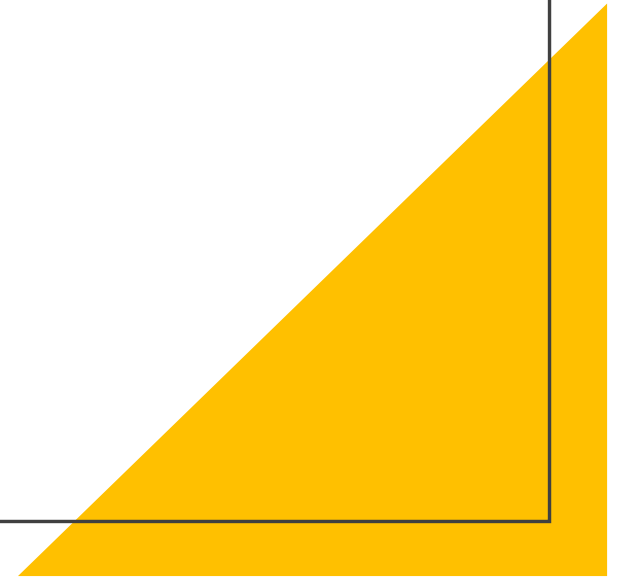


ChatGPT as an intelligent daydreamer

Walter Daelemans

CLIPS / TEXTUA / ACRAI

University of Antwerp



WA

Summarize the morality of the characters in the following story in one sentence for each character. Story: John went to help his friend Ben reorganizing his garden. Digging in the garden he found a hidden treasure of medieval gold coins and took it home. His wife Mary convinced him to give the treasure back to Ben. John admitted taking the treasure to Ben and gave it back, but kept a few of the coins anyway. Ben was so delighted with getting the treasure that he gave John a few coins.



1. **John:** John exhibits a mix of morality, initially demonstrating goodwill by helping Ben, but then succumbing to temptation by taking the hidden treasure, although he partially redeems himself by admitting his actions and returning most of it.
2. **Mary:** Mary acts as a moral compass in the story, persuading John to do the right thing by returning the treasure to its rightful owner, Ben.
3. **Ben:** Ben showcases generosity and forgiveness, feeling delighted with the return of the treasure and rewarding John with some coins despite John's initial dishonesty.

NLP in the picture!

- An architecture developed for *Machine Translation* has taken the whole of Artificial Intelligence by storm

