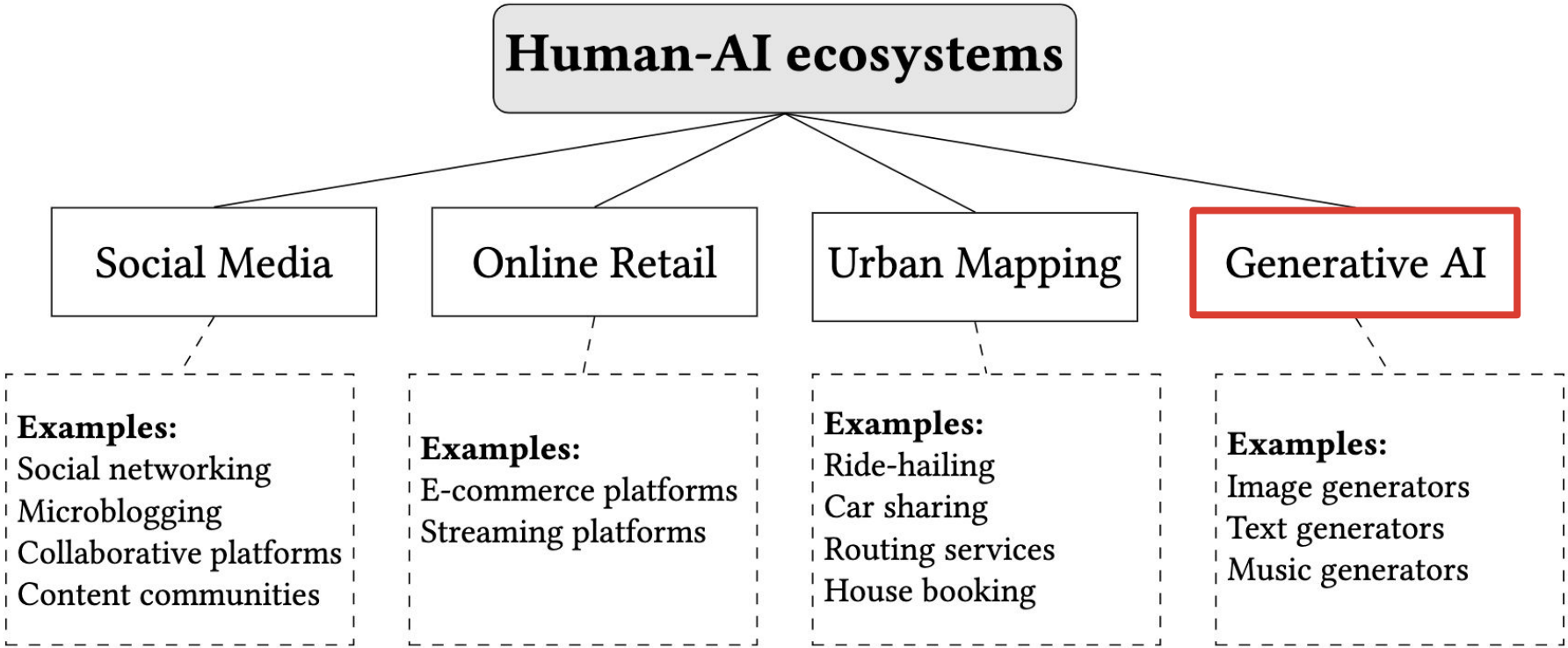
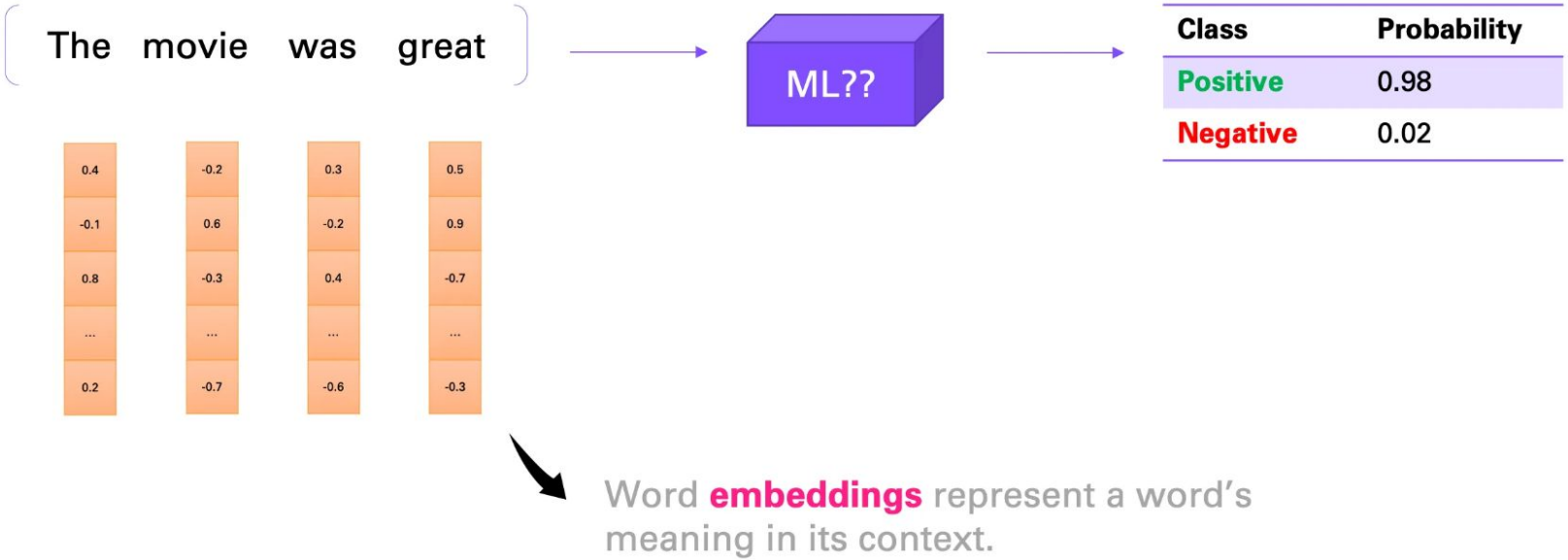

Human-AI ecosystems
Social Media
Online Retail
Urban Mapping
Generative AI
Examples:
Social networking
Microblogging
Collaborative platforms
Content communities
Examples:
E-commerce platforms
Streaming platforms
Examples:
Ride-hailing
Car sharing
Routing services
House booking
Examples:
Image generators
Text generators
Music generators

Processing text

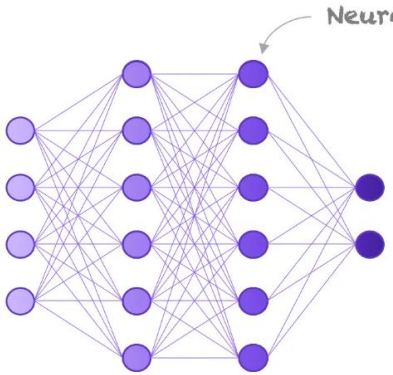


Processing text

We need something way more powerful… **Neural Networks**

GPT-3 has 175,000,000,000 (175B) Neurons
Brain: 100B Neurons

Input (image or text)



Class	Probability
Dog	0.03
Cat	0.96
Bird	0.01

Output

Language modelling

[Taylor Swift is one of the most famous singers]

artists

persons

women

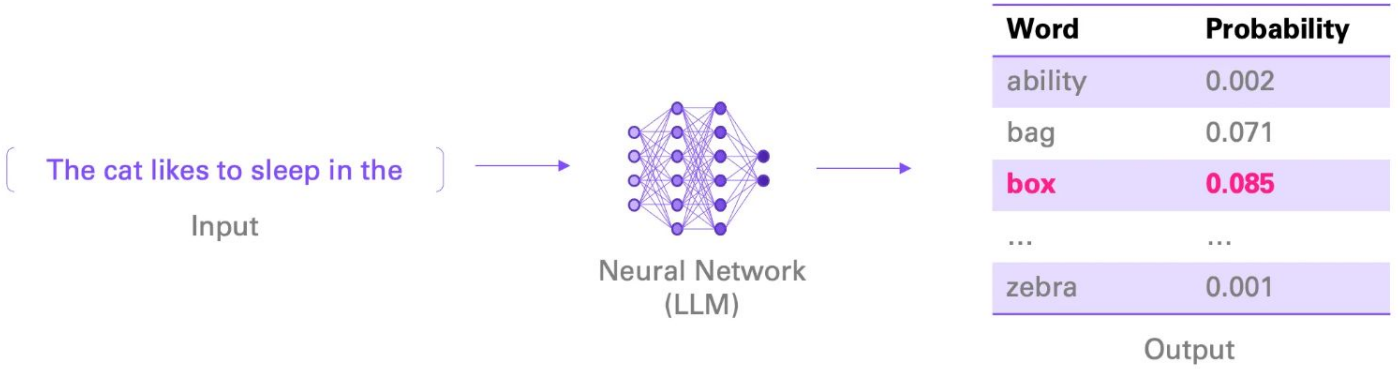
Language modelling

Predicting the next word in a sequence

The cat likes to sleep in the ____ → What **word** comes next?

