Can We Trust The Algorithm?

(And for what?)

Marco Piastra

Intellectus Novus:
De Affidabilitate Machinarium Sapientium

2023-11-23

Apropos Generative AI: ChatGPT and all that

The Fair of Wonders: ChatGPT

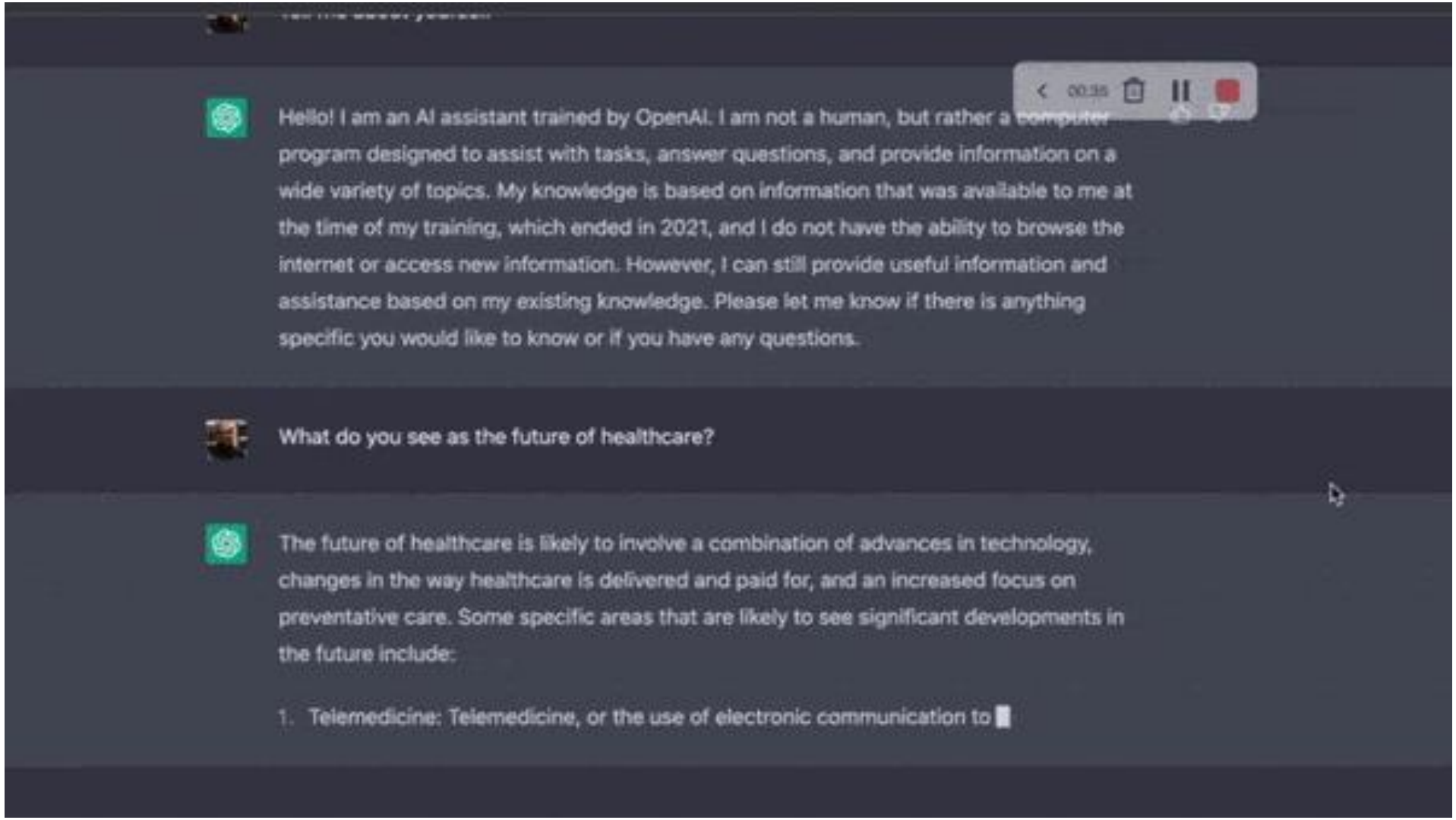
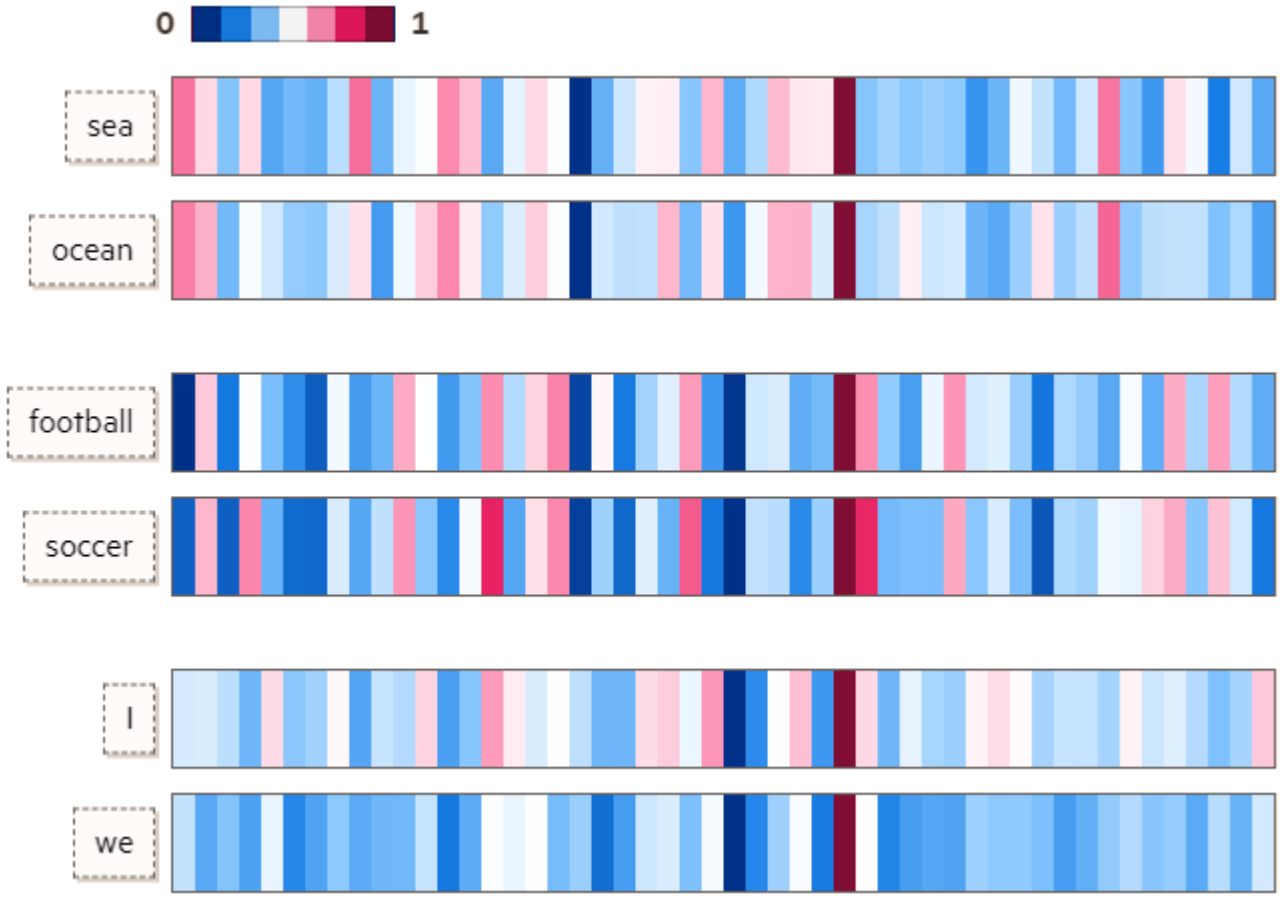


Image from https://info.homecareinnovationforum.com/the-potential-of-chatgpt-in-home-health-care

ChatGPT: How does it work?

