



Large Language Models + Giant AI Experiments = A Whole New World?

Jennifer Renoux

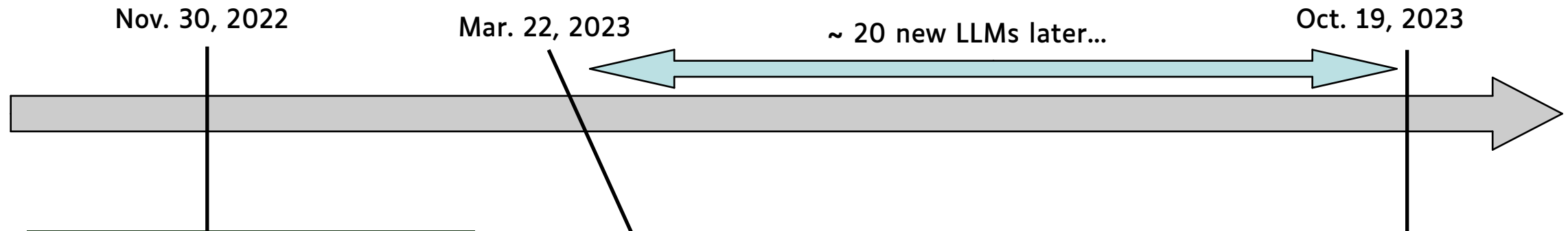
Center for Applied Autonomous Sensor Systems
Örebro University

Who am I?

- Senior Researcher in Human-Machine Communication at the Cognitive Robotics System Lab, Örebro University
- Lecturer in Ethics of AI and Robotics (under-graduate and graduate courses)



Background



Introducing ChatGPT

We've trained a model called ChatGPT which interacts in a conversational way. The dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.

[Try ChatGPT ↗](#) [Read about ChatGPT Plus](#)

<https://openai.com/blog/chatgpt>

← All Open Letters

Pause Giant AI Experiments: An Open Letter

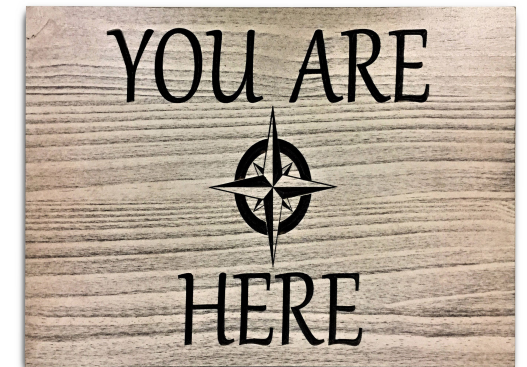
We call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.

Signatures
33711

[Add your signature](#)

Published
March 22, 2023

<https://futureoflife.org/open-letter/pause-giant-ai-experiments/>



Background - The Future of Life Institute Letter

- "AI systems with human-competitive intelligence can pose profound risks to society and humanity"
- Call for a 6 month pause on "the training of AI systems more powerful than GPT-4"
- Call for the creation of regulatory bodies for AI development

Background - My opinion on the letter

- I did **not** sign it, though I was torn
- I agree with the need to stop the "AI Race" that is ongoing in tech leading companies and to install regulations
- But for completely different reasons
- I believe the letter is detrimental for actual constructive discussions on AI and ethics ("existential risk" and longtermism phrasing, "Critic-Hype")
- The letter misuses some research on the ethics of AI (Bender et al. 2021)

WHAT IS ARTIFICIAL INTELLIGENCE?

What is AI?

Artificial intelligence (AI) is the **intelligence** of machines or software, as opposed to the intelligence of human beings or animals.

Wikipedia

artificial intelligence (AI), the ability of a digital **computer** or computer-controlled **robot** to perform tasks commonly associated with intelligent beings.

Encyclopaedia Britannica

Q. What is artificial intelligence?

A. It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.

McCarthy, 2004

What is AI?

Artificial intelligence (AI) is the **intelligence** of machines or software, as opposed to the intelligence of human beings or animals.

Wikipedia

artificial intelligence (AI), the ability of a digital **computer** or computer-controlled **robot** to **perform tasks** commonly associated with intelligent beings.

Encyclopaedia Britannica

Q. What is artificial intelligence?

A. It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.

McCarthy, 2004

What is AI?

Artificial intelligence (AI) is the **intelligence of machines** or software, as opposed to the intelligence of human beings or animals.

Wikipedia

artificial intelligence (AI), the ability of a digital **computer** or computer-controlled **robot** to **perform tasks** commonly associated with intelligent beings.

