

ChatGPT Usage in Research



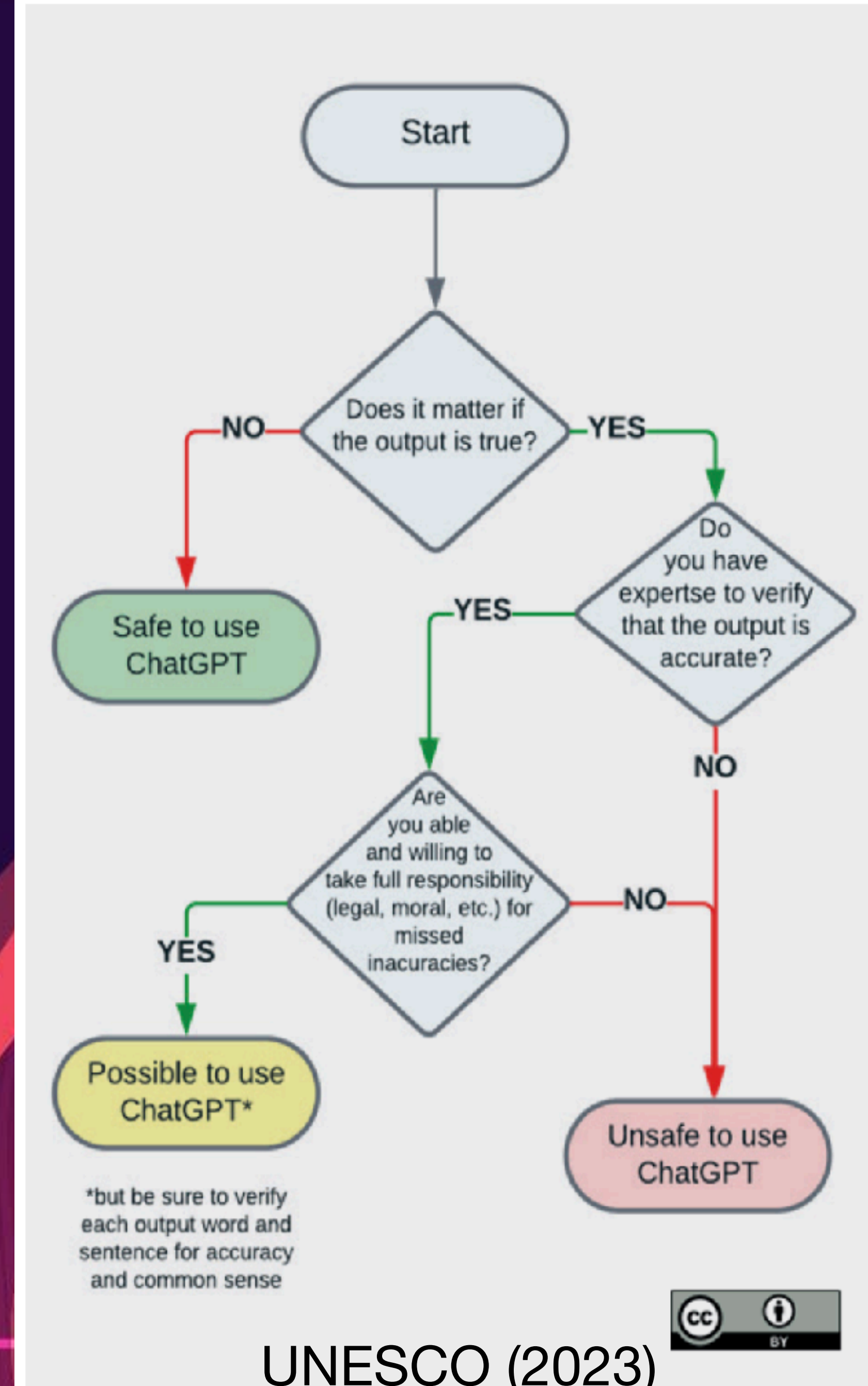
Chantal Schouwenaar
PhD Candidate
School of Business and Economics

How does it work?

The whole internet?

“Predict what comes next in text you’ve never seen before.”

Figure 1: When is it safe to use ChatGPT?⁵



Capabilities

Rewrite text in a certain style

Write & debug code (R/Python/etc.)

Make text longer/shorter

Provide roadmaps for complex tasks

Provide info from documentation

Integrated with

Deloitte

Wolfram Alpha

Expedia

Duolingo

Khan Academy

Github

Klarna

Notion

Open Table

Morgan Stanley

Stripe

openai.com

What is the best place to find reliable export data, specifically on the export of food by the Netherlands?