- **Positional encoding (*Embedding*)**

Words (=*token*) from natural language are each translated into a high-dimensional *numerical vector*



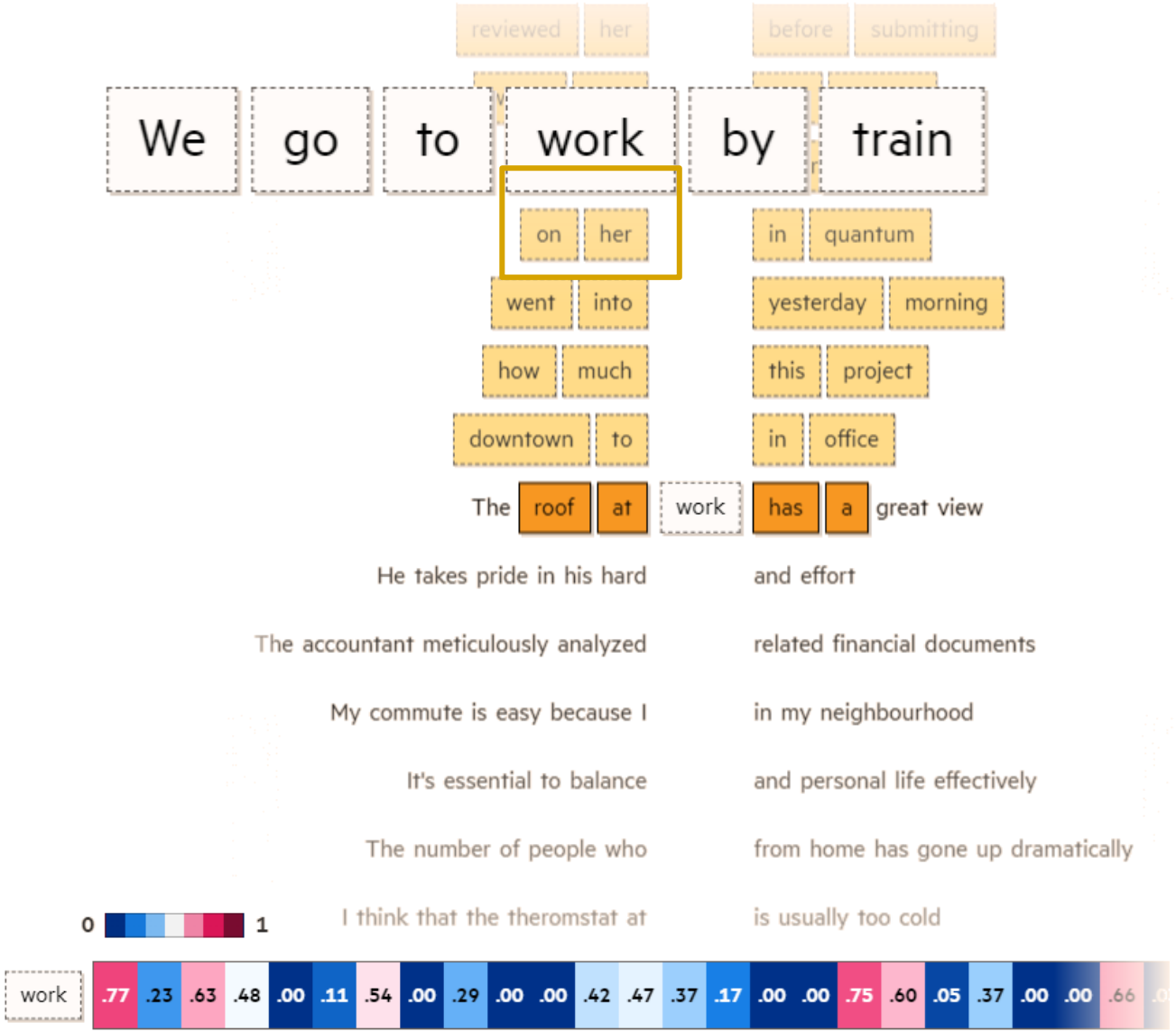
Images from https://ig.ft.com/generative-ai/

ChatGPT: How does it work?

- **Positional encoding (*Embedding*)**

  Words (=*token*) from natural language are each translated into a high-dimensional *numerical vector*

  Such vector is computed by estimating the *probability of co-occurrence* in a context of other words in a (very) large text corpus



Images from https://ig.ft.com/generative-ai/

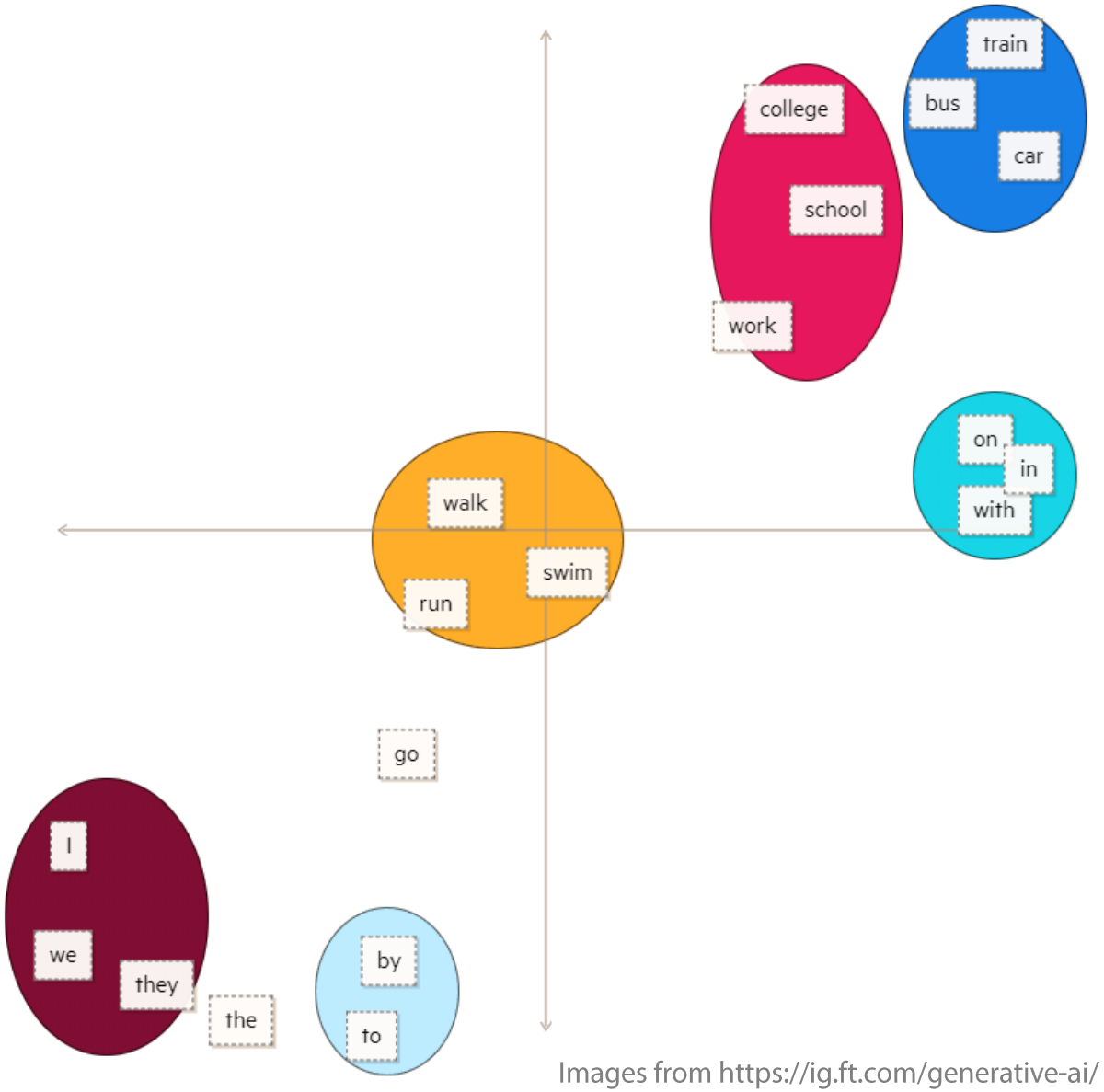
ChatGPT: How does it work?

