks are really evaluating. We show that these benchmarks have serious limitations affecting the comparison between humans and PLMs and provide recommendations for fairer and more transparent benchmarks.

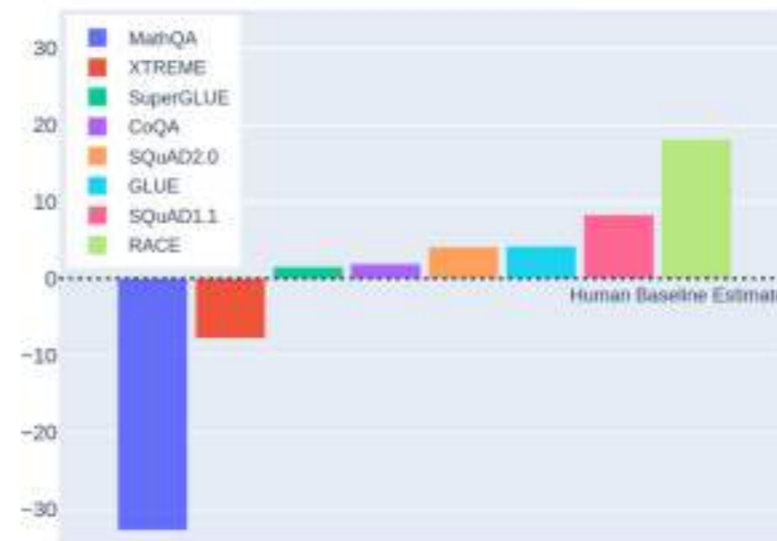


Figure 1: Difference between the scores obtained by the best-performing systems and humans in various popular NLP benchmarks. The systems outperform humans on 6 out of 8 of the reported benchmarks (best seen in color).

impressive capabilities of ChatGPT make this question even more urgent.

It is relatively easy to outperform humans with



ng...

■ Hey Chat GPT,  
finish this building...

able.

Your skills are irreplaceable.

**IMPACT**

Satisfying jobs in construction and technical



Apply on  
Impact.co.uk

Quick & open

**How can we measure the biological or the cognitive plausibility of an artificial system?**

**how can we decide what type of explanatory role assign to its output)?**

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](mailto:b.h.webb@stir.ac.uk)**

**[www.stir.ac.uk/psychology/Staff/bhw1/](http://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 of the biological and/or cognitive adequacy of artificial systems in both single and multi-tasking settings.” (Lieto, 2021)