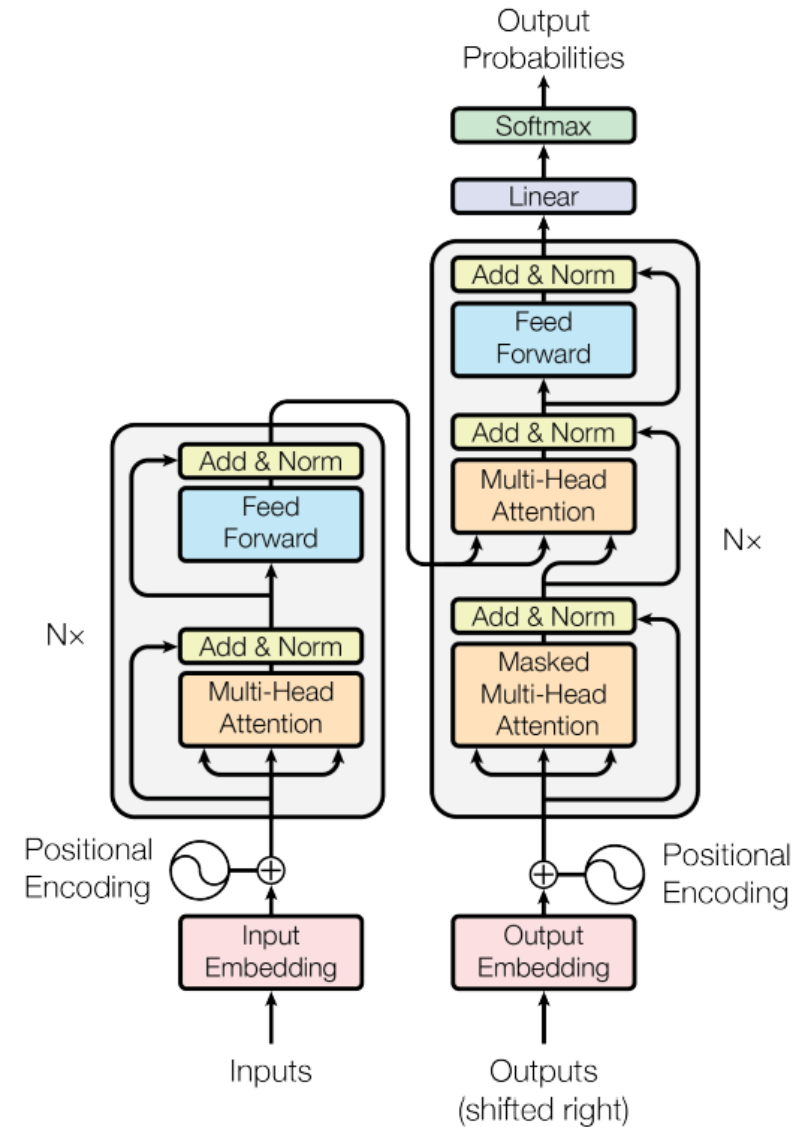


Figure 1: The Transformer - model architecture.

GenAI

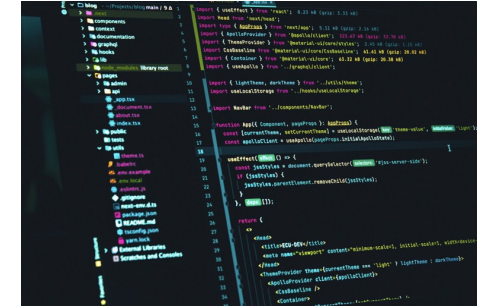


Image and Video generation and description

Text to speech and speech to text

Program code generation and understanding



Text to Music

William Caxton (b. 1415–24–1492) was the person who brought the technology of printing to England. Before Caxton set up his printing press in Westminster, London, in 1475 or 1476, books in England were copied out by hand, by scribes. These hand-written books are called manuscripts (the word 'manuscript' means 'written by hand' in Latin). Printing had a huge impact when it arrived – changing the way that information was produced and circulated in much the same way that the arrival of the internet has changed our world in more recent times.

Caxton's birth, education and apprenticeship
Caxton was born in Kent. He seems to have had some basic schooling, but he wasn't especially well educated. He became an apprentice to a mercer (a textile merchant) called Robert Large. We don't know precisely when he was born, but apprentices usually began their apprenticeships at 14 and this is how we calculate the approximate year of his birth. Training as a mercer was to give Caxton skills that would be useful in later life. Mercers dealt in haberdashery, cloth and silks, but the mercer's guild (a kind of club of tradespeople) was a powerful one. They were involved in politics and provided finance for the government. As a mercer, Caxton also had to spend time overseas doing business.

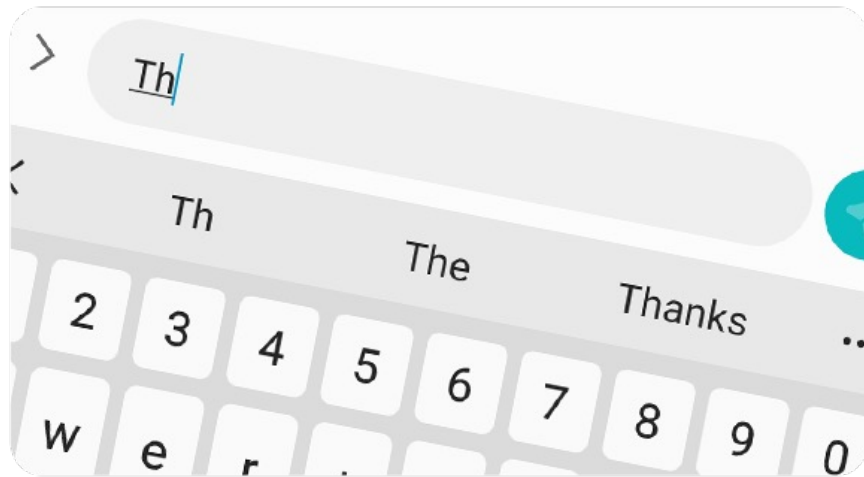
Caxton's life in Europe and growing involvement in print culture

