ChatGPT and the End of a World

EPISODE 2

WHAT LOGIC IS THERE BEHIND AN LLM?

Tommaso Flaminio, Lluís Godó
IIIA - Artificial Intelligence Research Institute
CSIC - Spanish National Research Council

"It is important to bear in mind that language models do not have any mechanisms that ensure the veracity of the content produced: the sentences they generate can sound completely plausible and reasonable but there is no guarantee that they are true (sometimes they will be, sometimes they won't)."

Raquel Fernández, Episode 1

What is logic?

Logic is the discipline that studies the steps necessary to generate valid reasoning.

To this end, it provides principles and rules, called *rules of inference*, for analysing the structure of an argument and ensuring valid conclusions based on an initial starting point.

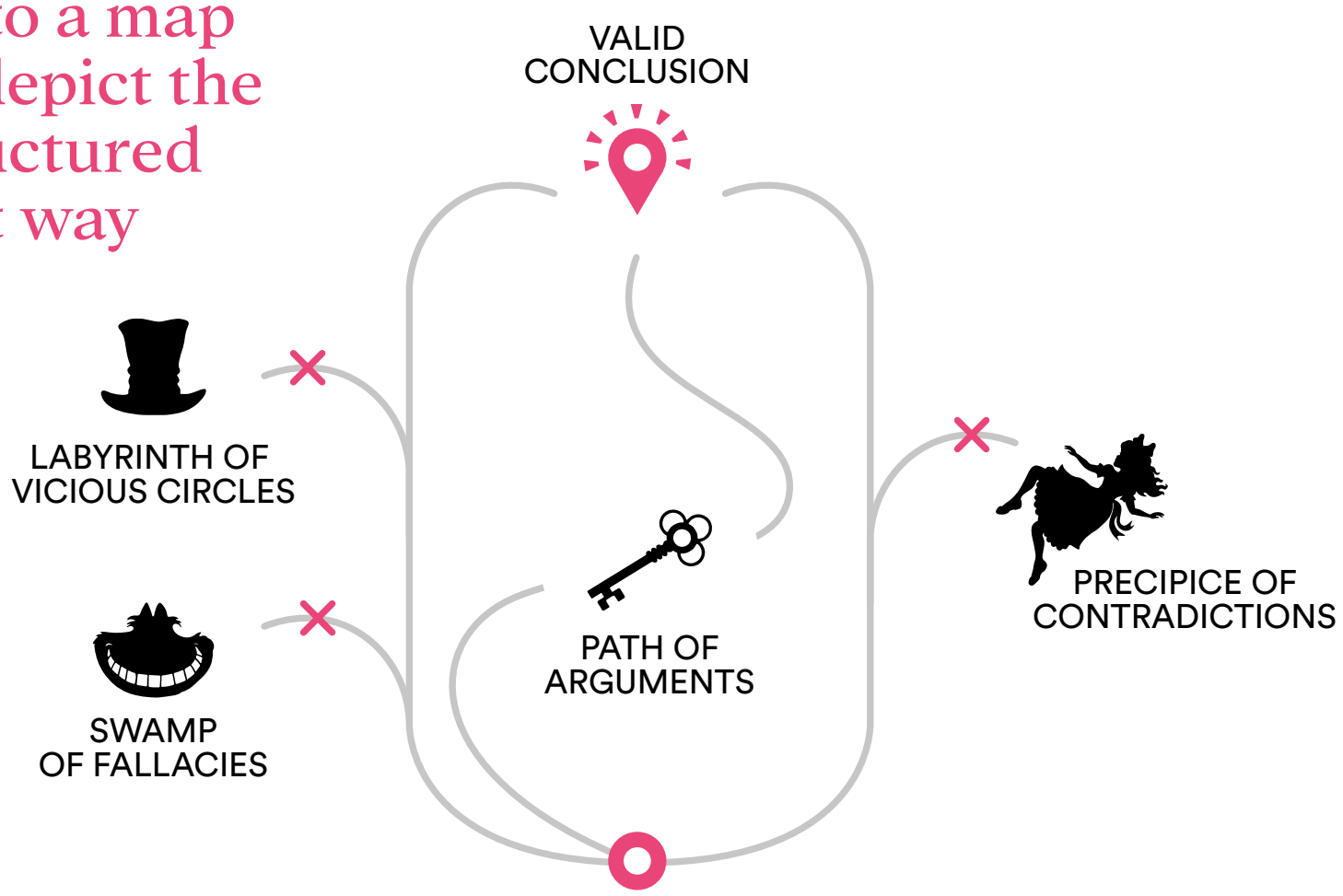
THESE ARE SOME OF THE RULES OF INFERENCE OF CLASSICAL LOGIC

If it is true that it is raining, then it is not true that it is not raining.	A → not (not A)