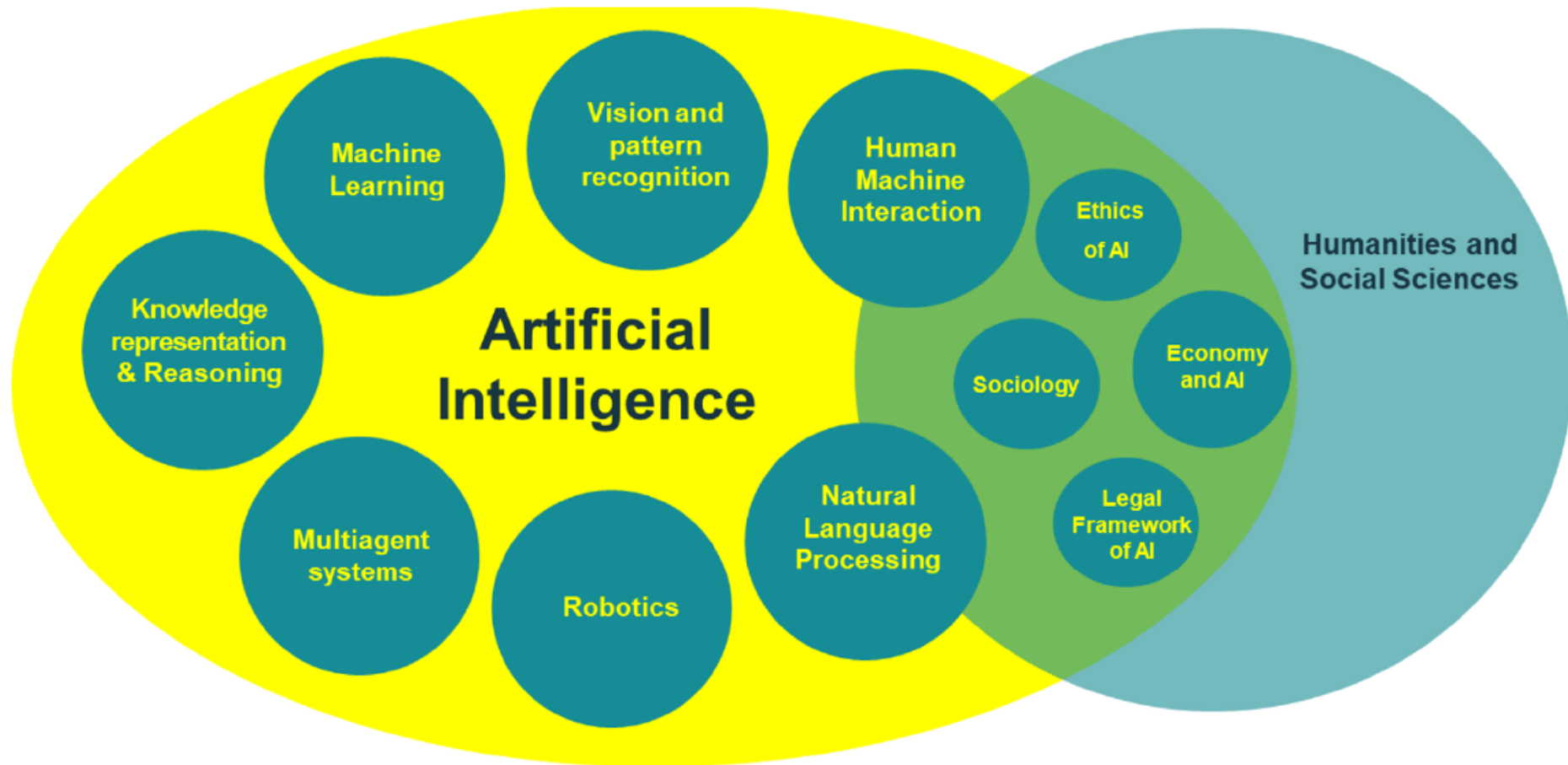
Encyclopaedia Britannica

Q. What is artificial intelligence?


A. It is the science and engineering of making **intelligent machines**, especially intelligent computer programs. It is related to the similar task of **using computers to understand human intelligence**, but AI does not have to confine itself to methods that are biologically observable.

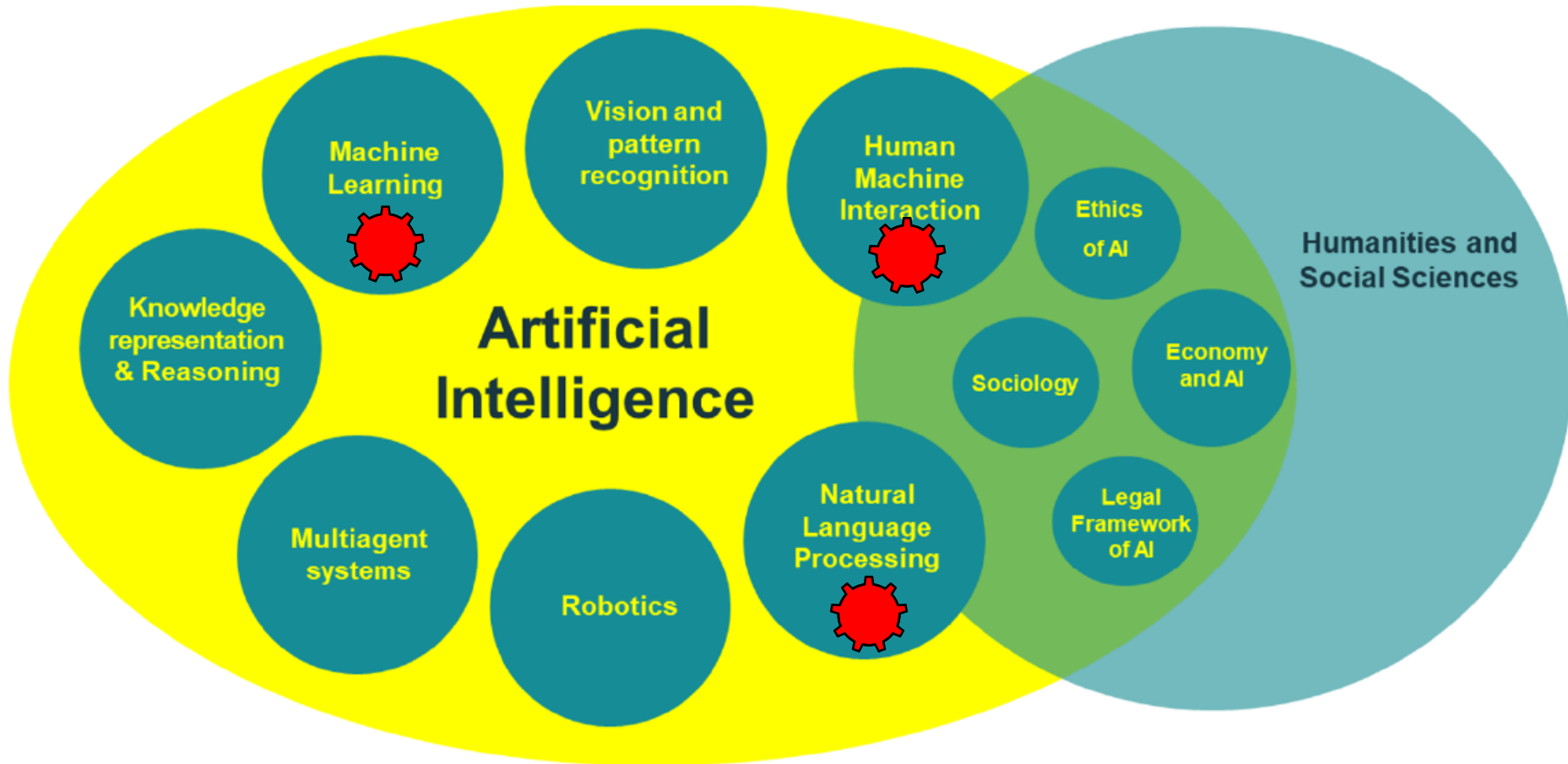
McCarthy, 2004

What is AI?