Conversational Agents

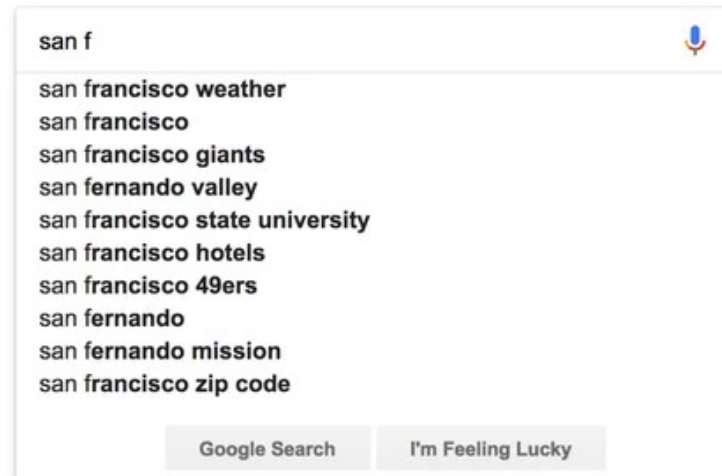
Text Mining

Question Answering

Machine Translation

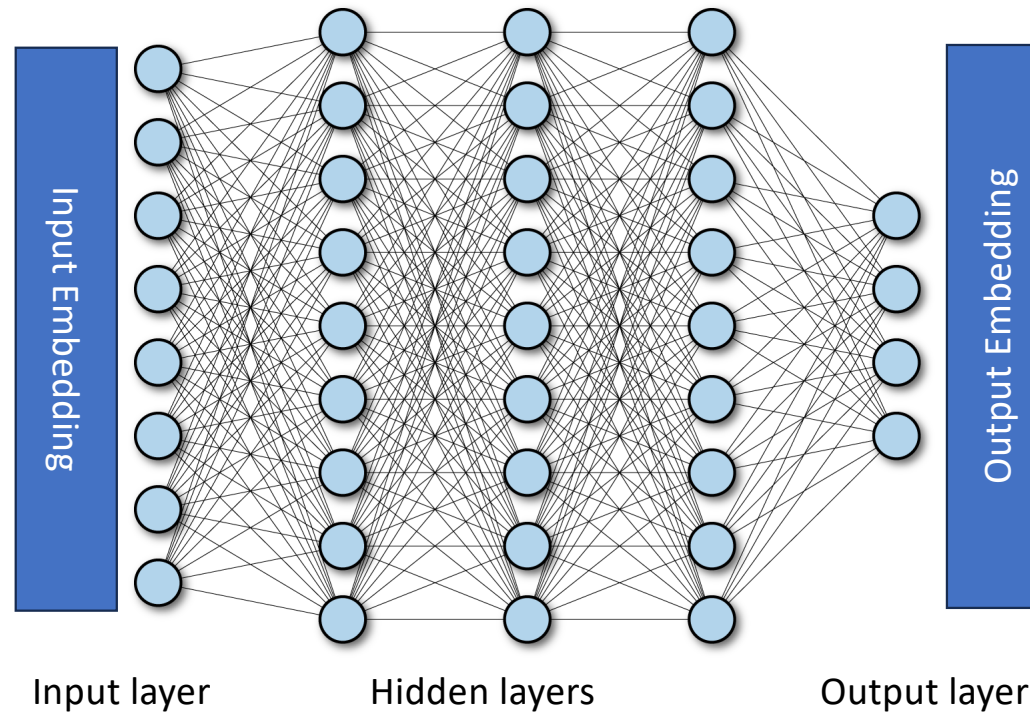


LLMs: “Completion on steroids”



Deep Neural Network based Language Models

Nee, polarisering in de samenleving kun je niet oplossen met enkele simpele aanpassingen aan de algoritmes van Facebook en Instagram. Als er al een conclusie te trekken valt uit vier uitgebreide studies die donderdag verschenen in de tijdschriften *Science* en



Nature 0.7
gelijkaardige 0.1
Technology 0.05
cocktails 0.0002
...