- **Positional encoding (*Embedding*)**

  Words (=*token*) from natural language are each translated into a high-dimensional *numerical vector*

  Such vector is computed by estimating the *probability of co-occurrence* in a context of other words in a (very) large text corpus

  In this way, the *numerical similarity* among vectors is representative of words' affinity in terms of role or meaning (or both)



Images from https://ig.ft.com/generative-ai/

ChatGPT: How does it work?

- **Input-output relations**

  Sentences in input (*prompt*) are fed word by word to a very sophisticated and complex *artificial neural network*

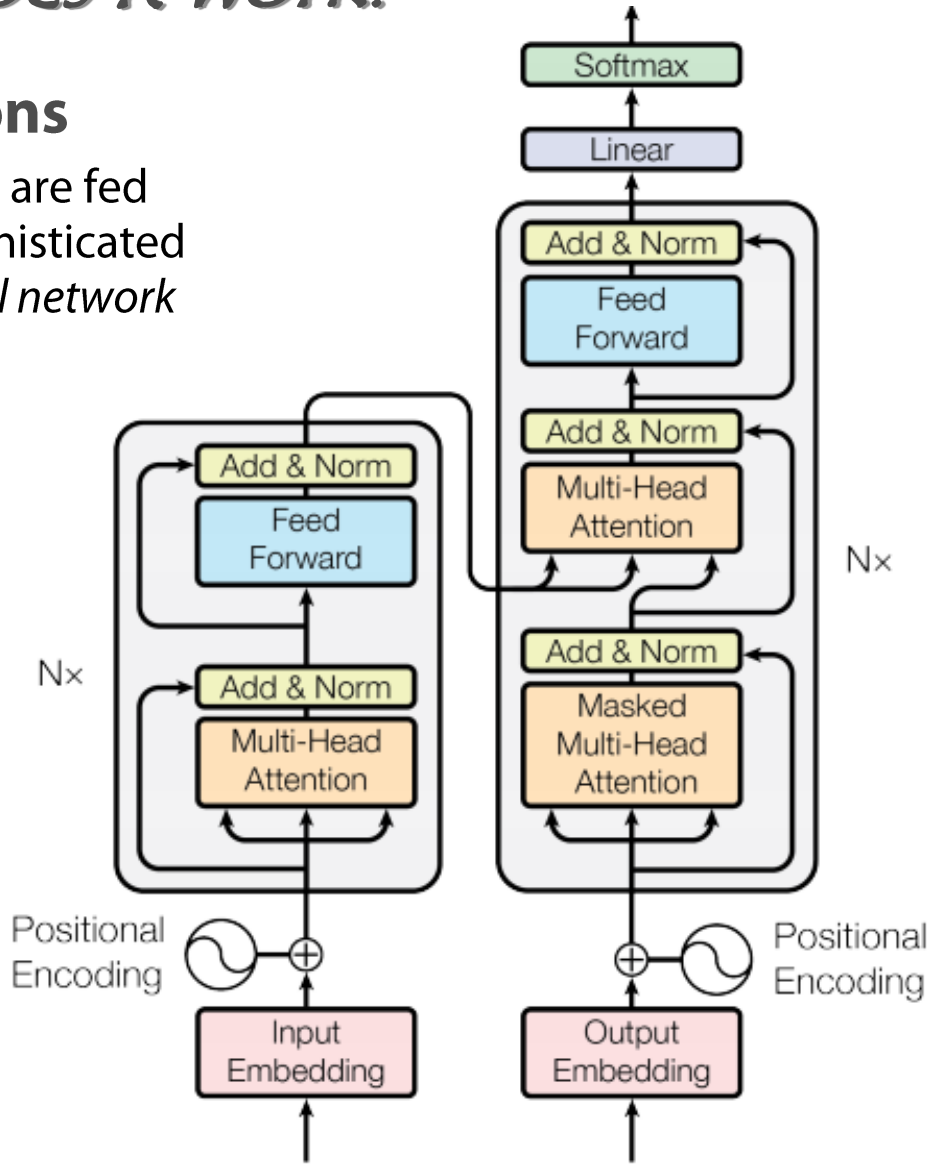


Image from https://arxiv.org/abs/1706.03762

ChatGPT: How does it work?

- **Input-output relations**

  Sentences in input (*prompt*) are fed word by word to a very sophisticated and complex *artificial neural network*

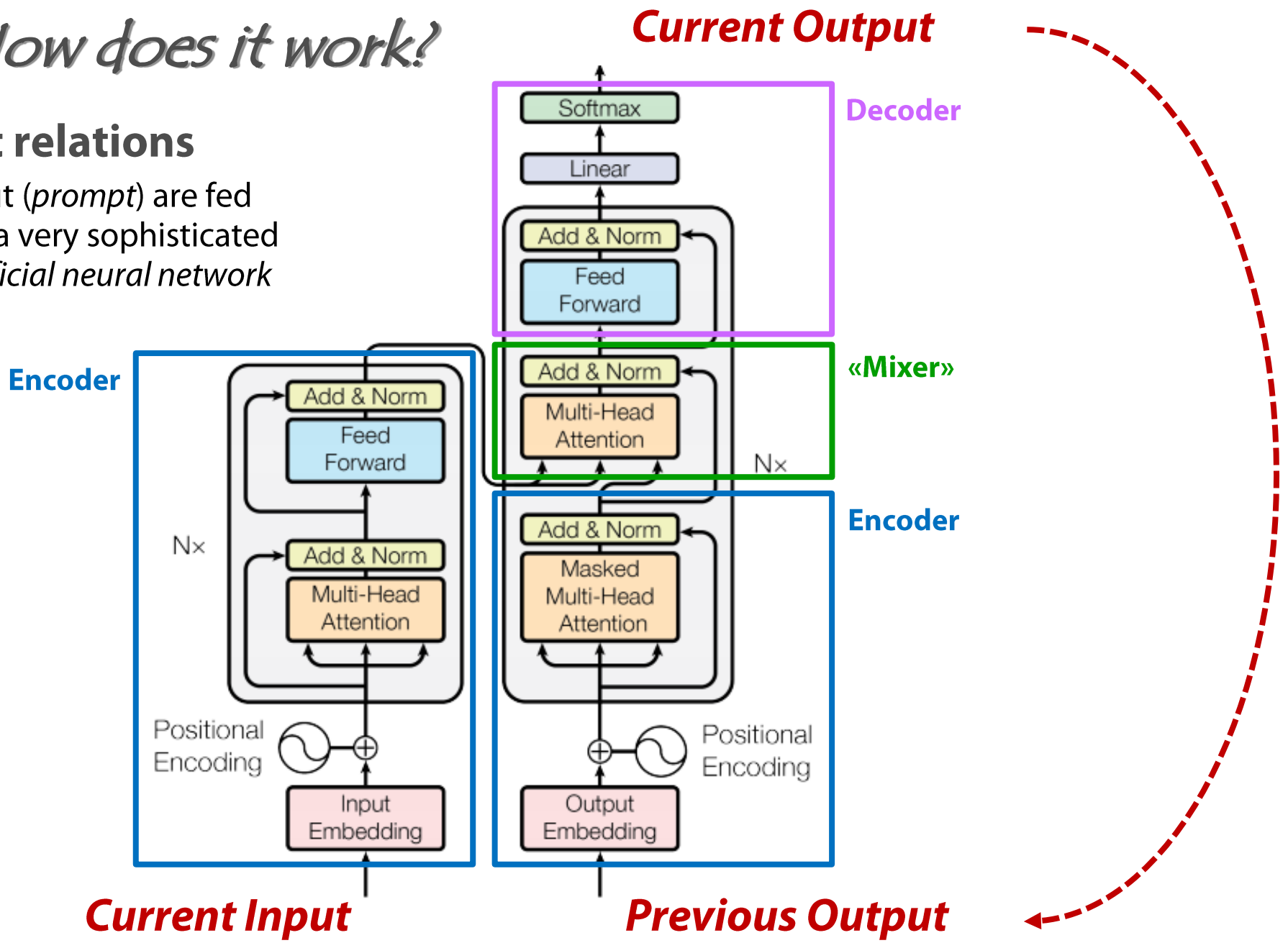


Image from https://arxiv.org/abs/1706.03762

ChatGPT: How does it work?