To find reliable export data on the export of food by the Netherlands, you can explore the following sources:


1. **Statistics Netherlands (CBS):** CBS is the official statistical agency of the Netherlands and provides comprehensive data on various aspects of the Dutch economy, including trade and exports. The agency publishes periodic reports and datasets that include information on food exports. You can access the data through their website: <https://www.cbs.nl/en-gb>
2. **Eurostat:** Eurostat is the statistical office of the European Union, which provides data on various aspects of the EU member states, including international trade. You can access their database to find information on the export of food products by the Netherlands: <https://ec.europa.eu/eurostat>
3. **UN Comtrade Database:** The United Nations Commodity Trade Statistics Database (Comtrade) provides comprehensive data on international trade for over 170 countries, including the Netherlands. You can search for specific food products or categories and filter the data by export and import. Access the database here: <https://comtrade.un.org/>
4. **World Bank's World Integrated Trade Solution (WITS):** WITS is a comprehensive database that provides access to international trade and tariff data. You can use this platform to access data on Dutch food exports and analyze trade patterns: <https://wits.worldbank.org/>
5. **Netherlands Enterprise Agency (RVO):** The RVO is a Dutch governmental organization that supports entrepreneurs in various fields, including international trade. They provide information on Dutch exports, trade regulations, and market opportunities. You can find relevant information on their website: <https://english.rvo.nl/>

These sources are considered reliable and authoritative, as they are maintained by official government agencies or international organizations. You can use the data from these sources for academic research, policy analysis, or business decision-making.

Settings



 General

 Beta features

 Data controls

Chat History & Training

Save new chats to your history and allow them to be used to improve ChatGPT via model training. Unsaved chats will be deleted from our systems within 30 days.

[Learn more](#)

Export data

Export

Delete account

Delete



Absolutely! I'd love to demonstrate some of the capabilities of LaTeX within Markdown, especially when it comes to mathematical expressions. Here are some examples, from algebra to calculus and linear algebra:

1. Quadratic equation: The solution to the quadratic equation $ax^2 + bx + c = 0$ is given by:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

2. The definition of a derivative in calculus: The derivative of a function $f(x)$ at a specific point x_0 is given by:

$$f'(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

3. The integral from a to b of a function $f(x)$ is denoted as:

$$\int_a^b f(x) dx$$

4. Euler's formula, a beautiful equation that connects five of the most important numbers in mathematics:

$$e^{ix} = \cos(x) + i \sin(x)$$

5. The formula for the determinant of a 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is:

$$\det(A) = ad - bc$$

6. The normal distribution, also known as the Gaussian distribution, is given by:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where μ is the mean and σ is the standard deviation.

7. The Fourier series of a function $f(x)$ is given by:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(nx) + b_n \sin(nx)]$$

where a_n and b_n are the Fourier coefficients.

I hope this gives you a good sense of the capabilities of LaTeX in Markdown for rendering mathematical equations.

LaTeX!



