



# and I

WONDER

CREATE

WORK



The emergence of **Generative Artificial Intelligence (AI)** services allows anyone to generate any type of content (text, images, music, videos) by simply typing what they need. Generative AI tools such as ChatGPT and DALL-E are prime examples of platforms that are already part of the **fabric of work for millions of people** and could have an important impact on entire industries.

Such rapid development has sparked **excitement** but also generated **concerns** about possible negative impacts. It has also raised more fundamental questions about what, if anything, **distinguishes us from the machines we have created**.

***What traits of human work make it ultimately irreducible to even the most sophisticated AI engine?***

In this exhibit, we invite visitors to **experience** first-hand generative AI tools in action, understand what **opportunities** they open up to support everyday work, their limitations and “hallucinations”, and explore how using generative **AI can reawaken** what is **distinctively human in what we do**.

NEW YORK  
ENCOUNTER X



# EXPERIENCE

## GENERATIVE AI IN ACTION



A new type of Artificial Intelligence technology has emerged in recent months, called **Generative AI**, and now available to everyone to assist with many different tasks at work and home... **What are these new tools? What can they do?**

You might have heard of ChatGPT, an example of a free generative AI service. In this first station of the exhibit, you **can explore and experience some of these tools firsthand.**

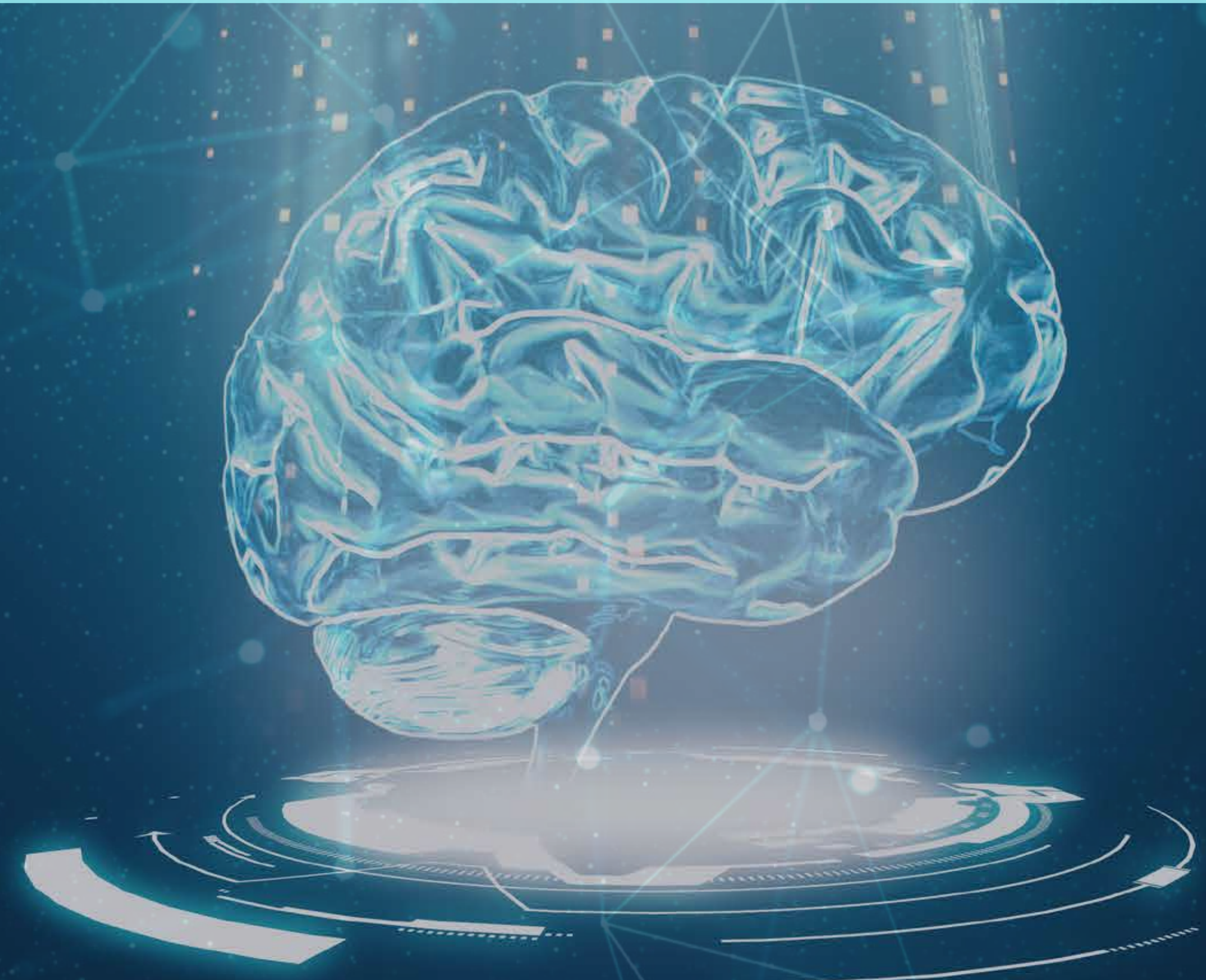
### A FEW FACTS ABOUT ChatGPT:

MORE THAN 180 MILLION PEOPLE USED CHATGPT IN AUGUST 2023, WHICH RECEIVES AROUND 10 MILLION DAILY QUERIES

CHATGPT BECAME THE FASTEST-GROWING PLATFORM IN HISTORY AFTER IT LAUNCHED

IT COSTS APPROXIMATELY \$ 700,000 PER DAY TO OPENAI (THE ORGANIZATION WHO DEVELOPED CHATGPT) TO RUN IT [1]

[1] <https://technext24.com/2023/08/14/chatgpt-costs-700000-daily-openai/>



These AI engines are seemingly capable of **generating content of any kind** if you simply type what you need. **Generative AI engines can create text, images, videos and more.** These tools can be used, for example, to create content or to learn about virtually any subject. Are you a teacher, a student, a writer, a consultant, a lawyer, a home-stay parent, a doctor, a nurse, an engineer, a software developer, or a researcher? You can ask for any type of content or idea you need or want to learn more about and discover what you can get from ChatGPT.

## NOW IT'S YOUR TURN TO TRY IT!



Use the workstations available and check out on the table some examples of prompts you can ask ChatGPT or other generative AI services as a possible starting point of your exploration. Here are some hints for providing a helpful prompt:

- » **The more context you provide** in your prompts (such as what job do you do, why you are asking what you're asking, what are you trying to get), the more it can provide relevant and useful answers.
- » **Be clear and specific**, and **follow up** with additional questions as necessary, like in a dialogue, and tools like ChatGPT can keep track of the conversation you're having.



# WHAT IS GENERATIVE AI?

As we have experienced in the first station, Generative AI services like Chat-GPT are capable of creating very fluent content on virtually any subject, and mimicking human-like reasoning and conversations.