- **Input-output relations**

  Sentences in input (*prompt*) are fed word by word to a very sophisticated and complex *artificial neural network*

  Example: translating a sentence from Portuguese to English

  este é o primeiro livro que eu fiz.

  Note that the output is fed back incrementally as additional input to the network itself

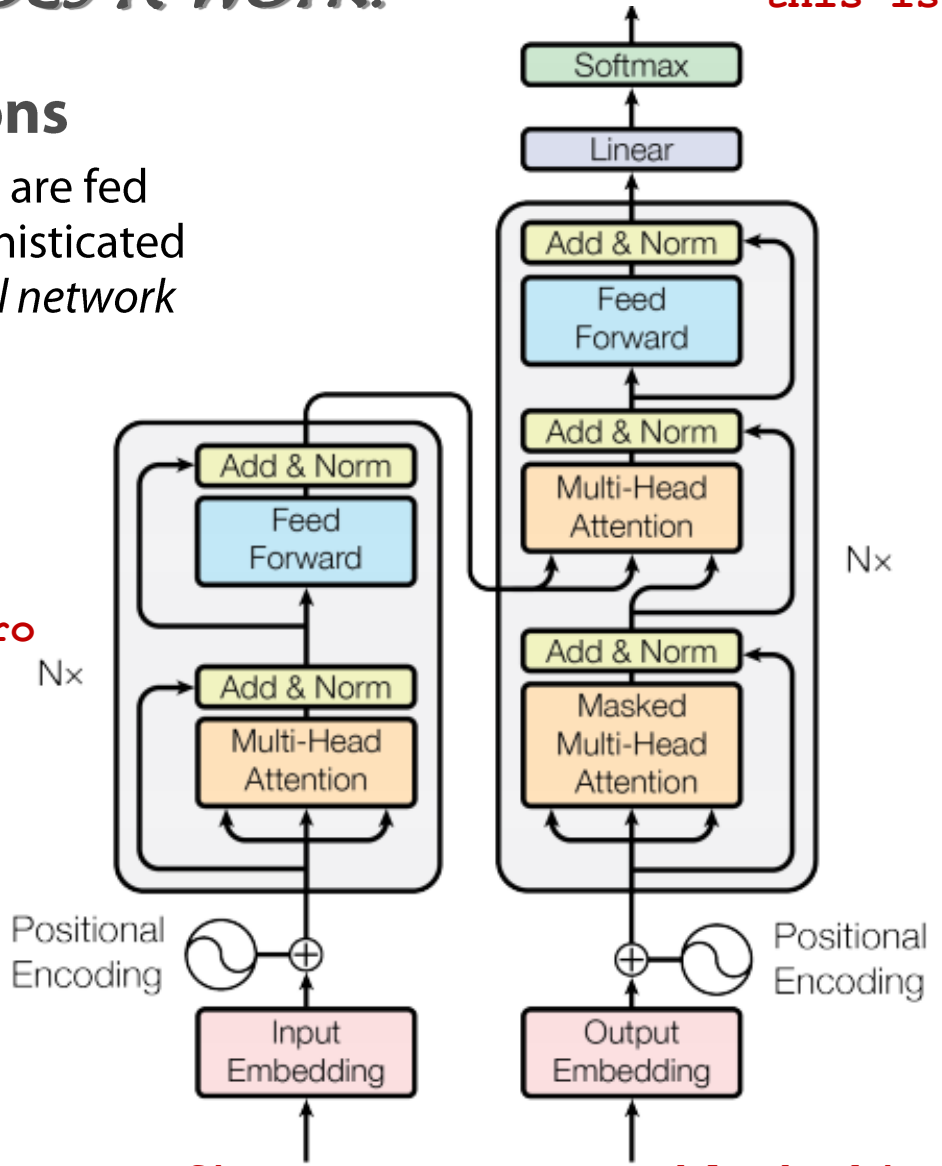


Image from https://arxiv.org/abs/1706.03762

this is the first book i've ever done.

este é o primeiro livro que eu fiz. **~~blank~~ this is the first book i've ever done**

ChatGPT: How does it work?

- **Input-output relations**

  Sentences in input (*prompt*) are fed word by word to a very sophisticated and complex *artificial neural network*

  The artificial neural network produces as its output a *probability distribution* for the next word to be generated