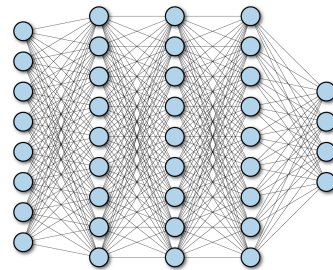
SELF-SUPERVISED LEARNING
Result: PRE-TRAINED MODEL

Measuring a model (GPT-3)

**Sustainability?
(Energy, water)**

Amount of training data
(500 billion)

Maximum prompt size
(4096)



Other measurements:

- Number of epochs
- Number of layers (blocks)

Number of connection weights (parameters)
(175 billion)

Why would this work? (Benchmarks)

WA

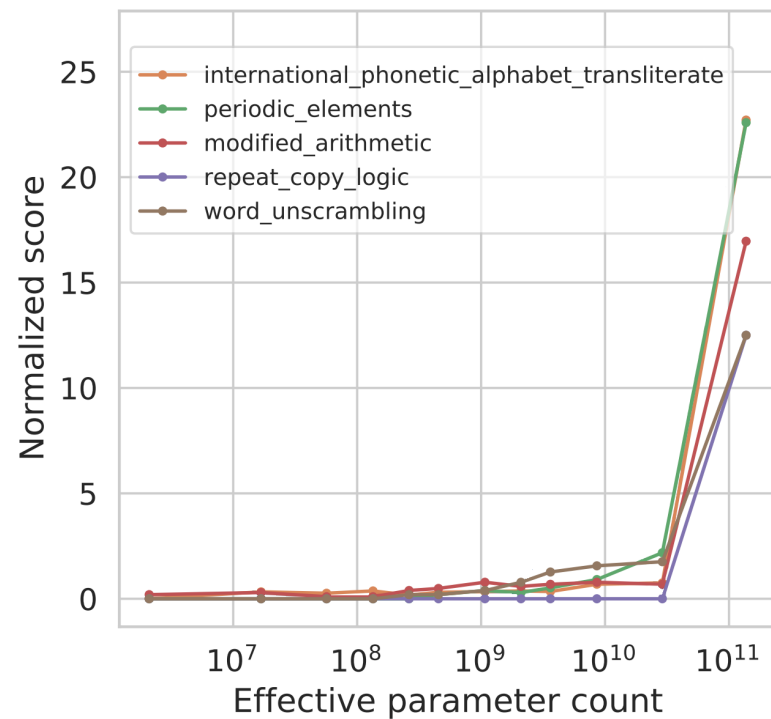
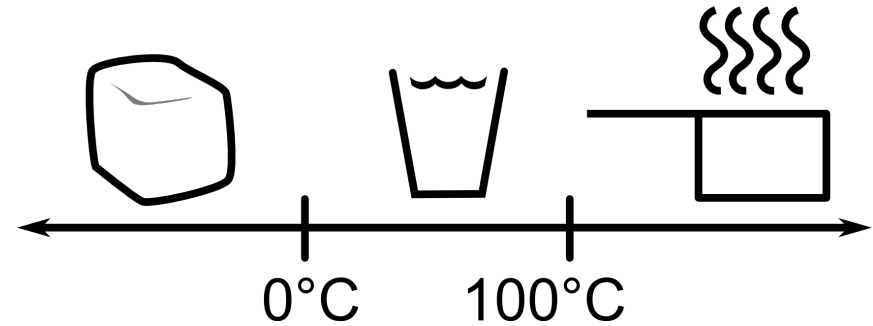
Continue the following with one word: Energy prices decreased a lot in Belgium in 2022.
Purchasing power of Flemish families has considerably



increased.