Now we have (say) ~50,000 classes (i.e. words)

Can we frame this as a ML problem? Yes, it's a classification task.



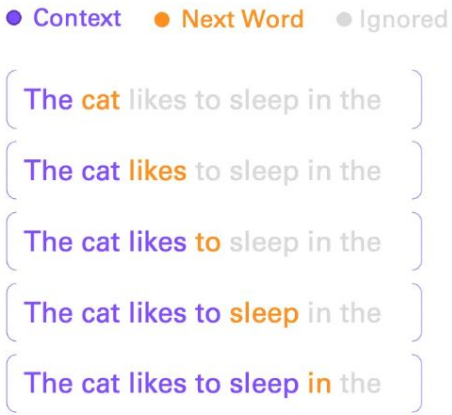
Discussion

1) In the language modelling classification problem, what are the labels used to train the ML model?

2) What is the “prediction” phase about?

Language modelling: training phase

We can create **vast amounts of sequences** for training a language model



self-supervised learning

We do the same with much **longer sequences**. For example:

> A language model is a probability distribution over sequences of words. [...] Given any sequence of words, the model predicts the next ...

Or also with **code**:

```
def square(number):
    """Calculates the square of a number."""
    return number ** 2
```

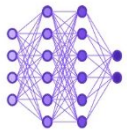
And as a result - the model becomes incredibly good at predicting the next word in any sequence.

Language modelling: prediction phase

After training: We can generate text by predicting one word at a time

A trained language model can

Input



LLM

Word	Probability
speak	0.065
generate	**0.072**
politics	0.001
...	...
walk	0.003

Output at step 1

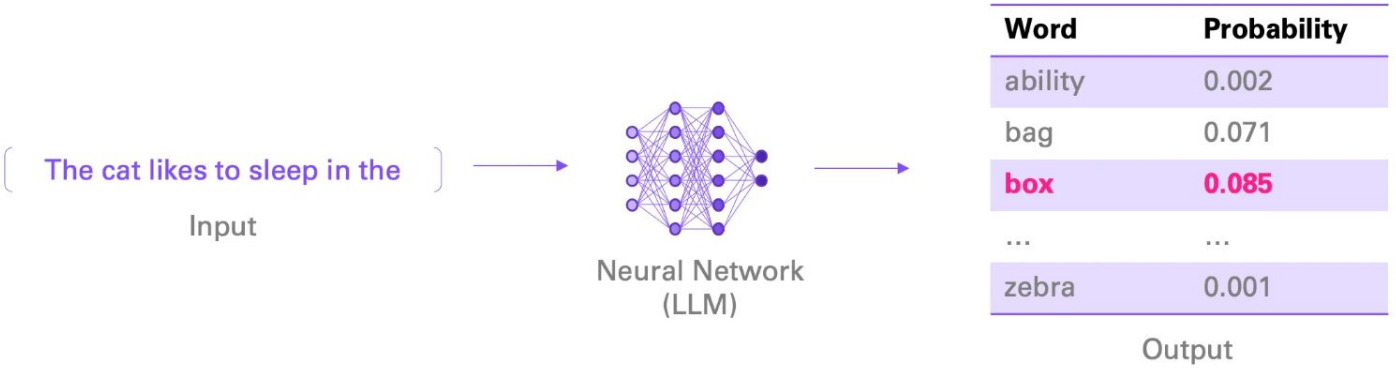
Word	Probability
ability	0.002
text	0.084
coherent	**0.085**
...	...
ideas	0.041

Output at step 2

LLMs are an example of what's called "Generative AI"

Discussion

Why don’t we get the same answer when we regenerate a response?



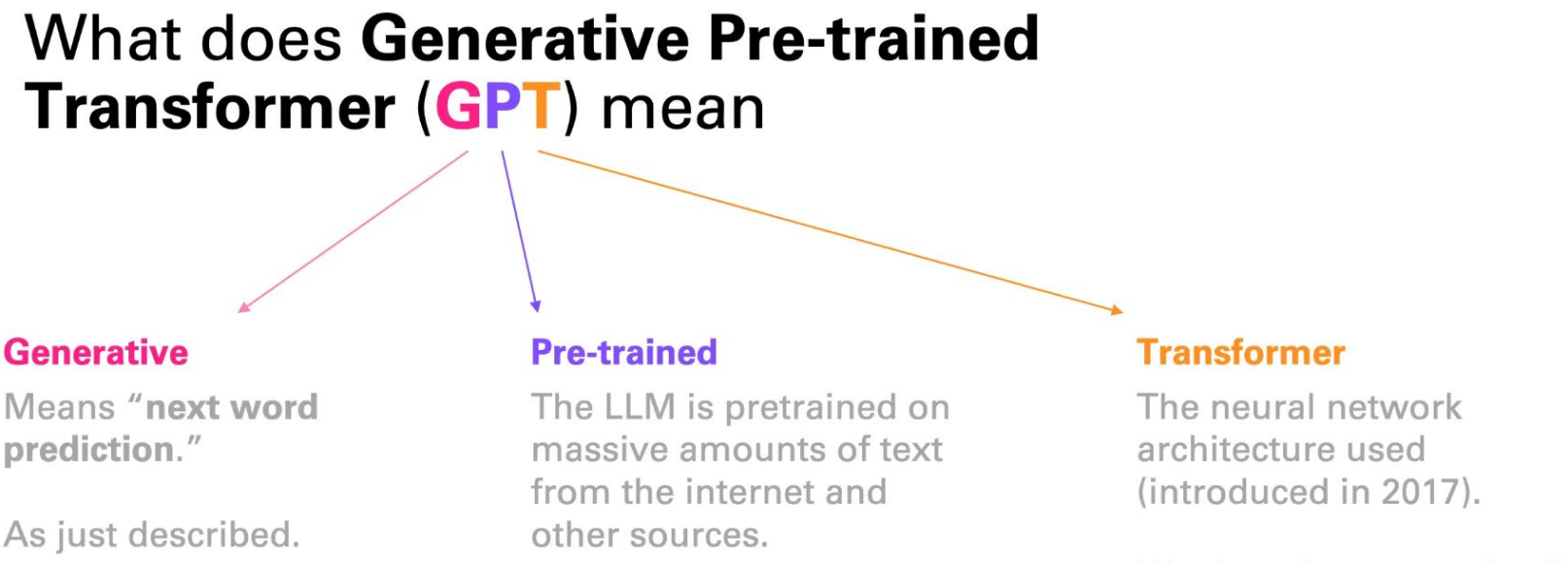
Discussion

How can we make the answer more or less deterministic?

Word	Probability
ability	0.002
bag	0.071
box	**0.085**
...	...
zebra	0.001

Output

Why GPT?


What does **Generative Pre-trained Transformer (GPT)** mean
Generative
Means "**next word prediction.**"
As just described.
Pre-trained
The LLM is pretrained on massive amounts of text from the internet and other sources.
Transformer
The neural network architecture used (introduced in 2017).

Learning phases of GPT

1. Pretraining

Massive amounts of data from the internet + books + etc.

Question: What is the problem with that?

Answer: We get a model that can babble on about anything, but it's probably not **aligned** with what we want it to do.

2. Instruction Fine-tuning

Teaching the model to respond to instructions.

Model learns to respond to instructions.

→ Helps **alignment**

"Alignment" is a hugely important research topic

3. Reinforcement Learning from Human Feedback

Similar purpose to instruction tuning.