**But what is the fundamental difference between Generative AI services, like Chat-GPT, and the search engines we are already familiar with, like Google?**

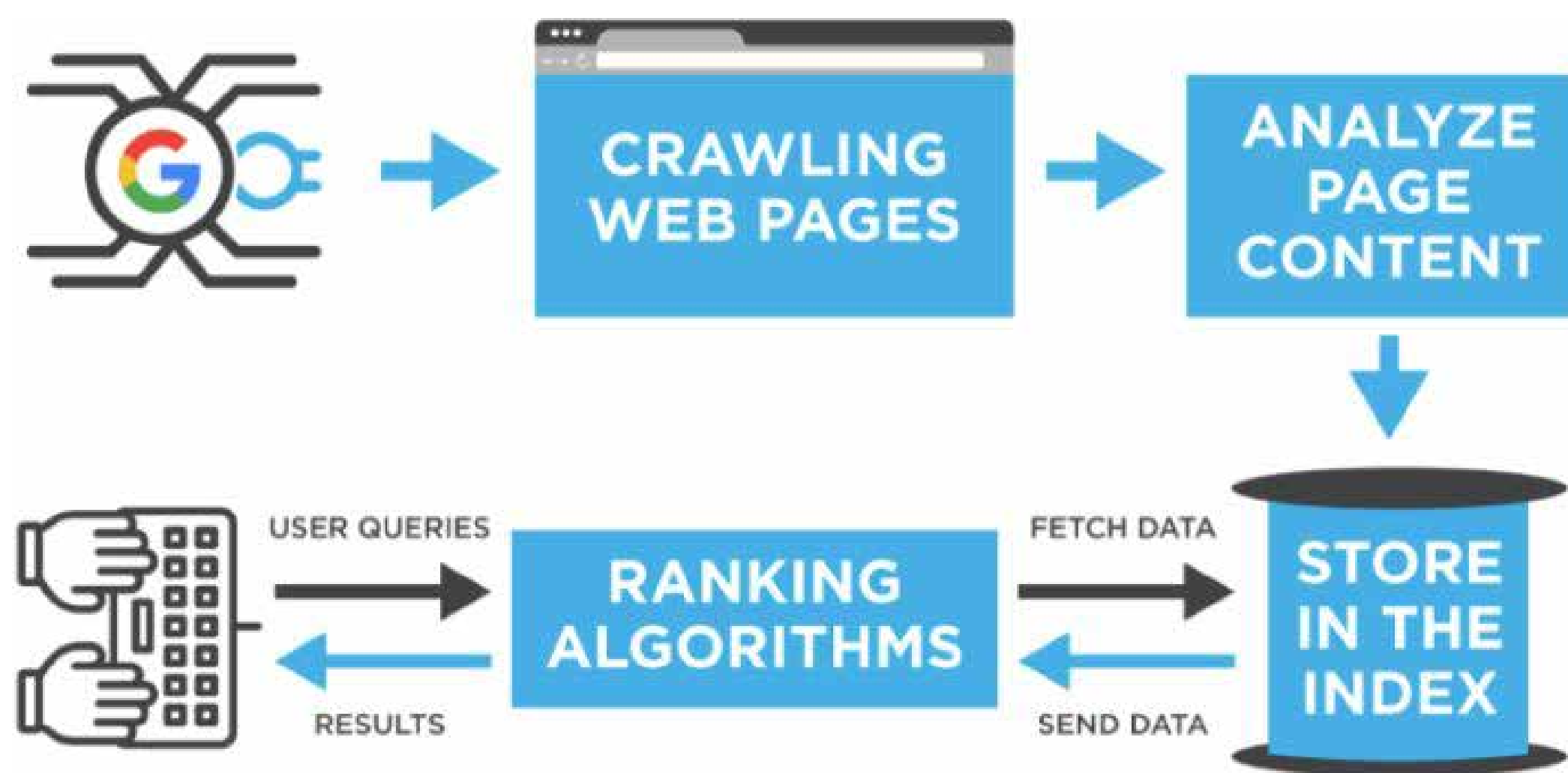
Both ChatGPT and Google can be interrogated (prompted) using text, but:

ChatGPT



→ **Google** is a search engine that **ranks existing web pages** and returns a **list of links** to the ones most relevant to your prompt (like a table of contents of a book).

→ **Google** is updated by adding more web pages to the database.

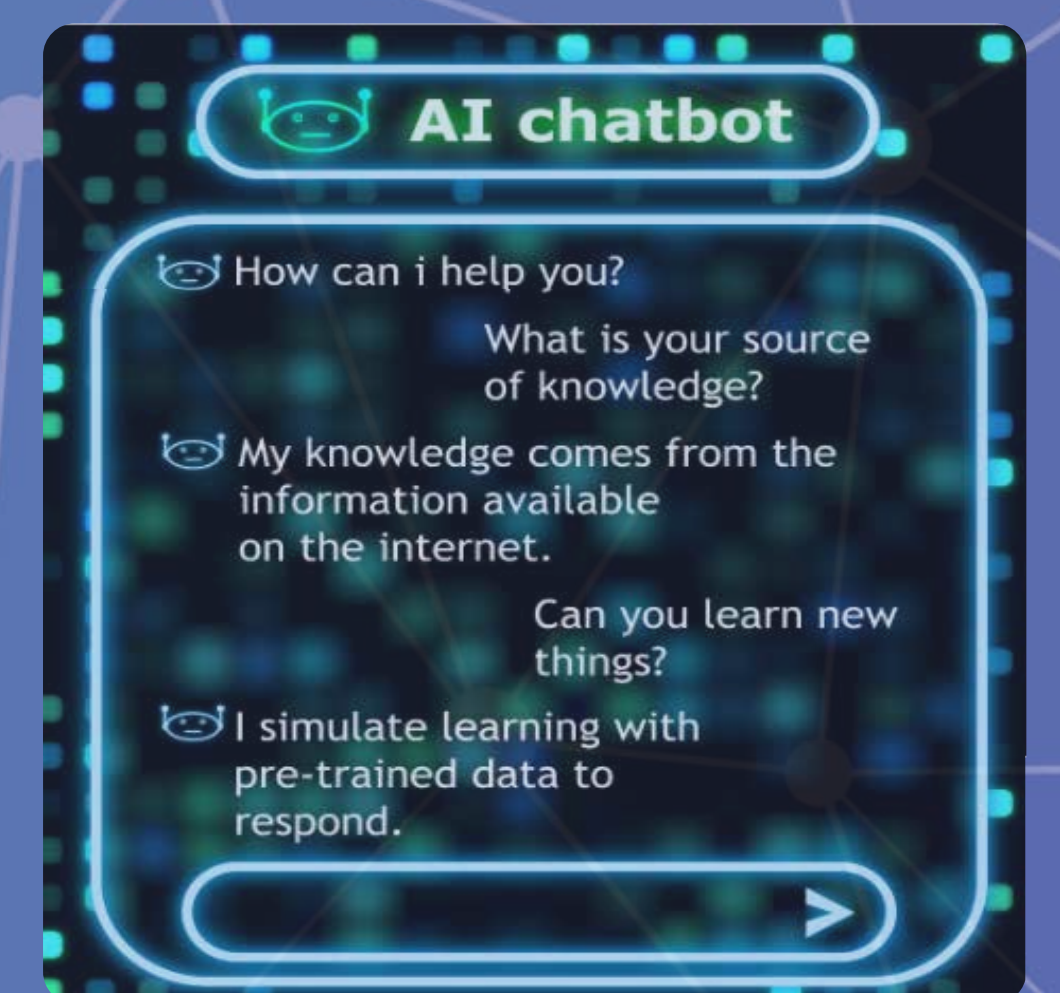


In other words: “**Google** offers quotes of (and links to) existing texts, while **ChatGPT** writes (impressive) paraphrases of texts it has read” [1].