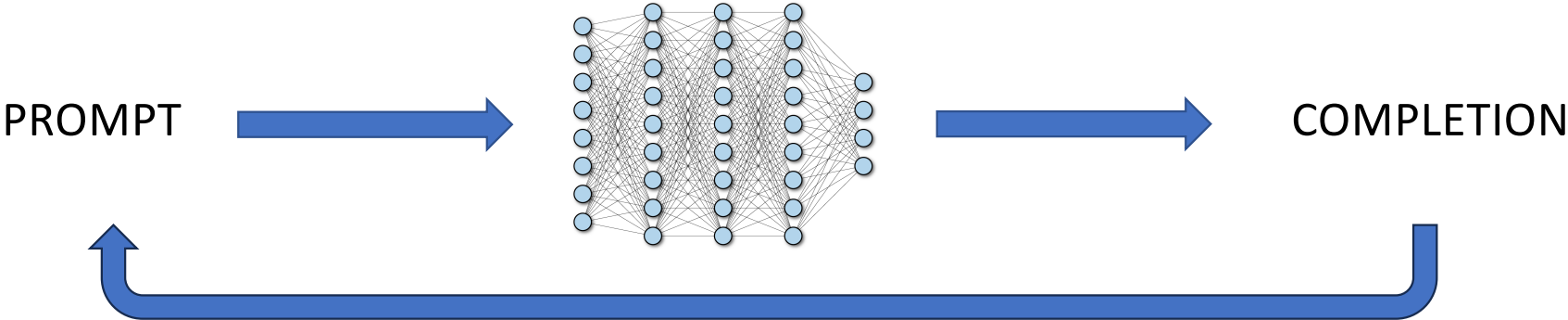
Emergence

- **No direct perception of reality**, only indirectly via text
- Language patterns
 - Form, contents, style
 - World knowledge and reasoning (?)
- Emergent patterns
 - Summarization
 - Translation
 - Conversation
 - Programming
 - ...
 - Interpreting morality
 - Multiple choice question generation



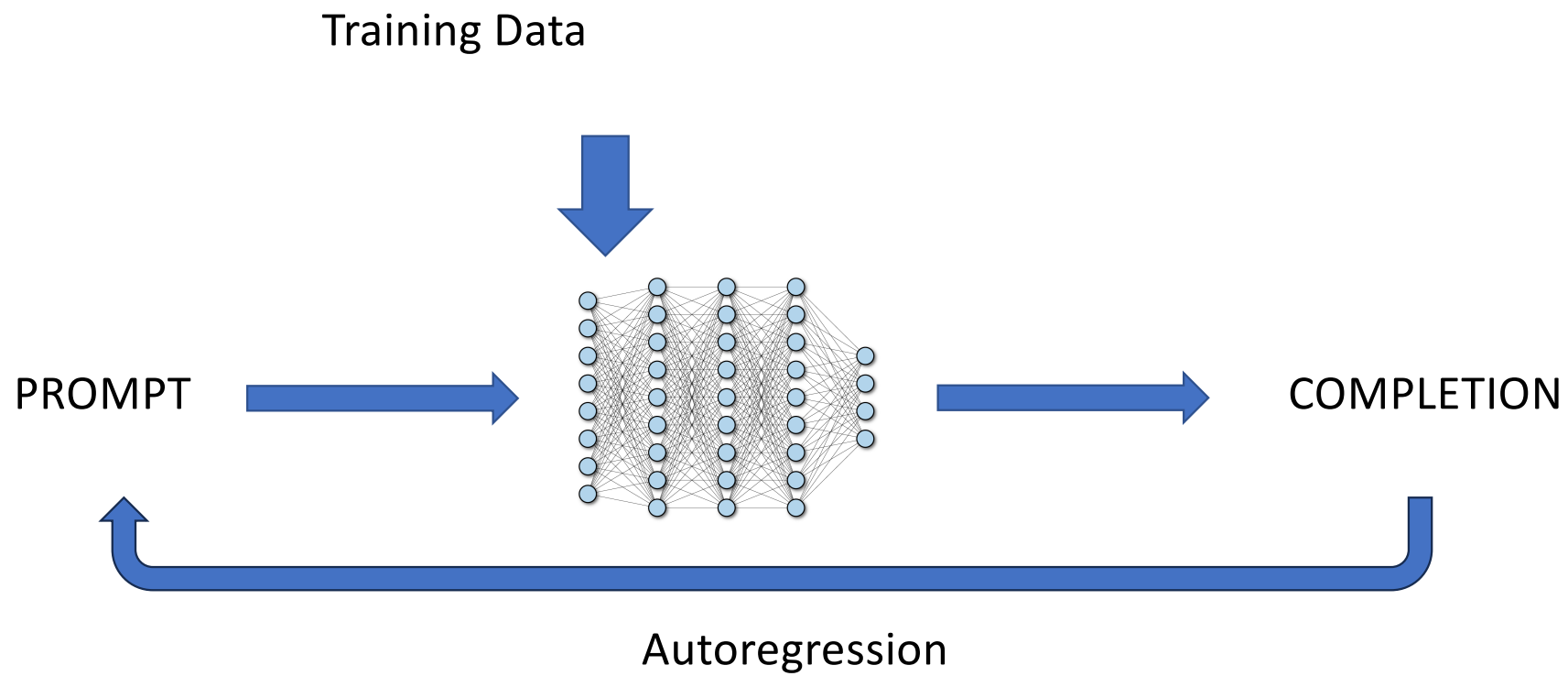
LLM ecology (September 2023)