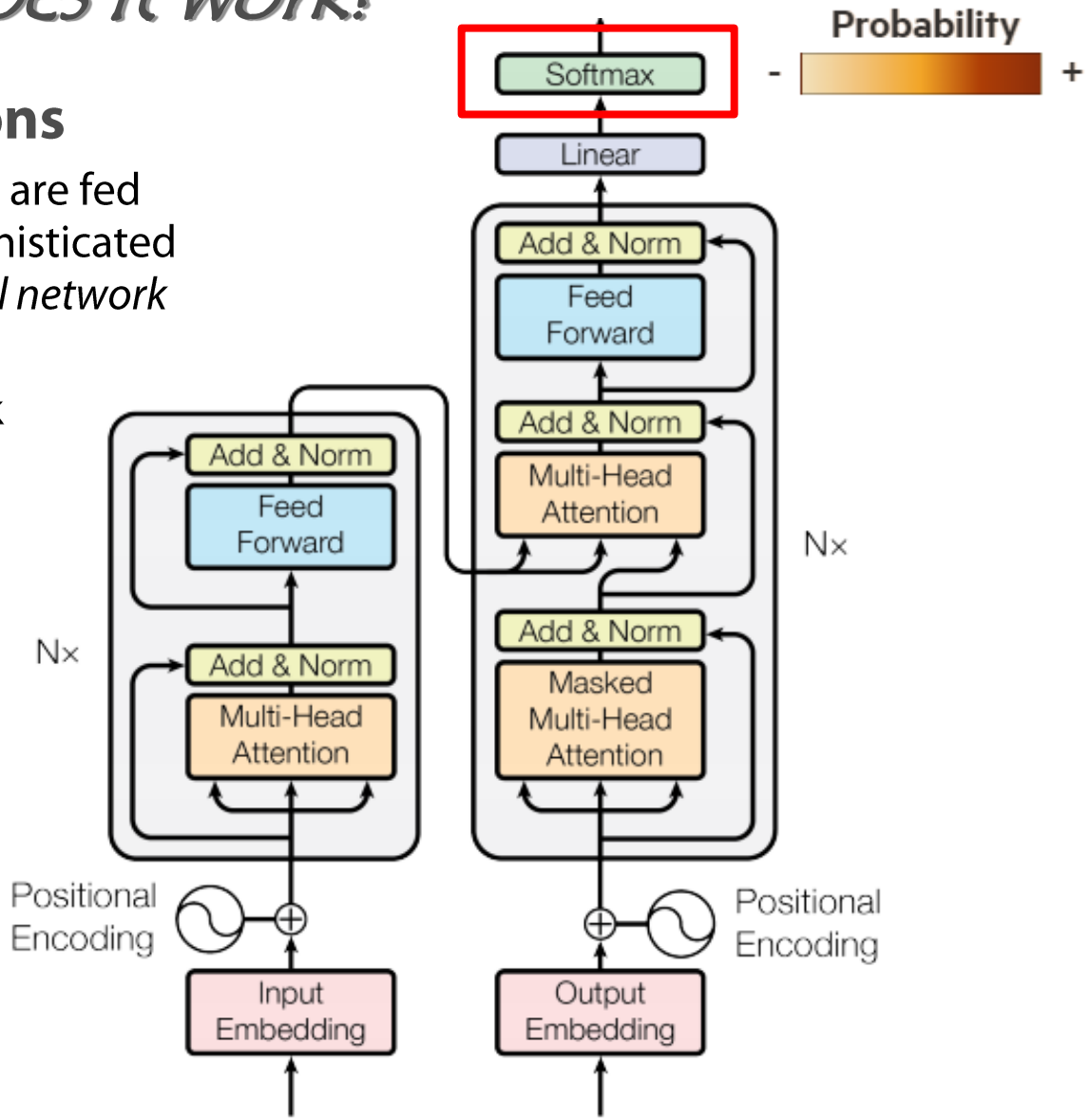
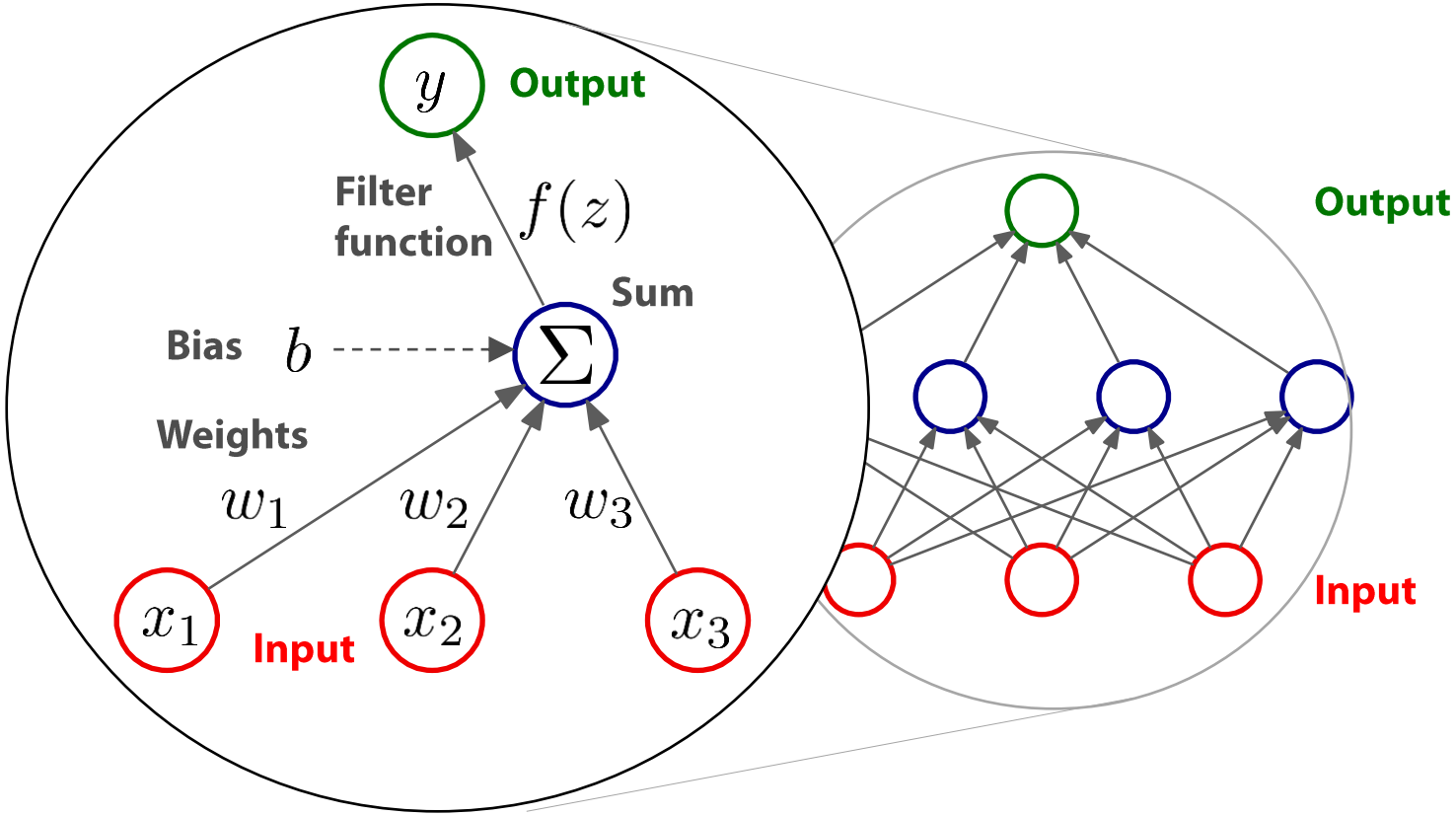


Image from https://arxiv.org/abs/1706.03762

Artificial Neural Networks

- **An assembly of simple computational units**

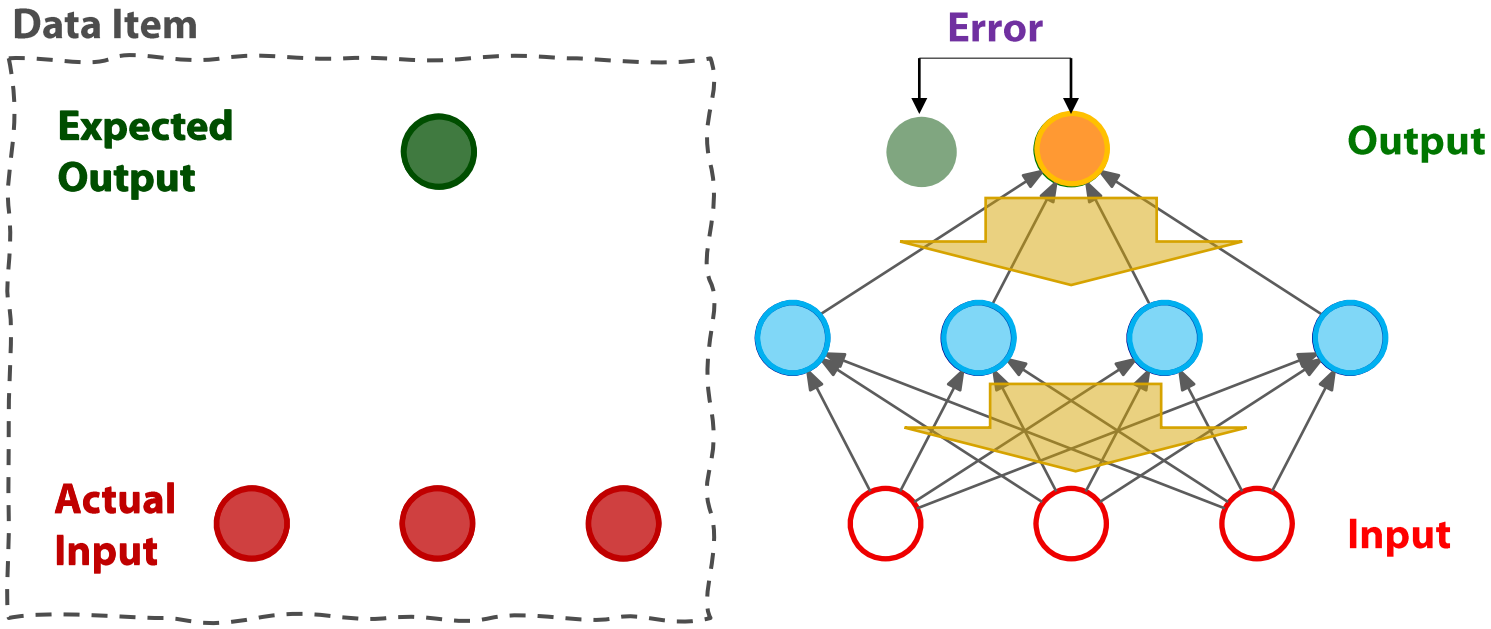
  Each unit performs numerical multiplications (of weights) and summations followed by the application of a non-linear filter function



Artificial Neural Networks

- **The learning process is an incremental *optimization of numerical parameters***

  Using a vast dataset of input-output pairs (*data items*)



Actual data are presented as **input**

The **input** is propagated upwards to compute the **output**

The **output** is compared with **expected output**

The **error** is propagated downwards to improve ***parameters***

General method:

- *show one data item*
- *improve*
- *repeat*

a huge number of times ...