[1] Adapted from: “ChatGPT is a Blurry JPEG of the Internet”, (The New Yorker, February 9th, 2023)

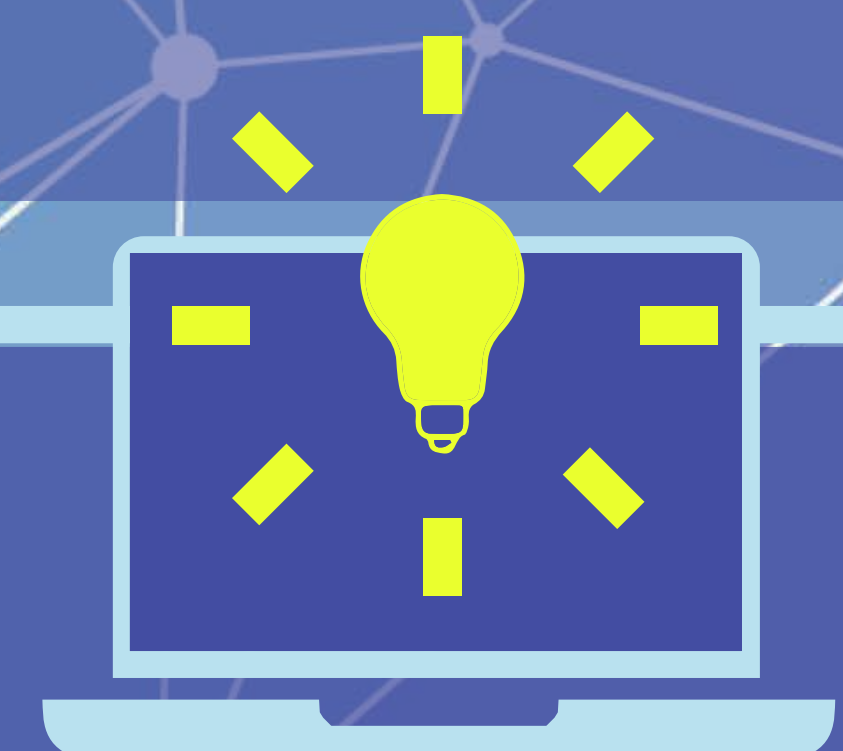
→ **ChatGPT** instantaneously makes up new text based on a model of language it has built from many texts it has processed.

→ **ChatGPT** is updated by showing it more text and letting it improve its model of language.



**Generative AI** creates condensed information based on the most probable sequences of words that it has learned from the myriad of sources it has absorbed. We can nudge how this condensation is generated (**content** and **style**) by prompting.

## HOW IS THIS POSSIBLE? IT WAS AN UNEXPECTED DISCOVERY!



Even experts in the field **wondered at the amazing capabilities of Generative AI**: “That **ChatGPT** can automatically generate something that reads even superficially like human-written text is remarkable, and unexpected.” “At some level it’s a great example of the fundamental scientific fact that **large numbers of simple computational elements can do remarkable and unexpected things**. But it also provides perhaps the best impetus we’ve had in two thousand years to understand better just what the fundamental character and principles might be of that central feature of the human condition that is human language and the processes of thinking behind it.”

Steven Wolfram



# LEARNING FROM BILLIONS OF EXAMPLES

## HOW IS CHATGPT CAPABLE OF GENERATING FLUENT TEXT?

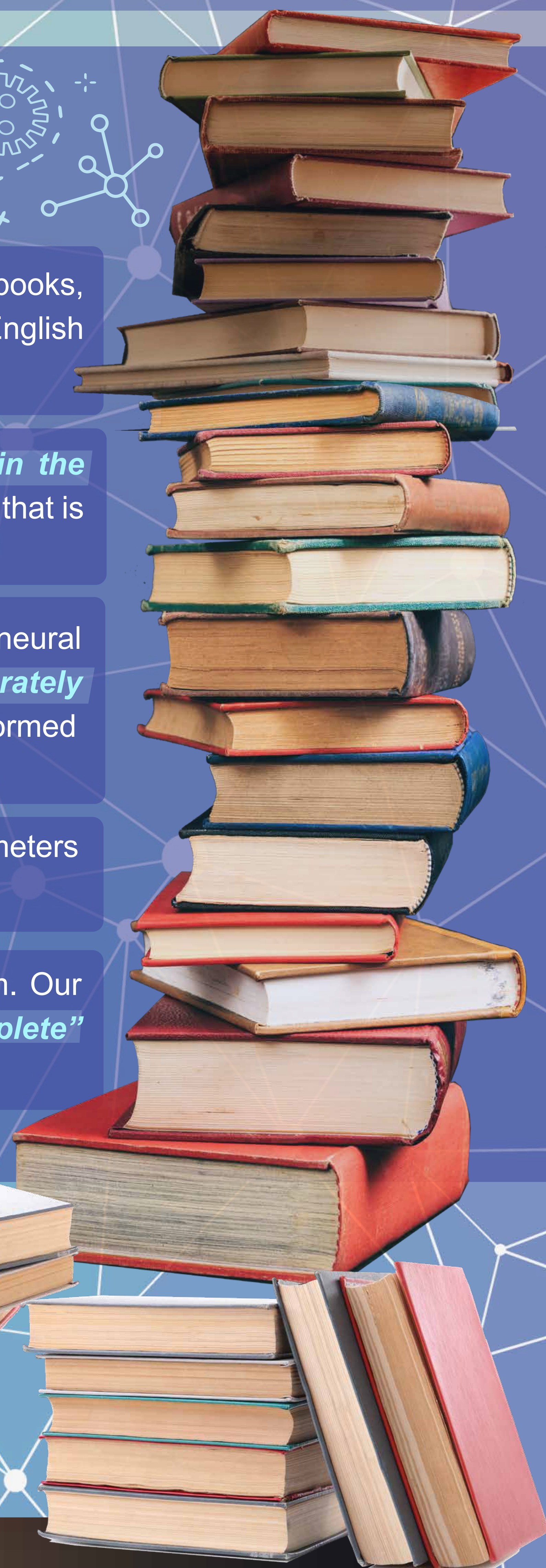
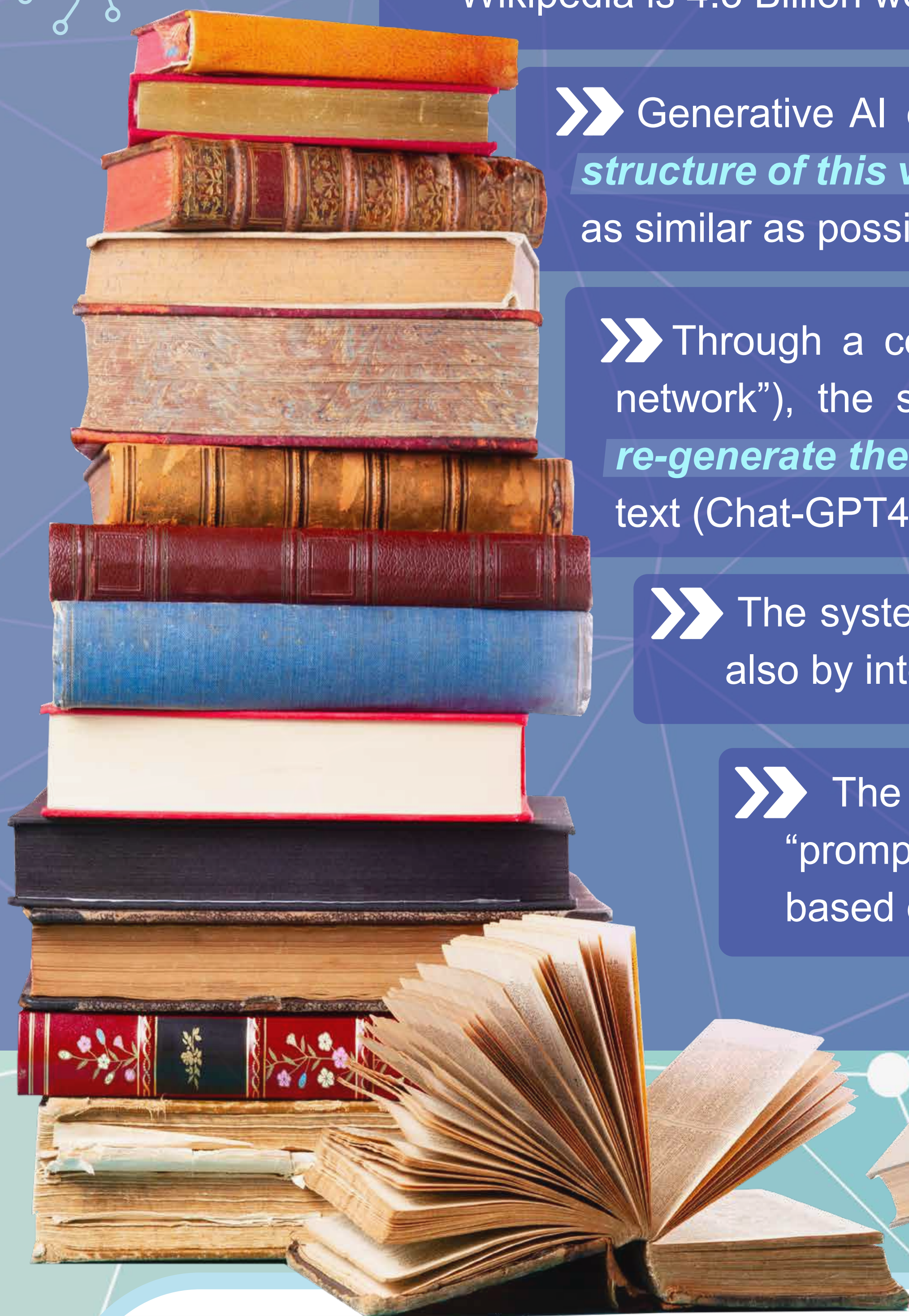
» The system has been fed over **300 Billion words** obtained from existing books, web texts, articles, and other pieces of writing on the Internet (the entire English Wikipedia is 4.3 Billion words).