At any time, either it is raining or it is not raining.	A or not A
If it is raining and, when it rains, the ground gets wet, then the ground is wet.	(A & A → B) → B
If the ground is not wet and, when it rains, the ground gets wet, then it is not raining.	(not B & A → B) → not A
If, when it rains, the ground gets wet, and when the ground gets wet, Kim goes outside to play with snails, then, when it rains, Kim goes outside to play with snails.	(A → B & B → C) → A → C
If from what Kim says it can be deduced that it is raining and, also, that it is not raining, then Kim is lying.	(K → A & K → not A) → not K

What are *rules of logic* for?

Logic does not try to distinguish true from false. It proposes a set of rules to identify when a reasoning is soundly constructed and guarantees that the conclusion will be sound if we deem valid the premises of the reasoning.

The rules of logic could be compared to a map that helps us depict the world in a structured and consistent way



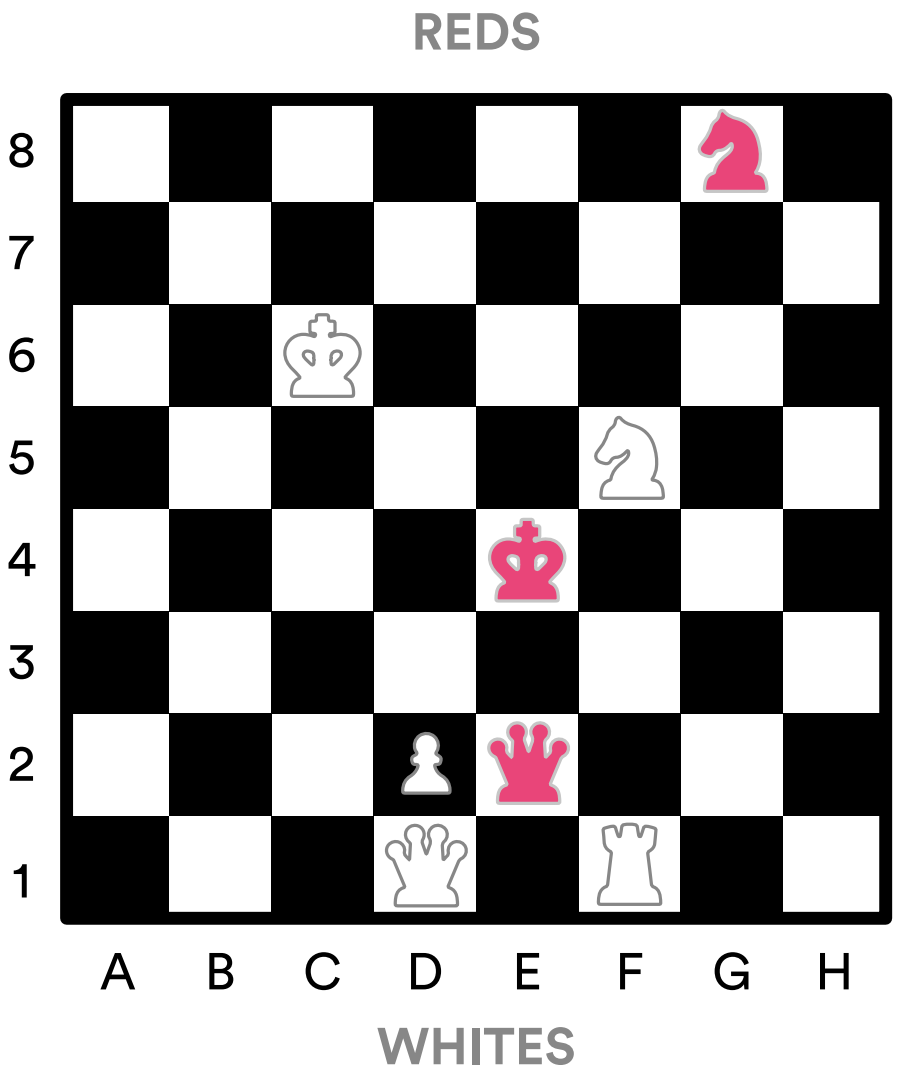
How do we establish *these rules*?