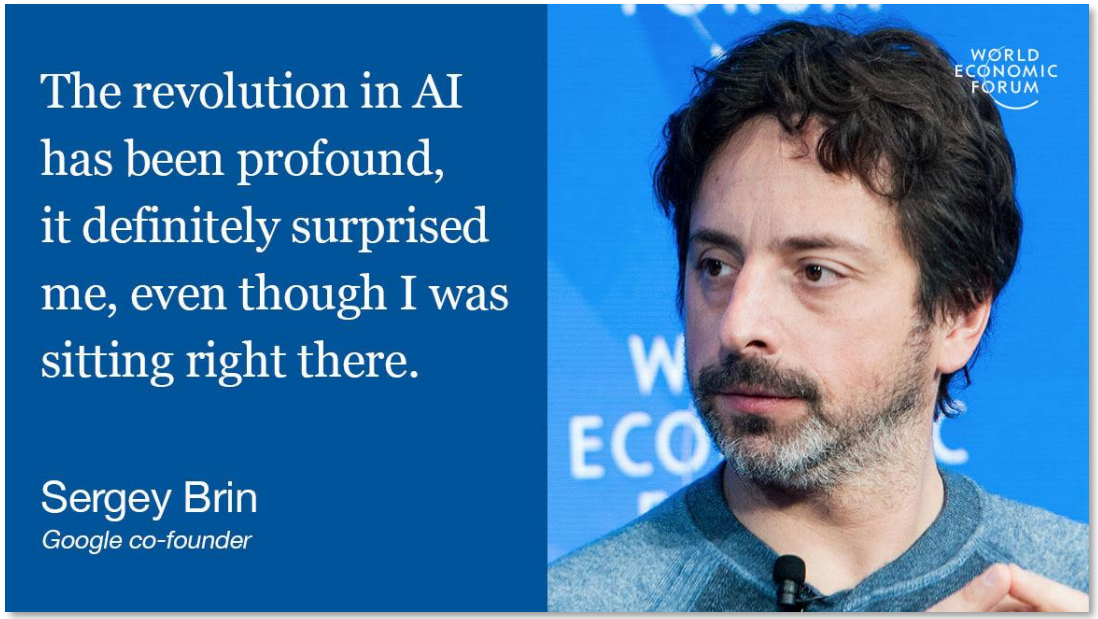
Generative AI: a Revolution?

Artificial Neural Networks

- **In short:**

  An artificial neural network represents a *deterministic associative function* (input -> output)
  which is *abstracted* from a (very) large amount of annotated data (input plus expected output)

A revolution? Since when?



- Sergey Brin [Google Co-Founder, January 2017]

  "I didn't pay attention to it [Artificial Intelligence] at all, to be perfectly honest."

  "Having been trained as a computer scientist in the 90s, everybody knew that AI didn't work. People tried it, they tried neural nets and none of it worked."

[Quote and image from https://www.weforum.org/agenda/2017/01/google-sergey-brin-i-didn-t-see-ai-coming/]

Artificial Neural Networks

- **In short:**

  An artificial neural network represents a *deterministic associative function* (input -> output)
  which is *abstracted* from a (very) large amount of annotated data (input plus expected output)

- ***Why do they work now and did not before***

  Recent progresses are due to:

  - better mathematical framework, for both representation and process
  - vast increase in the availability of parallel computational power
  - introduction of new network architectures, stratified and with self-adaptation capabilities
  - huge number of numerical parameters
    (ChatGPT has 175 billions)

ChatGPT in numbers

- ***Size***

  GPT 3.5, the software system on which ChatGPT is based, uses 175 billions numerical parameters

- ***Training***

  Estimated (*no actual data available*):

  - 570GB of data (>300 billions words) from books, internet texts and more

  OpenAI says that all data come from:

  - *"information that is publicly available on the internet"*
  - *"information that we license from third parties"*
  - *"information that our users or our human trainers provide"*

- ***User Base***

  - 1 million registered users in the first week
  - >100 millions users since January 2023