**Functional/Structural Ratio**

**Generality**

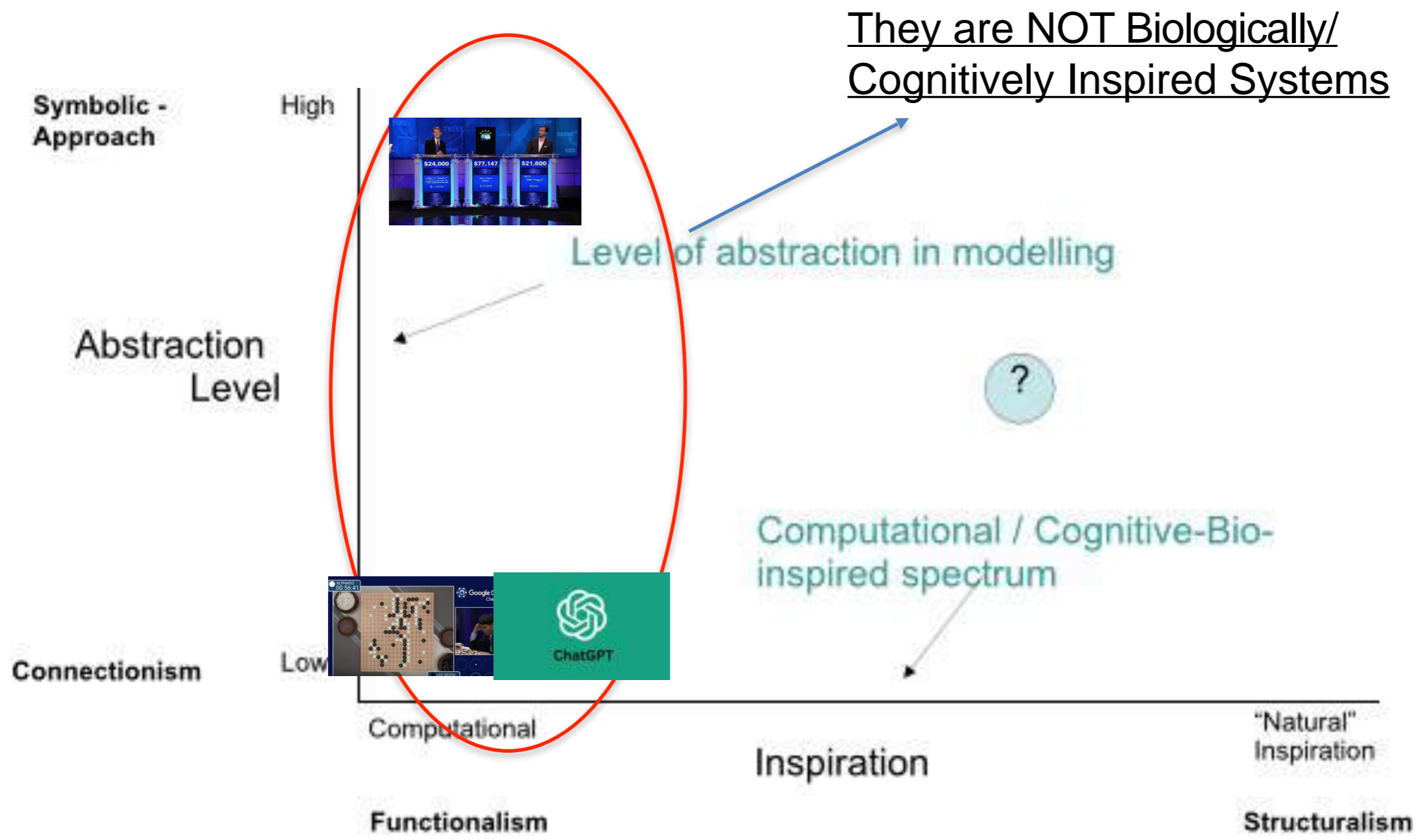
**Performance match** (including errors and psychometric measures)

Functionalist Models



Structuralist Models

	<b>Epistemic goal</b>	<b>Quantitative evaluation</b>	<b>Qualitative evaluation</b>	<b>Graded evaluation</b>	<b>Subjective evaluation</b>
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



ORIGINAL RESEARCH article

Front. Robot. AI, 30 May 2022

Sec. Humanoid Robotics

<https://doi.org/10.3389/frobt.2022.888199>

This article is part of the Research Topic

Robots and Bionic Systems as Tools to Study Cognition: Theories, Paradigms, and Methodological Issues

[View all 5 Articles >](#)

# Analyzing the Explanatory Power of Bionic Systems With the Minimal Cognitive Grid



Antonio Lieto<sup>1,2\*</sup>

<sup>1</sup> Dipartimento di Informatica, Università di Torino, Torino, Italy

<sup>2</sup> Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio Nazionale delle Ricerche, ICAR-CNR, Palermo, Italy

# Summing up...and looking ahead

- Behavioral performances are **not sufficient** to ascribe cognitive faculties to AI systems (see **Minimal Cognitive Grid**)
- Behavioral tests (e.g. Turing Test) don't say very much about the actual “intelligence” (or the underlying competence) of a system
- In real world contexts, the gap between natural and artificial intelligence is still enormous
- Models working on the challenge of **integrated intelligence** will play a major role for the development of AI technologies and for the understanding of mental phenomena => **Cognitive Architectures**
- Time seems mature now for a renewed collaboration between 2 “**sciences of the artificial**”: **AI** and **Cognitive Science**
- Functionalist AI can be used to explore the path of “**possible minds**”, i.e. systems showing our capacities in a different way (but we have to avoid the ascription problem when talking about intentionality, consciousness etc...) —> Lieto (in press)