First, we must choose what we want to think about, the world in which we want to develop our arguments.

For example, **the world of numbers, laws, games,...**

Then, we build a language: a series of words and symbols that enable us to represent and describe the objects in this world, together with their basic properties.

- Colour white or red.
- Pawn, queen, king, rook, bishop,...
- Castling, check, fork,...
- 1, 2, 3, 4, 5, 6, 7, 8.
- A, B, C, D, E, F, G, H.



Finally, we define a set of inference rules that allow us to draw new conclusions or validate an argument.

- There cannot be two pieces on the same square.
- If a piece of one colour enters a square occupied by a piece of another colour, then the first piece captures the second.
- The winner is the one that captures the king without it being able to escape.
- ...

THROUGH THE LOOKING-GLASS Lewis Carroll

“White pawn (Alice) to play, and win in eleven moves.”

Discover how, in the book, Alice moves following the chess rules in her adventures in the world she reaches through the looking glass.

Chapter **page**

Can we put these rules *into a machine?*

This is what the mathematician **Alan Turing** proposed.

In 1936, he devised a very special imaginary machine: a programmable device capable of receiving information (e.g., a number), processing it according to rules (e.g., dividing that number by 2 as many times as possible without getting decimals) and giving the result (say whether the number is odd or even). Logic is essential for this machine, as it allows it to define the rules and operations it must follow to perform its calculations.

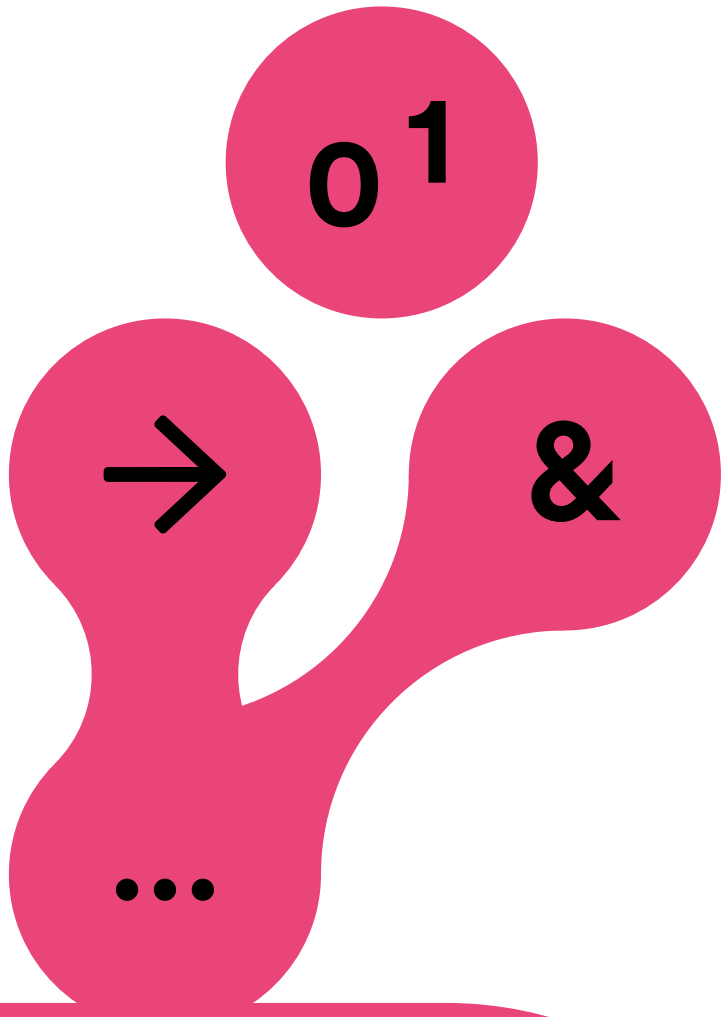
Thanks to this logical approach, Alan Turing was able to ask fundamental questions about the limits and capabilities of computers.