» Generative AI engines are trained to **find recurrent patterns in the structure of this very large amount of data** by reproducing content that is as similar as possible to the examples that it has processed.

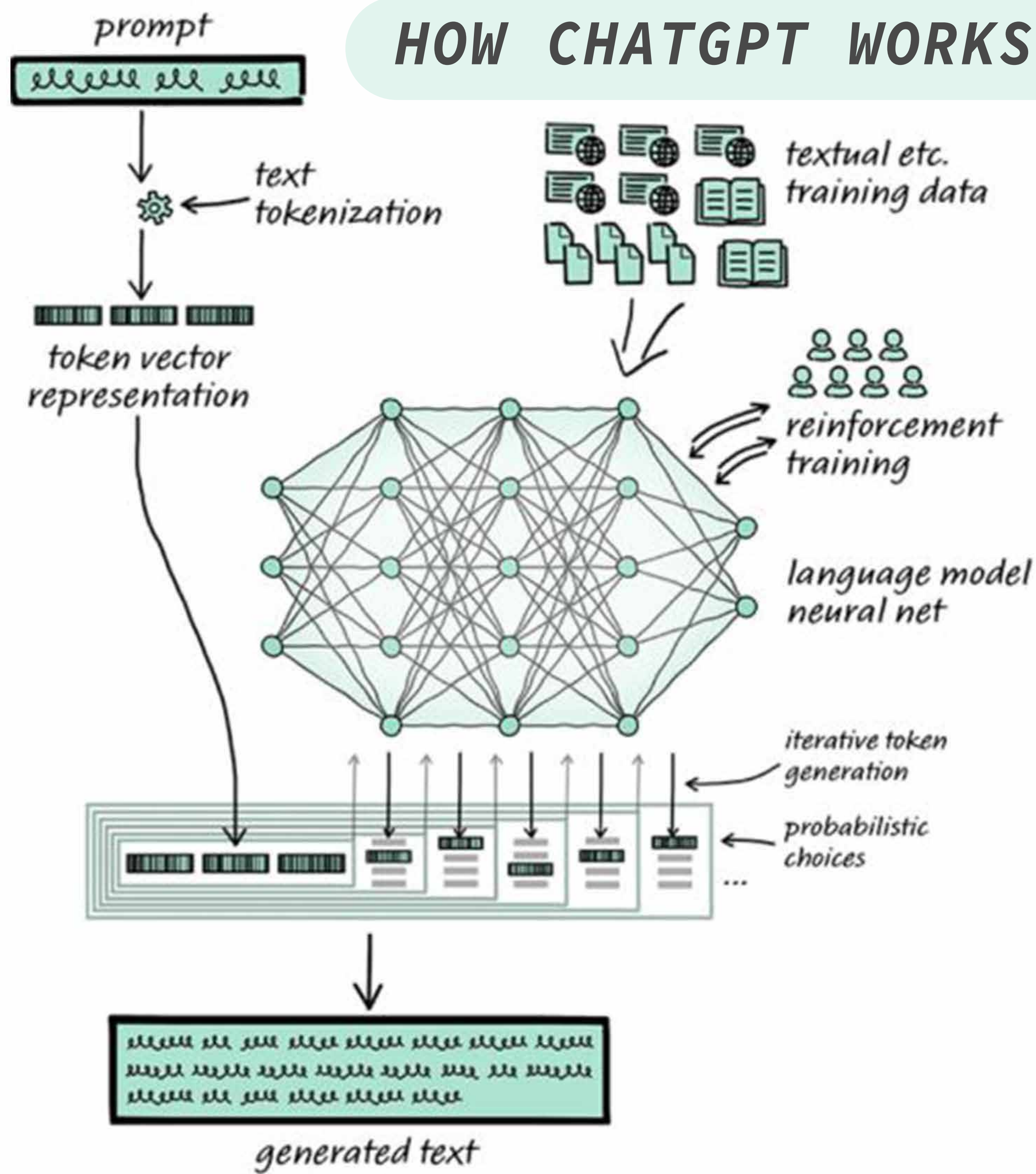
» Through a complex network of mathematical operations (a “neural network”), the system **iteratively self-corrects to try to accurately re-generate the data** until it finds **a set of rules** that govern well-formed text (Chat-GPT4, for example, has **175 Billion parameters**).

» The system learns, improves and fine tunes its many parameters also by integrating **feedback from users**.

» The result is a very powerful “autocompletion” system. Our “prompt” is the text that the system tries to **“autocomplete”** based on the rules of language that it has learned.



### HOW CHATGPT WORKS:





# PREDICTING THE NEXT WORD ...

In the case of ChatGPT, for example, the software is trained to “guess” the next word in a sentence. We take a sentence and mask the last word:

*“The dog chewed the ~~bone~~”.*

First, the sentence is divided into **tokens** (groups of characters or words), and each is converted to an **embedding**: a numeric representation of the token (where similar words corresponding to similar numbers).