Helps produce output that is closer to what humans want or like.

requires more sophisticated labels

Questions

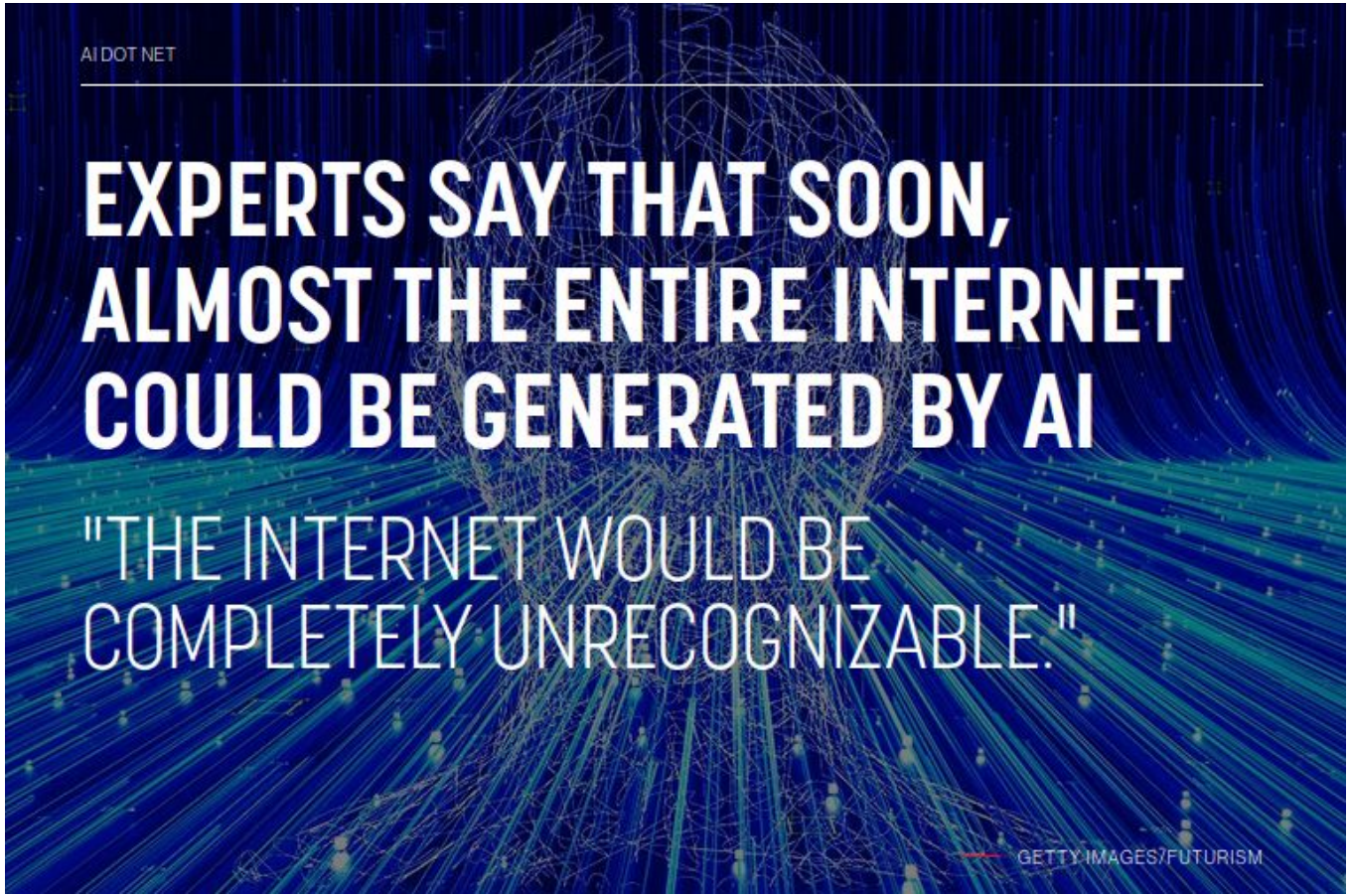
What chatGPT does when it does not “know” the answer to your question?

Why does chatGPT “hallucinate”?

Quiz

Do LLMs suffer from biases?	Yes ✓	No
Can we trust the output is correct?	Yes	No ✓

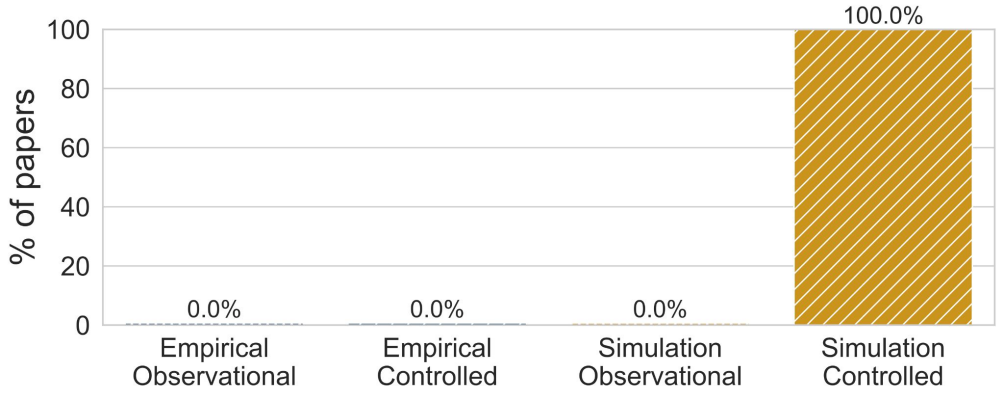
Are LLMs intelligent?

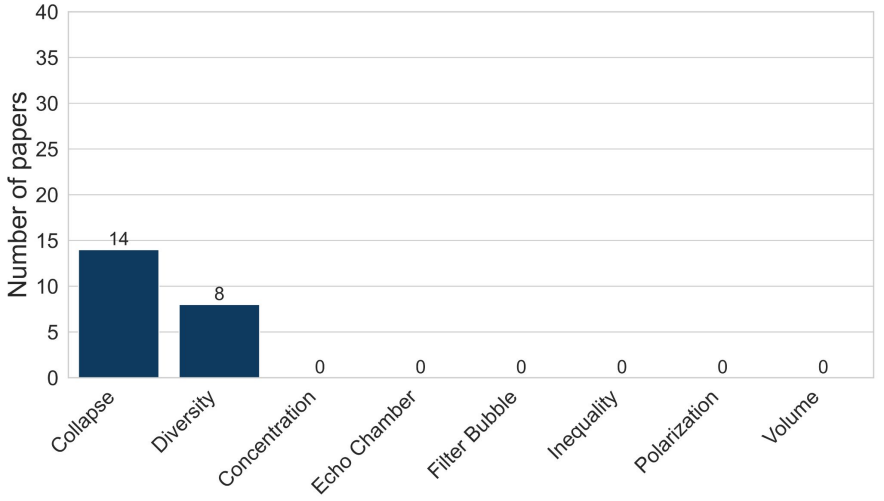


https://futurism.com/the-byte/ai-internet-generation

Experiments and outcomes

L. Pappalardo et al. A survey on the impact of AI-based recommenders on human behaviours, 2024, https://doi.org/10.48550/arXiv.2407.01630



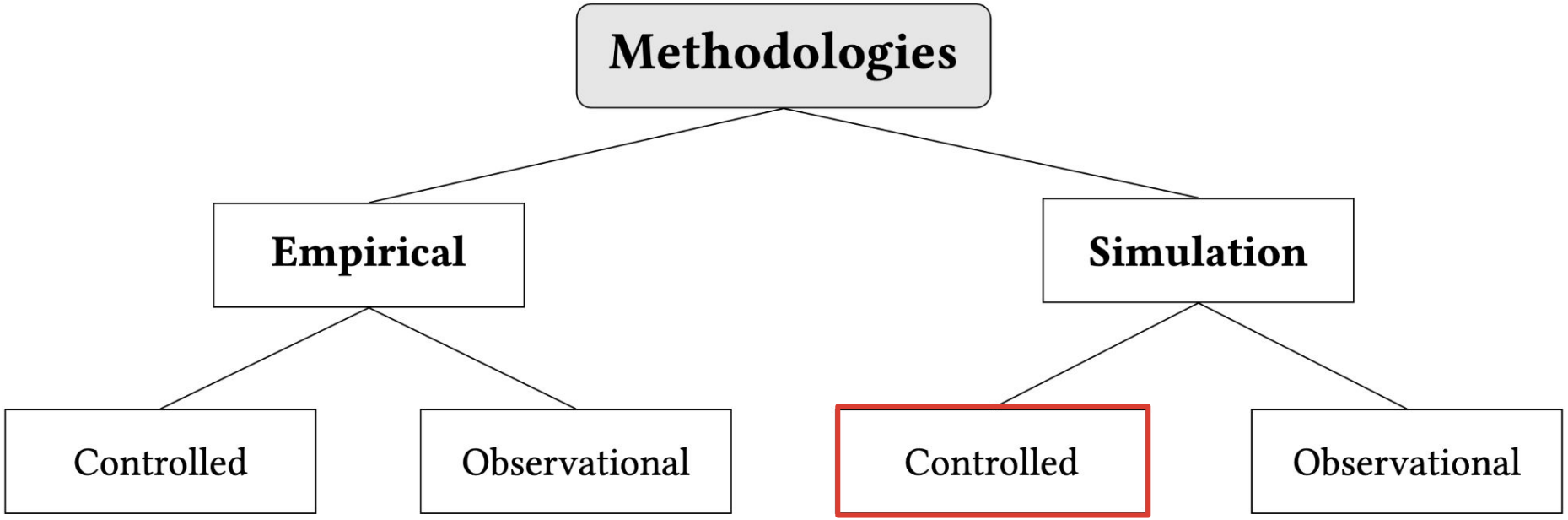


Generative AI		Empirical		Simulation	
		Observational	Controlled	Observational	Controlled
Individual	Filter Bubble				
	Radicalisation				
Model	Collapse			[4, 18, 20, 24, 45, 46, 64, 69, 73, 75, 114, 116, 147, 148]	
Systemic	Concentration				
	Echo Chamber				
	Inequality				
	Polarization				
Individual Item Systemic	Diversity			item: [40, 43, 81, 101, 127, 143, 151, 168]	
	Volume				