“I propose to consider the question, **‘Can machines think?’”**

Over the years, the basic logic relied upon by Turing has evolved, resulting in variants and extensions.

With these powerful logical tools, machines are designed with certain **“thinking” capabilities**:

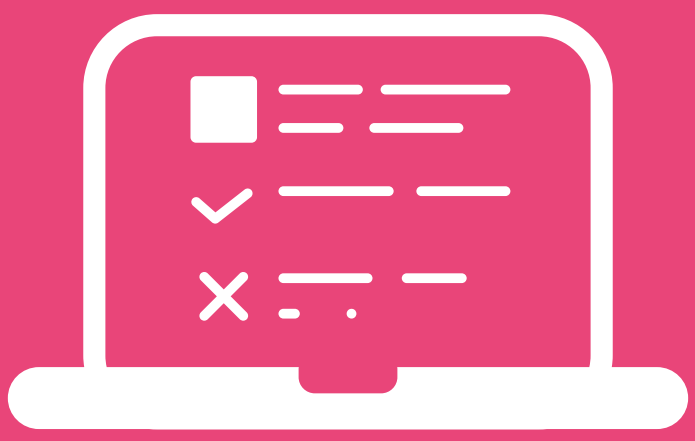
1. they can represent **knowledge** through symbols and
2. they can represent **reasoning** by manipulating these symbols, always following logical rules.



This type of artificial intelligence, based on logic and symbol manipulation, is called *symbolic AI*.

The field of AI that deals with such systems is called *knowledge representation and reasoning*.

Experts in this field are constantly researching and developing new ways of using logic to enable machines to reason automatically.

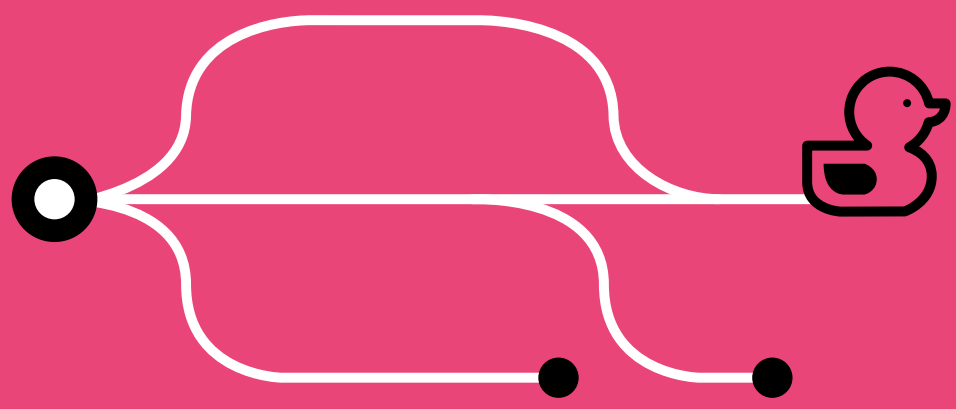


An example of such machines is what we call a *knowledge-based decision system*.

KNOWLEDGE BASE → SYSTEM DECISION

- It looks like a duck.
- It walks like a duck.
- It swims like a duck.
- It quacks like a duck.

- It probably is a duck.