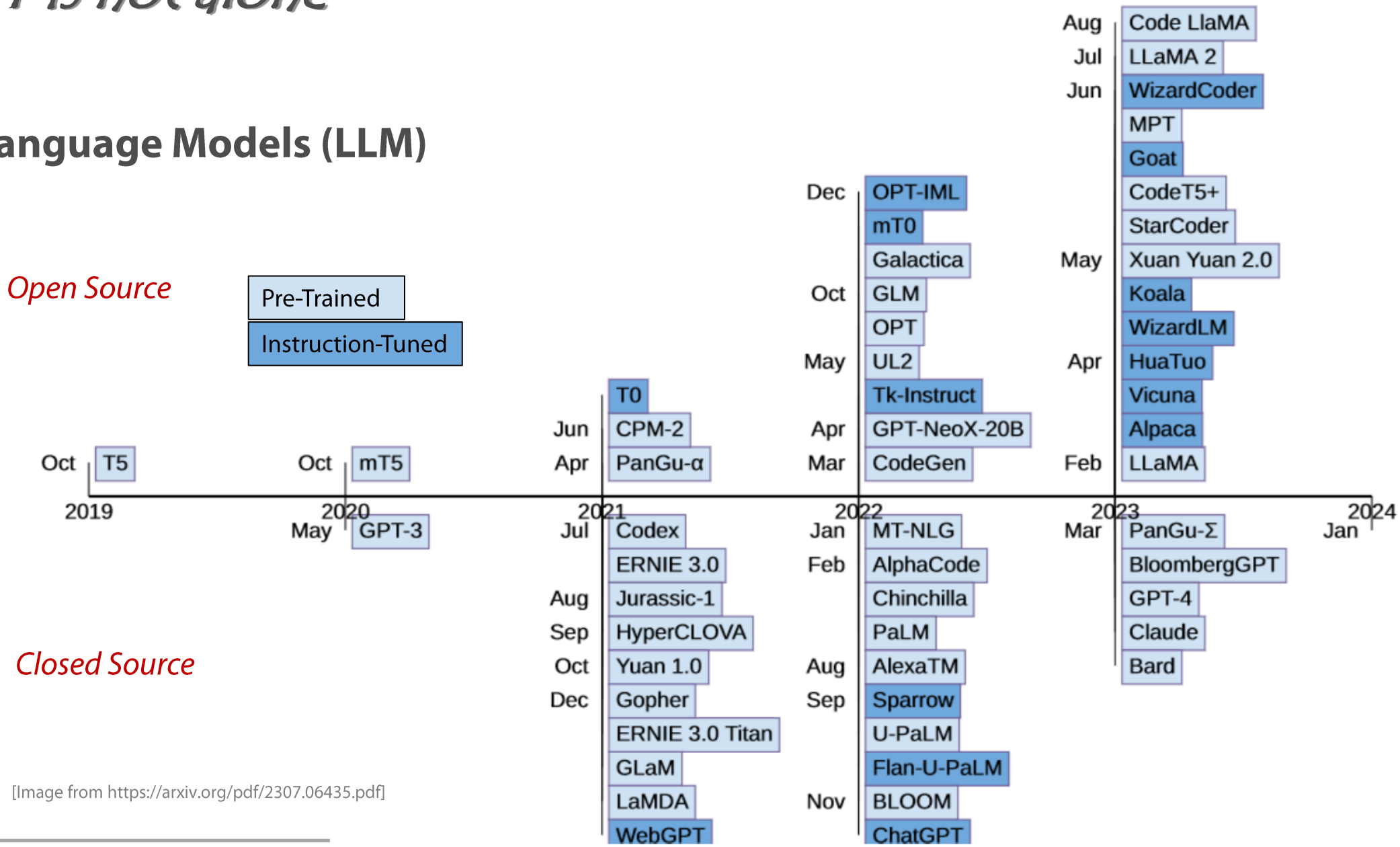
- ***Costs***

  Estimated (*no actual data available*):

  - $12 millions per each complete training process
  - $100,000 daily operational costs

ChatGPT is not alone

- **Large Language Models (LLM)**



[Image from https://arxiv.org/pdf/2307.06435.pdf]

Language is not the only generative way

- **DALL-E2**

  Image generation from text prompts

  «A teapot in the shape of an avocado»

[Image from https://www.nytimes.com/2022/04/06/technology/openai-images-dall-e.html]

Generative AI Makes Waves

An unexpected success

The New York Times

How ChatGPT Kicked Off an A.I. Arms Race

Even inside the company, the chatbot's popularity has come as something of a shock.

By Kevin Roose

Feb. 3, 2023

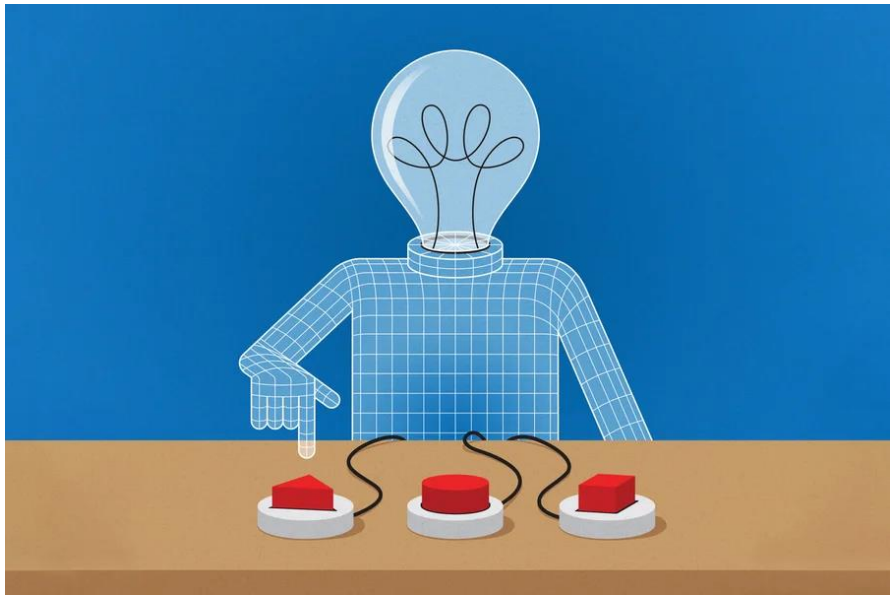
"[...] As ChatGPT has captured the world's imagination, Mr. Altman [*OpenAI's CEO*] has been put in the rare position of trying to downplay a hit product. He is worried that too much hype for ChatGPT could provoke a regulatory backlash or create inflated expectations for future releases, two people familiar with his views said."

[Quote and images from https://www.nytimes.com/2023/02/03/technology/chatgpt-openai-artificial-intelligence.html/]

Passing Tests: Yours or Mine?

"So what finally did it score overall?
Estimated on the basis of five subtests, the Verbal IQ of the ChatGPT was 155, superior to 99.9 percent of the test takers who make up the American WAIS III standardization sample of 2,450 people.
[...] so ChatGPT appears to be very intelligent by any human standards."

"Despite its high IQ, ChatGPT is known to fail tasks that require real humanlike reasoning or an understanding of the physical and social world.
ChatGPT easily fails at obvious riddles, such as *"What is the first name of the father of Sebastian's children?"*
(ChatGPT on March 21:
"I'm sorry, I cannot answer this question as I do not have enough context to identify which Sebastian you are referring to.")"



SCIENTIFIC AMERICAN

OPINION

MARCH 28, 2023 | 5 MIN READ

I Gave ChatGPT an IQ Test. Here's What I Discovered

The chatbot was the ideal test taker—it exhibited no trace of test anxiety, poor concentration or lack of effort. And what about that IQ score?

BY EKA ROIVAINEN

[Quote and images from https://www.scientificamerican.com/article/i-gave-chatgpt-an-iq-test-heres-what-i-discovered/

Passing Tests: Yours or Mine?

"A system like ChatGPT doesn't create, it imitates. [...] it instead copies, manipulates, and pastes together text that already exists."

"Even if ChatGPT isn't intelligent, couldn't it still take our jobs?"

"Although this ability can generate attention-catching examples, the technology is unlikely in its current form to significantly disrupt the job market. "

[Quote and images from https://www.newyorker.com/science/annals-of-artificial-intelligence/what-kind-of-mind-does-chatgpt-have

"[...] we discover that programs like ChatGPT don't represent an alien intelligence with which we must now learn to coexist; instead, they turn out to run on the well-worn digital logic of pattern-matching, pushed to a radically larger scale. "

Did ChatGPT Make Us Change Mind?

Not so long ago ... [2019]

AI isn't as scary as we imagine. AndreyZH/Shutterstock

The reality of AI is currently very different, particularly when you look at the threat of automation. Back in 2013, researchers estimated that, in the following ten to 20 years, 47% of jobs in the US could be automated. Six years later, instead of a trend towards mass joblessness, we're in fact seeing US unemployment at a historic low.

Current AI is good at finding patterns in large datasets, and not much else.

Good question:
where is the pattern?

[Quote and images from https://theconversation.com/ais-current-hype-and-hysteria-could-set-the-technology-back-by-decades-120514/