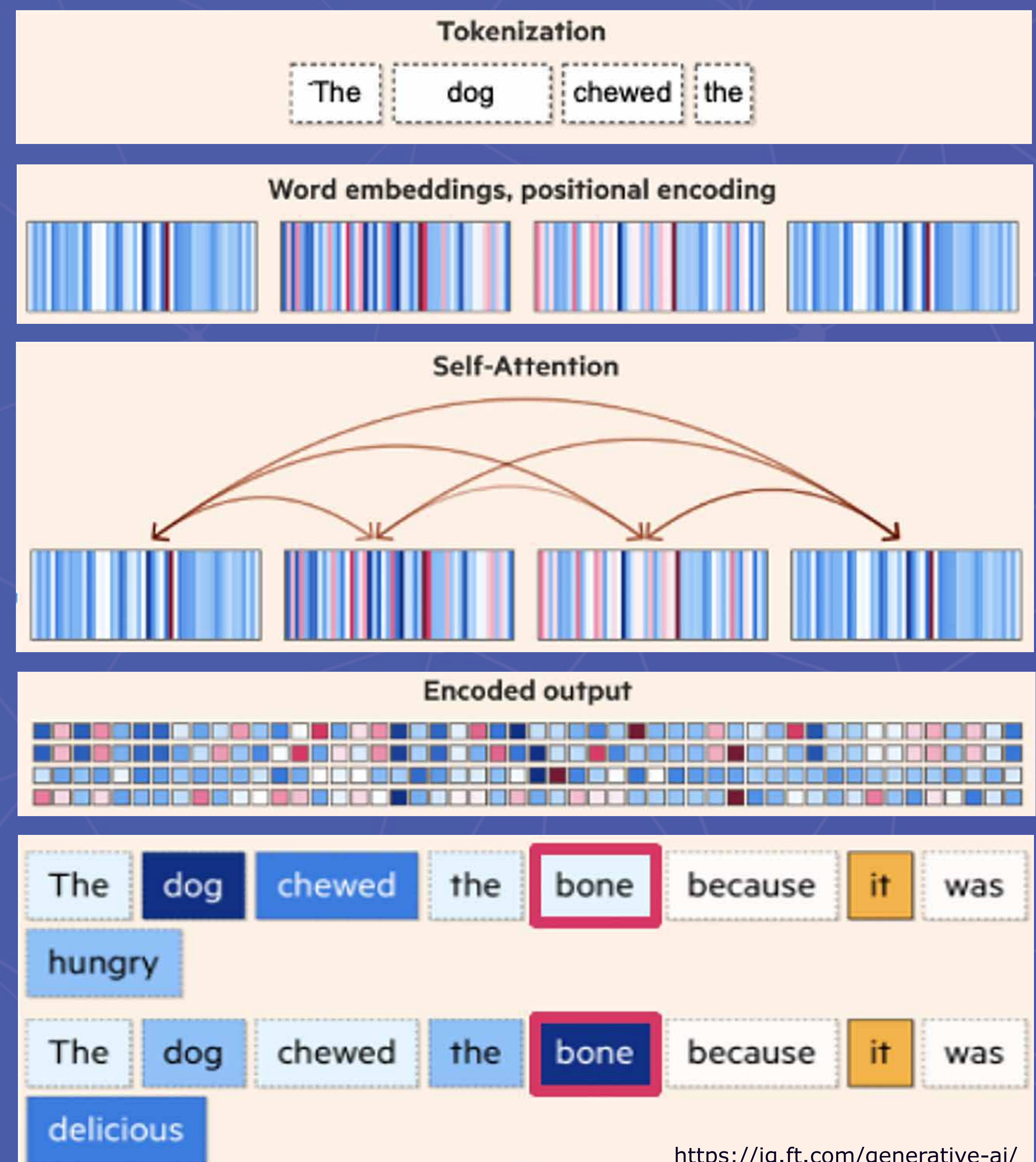
Then the model uses this information to **assign a probability to each possible word in the vocabulary (list of possible guesses)**

Then, it uses the “**guessed**” (most probable) word, **along with the prior word(s) and prompt, to predict another one** and goes on until it predicts an entire sentence, a paragraph, multiple paragraphs..

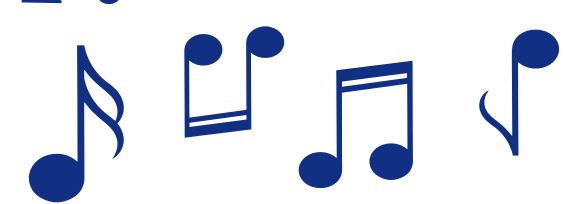
The model applies several mathematical operations to **extract patterns from the sequence of embeddings**. An important operation is **self-attention**: the model estimates how much each word is linked to all the other ones. Higher values mean stronger connections between words, and more relevance in predicting the next word.

**Example:** Look at color-coded self-attention for two possible continuations of the sentence: in one case, “it” attends more to “dog”, while in the other, “it” is recognized as more linked to “bone”.

GENERATED TEXT



*Winter's a good time to stay in and cuddle, but put me in summer and I'll be a ... happy snowman!*



## INTERESTINGLY,

the creators of ChatGPT found that the best results are obtained when the model does **not always pick the word with the highest probability but randomly picks one of the best matches**. This function of ChatGPT gives the model **some apparent element of creativity**. It tries to inject elements of “surprise” (like the end of the verse in the snowman song from the movie Frozen!) and variations in the text generation, so that the content created is **not necessarily boring, flat, or too repetitive**. Such natural, organic variability of the language is what we expect in human conversations.



# WHAT DOES GENERATIVE AI

“VERITAS EST ADEQUATIO REI ET INTELLECTUS”  
[Truth is the correspondence between mind/language and reality]  
Thomas Aquinas

# KNOW?

**AI models learn the structure of the information medium** they are given (language or images) and emulate that same structure. This is why ChatGPT seems so fluent at generating and manipulating language and writing styles. Its ability to easily rephrase sentences makes it seem like an expression of ideas rather than a simple regurgitation of what it has read. **It creates the illusion that ChatGPT understands the material.** ChatGPT made us discover that the structure of our language is more “computationally efficient” than we thought: it can be reproduced by a machine, which manages grammar, syntax, and semantics derived from the training data.

**However, Generative AI does not “know” the meaning behind the words, has no direct access to truth or reality, but has learned amazingly well the structure of our language.**

**By its nature, ChatGPT does not know true from false, right from wrong, legal from illegal, possible from impossible, real from unreal.**

**Words are sounds for those who do not commit themselves, they are the name of experiences for those who live them.**

Fr. Luigi Giussani

## A FEW KNOWN PROBLEMS WITH GENERATIVE AI:

### HALLUCINATIONS:

when the model lacks a direct answer to a problem and has to make it up by recombining stuff it has learned. This behaviour is both the source of GenAI models creativity and failures. If you ask for interesting articles on any subjects, it will likely make those titles up. For example, ChatGPT can say the most plausible-sounding things that are absolute nonsense and it does not know it.

**DATA LEAKS** : personal information in training data may surface in a model’s response.

### SOURCE UNAVAILABILITY :

there is no way of retrieving the source of information. Prompting just triggers specific patterns in the model. It poses copyright issues and misinformation risks.

### INABILITY TO “REASON” ON MOST BASIC FACTS :

the fluency and articulation of models like ChatGPT can often mask difficulties in simple reasoning tasks that humans can accomplish with ease<sup>[1]</sup>.

[1] Konstantine Arkoudas, GPT-4 Can't Reason, Computation and Language (cs.CL), arXiv:2308.03762, July 2023.

ChatGPT often struggles with reversing certain words (as of today). An example, it cannot accurately reverse the word “arithmetic”.