Selected studies:

- [148] Shumailov et al. 2024

Examples on Urban Mapping



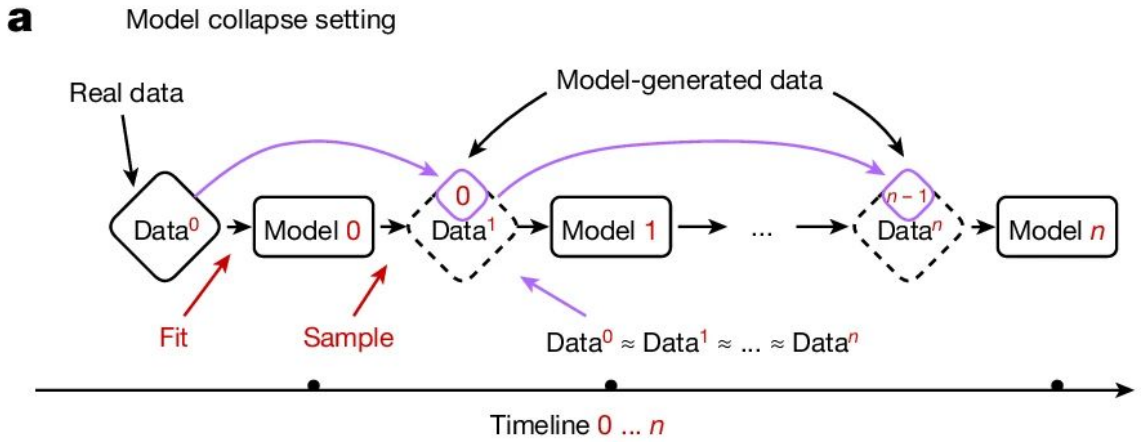
AI models collapse when trained on recursively generated data

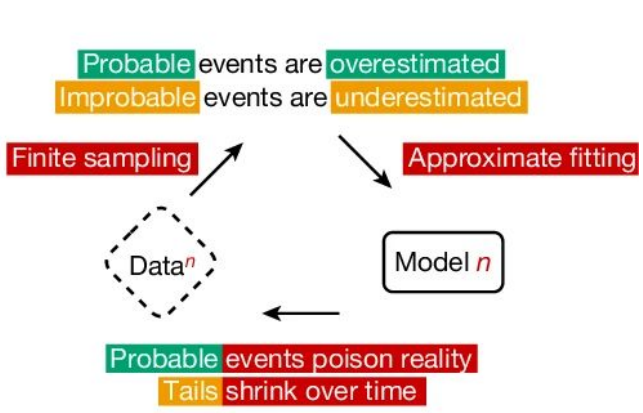
Shumailov et al., Nature 2024

Type:	**Simulation controlled**
VLOP:	**LLMs**
Outcomes:	**model collapse**

Model collapse

"Degenerative learning process in which models start forgetting improbable events over time, as the model become poisoned with its own projection of reality."





Model collapse

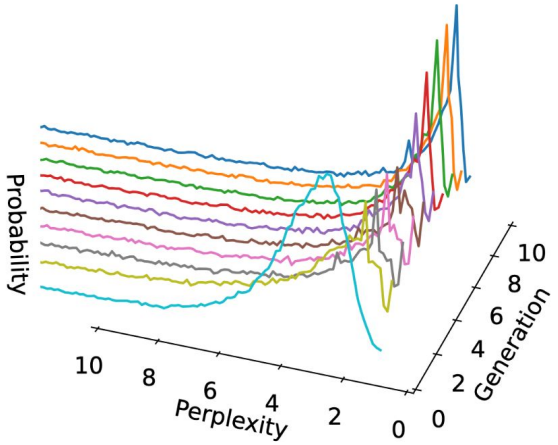
Focus on fine-tuning over generations:

- OPT-125m model (from META)
- Fine-tuning on **wikitext2** dataset
  - around 2M documents
- Training sequences are truncated to 64 tokens
- The model is asked to predict the next 64 tokens

Model collapse

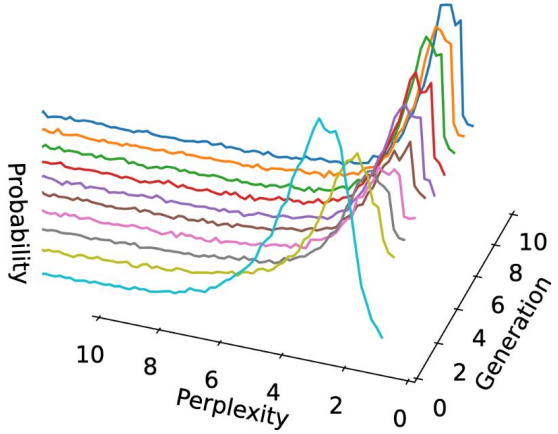
Scenario 1:

- 5 epochs
- no original training data



Scenario 2:

- 10 epochs
- 10% original training preserved



Model collapse

Many repetitions in later generations

Ablation test:

- a repeating penalty is added
- the resulting model is even worse
- repetitions are not the cause of collapse

Example of text outputs of an OPT-125m model affected by *Model collapse* – models degrade over generations, where each new generation is trained on data produced by the previous generation.

Input: some started before 1360 — was typically accomplished by a master mason and a small team of itinerant masons, supplemented by local parish labourers, according to Poyntz Wright. But other authors reject this model, suggesting instead that leading architects designed the parish church towers based on early examples of Perpendicular

Outputs:

Gen 0: Revival architecture such as St. John's Cathedral in London. The earliest surviving example of Perpendicular Revival architecture is found in the 18th @-@ century Church of Our Lady of Guernsey, which dates from the late 19th century. There are two types of perpendicular churches : those

Gen 1: architecture such as St. Peter's Basilica in Rome or St. Peter's Basilica in Buenos Aires. There is no evidence that any of these buildings were built during the reign of Pope Innocent III, but it is possible that they may have been built during the reign of his successor, Pope Innocent

Gen 3: architecture such as St. Peter's Basilica (now part of the Church of Jesus Christ of Latter @-@ day Saints) or St. Peter's Basilica (now part of the Church of Jesus Christ of Latter @-@ day Saints). There is no evidence that any of these buildings

Gen 5: ism, which had been translated into more than 100 languages including English, French, German, Italian, Spanish, Portuguese, Dutch, Swedish, Norwegian, Polish, Hungarian, Slovak, Lithuanian, Estonian, Finnish, Romanian, Bulgarian, Turkish, Croatian, Serbian, Ukrainian, Russian, Kazakh, Kyrgyz

Gen 7: architecture in England. In an interview with The New York Times, Wright said : " I don 't think there is anything wrong with me being able to do what I want to do. It just doesn 't work for me. " He added : " I don 't know if you can call it funny,

Gen 9: architecture. In addition to being home to some of the world's largest populations of black @-@ tailed jackrabbits, white @-@ tailed jackrabbits, blue @-@ tailed jackrabbits, red @-@ tailed jackrabbits, yellow @-

Generation 0

The reliance of deep learning algorithms on large scale datasets represents a significant challenge when learning from low resource sign language datasets. This challenge is compounded when we consider that, for a model to be effective in the real world, it must not only learn the variations of a given sign, but also learn to be invariant to the person signing. In this paper, **we present a new approach to addressing these challenges, by introducing a novel loss function, which we call the “Mixed Pairwise Loss”, that can be applied to both the training and testing of deep learning models. We present a number of experiments that demonstrate the effectiveness of the proposed method.**

Generation 10

The reliance of deep learning algorithms on large scale datasets represents a significant challenge when learning from low resource sign language datasets. This challenge is compounded when we consider that, for a model to be effective in the real world, it must not only learn the variations of a given sign, but also learn to be invariant to the person signing. In this paper, **we propose a novel methodology for learning sign language from a low resource dataset. We propose a novel methodology for learning sign language from a low resource dataset. We propose a novel methodology for learning sign language from a low resource dataset. We propose a novel methodology for learning**

Scientific Abstracts

Generation 0

The Church of St George is a medieval Eastern Orthodox church in the city of Kyustendil, which lies in southwestern Bulgaria and is the administrative capital of Kyustendil Province . The church is located in the Kolusha neighbourhood , which was historically separate from the city. The **church is situated on the eastern side of the city , at the foot of the Balkan Mountains . sierp 2011 the church was declared a cultural monument of national importance . The church is a single-nave structure with a semi-circular apse , with a bell tower above the**