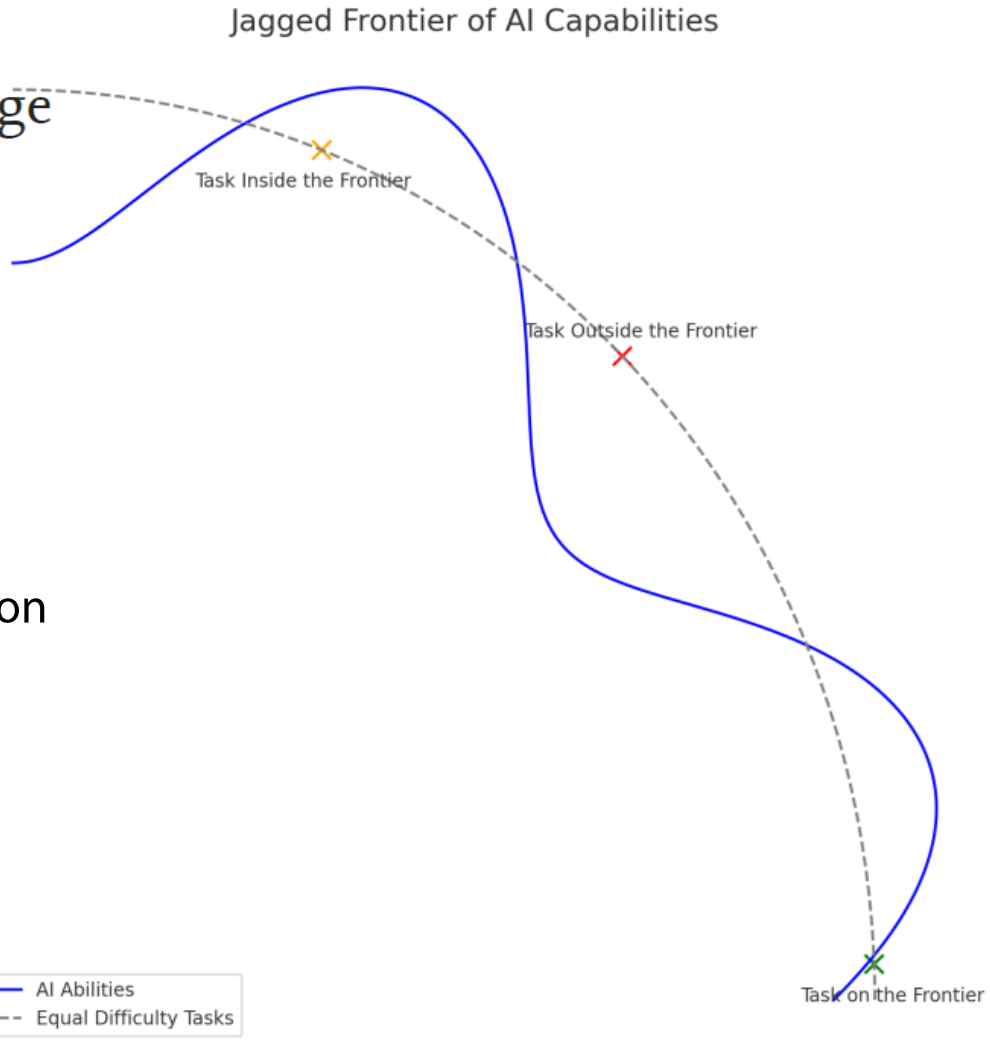
Does Generative AI Really Pay Back? (It depends …)

Generative AI and People At Work

Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality

Harvard Business School Technology & Operations Mgt. Unit Working Paper No. 24-013

58 Pages • Posted: 18 Sep 2023 • Last revised: 27 Sep 2023



- ***Comprehensive study***

  Harvard Business School + Boston Consulting Group

  758 participants, 18 realistic tasks, quantitative performance evaluation

  Three groups, with random assignment:

  1) no AI access
  2) GPT-4 access
  3) GPT-4 access with a prompt engineering overview

- ***The result is***

  "The actual performance improvement depends on the task"

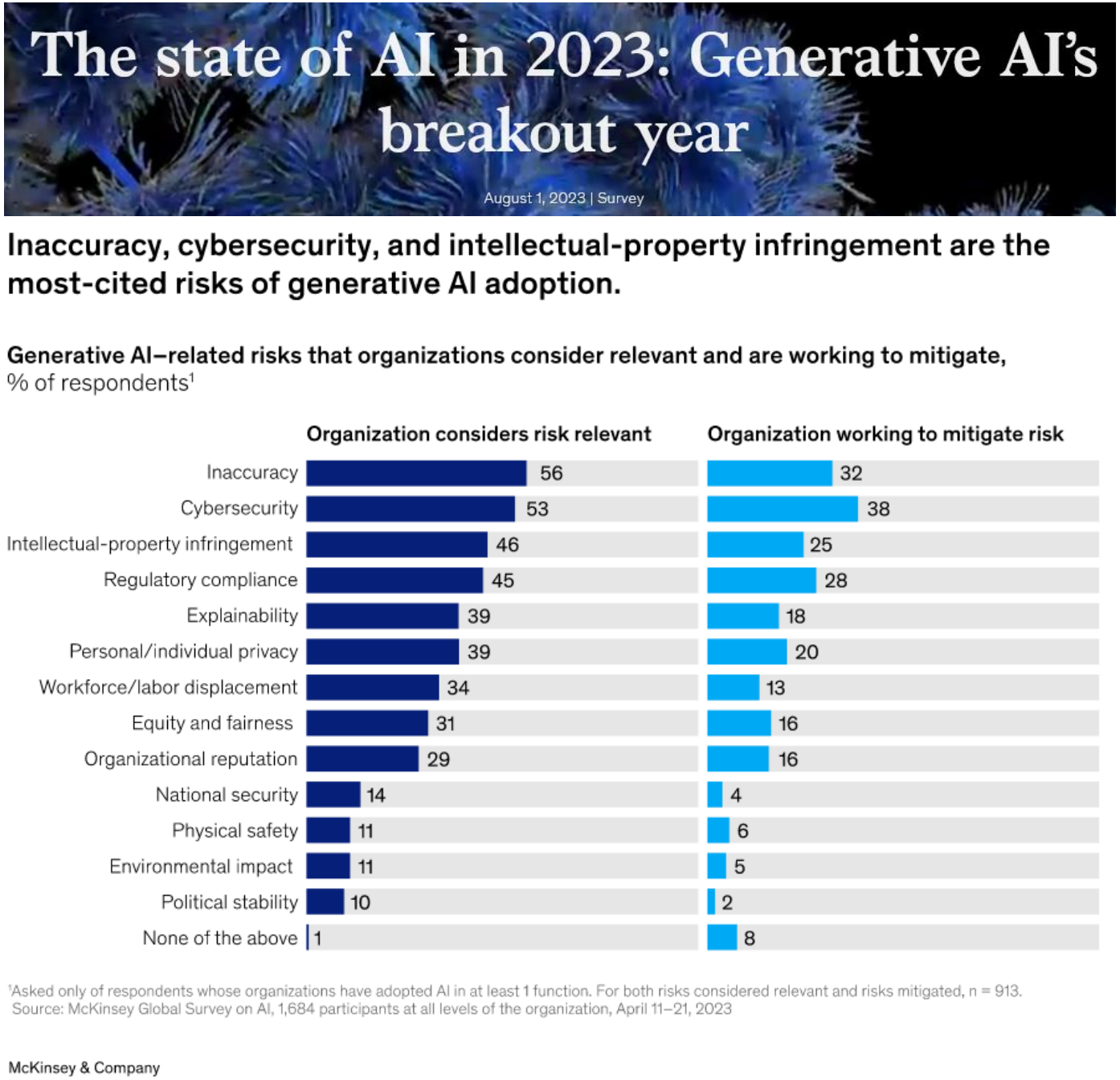
[Image and quotes from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4573321]

Using AI For Business

QuantumBlack
AI by McKinsey

- ***Comprehensive survey***

  1658 respondents world-wide,
  with different roles,
  in companies that have adopted
  generative AI



[Image https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year]

AI Rules & Laws

The Artificial Intelligence Act

What is the EU AI Act?

The AI Act is a proposed European law on artificial intelligence (AI) – the first law on AI by a major regulator anywhere. The law assigns applications of AI to three risk categories. First, applications and systems that create an **unacceptable risk**, such as government-run social scoring of the type used in China, are banned. Second, **high-risk applications**, such as a CV-scanning tool that ranks job applicants, are subject to specific legal requirements. Lastly, applications not explicitly banned or listed as high-risk are largely left unregulated.

https://artificialintelligenceact.eu/

A risk-based approach to regulation

Unacceptable risk
e.g. social scoring

Prohibited

High risk
e.g. recruitment, medical devices

Permitted subject to compliance with AI requirements and ex-ante conformity assessment

*Not mutually exclusive

AI with specific transparency obligations
'Impersonation' (bots)

Permitted but subject to information/transparency Obligations

Minimal or no risk

Permitted with no restrictions

European Commission

https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf

Thank You!

(If you believe, this presentation
was NOT made by generative AI)