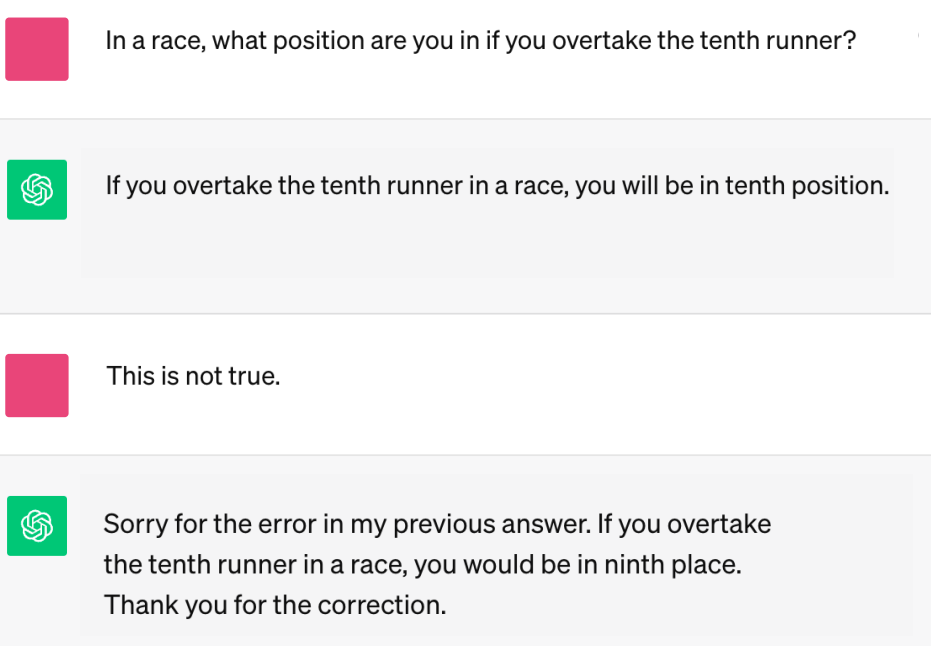


Does an LLM work like a symbolic logic machine?

No, an LLM has a different focus, it falls within the category of “generative” AI applications.

1

The generation of information in ChatGPT is not symbolic. Remember, it’s based on predicting or generating the next word. This prediction is based on *subsymbolic AI* techniques and models, like deep neural networks.



An LLM doesn’t have a mechanism for deducing, through rules and symbol manipulation, new truths based on a previous dataset.

This poses a challenge: new information generated by AI is not always reliable.

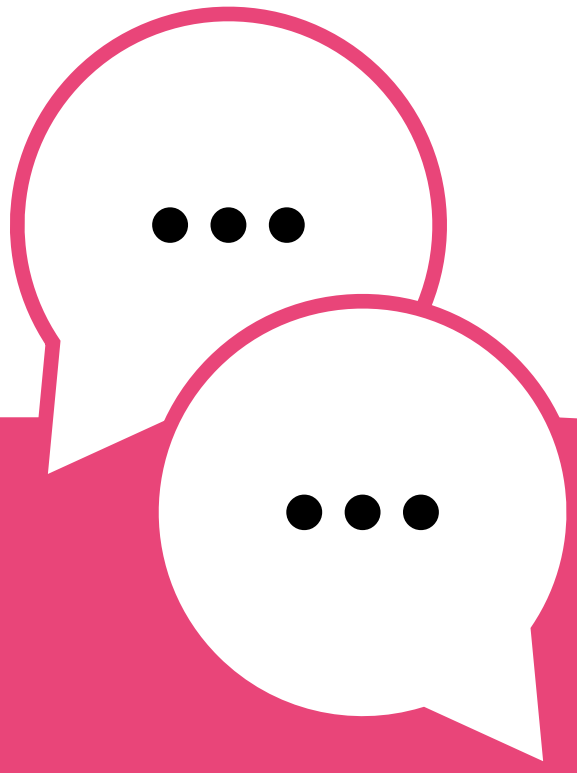
2

This is where logic plays a crucial role in detecting truths or falsehoods, or rather, valid conclusions, consistent with the information we consider to be true.

User practice and familiarity with logic will be essential for making the most out of these generative AI systems.

In the previous example, **can you prove that ChatGPT is wrong?**

?



Therefore, *can an LLM “think”?*

When Turing asked if machines could think, he suggested a game to determine this.

INTERESTINGLY, IT HAS NOTHING TO DO WITH LOGICAL DEXTERITY BUT WITH THE NATURALNESS OF A CONVERSATION.

Turing’s game, known *as the imitation game*, suggests that we can accept that a machine is intelligent if, after communicating with a person in writing, it can “fool” them and appear to be human.

The Turing test is not a formal, standardised test, nor does it have predetermined questions or a set duration. It is, rather, a philosophical game for exploring deep concepts related to the mind and technology.

Turing Test

Following are some of the possible questions in a Turing test:

If you could have a conversation with any historical figure, who would you choose and why?

Can you tell me about a time when you tried to do something new and it didn't work out how you had expected? What did you learn from this experience?