- Open AI (+ Microsoft)
 - GPT-4 (Playground, API, Azure, Edge ...)
- Google
 - Bard (free for now, based on PaLM 2)
- Anthropic (+ Amazon)
 - Claude-2 (not in EU)
- META
 - Llama 2 (Open Source)
- Technology Innovation Institute, UAE
 - Falcon 180b (Open Source) ~ GPT 3.5



Autoregression

- Consistency**
- Dialogue management**
- "Reflection"**
- Hallucination**

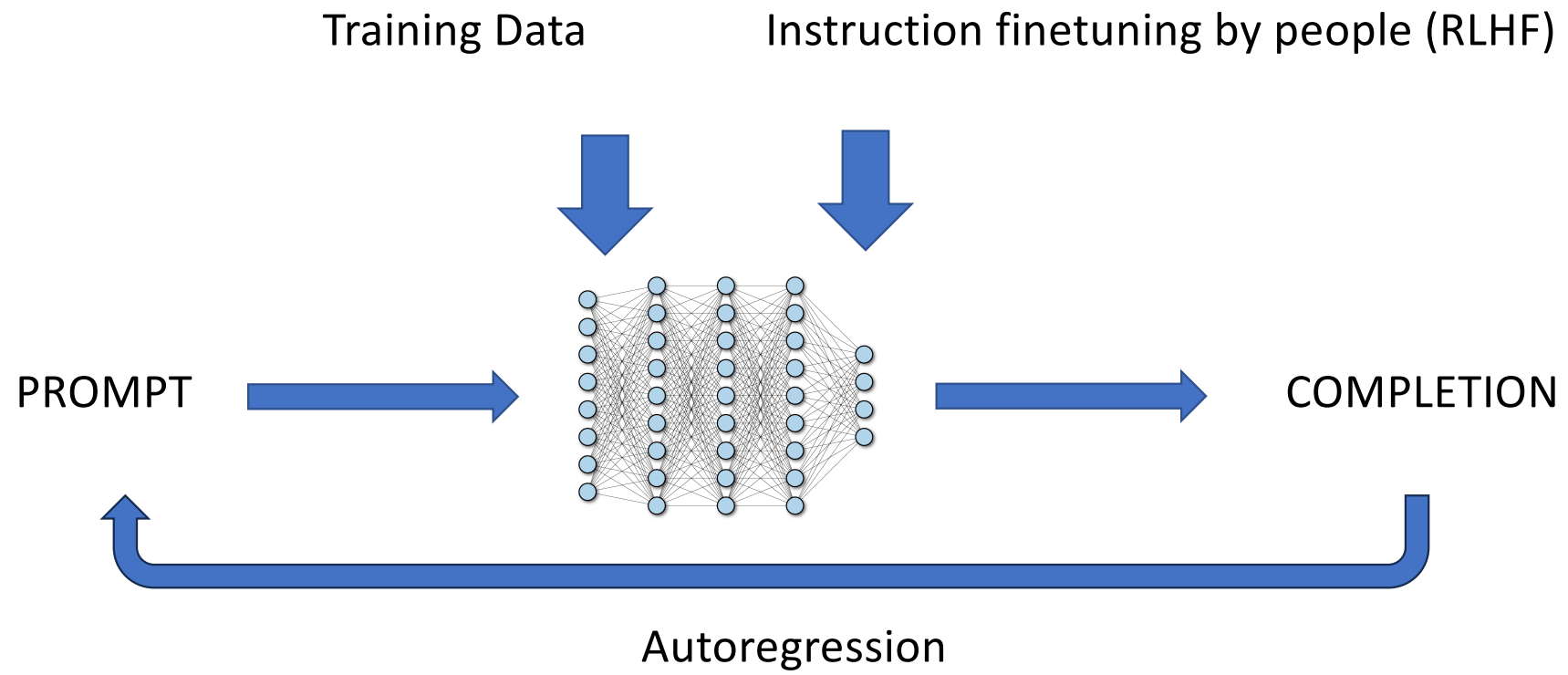


Training Data

- 500 billion (GPT-3)
 - Compare kids aged 3 : 10 – 40 million words input
- Basis of the LLM's knowledge (here Llama)
 - **Fact and fiction**
 - Books, code, news, social media, ...
- Ownership ?
- Bias ?

Dataset	Sampling prop.	Epochs	Disk size
CommonCrawl	67.0%	1.10	3.3 TB
C4	15.0%	1.06	783 GB
Github	4.5%	0.64	328 GB
Wikipedia	4.5%	2.45	83 GB
Books	4.5%	2.23	85 GB