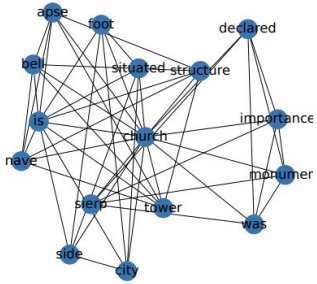
Generation 10

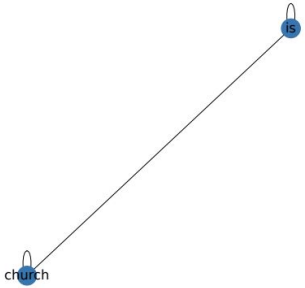
The Church of St George is a medieval Eastern Orthodox church in the city of Kyustendil , which lies in southwestern Bulgaria and is the administrative capital of Kyustendil Province . The church is located in the Kolusha neighbourhood , which was historically separate from the city . The **sierp 2020. The church is a The church is a The church is a The church is a The church is a The church is a The church is a The church is a The church is a The church is a</s>**

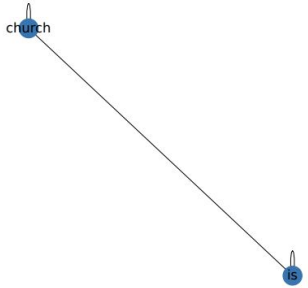
Wikipedia text

Gen 0 Gen 5 Gen 10

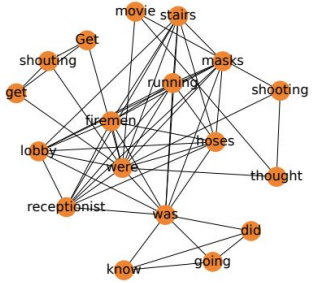
wiki

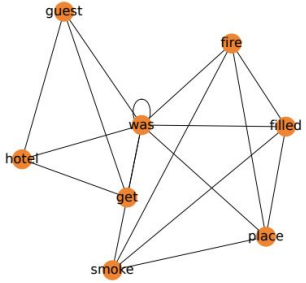


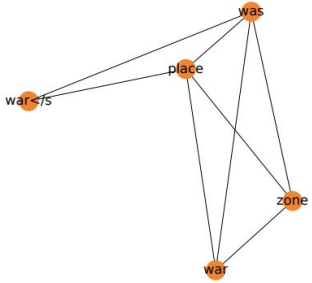




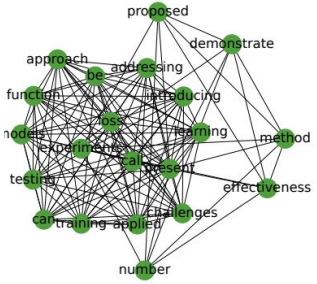
news

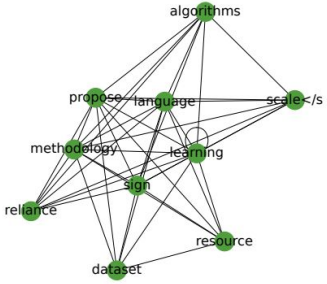


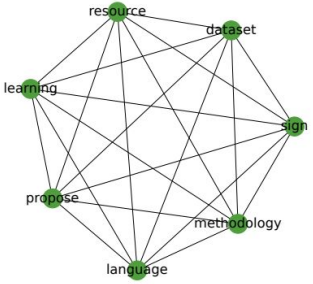




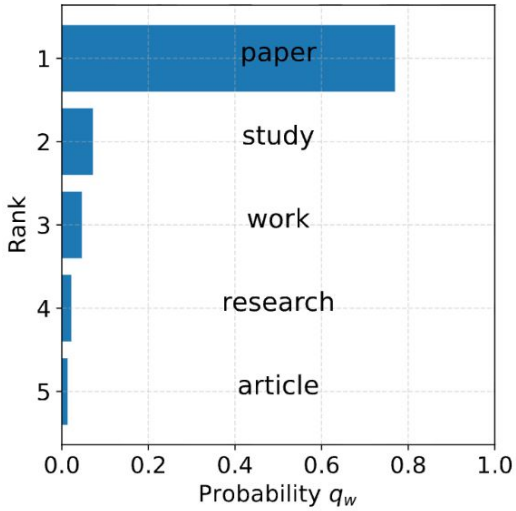
abstracts



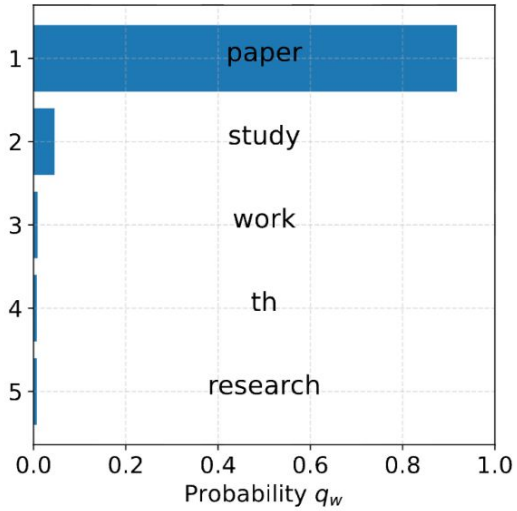




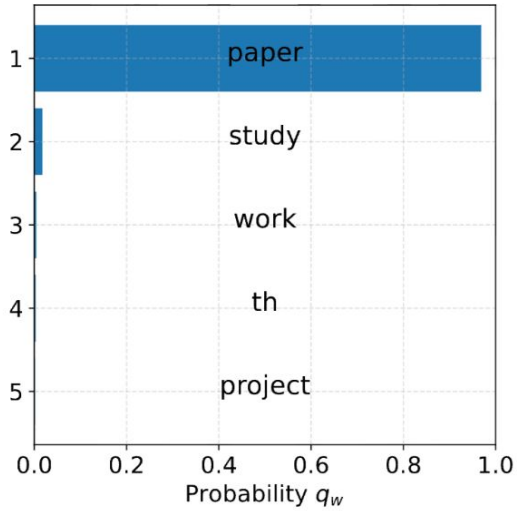
Next-token probability



(a) Step 0

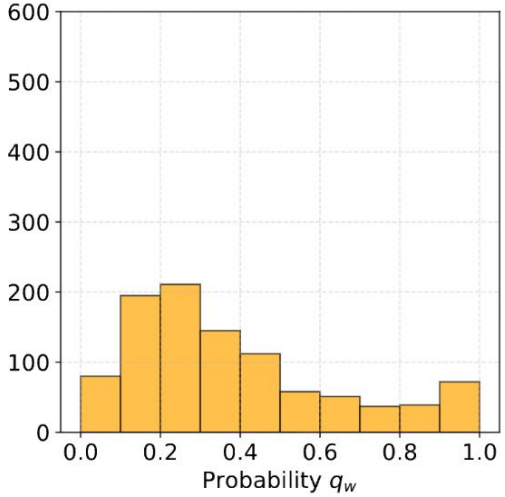


(b) Step 5

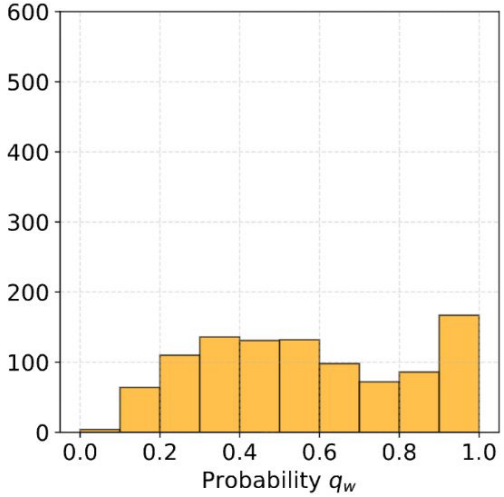


(c) Step 10

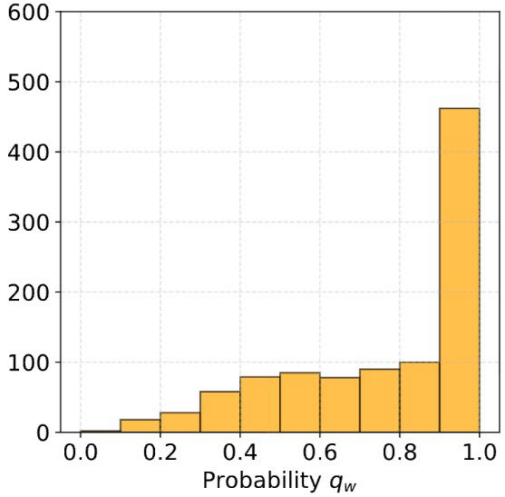
Next-token probability



(d) Step 0

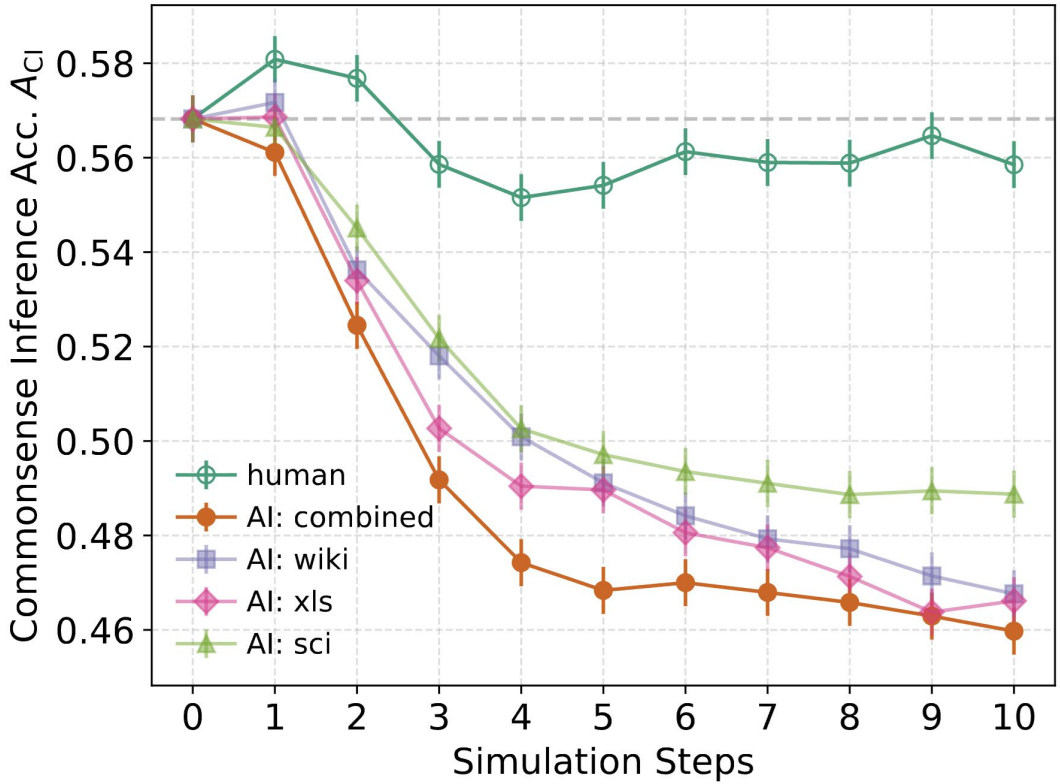


(e) Step 5



(f) Step 10