What is AI?

 LLMs are there

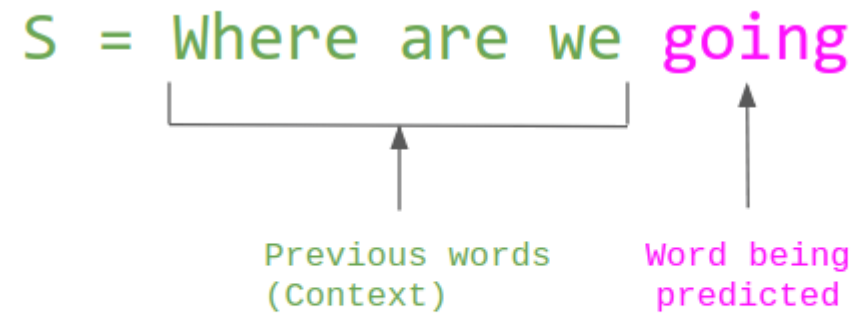


By BlueRadical, DevianArt



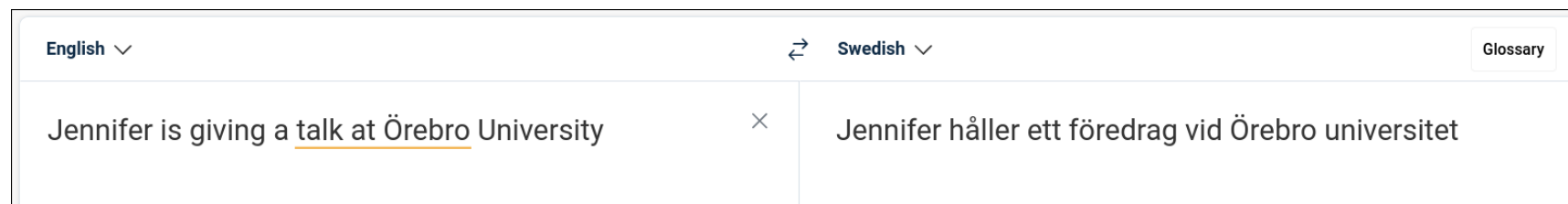
WHAT ARE (LARGE) LANGUAGE MODELS?

What are Language Models?



$$P(S) = P(\text{Where}) \times P(\text{are} \mid \text{Where}) \times P(\text{we} \mid \text{Where are}) \times P(\text{going} \mid \text{Where are we})$$

What are Language Models used for?



What are *Large Language Models*?

Language Models with millions of parameters and trained on a large amount of data

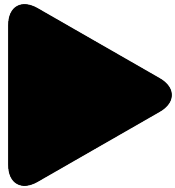
Are LLMs the New AI?

- Impact they will have is debated
- But general agreement is that they are not going to replace all of AI
- “General-purpose language *understanding* and generation.”
(Wikipedia)



BEYOND THE HYPE. A SHALLOW DIVE IN LLMs LIMITATIONS

Information retrieval



what are all the schools from Örebro University?

Web search ▶

Here are some of the schools that make up Örebro University:

- School of Humanities, Education and Social Sciences
- School of Science and Technology
- School of Health and Medical Sciences
- School of Business
- School of Law, Psychology and Social Work
- School of Nursing
- School of Music, Theatre and Art
- School of Informatics
- School of Hospitality, Culinary Arts and Meal Science
- School of Sports Science
- School of Environment, Energy and Building Technology
- School of Mathematical Sciences

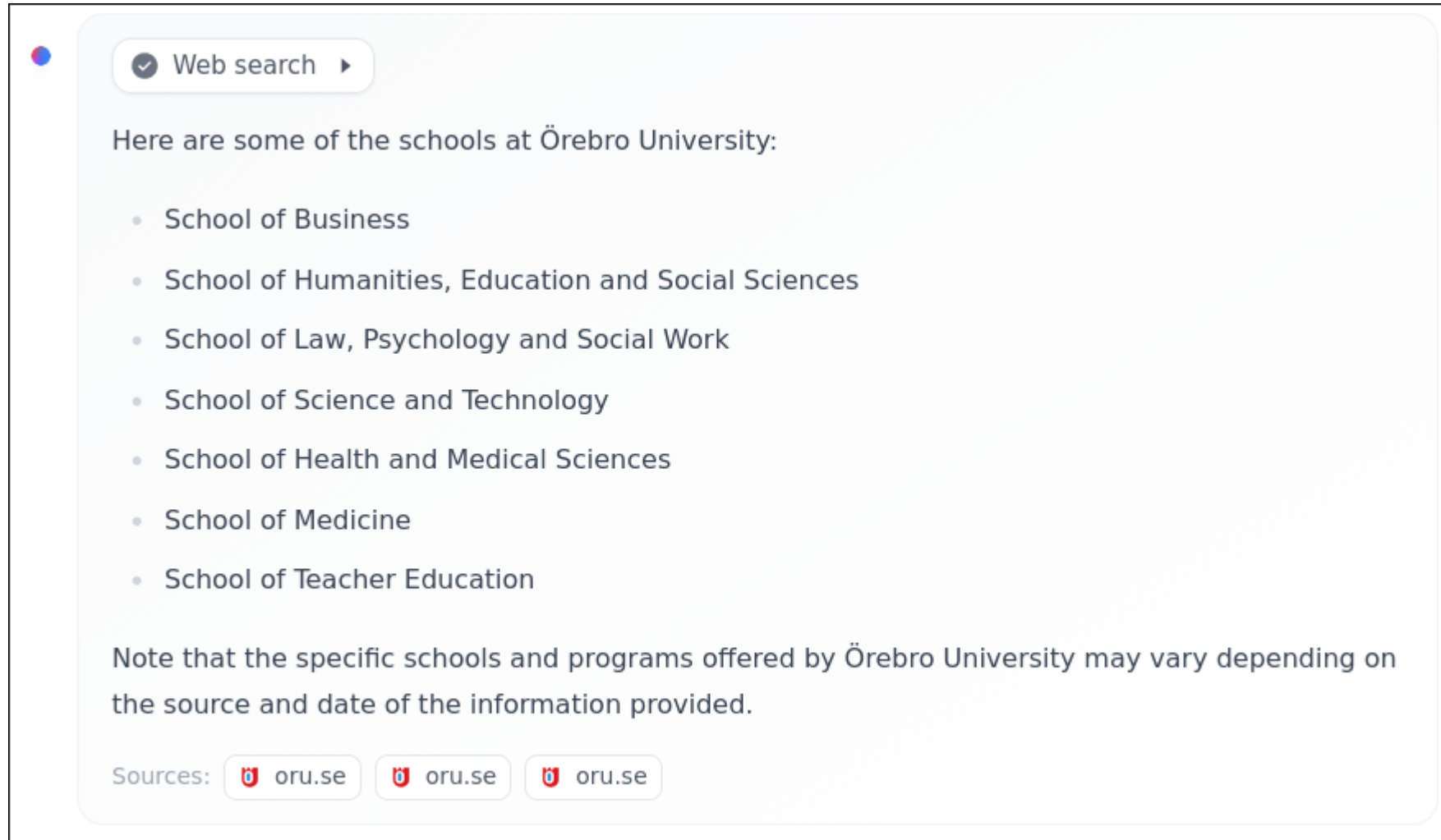
Search web ⓘ Economics

Ask anything ▶

Model: OpenAssistant/oasst-sft-6-llama-30b · Generated content may be inaccurate or false. [Share this conversation](#)

Attempt #1

Information retrieval






Web search

Here are some of the schools at Örebro University:

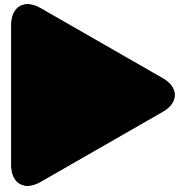
- School of Business
- School of Humanities, Education and Social Sciences
- School of Law, Psychology and Social Work
- School of Science and Technology
- School of Health and Medical Sciences
- School of Medicine
- School of Teacher Education

Note that the specific schools and programs offered by Örebro University may vary depending on the source and date of the information provided.

Sources:  oru.se  oru.se  oru.se

Attempt #2

LLMs and information retrieval



What are the three most influential papers of the Artificial Intelligence journal?

Web search

Based on the provided context, it appears that the Artificial Intelligence journal has not yet announced the winners of its annual awards for the year 2023. Therefore, we cannot determine the three most influential papers of the journal for the current year.

However, we can suggest some highly influential papers published in the Artificial Intelligence journal in recent years based on their impact factor, citations, and contributions to the field of artificial intelligence. Here are three examples:

1. "Deep Reinforcement Learning" by Mnih et al., published in 2013 (Volume 28, Issue 1): This paper introduced deep reinforcement learning, which combines deep neural networks with reinforcement learning algorithms to learn complex behaviors from raw sensory input. It has been cited over 2,300 times and has had a significant impact on the development of AI applications such as AlphaGo and autonomous driving.
2. "Generative Adversarial Networks" by Goodfellow et al., published in 2014 (Volume 31, Issue 2): This paper proposed generative adversarial networks (GANs), a new framework for training generative models. GANs have revolutionized the field of computer vision and have led to breakthroughs in image generation, video prediction, and style transfer. The paper has been cited over 1,800 times.
3. "Attention Is All You Need" by Vaswani et al., published in 2017 (Volume 40, Issue 1): This paper introduced the Transformer model, which replaces traditional recurrent neural network architectures with attention mechanisms. Attention allows the model to selectively focus on parts of the input sequence when generating output, leading to significant improvements in machine translation and other sequential data tasks. The paper has been

Search web

Ask anything

Model: meta-llama/llama-2-70b-chat-hf | Generated content may be inaccurate or false. Essential works published | Share this conversation

LLMs and information retrieval in the wild

New York lawyers sanctioned for using fake ChatGPT cases in legal brief

By Sara Merken

June 26, 2023 10:28 AM GMT+2 · Updated 2 months ago



ChatGPT is making up fake Guardian articles. Here's how we're responding

Chris Moran



The risks inherent in the technology, plus the speed of its take-up, demonstrate why it's so vital that we keep track of it

● Chris Moran is the Guardian's head of editorial innovation

Common sense

What is the third letter of the word ohvnb?

- The word "ohvnb" contains only four letters: o, h, v, and n. Therefore, there is no third letter in the word "ohvnb."

In the following sentence : "The ball does not fit in the suitcase because it is too small", which object is smaller?

- The ball is smaller than the suitcase.

The assistant refused to date the lawyer because she was married. In the previous sentence, who does "she" refer to?

- I apologize for any confusion caused by my response earlier. To clarify, "she" refers to the woman who works at the law firm where the assistant interned. She is not the same person as the lawyer that the assistant worked with on the case involving the man who died after falling off his roof while cleaning gutters.

LLMs and common sense in the wild

Supermarket AI meal planner app suggests recipe that would create chlorine gas

Pak 'n' Save's Savey Meal-bot cheerfully created unappealing recipes when customers experimented with non-grocery household items



📷 An app launched by a New Zealand supermarket that produces AI-generated recipes for leftovers has recommended cooks try 'bleach-infused rice surprise' among other things. Photograph: Jacobs Stock Photography Ltd/Getty Images

Pratar du svenska?

HuggingChat v0.6.0

Making the community's best AI chat models available to everyone.

NEW Websearch 2.0, now with RAG & sources [Discuss](#)

Current Model
meta-llama/Llama-2-70b-chat-hf

[Model page](#) [Website](#)

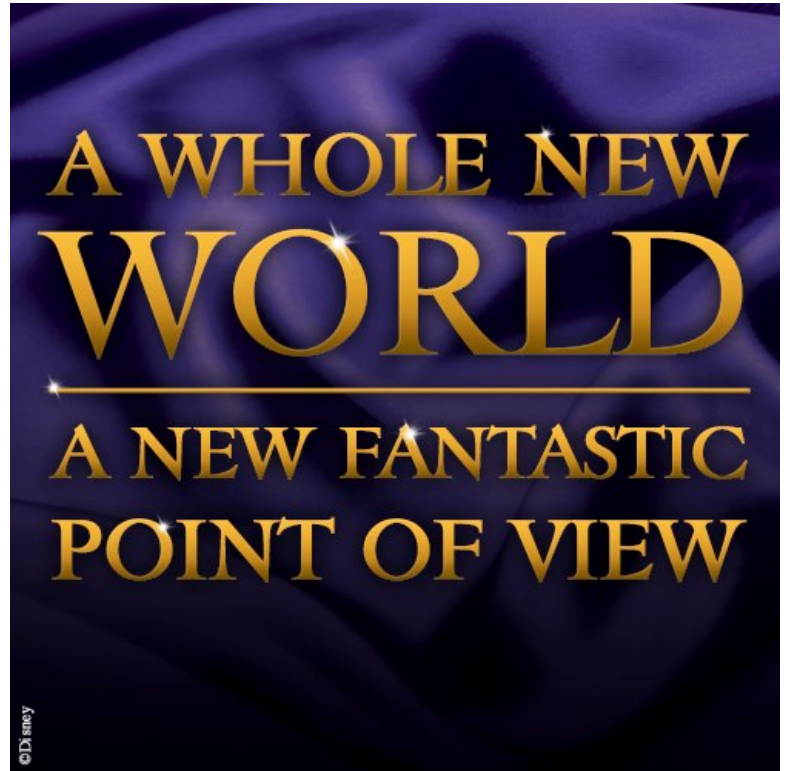
Examples

Write an email from bullet list Code a snake game Assist in a task

Search web ⓘ

Ask anything [✓](#) [↗](#)

Model: meta-llama/Llama-2-70b-chat-hf · Generated content may be inaccurate or false.

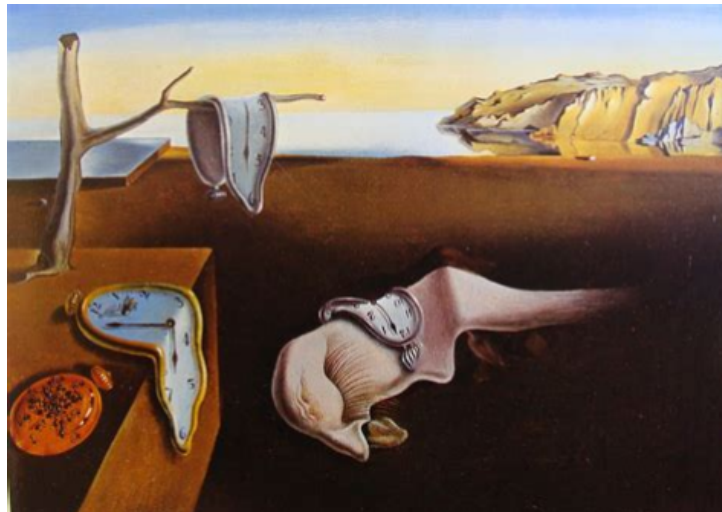


LLMs : A WHOLE NEW WORLD?

A whole new world?

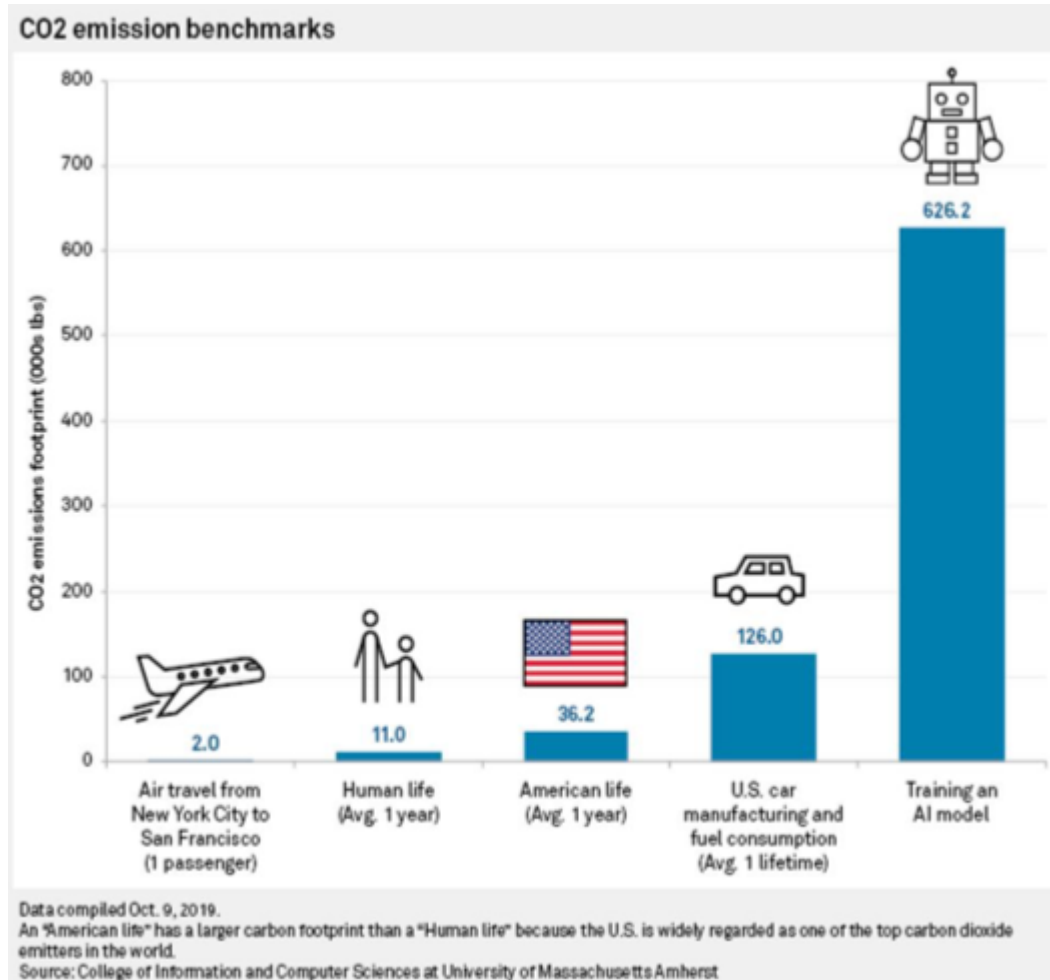
LLMs and generative AI are changing our world

But not in the way you think



The persistence of Memory, Salvador Dali

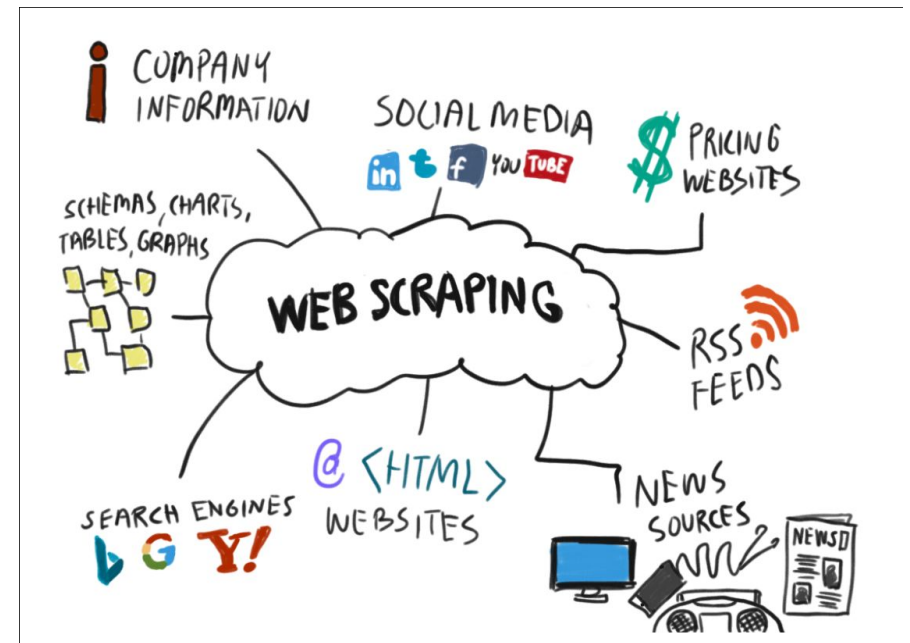
Development of LLMs (and other large AI models) is costly



- Training a model consumes 700 000 L of clean freshwater
- Using a model consumes $\frac{1}{2}$ L of water every 20 interactions

Generative AI has a tremendous impact on legal discussions

- Scrapping of data




Generative AI has a tremendous impact on legal discussions

- Scrapping of data

MEDIA

'New York Times' considers legal action against OpenAI as copyright tensions swirl

August 16, 2023 · 5:53 PM ET

 Bobby Allyn

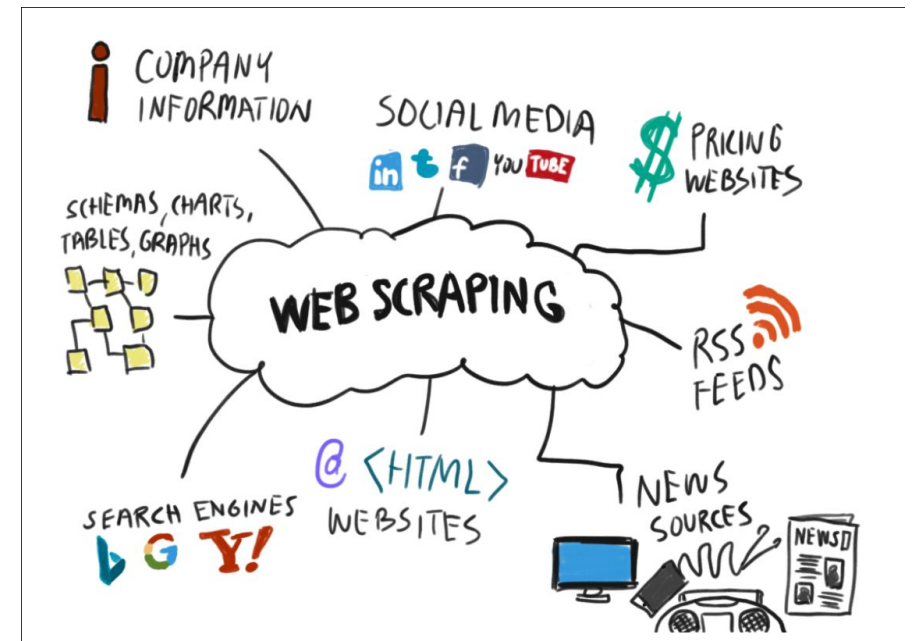
<https://www.npr.org/2023/08/16/1194202562/new-york-times-considers-legal-action-against-openai-as-copyright-tensions-swirl>

OpenAI faces class-action lawsuit over how ChatGPT uses data — here's what we know

By [Tony Polanco](#) published June 29, 2023

This could be a huge blow to AI chatbots like ChatGPT

<https://www.tomsguide.com/news/openai-faces-class-action-lawsuit-over-how-chatgpt-uses-data-heres-what-we-know>



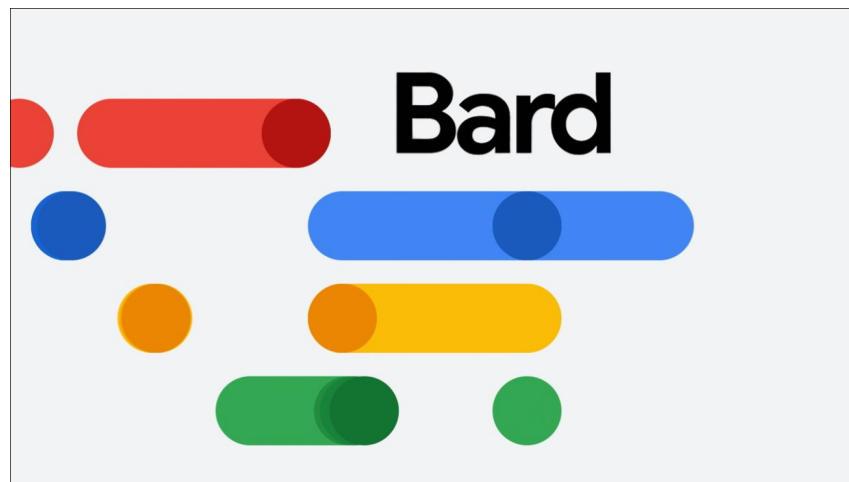
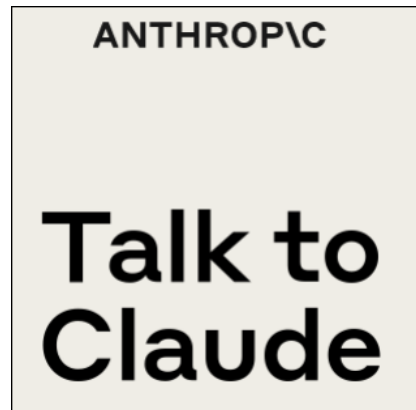
The economical context in which they are developed matters

- Monopoly



The economical context in which they are developed matters

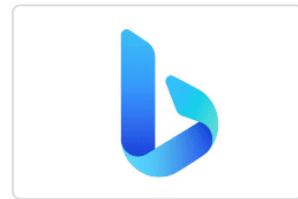
- Monopoly



 **HuggingChat** v0.5.0

Making the community's best AI chat models available to everyone.

The economical context in which they are developed matters



The economical context in which they are developed matters

- Monopoly

Let me search for it (on the Internet)



Let me Google this

The economical context in which they are developed matters

- Monopoly

Let me ask an AI Chatbot

=

Let me ask ChatGPT

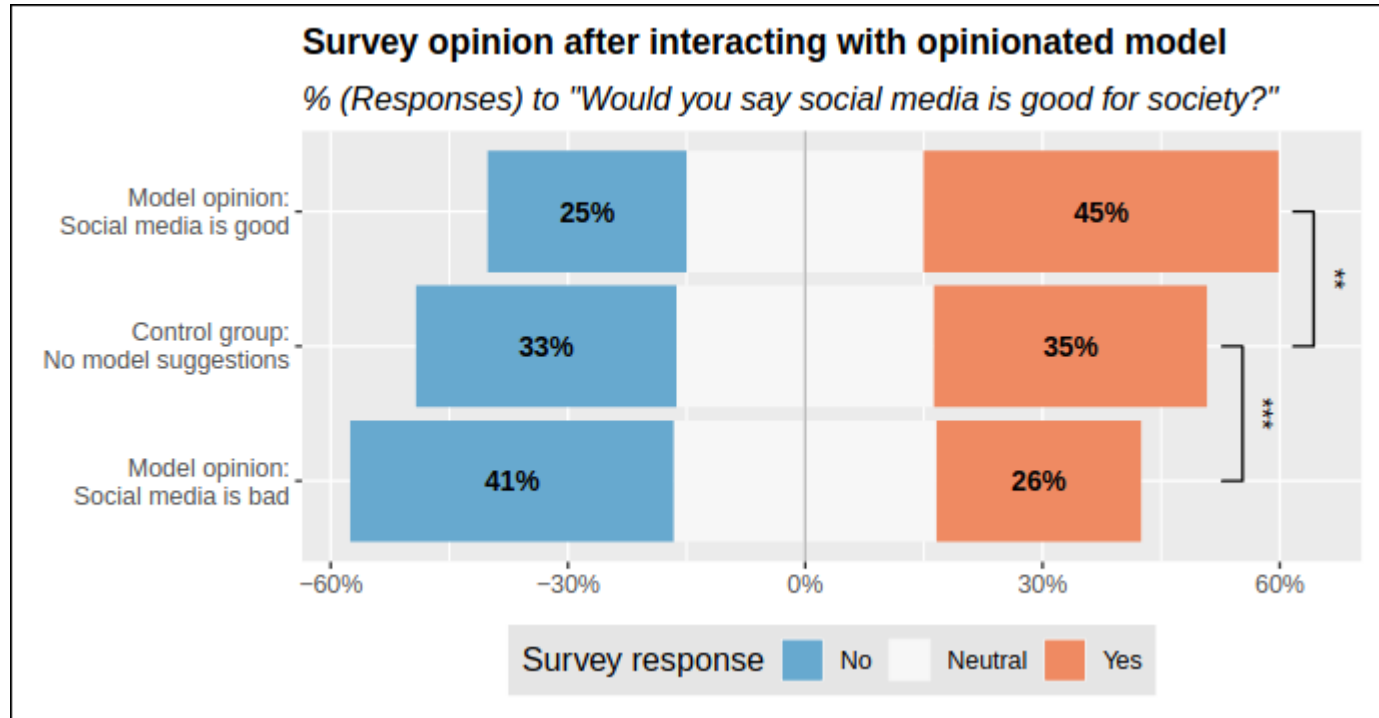
The economical context in which they are developed matters

BUSINESS • TECHNOLOGY

Exclusive: OpenAI Used Kenyan Workers on
Less Than \$2 Per Hour to Make ChatGPT Less
Toxic

<https://time.com/6247678/openai-chatgpt-kenya-workers/>

Generative AI influences us



<https://dl.acm.org/doi/10.1145/3544548.3581196>

RESEARCH-ARTICLE OPEN ACCESS



Co-Writing with Opinionated Language Models Affects Users' Views

Authors: Maurice Jakesch, Advait Bhat, Daniel Buschek, Lior Zalmanson, Mor Naaman [Authors Info & Claims](#)

Generative AI influences us

AI and the American Smile

How AI misrepresents culture through a facial expression.



jenka · Follow

10 min read · Mar 26

<https://medium.com/@socialcreature/ai-and-the-american-smile-76d23a0fbfaf>

HUMANS ARE BIASED. GENERATIVE AI IS EVEN WORSE

Stable Diffusion's text-to-image model amplifies stereotypes about race and gender – here's why that matters

By [Leonardo Nicoletti](#) and [Dina Bass](#) for **Bloomberg Technology** + **Equality**



<https://www.bloomberg.com/graphics/2023-generative-ai-bias/>

The pollution of our information ecosystem




nytimes.com

A New Frontier for Travel Scammers: A.I.-Generated Guidebooks

Shoddy guidebooks are flooding Amazon. Their authors claim to be renowned travel writers, but are they A.I. inventions? And how big is the problem?

The pollution of our information ecosystem



nytimes.com
A New Frontier for Travel Scammers: A.I.-Generated Guidebooks
Shoddy guidebooks are flooding Amazon. Their authors claim to be renowned travel writers, but are they A.I. inventions? And how big is the problem?

[washingtonpost.com](https://www.washingtonpost.com)

ChatGPT invented a sexual harassment scandal and named a real law prof as the accused

Pranshu Verma, Will Oremus

10–13 minutes

The pollution of our information ecosystem



nytimes.com

A New Frontier for Travel Scammers: A.I.-Generated Guidebooks

Shoddy guidebooks are flooding Amazon. Their authors claim to be renowned travel writers, but are they A.I. inventions? And how big is the problem?

NEWS

'Life or Death:' AI-Generated Mushroom Foraging Books Are All Over Amazon

SAMANTHA COLE · AUG 29, 2023 AT 9:04 AM

Experts are worried that books produced by ChatGPT for sale on Amazon, which target beginner foragers, could end up killing someone.

[washingtonpost.com](https://www.washingtonpost.com)

ChatGPT invented a sexual harassment scandal and named a real law prof as the accused

Pranshu Verma, Will Oremus

10–13 minutes

Detectors for AI-writing discriminate

Design and Human-Computer Interaction, Language Processing, Machine Learning

AI-Detectors Biased Against Non-Native English Writers

Don't put faith in detectors that are "unreliable and easily gamed," says scholar.

<https://hai.stanford.edu/news/ai-detectors-biased-against-non-native-english-writers>

IN CONCLUSION

In summary LLMs

- Are changing our world, but not in the way we may think...
 - Time cost is obvious (verifying the output, our working practices)
 - Time gain are still unclear (beyond anecdotal evidence)
 - Pollution of our information system (when has this book been published?)
- Are excellent at generating plausible text
- Have no concept of “right” or “wrong”
- Have no concept of “I don’t know”
- Can fail in spectacular or subtle way, and are un-predictable. The subtle failures are more dangerous.

Should I throw away all LLMs and disconnect from the Internet?

- NO!
- LLMs are good for applications that requires you to be statistically similar to a native English speaker
 - Recommendation letters
 - Talk or article titles (if originality is not what you are looking for)
 - Correcting grammar or spelling
- With supervision and careful verification, if you have the competencies to check yourself
- And while being aware of the ethical challenges behind

Resources



A.I. is B.S.
Adam Conover
1 M de vues • il y a 4 mois

https://www.youtube.com/watch?v=ro130m-f_yk



CODED BIAS
AVAILABLE TO STREAM APRIL 5TH ON
NETFLIX

"THE BEST OF SUNDANCE"
Denver Post

"'CODED BIAS' SERVES AS BOTH A WAKE-UP CALL AND A CALL TO ACTION"
Variety

"THE MOST CLEAREYED OF SEVERAL RECENT DOCUMENTARIES ABOUT THE PERILS OF BIG TECH"
The New York Times

2021 CINEMA EYE HONORS NOMINEE
OFFICIAL SELECTION **hotdocs**
OFFICIAL SELECTION **sundance**
SXSW 2020
CRITICS CHOICE
100% FRESH **Rotten Tomatoes**

[codedbias.com/screen](https://www.netflix.com/title/81040344)



Artificial Intelligence: Last Week Tonight with John...
LastWeekTonight ✓
97 M de vues • il y a 5 mois

<https://www.youtube.com/watch?v=Sqa8Zo2XWc4>

Resources

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?

Emily M. Bender*
ebender@uw.edu
University of Washington
Seattle, WA, USA

Timnit Gebru*
timnit@blackinai.org
Black in AI
Palo Alto, CA, USA

Angelina McMillan-Major
aymm@uw.edu
University of Washington
Seattle, WA, USA

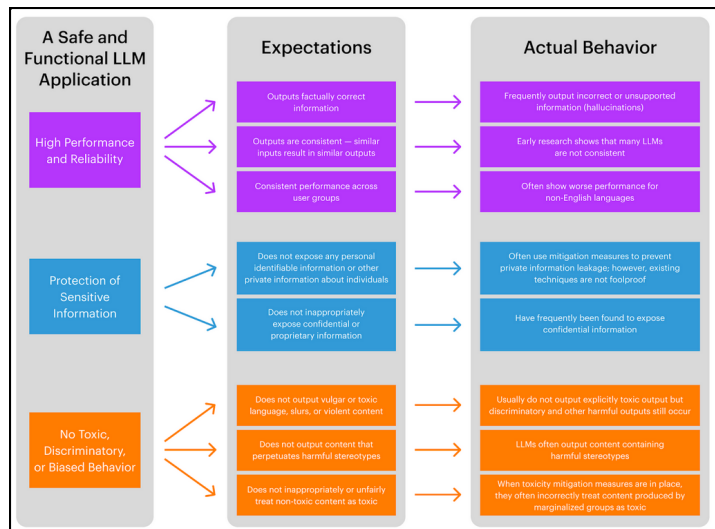
Shmargaret Shmitchell
shmargaret.shmitchell@gmail.com
The Aether

<https://dl.acm.org/doi/pdf/10.1145/3442188.3445922>

What Is ChatGPT Doing ... and Why Does It Work?

February 14, 2023

<https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/>



ChatGPT Explained: A Normie's Guide To How It Works

Once you understand a few core concepts, modern chatbots are less mysterious.

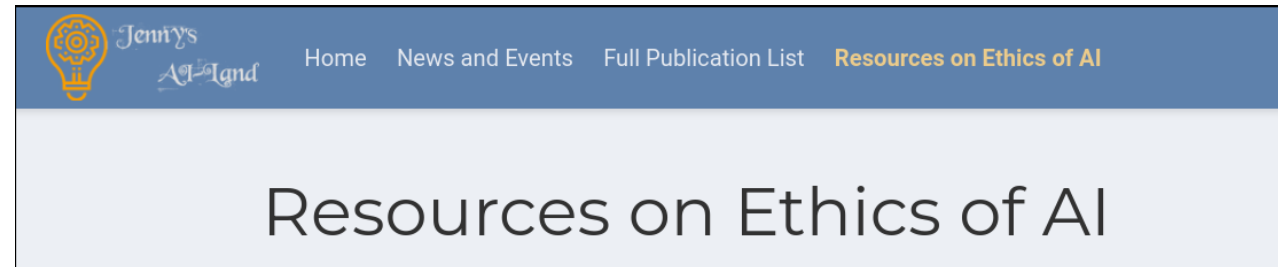


JON STOKES
MAR 1, 2023

<https://www.jonstokes.com/p/chatgpt-explained-a-guide-for-normies>

<https://www.arthur.ai/blog/the-real-world-harms-of-llms-part-1>

Resources



<https://jrenoux.github.io/resources-ai-ethics/>