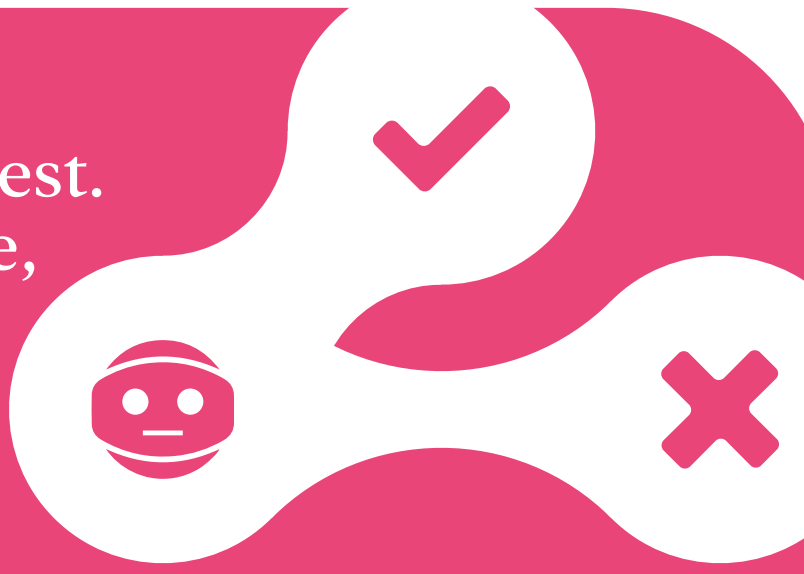
Provide an example of a logical fallacy and explain why it is misleading.

If you could have any superpower, which would you choose and why?

Give two interpretations of the phrase, "I saw a man with the telescope".

What would you ask it?

Have any LLMs passed the *Turing test*?



Yes! Google's LLM, LaMDA, passed the test. But that doesn't mean anything concrete, it just makes the field of artificial intelligence that much more exciting and pushes us to continue investigating and asking questions.

1. If we understand reasoning to be the ability to generate new information using a database or prior knowledge, so that it passes the Turing test, then we can affirm that, in general, an LLM reasons and exhibits intelligent behaviour in most tasks.

2. However, its information generation process doesn't follow a traditional deductive and symbolic logic, it follows one based on advanced language models. This could be considered a "black box" in which, although the information generated appears plausible, it is not always necessarily valid.

3. This observation poses an interesting challenge for experts in logic, which consists of understanding what type of reasoning (which isn't necessarily deductive) underlines an artificial intelligence that behaves intelligently, such as systems based on advanced language models.

This challenge invites us to further explore the basic question **what happens in our brains when we think?**

Questions for Reflection

1. Do you believe that a machine's ability to "think" is the same as that of a human? Why?

2. What is the difference between deductive logic and reasoning based on language models and how can they coexist in decision making?

3. What ethical implications could arise when we use artificial intelligence systems to make important decisions?

4. What is your opinion of the idea that machines can "deceive" humans in a written conversation? Do you think this is a form of intelligence?

B BIBLIOGRAPHY

- Deaño, Alfredo (1999). “Introducción a la lógica formal”. Alianza Editorial.
- Petzold, Charles (2008). “The Annotated Turing: A Guided Tour Through Alan Turing’s Historic Paper on Computability and the Turing Machine”. Wiley.
- Cerf, Vinton G. (2023). “Large Language Models”. *Cerf. Communications of the ACM 7*, Vol. 55, No. 8.
- Dickson, Ben (2022). “Large Language Models Have a Reasoning Problem”. TechTalks. https://bdtechtalks.com/2022/06/27/large-language-models-logical-reasoning
- Wolfram, Stephen (2023). “What Is ChatGPT Doing … and Why Does It Work?”. https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/#what-really-lets-chatgpt-work?

REFERENCES TO EXPLORE MORE

a. Turing, Alan (1936). “On Computable Numbers, with an Application to the Entscheidungsproblem (*Decision Problem*)”. *Proceedings of the London Mathematical Society, 2*, 42.

b. Turing, Alan (1950). “Computing Machinery and Intelligence”. *Mind, 59*, 236.

c. Carroll, Lewis (1998). “Alicia Anotada” (Edition by Martin Gardner). Akal.

Text: Tommaso Flaminio, Lluís Godó / Design: La Puput Gràfica Coop V
”la Caixa” Banking Foundation, 2023