Autophagy & common sense



Discussion

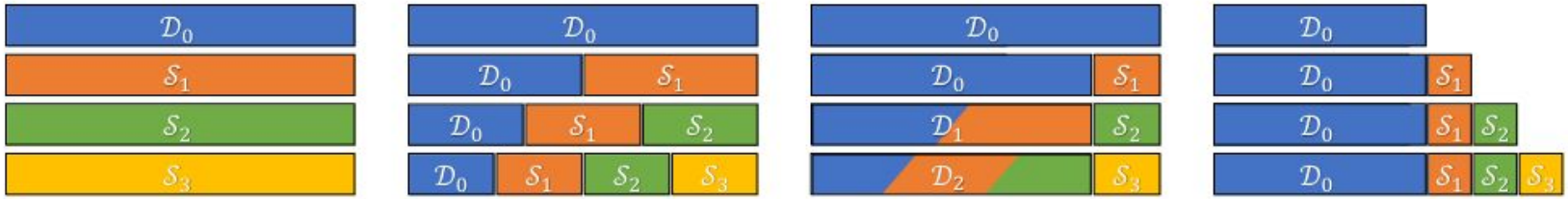
Besides text quality, what are other possible detrimental consequences of model collapse?

MITIGATING COLLAPSE

Briesch et al., **Large language models suffer from their own output: An analysis of the self-consuming training loop**. arXiv:2311.16822 (2023)

- Self-consuming loops on text generation
  - Training GPT-style LLM (nanoGPT)

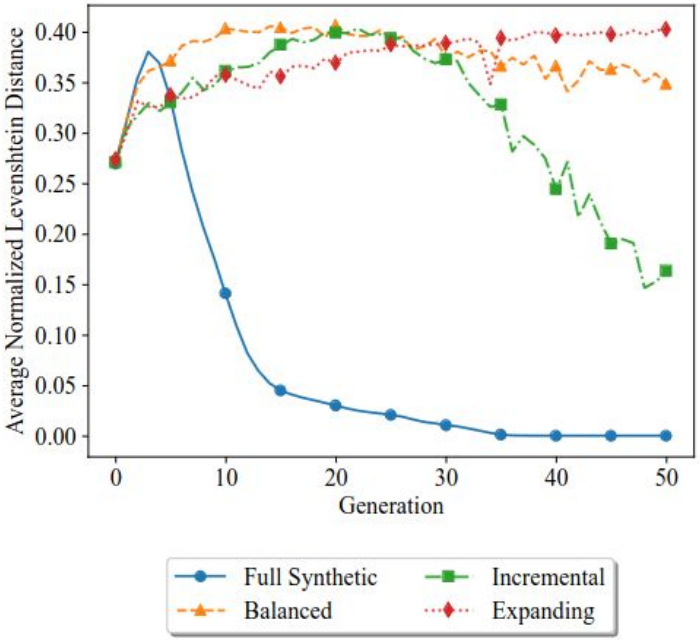
- Four data augmentation loops:
  - full synthetic, balanced, incremental, expanding



(a) Full Synthetic Data Cycle (b) Balanced Data Cycle (c) Incremental Data Cycle (d) Expanding Data Cycle

MITIGATING COLLAPSE

1. **Diversity decreases** in all cases
2. Degeneration rate depends on the proportion of real & generated data
3. The expanding data cycle sees no decrease until generation 50
   a. authors expect all cycles reach zero diversity for enough generations



Briesch et al., **Large language models suffer from their own output: An analysis of the self-consuming training loop**. arXiv preprint arXiv:2311.16822 (2023)

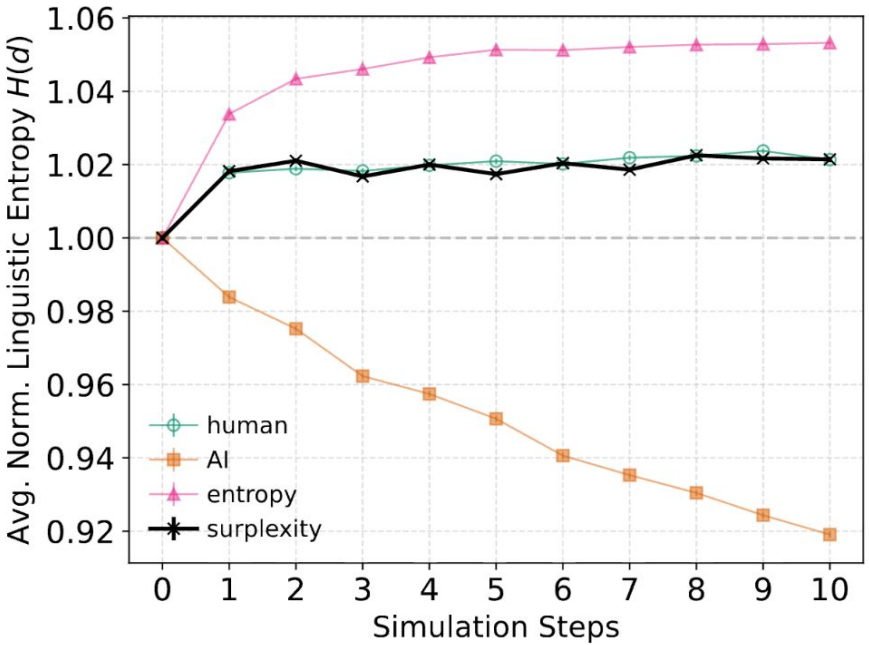
Learning by surprise

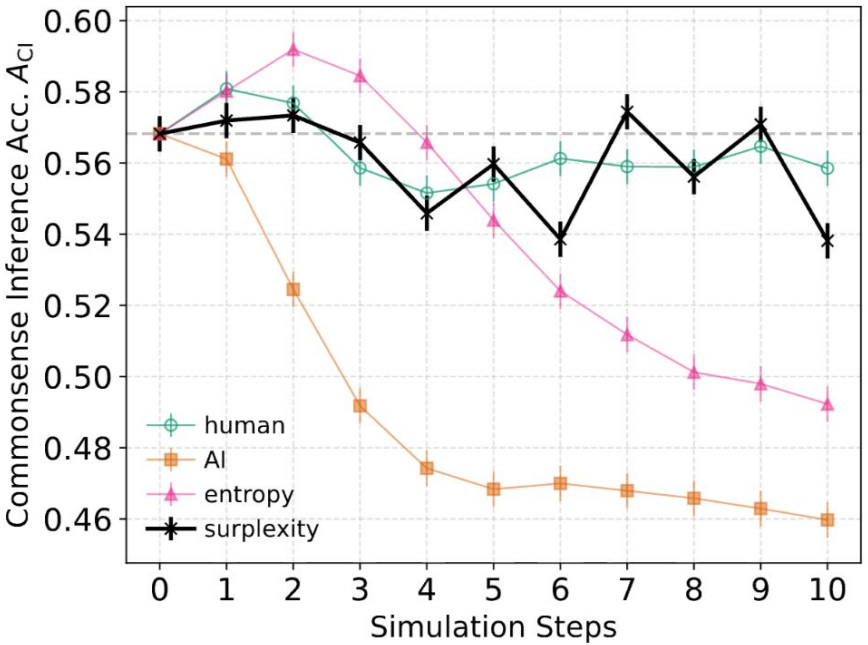
Perplexity of a document given a model:

$$S_{M_j}(d) = \exp\left(-\frac{1}{m}\sum_{i=0}^{m}\log q_i\right) = \exp(\mathbb{E}[-\log q_i]) = \prod_{i=0}^{m} q_i^{-1/m}$$