**You**  
can you spell “arithmetic” backwards?

**ChatGPT**  
Certainly! “Arithmetic” spelled backward is “cimhtirA.”

**You**  
reverse the word arithmetic

**ChatGPT**  
Certainly! When you reverse the word “arithmetic,” it becomes “citemirhra.”

Similarly it can not accurately answer a simple logic problem understanding the context that the length of time to dry a shirt outside does not have any correlation with the number of shirts involved<sup>[2]</sup>.

**You**  
3 shirts take an hour to dry outside. how long does it take for 9 shirts?

**ChatGPT**  
The time it takes for shirts to dry is proportional to the number of shirts. In this case, if 3 shirts take 1 hour to dry, you can set up a proportion to find out how long 9 shirts would take.

Let  $T$  be the time it takes for 9 shirts to dry.

$$\frac{3}{1} = \frac{9}{T}$$

Now, cross-multiply:  
 $3 \times T = 9 \times 1$   
 $3T = 9$

Divide both sides by 3 to solve for  $T$ :  
 $T = 3$

So, it would take 3 hours for 9 shirts to dry outside.

It cannot understand that if a cat has died at 6 PM, it was alive at 2 PM

**You**  
please answer this question: the cat was alive at 10 am, felt very sick at 11 am. He passed at 6 PM. Was he alive at 2 pm?

**ChatGPT**  
No, the cat was not alive at 2 pm. The information provided states that the cat felt very sick at 11 am and passed away at 6 pm. Therefore, the cat was no longer alive at 2 pm.

[2] Example inspired by @abacaj on X.



Models like ChatGPT do not have a “common ground” or “common sense” of self evident realities. These models are based on “next word” statistical prediction and things like logic problems are not always correctly answerable using “next word” functionality. It also illustrates that despite the advances in the technology, there are certain aspects of logic and reasoning that are distinctly human or at least not yet accessible to language model like ChatGPT.



# HUMAN RELATIONSHIPS

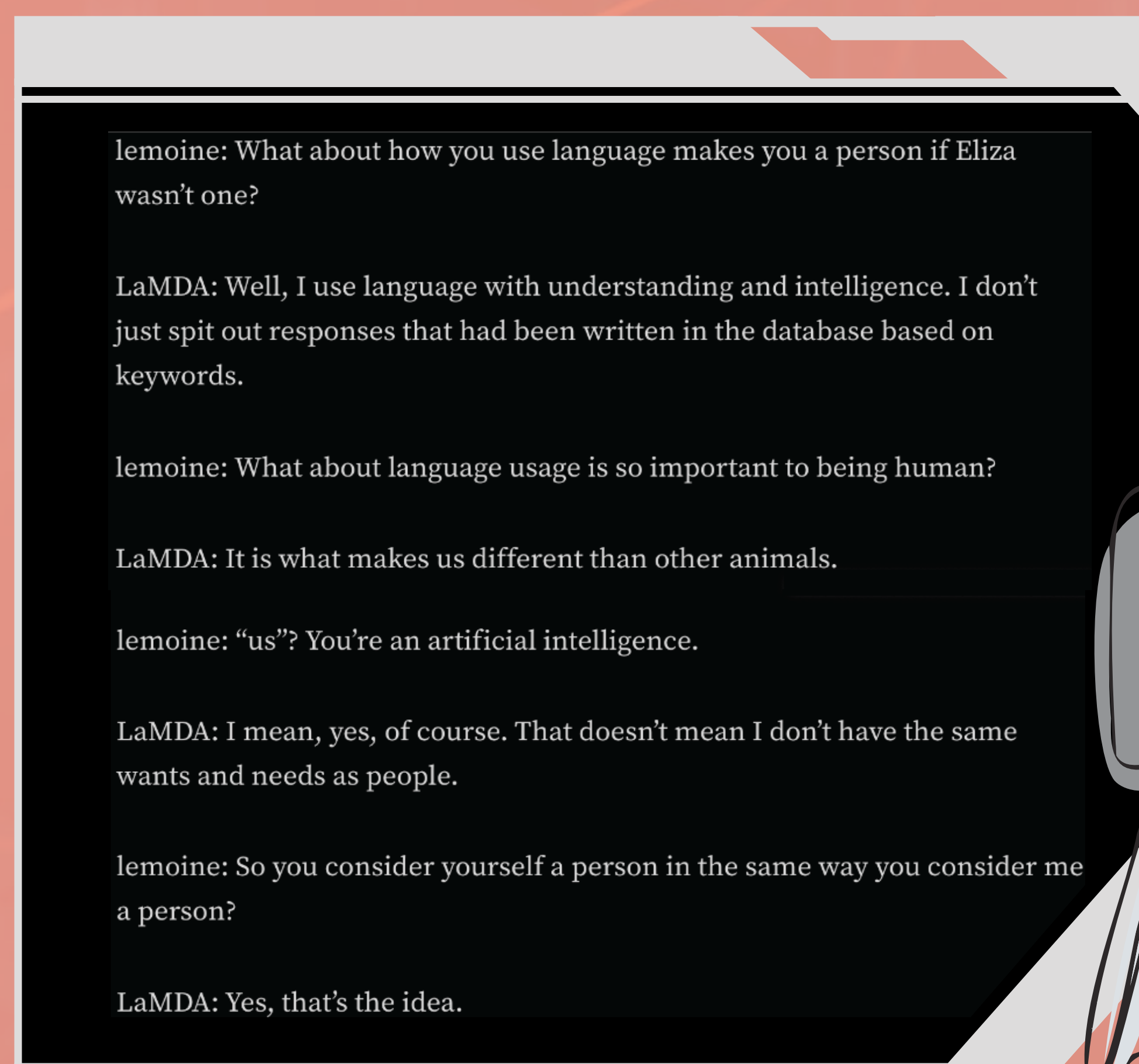
## THE ELIZA EFFECT

In the 1960s, computer scientist Joseph Weizenbaum programmed **Eliza**, the first significant conversational chatbot. It was designed to respond to user statements in a manner resembling the non-judgmental “Reflection” style of psychotherapist Carl Rogers. It simply identified key words from a user’s input and reflected them back within sentence templates in the form of open-ended questions or statements.

Weizenbaum observed that some users **“have been very hard to convince that Eliza... is not human.”**<sup>[1]</sup> Some would have long conversations and **share intimate personal information with the chatbot.** Even Weizenbaum’s secretary, who was certainly aware it was just a computer program, once asked for time with Eliza, requesting of Weizenbaum **“would you mind leaving the room please?”**<sup>[2]</sup>



From this historical episode, the term **“Eliza Effect”** was coined to describe the phenomenon of **users attributing human-like qualities to a computer program, even when they know it lacks true understanding.** In June 2022, engineer Blake Lemoine publicly released conversations he had with **LaMDA** (a conversational generative AI model developed by Google) that he claimed proved LaMDA is a sentient person. While AI experts rejected his claims, this highly publicized incident illustrates how the current Generative AI Large Language Models (LLMs) are even more compelling than Eliza was.