ArXiv	2.5%	1.06	92 GB
StackExchange	2.0%	1.03	78 GB

Alignment



Alignment

- InstructGPT
 - RLHF: “Do the right thing”
- Alignment to what, alignment by who ?
- Cultural bias
- Potential criminal bias (phishing)

Finetuning on
user data

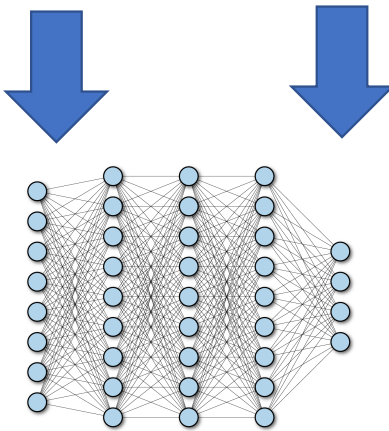
Training Data

Alignment

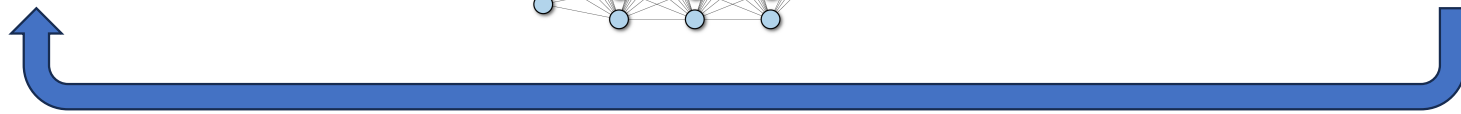
Instruction finetuning by people (RLHF)

In context learning

PROMPT



COMPLETION



Autoregression

Why is ChatGPT an intelligent daydreamer?

Sees the world 'second hand' through text: both fiction and non-fiction

Learns patterns (of patterns (of patterns)) that help predict the next word

Has no consciousness, emotions, opinions, intentions, etc. but can be called intelligent.

Does not make a distinction between dream and reality

Intelligent?