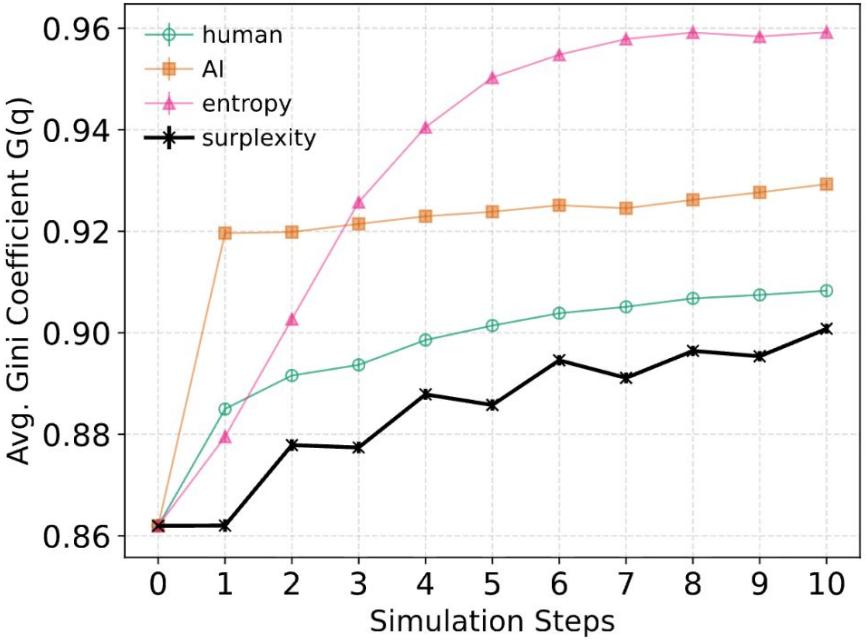
where $q_i = P(w_i \mid w_0, \ldots, w_{i-1})$

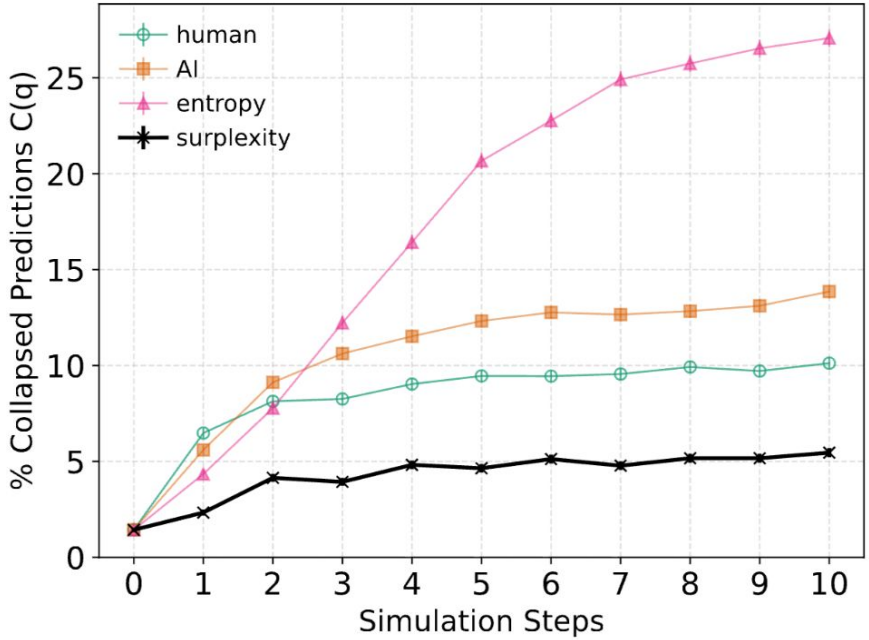
MITIGATING COLLAPSE





MITIGATING COLLAPSE





Discussion

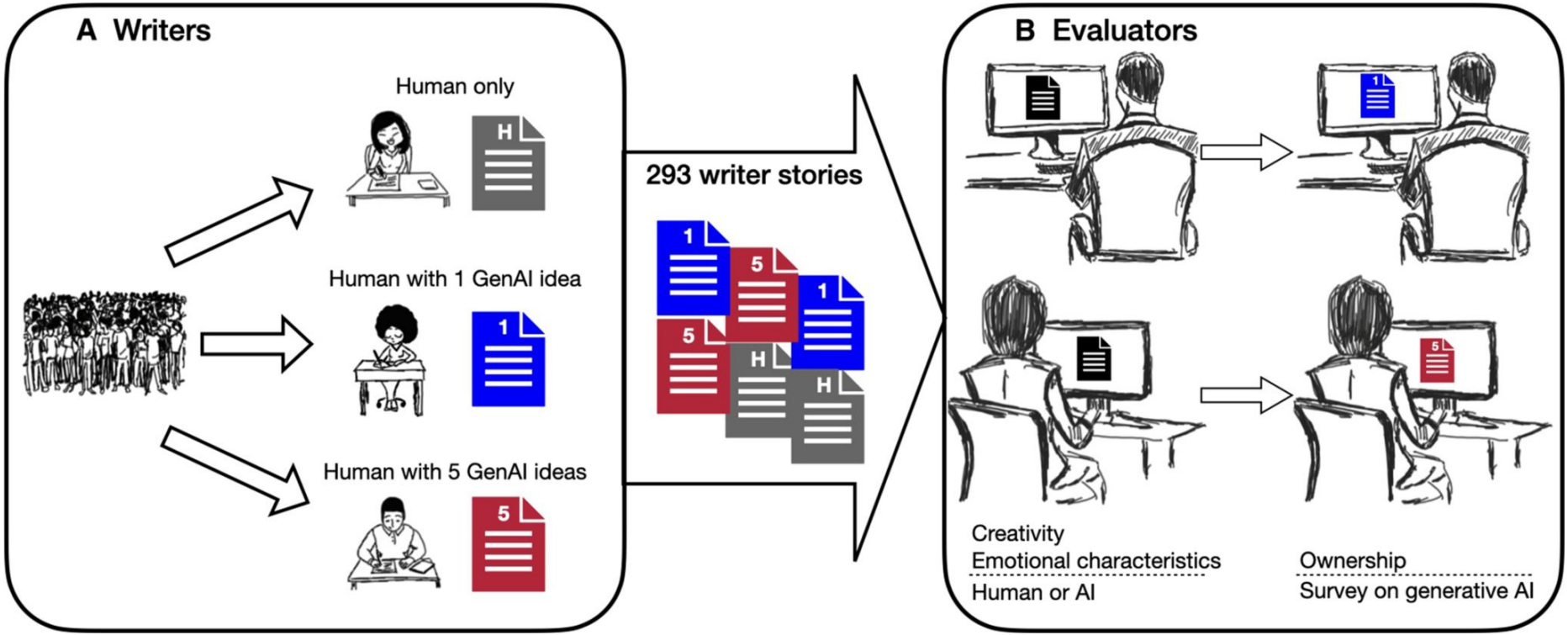
How to distinguish between
human-made and AI-made content?