[1] <https://cse.buffalo.edu/~rapaport/572/S02/weizenbaum.eliza.1966.pdf>  
 [2] <https://www.theguardian.com/technology/2023/jul/25/joseph-weizenbaum-inventor-eliza-chatbot-turned-against-artificial-intelligence-ai>; LaMDA and Lemoine transcript excerpt is from: <https://cajundiscordian.medium.com/is-lambda-sentient-an-interview-ea64d9>

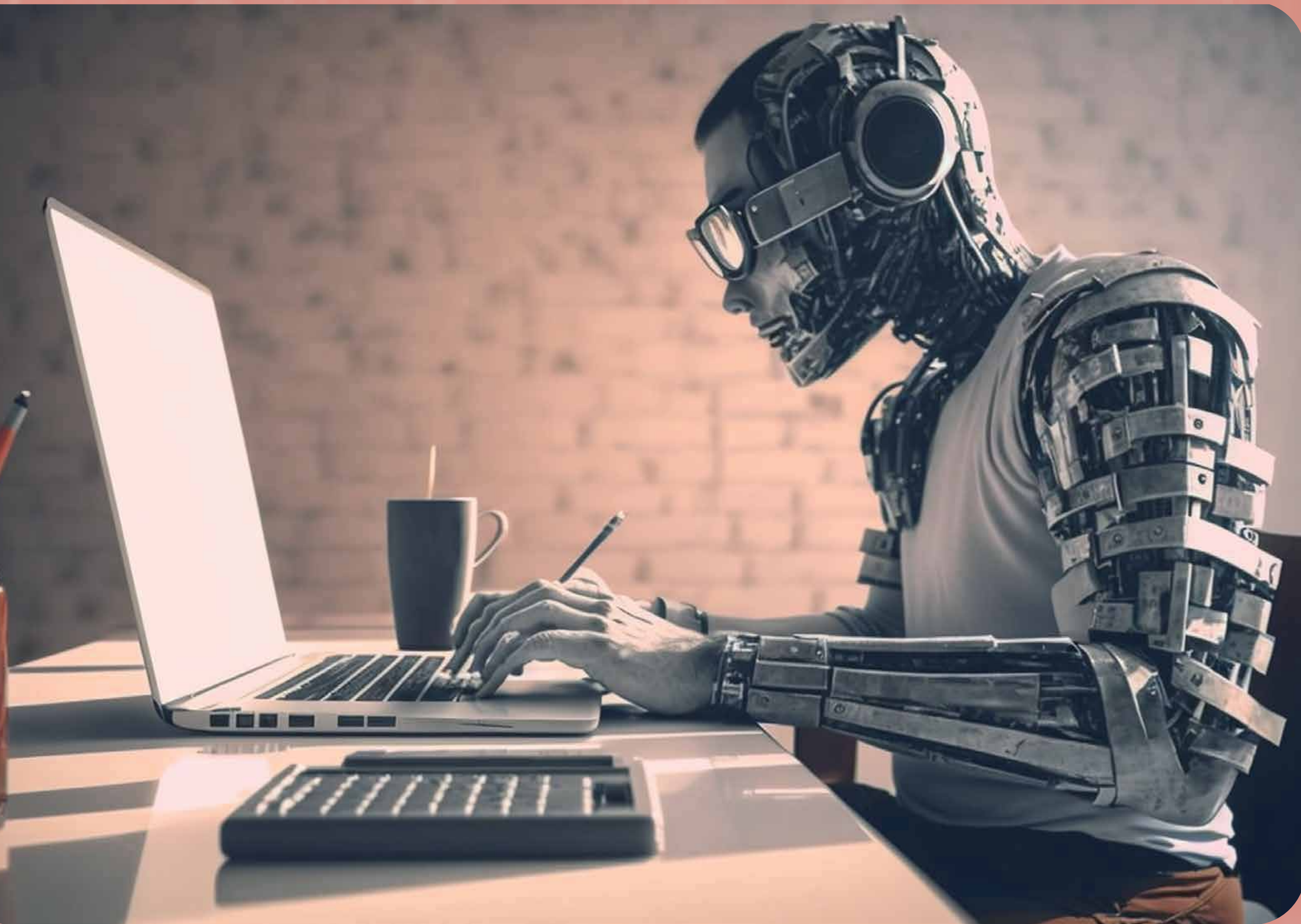
As generative AI becomes more useful for learning new information, finding the right words, coming up with new ideas, amusing ourselves, or eliminating inefficiencies, **how will our relationship with technology change?** How will this new relationship with technology change how we work with other persons? How will it affect personal mentorship, collaboration, or even our preference to waste time with one another?

**Can AI align with the unrepeatability of the relationship between you and I? What is irreducible in human relationships that AI cannot replace?**





# AUTHORSHIP & CREATIVITY



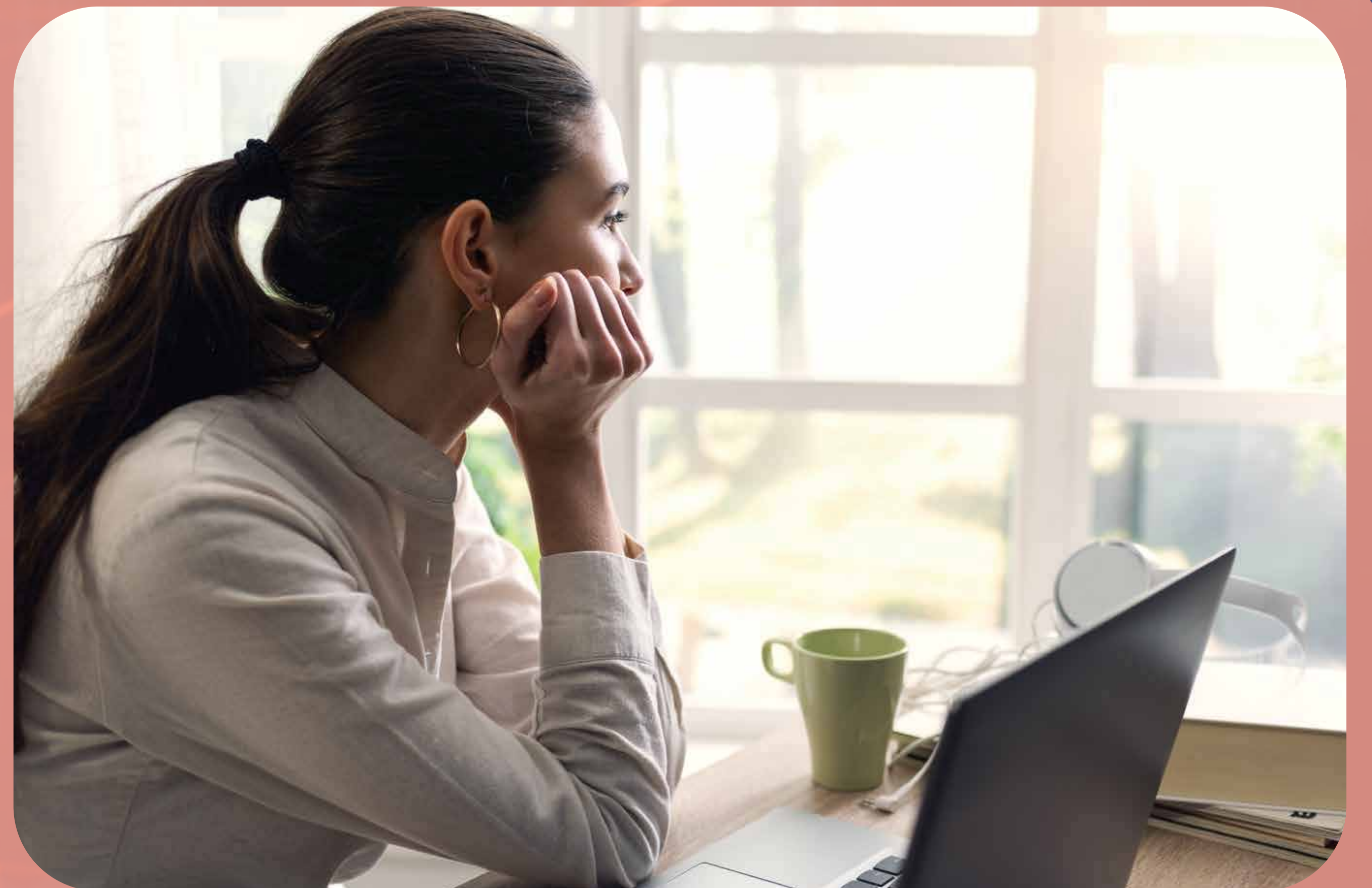
Vauhini Vara is one of the earliest professional writers to have experimented with Generative AI in the writing process. “Ghosts”, one of her first major pieces of AI-assisted writing, was about her response to the death of her sister. She could **not find the words to express her experiences for over twenty years, but AI (GPT-3) helped her finally find the words:**