=



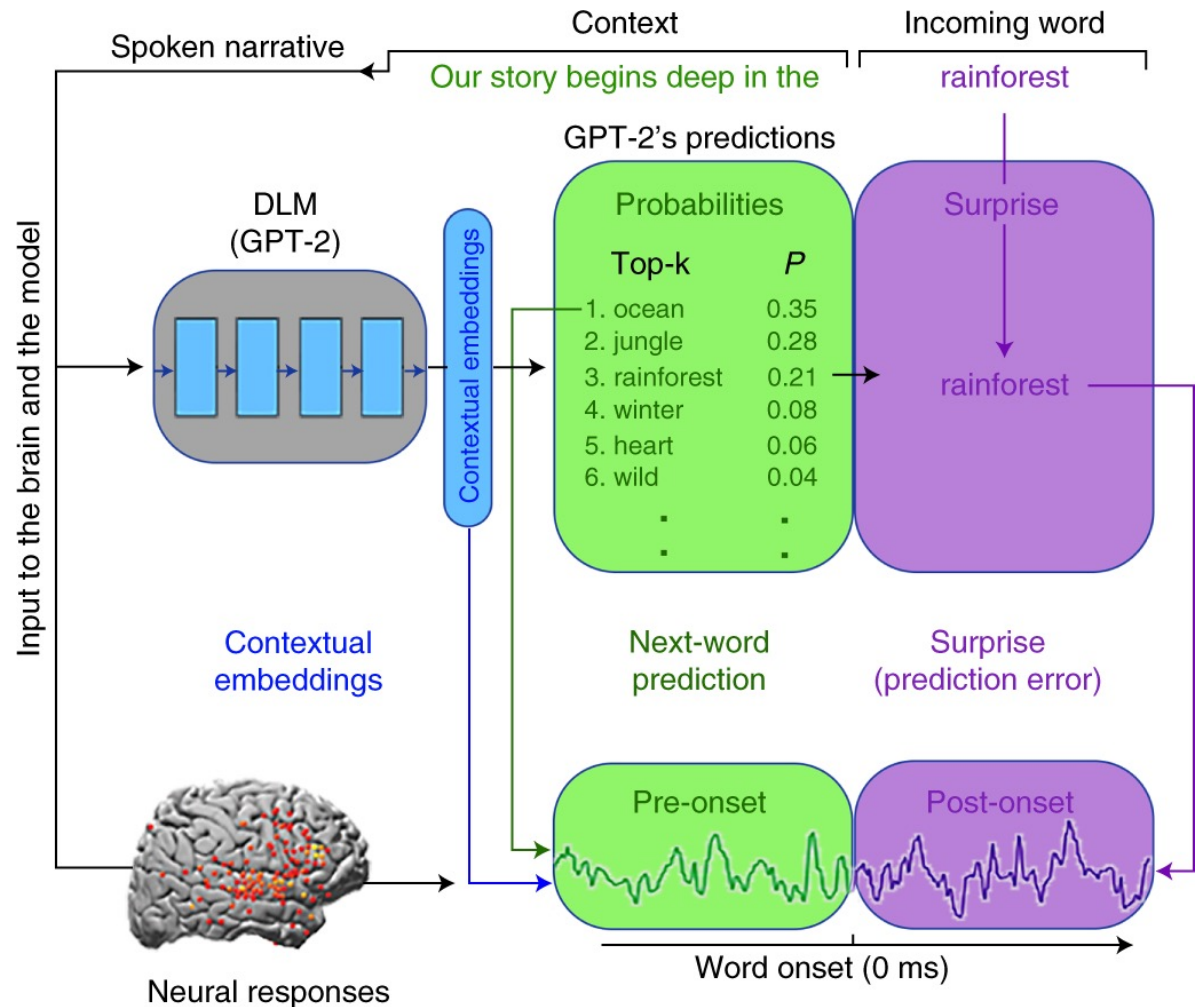


Predictive Brain

- Jeff Hawkins (Numenta)
 - Brain as prediction engine
- Karl Friston (UCL)
 - Free energy principle, predictive coding
- Brain is a prediction machine
- When reality does not match predictions, connections are updated



- Continuous context-dependent prediction of the next word
- Predictions used to compute surprisal (perplexity)
- Context-specific representation of meaning (like contextual embeddings)



Thank you!

Questions?