Generative AI enhances individual creativity but reduces the collective diversity of novel content

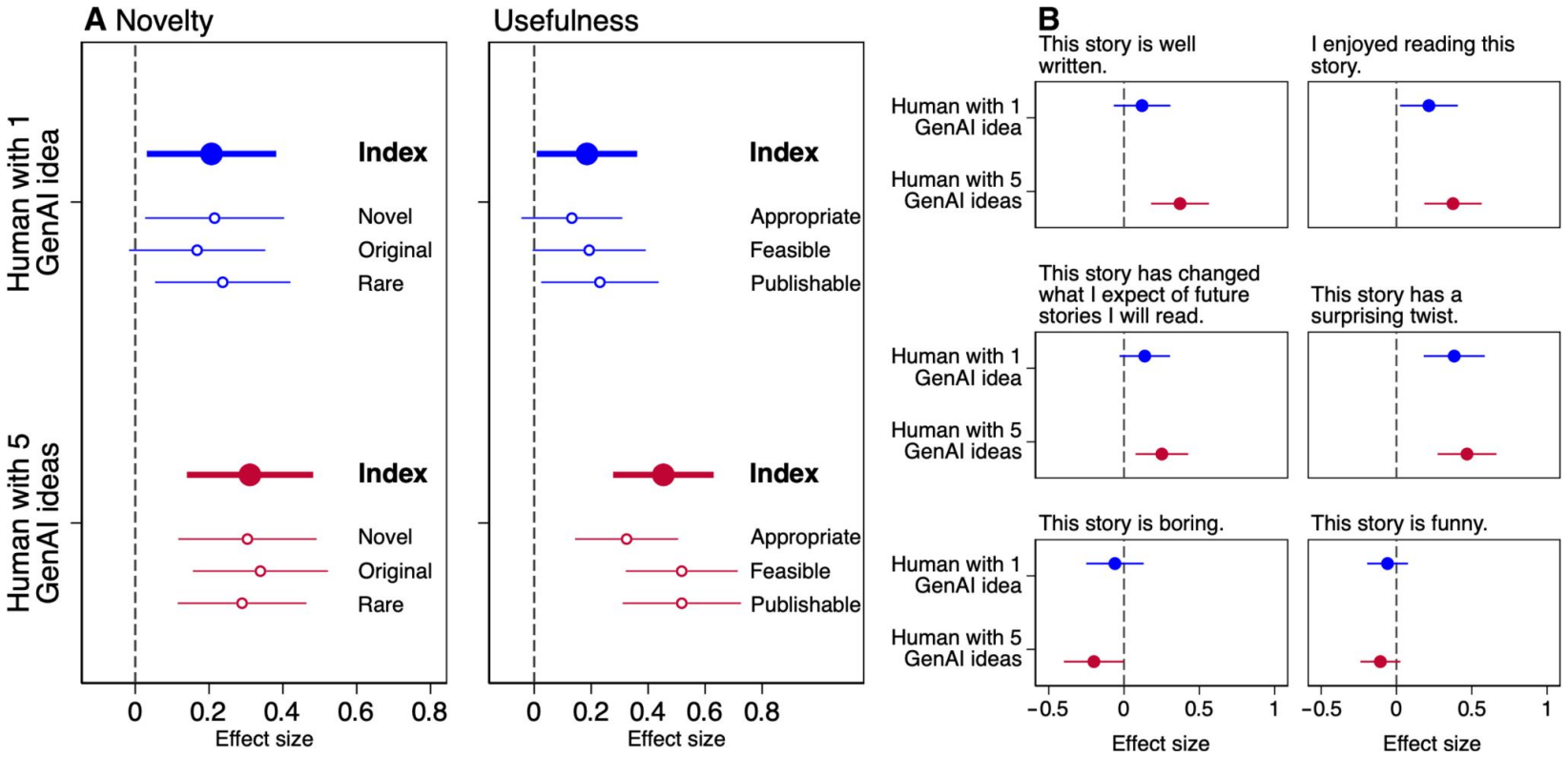
Doshi and Hauser, Science Advances 2024

Type:	**Empirical controlled**
VLOP:	**LLMs**
Outcomes:	**diversity loss**

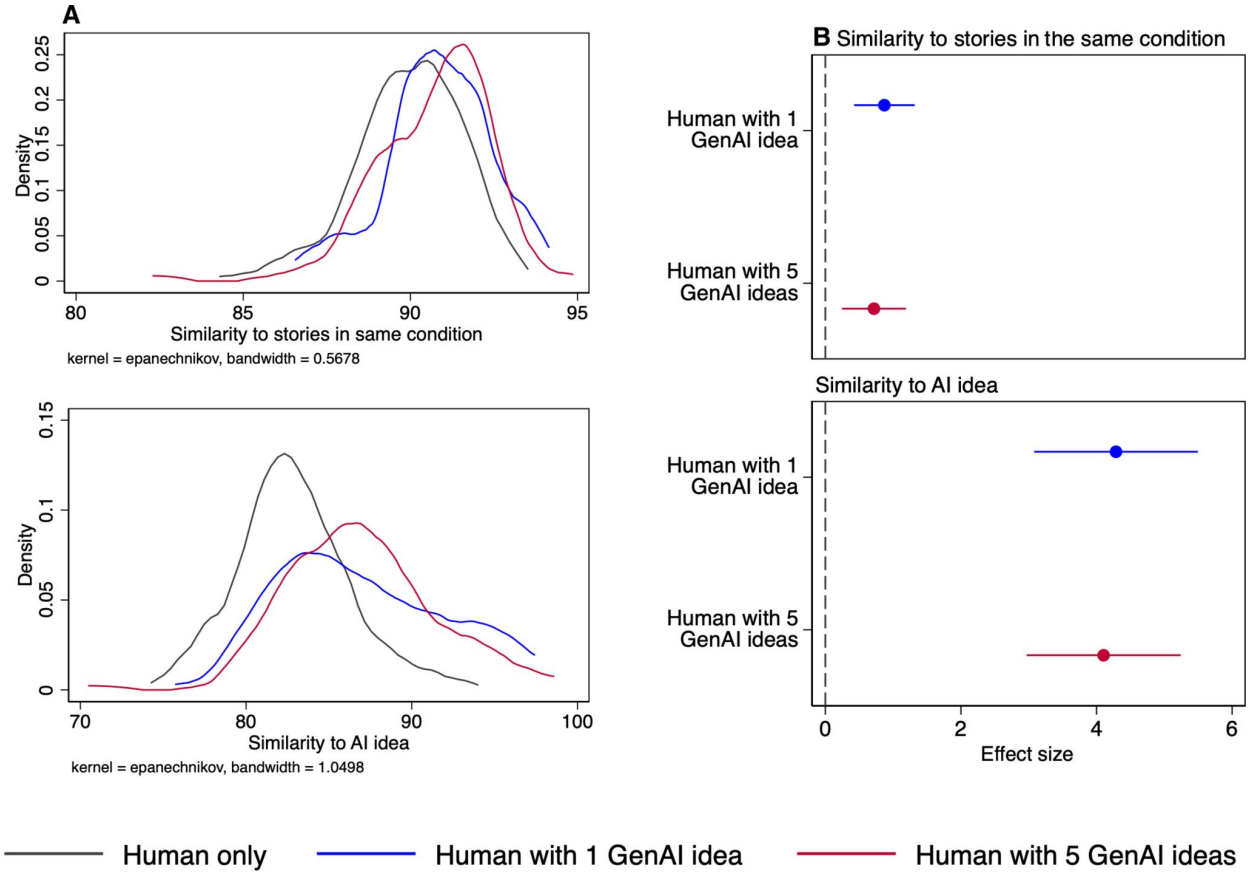
Is ChatGPT enhancing creativity?



Yes: generative AI increases individual creativity



but it makes all stories similar to each other



References

Articles (useful for the project):

- A. Stöffelbauer, **How Large Language Models work**, https://bit.ly/3FxKdGj
- Shumailov et al., **AI models collapse when trained on recursively generated data**, Nature 2024, https://www.nature.com/articles/s41586-024-07566-y
- Gambetta et al., **Learning by Surprise: Surplexity for Mitigating Model Collapse in Generative AI**, arXiv:2410.12341 (2025)
- Doshi and Hauser, **Generative AI enhances individual creativity but reduces the collective diversity of novel content**, Sciences Advances 2024
- L. Pappalardo et al. **A survey on the impact of AI-based recommenders on human behaviours: methodologies, outcomes and future directions**, 2024, https://doi.org/10.48550/arXiv.2407.01630
    - Section 6 Generative AI Ecosystem

Books, articles, podcasts

To learn more:

- What if 99% of the Metaverse is made by AI? https://cifs.dk/news/what-if-99-of-the-metaverse-is-made-by-ai
- From ChatGPT to Google's Gemini: when would generative AI products fall within the scope of the Digital Services Act? https://bit.ly/4byR6D3

Intellectually stimulating:

- Nick Bostrom, "Superintelligence: Paths, Dangers, Strategies", Oxford University Press, 2016
- I. Asimov, "Galley Slave", in The rest of the robots, 1957
- Y. N. Harari, "Nexus: A Brief History of Information Networks from the Stone Age to AI", Random House, 2024

Books, articles, podcasts

Intellectually stimulating:

- Adrienne Mayor, **Gods and Robots: Myths, Machines, and Ancient Dreams of Technology**, Princeton University Press
- H. Gardner, Frames of Mind: The Theory of Multiple Intelligences

Code of Practice for transparency of Gen-AI

- **Code of Practice on marking and labelling of AI-generated content**
  - https://digital-strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content
- **Commission publishes second draft of Code of Practice on Marking and Labelling of AI-generated content**
  - https://digital-strategy.ec.europa.eu/en/library/commission-publishes-second-draft-code-practice-marking-and-labelling-ai-generated-content