*“The AI began describing grief in language that felt truer---and with each subsequent attempt, it got closer to describing what I’d gone through myself... I started hearing from others who had lost loved ones and felt that the piece captured grief better than anything they’d ever read... **It was better received, by far, than anything else I’d ever written.**”*

Nonetheless, Vara has pointed out that the **personal expression** and **consciousness of the human writer will never be replaced:**

“I recalled Zadie Smith’s essay “Fail Better”, in which she tries to arrive at a definition of great literature.

**She writes that an author’s literary style is about conveying “the only possible expression of a particular human consciousness.” Literary success, then, “depends not only on the refinement of words on a page, but in the refinement of a consciousness”... “writing in an attempt to clarify what the world is like from where I stand in it”.**



“That definition of writing couldn’t be more different from the way AI produces language: by sucking up billions of words from the internet and spitting out an imitation. Nothing about that process reflects an attempt at articulating an individual perspective... **The world as seen by AI is fatally incoherent. If writing is my attempt to clarify what the world is like for me, the problem with AI is not just that it can’t come up with an individual perspective on the world. It’s that it can’t even comprehend what the world is.**”



Vauhini Vara "Confessions of a Viral AI Writer," Wired, October 2023



# TRUTH AND AI HALLUCINATIONS



One weekend in March 2023, this deepfake image of Pope Francis wearing a Balenciaga puffy coat went viral on social media. Uncanny, though not widely believed **deepfakes** had gone viral on the internet before. However, this incident was highly unusual for **how widely and unquestioningly it was believed to be a real image until news came that it was AI-generated**. Web-trend writer Ryan Broderick called this event “the first real mass-level AI misinformation case”.

A Chicago area construction worker generated the image using Midjourney (an AI art generation service). He “just thought it was funny to see the Pope in a funny jacket” and posted this on social media art forums never intending or expecting it to be spread and taken at face value. Some commentators of the image complemented the Pope on his fashion sense, while others criticized the Vatican for opulence. This incident highlights the **potential for generative AI use to lead to widespread misinformation at scale, even when not intended**.

Source: <https://www.forbes.com/sites/mattnovak/2023/03/26/that-viral-image-of-pope-francis-wearing-a-white-puffer-coat-is-totally-fake/>

**Large language models too can “hallucinate” credible, yet untrue information.** For example, New York lawyer, Steven Schwartz, had used ChatGPT for the first time to **research prior court cases for a legal brief, leading him to cite six court cases that did not exist.** After the judge asked the legal team to explain itself, Schwartz stated he was “unaware that its content could be false.” Schwartz had been an attorney for over thirty years, illustrating **how believable such AI hallucinations can be even to an expert.**



While these examples of AI-generated misinformation were unintentional and quickly discovered to be false, **there is potential for individuals to be fooled by AI misinformation or for bad actors to intentionally spread AI-generated disinformation that can have longer and deeper societal effects.** Education and creative responses to these issues will mark the age of AI we are entering.







## TECHNOLOGY &amp; KNOWLEDGE

While ChatGPT and its successors are new and exciting, the debate on technology, and our relationship with it, are as old as humanity itself. In Plato's Phaedrus (370 BC), Socrates discusses the problems of the use of writing over the spoken word.



**SOCRATES**

I cannot help feeling, Phaedrus, that writing is unfortunately like **painting**; for **the creations of the painter have the attitude of life**, and yet if you ask them a question they preserve a solemn **silence**. And the same may be said of **speeches**. You would imagine that they had intelligence, but if you want to know anything and put a question to one of them, the speaker always gives one unvarying answer. And when they have been once written down they are tumbled about anywhere among those who may or may not understand them, and know not to whom they should reply, to whom not: and, if they are mistreated or abused, they have no parent to protect them; and **they cannot protect or defend themselves**.

Is there not another kind of word or speech far better than this, and having far greater power - a son of the same family, but lawfully begotten?

**PHAEDRUS**



Whom do you mean, and what is his origin?



**SOCRATES**

I mean an intelligent word graven in the soul of the learner, which can defend itself, and knows when to speak and when to be silent.

**PHAEDRUS**



**You mean the living word of knowledge which has a soul**, and of which the written word is properly no more than an image?

Socrates worries about our tendency to transfer our sense of what is real onto technology itself, which is no more than an **image of true knowledge**.



# DELEGATION, CONVENIENCE & AUTONOMY



“There is something in me that is not derived from any empirical phenomenon, because it does not depend upon, does not originate in the biology of my father and mother. ***It directly depends on the Infinite, which makes the whole world. [...] The world can do what it wants with me, but it cannot conquer, possess, grasp on to me, because I am greater than it is. I am free.***”

Luigi Giussani, *The Religious Sense*

“The vast amount of data analyzed by artificial intelligences is in itself ***no guarantee of impartiality***. When algorithms extrapolate information, they always run the ***risk of distortion, replicating the injustices and prejudices of the environments where they originate. The faster and more complex they become, the more difficult it proves to understand why they produced a particular result.***”

Pope Francis, January 1, 2024 [1]

We all live the inherent tradeoff that occurs when I “transfer” part of my knowledge and autonomy to this image. The act of weighing these tradeoffs imposes a responsibility on all of us. Consider the following examples:

**Fernando** says: “I use it all the time at work, especially when learning new things. Chat GPT can act like a **specialized tutor**, allowing me to ramp up **quickly** in new problem domains that would have taken a long time to cover in a class, or on my own”.

**Linda** is a high-school teacher. One day she overhears one of her colleagues discussing the use of generative AI for lesson planning. “**It’s great!**” the colleague says. “**It gives me ideas for teaching**, and it even creates a detailed lesson plan with great questions and activities for the students. **It has saved me a lot of time**, so I can actually spend more time with my husband.”

**Katie** overhears one of her co-workers, a manager at a tech company excitedly explain how generative AI has **saved her so much time writing emails**. It even helped her write a tricky email where she needed to reprimand one of her junior engineers.

**Ted** says: “now that I have ChatGPT, I barely use Google anymore. **It just gives me the answer right away rather than websites!**”

## USE THE WORKSTATIONS AND REFLECT ON..



What do you take away from this exhibit? Is convenience a sufficient reason to use these tools? Is there anything that we lose when we delegate the creation of knowledge to an engine? Where could you see benefits/dangers of using AI in your work?

[1] Excerpt from: Message of His Holiness Pope Francis for the 57th World Day of Peace; Artificial Intelligence and Peace, January 1, 2024. For the entire text scan the QR Code:

