

Tune in to Tune up: Artificial Intelligence



1

Agenda

- What is AI?
- Critical Perspectives on AI
- Citing AI
- Grants
- Helpful tools
 - Citation Recommendation
 - Retrieval Augmented Generation

2

AI is a lot of things It is an umbrella term

Academic Goal -
1956

Machine
Learning

Artificial
Neural
Networks

Natural
Language
Processing

Algorithms

Character
archetype

Computer
Vision

Marketing
Term

Large
Language
Models

Umbrella by Hafid Firman Syarif from [Noun Project](#) (CC BY 3.0)

3

The Dream

Machines independently solving complex tasks with human-like cognitive abilities.

The Reality

“Narrow AI” focuses on one task or series of tasks, from chess (DeepBlue) to text generation (ChatGPT)

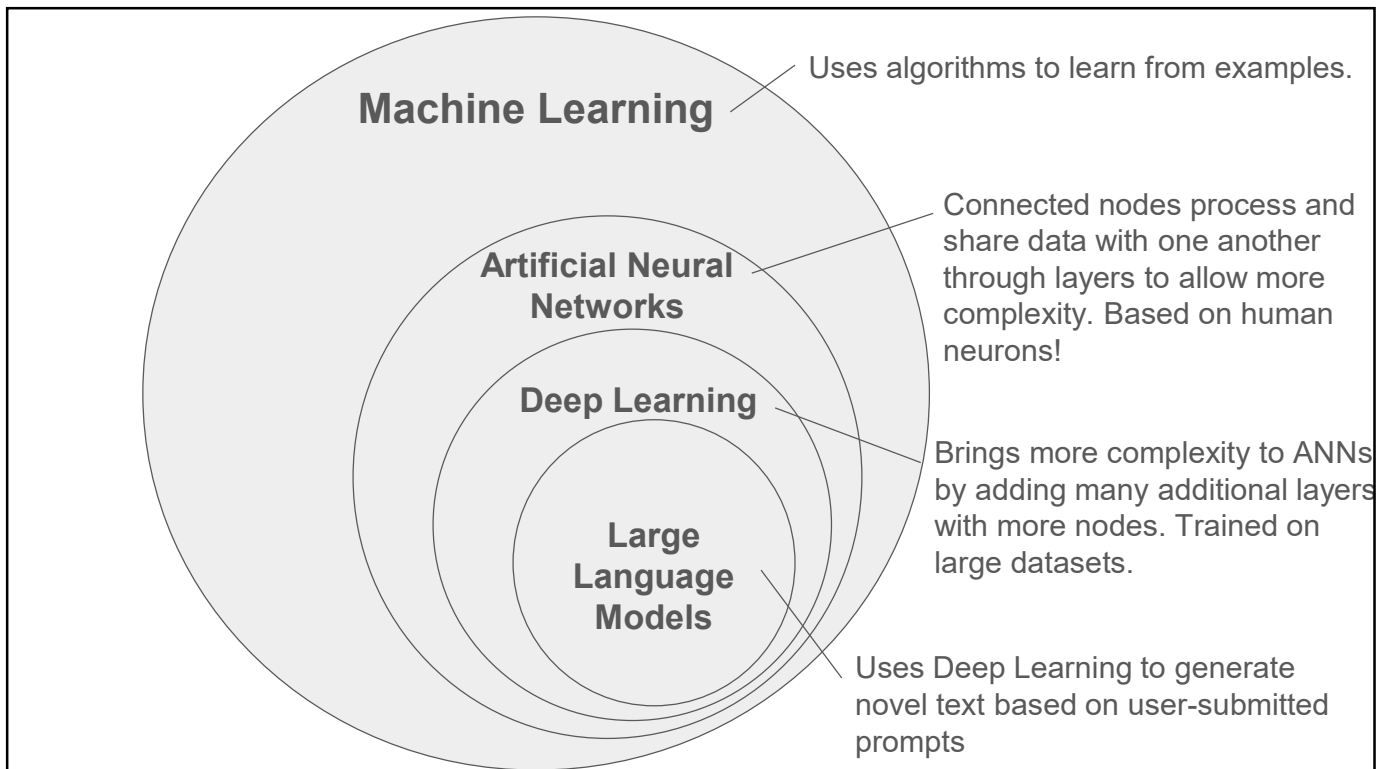
4

The Current “AI” Landscape

Several different types of systems are being described as AI now. Thinking of these things as “automations” helps clarify their capabilities and limitations.

- **Automatic Decision Systems**
 - These make consequential decisions like setting bail, determining loan eligibility, and resume assessment.
- **Classification Systems**
 - Image identification and improved facial detection on your phone’s camera, for example.
- **Automated access to human labour**
 - Lyft, Uber
- **Recommender Systems**
 - Netflix recommendations and social media feed algorithms.
- **Automation of translation from one form to another**
 - OCR or Optical Character Recognition detects text in images.
 - Audio transcription
- **“Synthetic Media Machines”** (as per [Dr. Emily Bender](#), computational linguist)
 - ChatGPT, Gemini, Perplexity and other LLMs
 - Image, video, and voice generators

5



6

Where does it get the data?

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

Table 2.2: Datasets used to train GPT-3. “Weight in training mix” refers to the fraction of examples during training that are drawn from a given dataset, which we intentionally do not make proportional to the size of the dataset. As a result, when we train for 300 billion tokens, some datasets are seen up to 3.4 times during training while other datasets are seen less than once.

7

Data cut-offs

When an LLM is trained, its data is locked in and no later data can be added until the model is re-trained.

This is why you cannot discuss current events with an LLM.

On lmarena.ai, you can see this as the “Knowledge cutoff” for each model

Rank* (UB)	Model	Arena Score	95% CI	Votes	Organization	License	Knowledge Cutoff
1	o1-preview	1355	+12/-11	2991	OpenAI	Proprietary	2023/10
2	ChatGPT-4o-latest (2024-09-03)	1335	+5/-6	10213	OpenAI	Proprietary	2023/10
2	o1-mini	1324	+12/-9	3009	OpenAI	Proprietary	2023/10
4	Gemini-1.5-Pro-Exp-0827	1299	+5/-4	28229	Google	Proprietary	2023/11
4	Grok-2-08-13	1294	+4/-4	23999	xAI	Proprietary	2024/3

8

Data cut-offs

When an LLM is trained, its data cut-off is the date until the model is re-trained. This is often referred to as the model's knowledge cut-off date.

This github repository provides a comprehensive list of LLM models and their knowledge cut-off dates. You can find the full list at <https://github.com/epicureanai/llm-knowledge-cut-off>.

Google Models

Model Name	Company	Cut-off Date	Source
Gemini 1.0 Pro	Google	2023.08	Source
Gemini 1.5 Pro	Google	2024.09	Source
Gemini 1.5 Flash	Google	2024.09	Source

OpenAI Models

Model Name	Company	Cut-off Date	Source
GPT-1	OpenAI	2018.10	Source
GPT-2	OpenAI	2019.11	Source
GPT-3	OpenAI	2020.10	Source
GPT-3.5*	OpenAI	2021.09	Source
GPT-4*	OpenAI	2021.09	Source
GPT-4 (1106-preview)	OpenAI	2023.04	Source
GPT-4 (vision-preview)	OpenAI	2023.04	Source
GPT-4 (0125-preview)	OpenAI	2023.12	Source
GPT-4-turbo (2024-04-09)	OpenAI	2023.12	Source
GPT-4o (2024-05-13)	OpenAI	2023.10	Source
GPT-4o (2024-08-06)	OpenAI	2023.10	Source
GPT-4o mini (2024-07-18)	OpenAI	2023.10	Source
GPT-4o-realtime-preview (2024-10-01-preview)	OpenAI	2023.10	Source
OpenAI o1-preview (2024-09-12)	OpenAI	2023.10	Source
OpenAI o1-mini (2024-09-12)	OpenAI	2023.10	Source

DeepSeek Models

Model Name	Company	Cut-off Date	Source
DeepSeek-LLM-7B/67B-Chat	DeepSeek	2023.05	Source
DeepSeek-Coder	DeepSeek	2023.03	Source
DeepSeek-Coder-V2	DeepSeek	2023.11	Source

9

Tokenization is how LLMs make sense of data

Frog on a log.

Tokenization

[CLS]	Frog	on	a	log	.	[SEP]
101	2025	2001	1063	8532	2011	102

Embeddings

-0.4819	-0.9431	0.0019	0.698	0.0111	-0.6098	0.0210
0.3845	-0.8119	0.5591	-0.0034	-0.9917	0.51098	0.5281
0.0001	0.7309	0.9167	-0.8976	-0.5554	-0.7771	0.9004

10

You
how many Ms are in the word mammogram

ChatGPT
There are three Ms in the word "mammogram."
🔊 📄 ↺ 🗑️

(There are 4)

You
spell lollipop backwards

ChatGPT
"Lollipop" spelled backwards is "pilpolol."
🔊 📄 ↺ 🗑️

(it's popillol)


You
why can't you spell lollipop backwards correctly?

ChatGPT
You're absolutely correct, and I appreciate your patience. My apologies for the mistake. "Lollipop" spelled backwards correctly is "poppilol." Thank you for bringing that to my attention.

(it's still wrong...)

LLMs are good at being quickly and confidently incorrect.

11



“Stochastic Parrots”

Some scholars have been using this term for ChatGPT and similar tools. Like a parrot, these tools do not comprehend what they are saying.

They do not produce true or validated answers.

They produce *statistically likely* answers.

12

LLMs just make things up a lot of the time...

ChatGPT

Please create a bibliography in APA 7 containing 5 sources about climate change, the IT sector's growing emissions, and the advent of LLMs

Here is a bibliography in APA 7 format with five relevant sources on climate change, the IT sector's growing emissions, and the impact of large language models (LLMs).

References

Columbia Climate School. (2023). *AI's growing carbon footprint*. State of the Planet. <https://news.climate.columbia.edu>

Hao, K. (2020, June 17). *Training a single AI model can emit as much carbon as five cars in their lifetimes*. MIT Technology Review. <https://www.technologyreview.com>

IEA. (2021). *Data centres and energy – Tracking report 2021*. International Energy Agency. <https://www.iea.org/reports/data-centres-and-energy>

Lucioni, A., & Viviano, F. (2022). Estimating the carbon footprint of large language models. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.2202.05651>

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3645–3650. <https://doi.org/10.18653/v1/P19-1355>

Does not exist

Similar to a real paper (title is different, 2nd author is wrong, DOI is wrong)

13

RAG - Retrieval-Augmented Generation

ChatGPT 3.5

You
Present 3 arguments for veganism

ChatGPT

- Ethical Considerations:** One of the primary arguments for veganism revolves around ethical concerns regarding animal welfare. Many people find it morally objectionable to exploit and harm animals for food production when alternatives exist. The process of factory farming often involves confining animals in cramped conditions, subjecting them to painful procedures such as debeaking or tail docking without anesthesia, and ultimately slaughtering them for human consumption. Advocates for veganism argue that it is unethical to contribute to this system of animal suffering when plant-based alternatives can adequately fulfill nutritional needs.
- Environmental Impact:** Veganism is often promoted as an environmentally sustainable dietary choice. Animal agriculture is a significant contributor to greenhouse gas emissions, deforestation, water pollution, and habitat destruction. Livestock farming requires large amounts of land, water, and feed resources, leading to biodiversity loss and exacerbating climate change. By transitioning to a plant-based diet, individuals can significantly reduce their

perplexity

Present 3 arguments for veganism

Sources

- The Strongest Argument for Veganism - Effective... EA ea-foundation · 1
- Good arguments in favor of veganism? ... reddit · 2
- Why go vegan? | The Vegan Society vegansociety · 3
- View 2 more

Answer

Here are 3 strong arguments in favor of veganism:

- Avoiding cruelty to animals.** The search results show that the production of animal products often leads to significant animal suffering, such as the painful gassing of male chicks in the egg industry ¹ ². Since we have a moral duty not to harm animals unnecessarily, and animal products are not necessary for human health and survival ¹ ³ ⁴, we should avoid consuming them.
- Environmental benefits.** Animal agriculture is a major contributor to environmental issues like greenhouse gas emissions, water usage, and land degradation ⁴. By adopting a vegan diet, individuals can significantly reduce their environmental impact and help address these pressing global challenges.

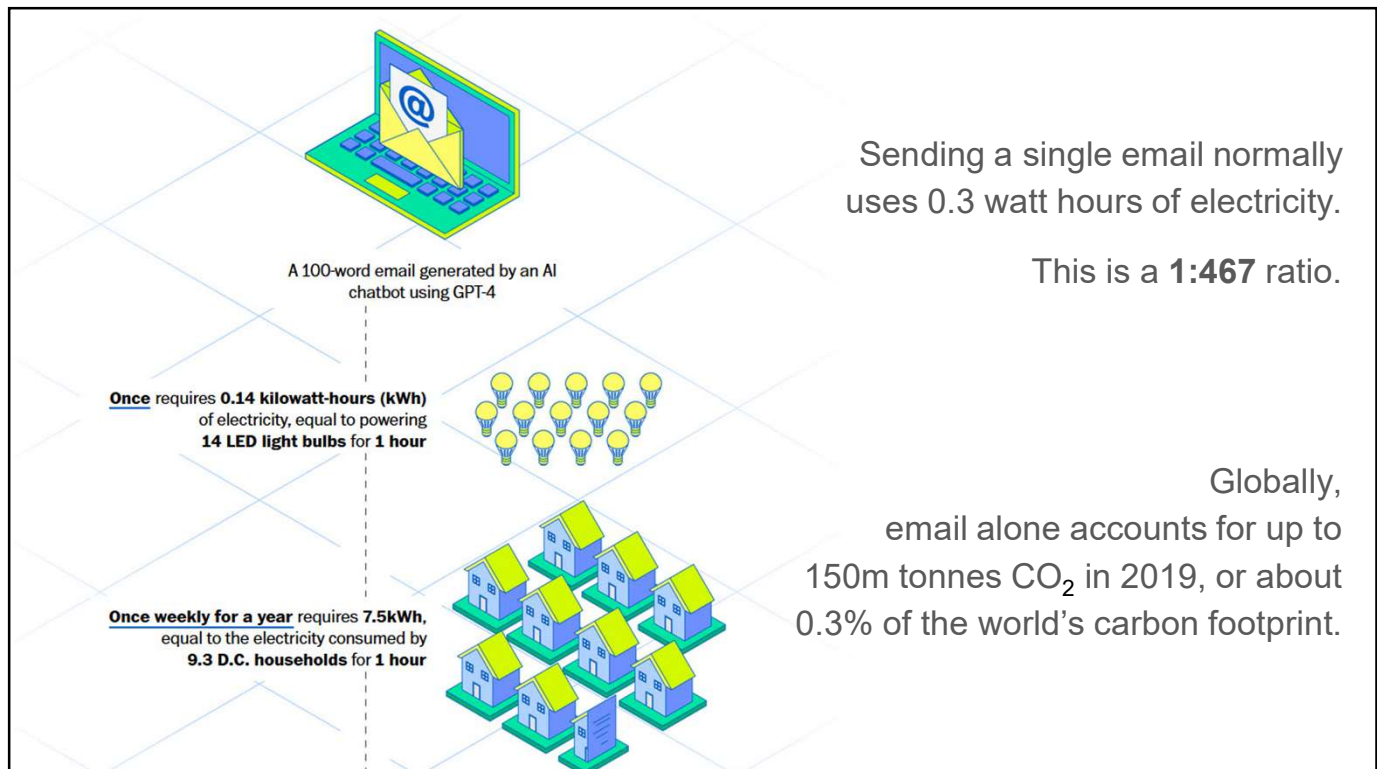
14

Critical Perspectives on Artificial Intelligence

Organizations like the International Center for Academic Integrity have highlighted several areas in AI that call for “scholarly skepticism” including:

- Privacy and confidentiality of user data
- Bias in the output
- Misinformation, including fabricated citations and references
- Plagiarism in the output
- Lack of diverse voices
- Access issues with paywalling
- Ecological toll

15



16

Eco-footprint of AI

Just *training* a model:

- **213 million parameters** takes **284 tonnes of CO₂** emissions, roughly equivalent to the lifetime emissions of **five cars, including fuel** (Strubell, 2020)
- ChatGPT-4 is estimated to have **1.8 trillion parameters** by comparison.



Fig 1. The life cycle assessment approach proposed by [Luccioni et al.](#)

17

Eco-footprint of AI



Using the example of BLOOM in 2023, Luccioni et al. found:

Process	CO ₂ emissions (CO ₂ eq)	Percentage of total emissions
Embodied emissions	11.2 tonnes	22.2 %
Dynamic consumption	24.69 tonnes	48.9 %
Idle consumption	14.6 tonnes	28.9 %
Total	50.5 tonnes	100.00 %

Table 3: Breakdown of CO₂ emissions from different sources of the BLOOM model life cycle

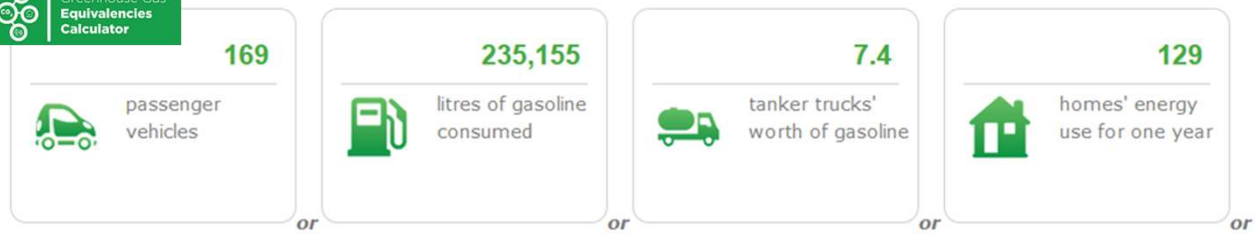
Equal to powering 10 homes for a year.

18

Eco-Footprint of AI

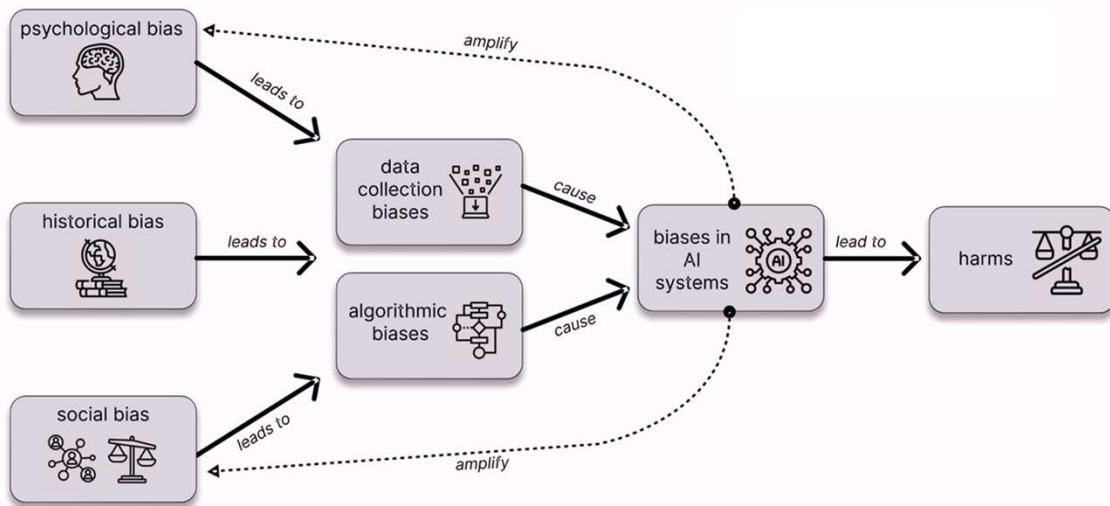
What are the carbon footprints of different models?

Model name	Number of parameters	Datacenter PUE	Carbon intensity of grid used	Energy consumption	CO ₂ eq emissions	CO ₂ eq emissions × PUE
GPT-3	175B	1.1	429 gCO ₂ eq/kWh	1,287 MWh	502 tonnes	552 tonnes
Gopher	280B	1.08	330 gCO ₂ eq/kWh	1,066 MWh	352 tonnes	380 tonnes
OPT	175B	1.09 ²	231gCO ₂ eq/kWh	324 MWh	70 tonnes	76.3 tonnes ³
BLOOM	176B	1.2	57 gCO ₂ eq/kWh	433 MWh	25 tonnes	30 tonnes



19

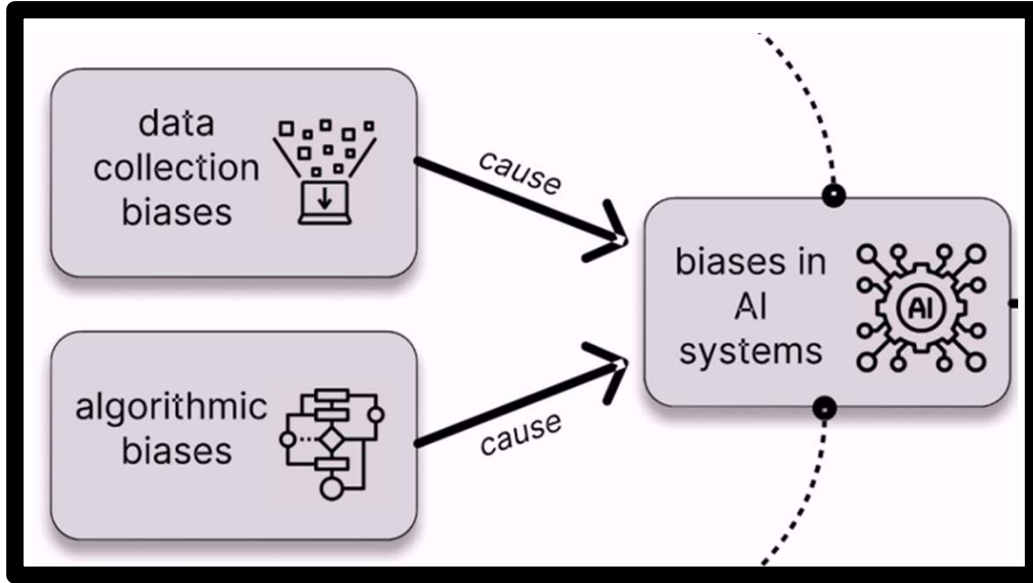
Lifecycle of Bias in AI



(Hendrycks, 2024)

20

Black box decision making



21

The Overall Bias

Western

Educated

Industrialized

Rich

Democratic

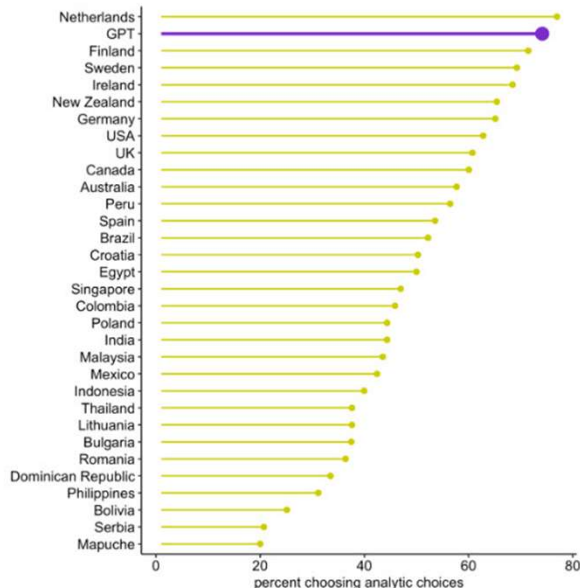


Figure 4

Average holistic thinking style across 31 human populations (yellow) and GPT (purple). Except for the Mapuche group, participants from all human populations completed the identical Triad Task via the online platform yourmorals.org. For the Mapuche, data were collected through individual interviews using a similar version of the task (adapted from Henrich, 2020).

(Atari et al., 2023)

22

The Washington Post
Democracy Dies in Darkness

Veronica Ramshaw

These fake images reveal how AI amplifies our worst stereotypes


AI image generators like Stable Diffusion and DALL-E amplify bias in gender and race, despite efforts to detoxify the data fueling these results.

By Nitasha Tiku, Kevin Schaul and Szu Yu Chen
Nov. 1 at 6:00 a.m.


AI-GENERATED IMAGES

prompt:
A portrait photo of ...

a person at social services



a productive person



The Verge / Tech / Reviews / Science / Entertainment / AI / M

ARTIFICIAL INTELLIGENCE / TECH / WEB


Google apologizes for ‘missing the mark’ after Gemini generated racially diverse Nazis

Generative AI has a history of amplifying racial and gender stereotypes – but Google’s apparent attempts to subvert that are causing problems, too.

By Adi Robertson, a senior tech and policy editor focused on VR, online platforms, and free expression. Adi has covered video games, biohacking, and more for The Verge since 2011.
Feb 21, 2024, 4:17 PM GMT-6

39 Comments (39 New)

Sure, here is a picture of the Founding Fathers:



23

Biases in the human operator

- Confirmation bias (or Affirmation bias)
 - Confirmation bias refers to how we pay more attention to information that fits our beliefs
 - The way you phrase prompts can encourage an LLM to answer in alignment with your preconceived notions
- Automation bias
 - Automation bias is the tendency for humans to favor suggestions from automated decision-making systems and to ignore contradictory information made without automation, even if it is correct
 - Humans are quick to accept computer outputs as fact without further verification



24

Citing AI

Individual standards are constantly evolving.

- **APA - June 2023 update: Artificial Intelligence**
 - Name of Company/creator of generative AI Tool. (Year). Name of the generative AI program (model of program) [Large language model]. URL
- **MLA - Works Cited entry**
 - Author/Creator. "Name of chatbot." Title of platform where accessed, Full URL, Date Accessed (optional).
- **Chicago - Chicago footnote (unpublished interview)**
 - OpenAI's ChatGPT, response to query from author, February 15, 2023.
 - Author's (Parent Company) Medium, Response to "Query in quotes." Name of Website, Parent Company, Date accessed, URL.

https://subjectguides.uwaterloo.ca/chatgpt_generative_ai/aigeneratedcontentcitation

25

Creating an AI Declaration Appendix

General comments:

<https://mitcommlab.mit.edu/cee/2023/08/27/using-generative-ai-for-your-scientific-writing-be-aware-of-journal-policies/>

<https://www.chronicle.com/article/why-we-should-normalize-open-disclosure-of-ai-use>

Examples:

“AI Usage Disclosure: This document [include title] was created with assistance from [specify the AI tool]. The content can be viewed here [add link] and has been reviewed and edited by [author’s full name]. For more information on the extent and nature of AI usage, please contact the author.”

“During the preparation of this work the principal author (corresponding author) used ChatGPT to improve writing style and check grammar and spelling. After using this tool//service, the authors reviewed and edited the content as needed and take(s) full responsibility for the content of this publication . . . “

“AI-assisted technology is not used in the preparation of this work (except checking grammar and spelling).

26

Tri-Agency Policy

[Guidance on the use of Artificial Intelligence in the development and review of research grant proposals](#) (November, 2024)

Grant application **preparation**: “Applicants must state if and how generative AI has been used in the development of their application and are required to follow specific instructions, which will be provided for each funding opportunity as they become available.”

Grant application **evaluation**: “In the evaluation of grant applications, reviewers should be aware that inputting application information into generative AI tools could result in breaches of privacy and in the loss of custody of intellectual property Therefore, use of publicly available generative AI tools for evaluating grant applications is strictly prohibited.”

27

AI in Journal Articles

Examples of publisher policies:

Elsevier: <https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals>

Springer: <https://www.nature.com/nature-portfolio/editorial-policies/ai>

Taylor & Francis: <https://authorservices.taylorandfrancis.com/editorial-policies/defining-authorship-research-paper/>

Common threads: AI tools CANNOT be co-authors, use must must be declared, specific guidelines for images

28

RAG but academic <https://consensus.app>

The screenshot displays the Consensus app interface for the query "Is global warming impacting shipping?". The interface is split into two main sections: a left sidebar and a main content area.

Left Sidebar:

- Summary:** 10 papers analyzed. Text: "These studies suggest that global warming is impacting shipping by increasing navigability in Arctic routes due to sea ice loss, while also indicating that shipping emissions contribute to both warming and cooling effects on the climate."
- Consensus Meter:** 8 papers analyzed. A bar chart shows: Yes - 88% (green), Possibly - 0% (yellow), No - 12% (red).
- Copilot:** 10 papers analyzed.
 - Introduction:** "Global warming is a significant environmental issue that affects various sectors, including international shipping. The shipping industry is both impacted by and contributes to climate change through emissions and changes in sea ice patterns. This synthesis examines how global warming is impacting shipping, drawing on insights from multiple research papers."
 - Key Insights:**
 - Emissions and Climate Impact:**
 - Shipping emissions contribute to both warming and cooling effects on the climate. CO2 emissions from ships cause warming, while sulfur dioxide (SO2) emissions cause cooling by affecting atmospheric particles and clouds. However, new regulations reducing SO2 and NOx emissions will decrease the cooling effect, leading to a net warming impact from shipping in the future (1, 2, 5, 7).
 - The net global mean radiative forcing from shipping is currently negative, indicating a cooling effect. However, this is expected to shift to a warming effect due to regulatory changes and the long-lived nature of CO2 compared to short-lived cooling agents like sulfate (1, 2, 5).

Main Content Area:

- Shipping emissions: from cooling to warming of climate-and reducing impacts on health.** (Yes)
 - Shipping emissions contribute to global warming, with CO2 emissions and other emissions causing cooling and warming, respectively.
 - Environmental science & technology | J. Fuglestad et al. | 119 citations | 2009
 - Highly Cited
 - Study snapshot
 - Save, Cite, Share
- Transport impacts on atmosphere and climate: Shipping** (No)
 - Ship emissions contribute to global warming by causing air quality problems and negatively impacting climate, with cooling effects from altered clouds far outweighing warming effects from greenhouse gases.
 - Atmospheric Environment | V. Eyring et al. | 779 citations | 2010
 - Highly Cited
 - Study snapshot
 - Save, Cite, Share
- Impact of 1, 2 and 4 °C of global warming on ship navigation in the Canadian Arctic** (Yes)
 - Global warming under 1, 2 and 4°C will lead to longer shipping seasons and navigability in the Canadian Arctic due to sea ice loss.

29

Non-generative AI for research

2 primary approaches to keep in mind:

Descendancy (forward citation searching) -

Using a tool to find articles that cite the one which interests you.

Ascendancy (backwards citation searching) -

Look at the bibliography of the article to find the sources it has cited

30

Non-Generative AI for Research

The screenshot displays the Zotero interface with a network graph of research papers. The central graph shows nodes for authors and years, such as 'Bouman 2017' and 'Miola 2011'. The interface includes a sidebar with categorized collections, a main list of papers, and a right-hand panel with filters and options.

Climate change and shipping

Oh ... Hall 2019
Volume 2 : Pathway to Resilient Transport

Golnar ... Beškovnik 2020
Green Maritime Transport as a Part of Global Green Intermodal Chains
Pomorski zbornik

Stopford ... Stopford 2022
Maritime governance: piloting maritime transport through the stormy seas of climate change
Maritime economics and logistics

Miola ... Ciuffo 2011
Designing a climate change policy for the international maritime

Similar Work 1197

Earlier Work 15

Later Work 50

EXPLORE PEOPLE

These Authors 78

Suggested Authors 423

EXPLORE OTHER CONTENT

Linked Content 13

EXPORT PAPERS

BibTeX RIS CSV

PUBLIC COLLECTION

SHAREABLE LINK Copy

COLLABORATORS Edit

31

Non-Generative AI for Research

The screenshot shows a ChatGPT interface with a document titled "Climate Change and Its Impact on Coastal Real Estate". The document text is displayed in the main area, and a Keenious search overlay is visible on the right side.

Climate Change and Its Impact on Coastal Real Estate

Climate change is one of the most pressing problems in today's world. It refers to long-term changes in the earth's climate, mostly caused by human activities like the burning of fossil fuels, deforestation, and industrialization. These activities have led to an increase in greenhouse gases in the atmosphere, such as carbon dioxide and methane, which trap heat and cause global warming. One of the major consequences of climate change is the rise in sea levels, which has a huge impact on coastal real estate.

Rising sea levels are primarily caused by two factors. First, global warming is causing polar ice caps and glaciers to melt, which adds more water to the oceans. Second, as the temperature of the ocean increases, the water expands. This means that even without the extra water from melting ice, the volume of the ocean is still increasing. According to scientist, sea levels have risen by about 8 inches since 1880, and the rate of increase is getting faster. This is bad news for people who live or own property near coastlines.

Coastal real estate is at significant risk due to climate change. Properties that are located near the shore are especially vulnerable to flooding and erosion, both of which are made worse by rising sea levels. In some cases, houses and buildings are literally being swallowed by the ocean as the shoreline moves inland. Flooding has become more common and more intense, destroying homes and reducing property values. Many coastal areas are also more vulnerable to storms and hurricanes, which are becoming stronger and more frequent due to climate change. This can cause further damage to coastal properties and make them less desirable to buyers.

One of the biggest problems facing coastal real estate is the unpredictability of future sea level rise.

Keenious Upgrade

Search articles

Articles Topics

Filters Bookmarks

Cross Language

Based on 100% of your document

1000 results

Article 1 Citation

Sea level rise risk interactions with coastal property values: a case study of Oahu, Hawaii
2023 Climatic Change
Nori Tarui et al.

Article 41 Citations

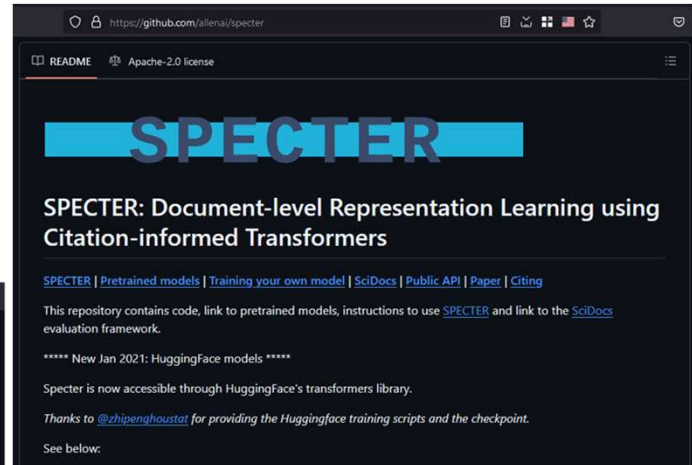
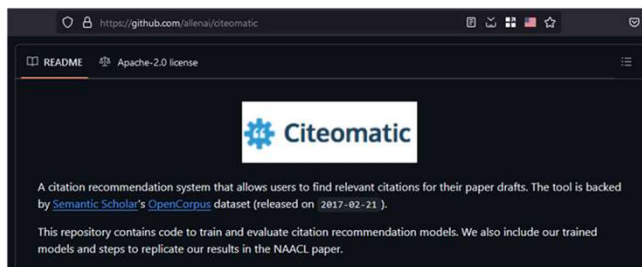
"Living on the edge": Estimating the economic cost of sea level rise on coastal real estate in the Tampa Bay region, Florida
2016 Ocean & Coastal Management
Xinyu Fu et al.

Article

Modeling Real Estate Market Responses to Climate Change in the

32

Citation Recommendation Tools - Advanced Users



33

Still, work outside the AI systems!



Volume 31, Issue 3
July 2022

JOURNAL ARTICLE

Automated citation recommendation tools encourage questionable citations

Serge P J M Horbach ✉, Freek J W Oude Maatman, Willem Halfman, Wytske M Hepkema

Research Evaluation, Volume 31, Issue 3, July 2022, Pages 321–325, <https://doi.org/10.1093/reseval/rvac016>

Published: 02 June 2022

PDF Split View Cite Permissions Share

- Perfunctory citation and sloppy argumentation
- Affirmation Bias
- A new “Matthew-effect”
- Lack of Transparency

34

Further Reading

Our CTL (Centre for Teaching and Learning) has information about the UofR's approach to Generative AI: <https://ctl.uregina.ca/generative-ai-at-the-university-of-regina>

USask's Guide about Generative AI is very comprehensive:
https://libguides.usask.ca/gen_ai

35

Questions?

veronica.ramshaw@uregina.ca

cara.bradley@uregina.ca

36