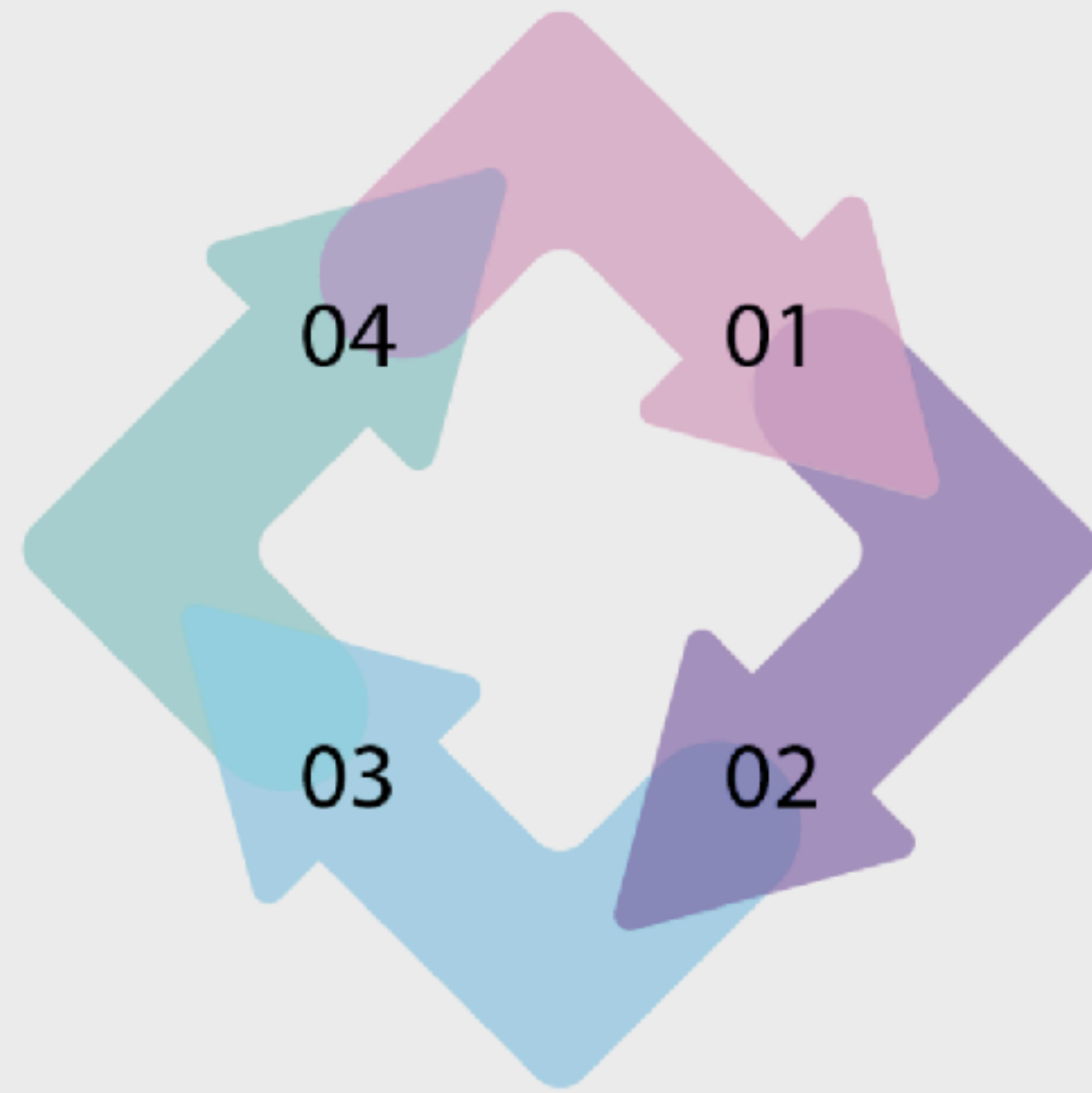
Writing up

- Improve writing quality
- Reformat citations and references
- Translate writing



Data analysis

- Code data
- Suggest themes or topics for analysis



Research design

- Generate ideas for research questions or projects
- Suggest data sources



Data collection

- Search archives and datasets
- Translate sources into other languages

Image created by UNESCO IESALC

AI SCIENCE SEARCH ENGINES EXPLODE IN NUMBER

Tools powered by large language models are intended to help researchers digest or do science.

By Katharine Sanderson

As large language models (LLMs) gallop ever onwards — among them GPT-4, OpenAI's latest incarnation of the technology behind ChatGPT — scientists are beginning to make use of their power. The explosion of tools powered by artificial intelligence (AI) includes several search engines that aim to make it easier for researchers to grasp seminal scientific papers or summarize a field's major findings. Their developers claim the apps will democratize and streamline access to research.

But some tools need more refinement before researchers can use them to help their studies, say scientists who have experimented with them. Clémentine Fourrier is a Paris-based researcher who evaluates LLMs at Hugging Face, a company in New York City that develops open-source AI platforms. She used an AI search engine called Elicit, which uses an LLM to craft its answers, to help find papers for her PhD thesis. Elicit searches papers in the Semantic Scholar database and identifies the top studies by comparing the papers' titles and abstracts with the search question.

Variable success

Fourrier says that, in her experience, Elicit didn't always pick the most relevant papers. The tool is good for suggesting papers "that you probably wouldn't have looked at", she says. But its paper summaries are "useless", and "it's also going to suggest a lot of things that are not directly relevant", she adds. "It's very likely that you're going to make a lot of mistakes if you only use this."

Jungwon Byun, chief operating officer at Ought, the company in San Francisco, California, that built Elicit, says: "We currently have hundreds of thousands of users with diverse specializations so Elicit will inevitably be weaker at some queries." The platform works differently from other search engines, says Byun, because it focuses less on keyword match, citation count and recency. But users can filter for those things.

Other researchers have had more positive experiences with the tool. "Elicit.org is by far my favourite for search," says Aaron Tay, a librarian at Singapore Management University. "It is close to displacing Google Scholar as my

first go-to search for academic search," he says. "In terms of relevancy, I had the opposite experience [to Fourrier] with Elicit. I normally get roughly the same relevancy as Google Scholar — but once in a while, it interprets my search query better."

These discrepancies might be field-dependent, Tay suggests. Fourrier adds that, in her research area, time is crucial. "A year in machine learning is a century in any other field," she says. "Anything prior to five years is completely irrelevant," and Elicit doesn't pick up on this, she adds.

Full-text search

Another tool, scite, whose developers are based in New York City, uses an LLM to organize and add context to paper citations — including where, when and how one paper is cited by

"A year in machine learning is a century in any other field."

another. Whereas ChatGPT is notorious for 'hallucinations' — inventing references that don't exist — scite and its Assistant tool remove that headache, says scite chief executive Josh Nicholson. "The big differentiator here is that we're taking that output from ChatGPT, searching that against our database, and then matching that semantically against real references," Nicholson says that scite has partnered with more than 30 scholarly publishers, including major firms such as Wiley and the American Chemical Society, and has signed a number of indexing agreements — giving the tool access to the full text of millions of scholarly articles.

Nicholson says that scite is also collaborating with Consensus — a tool that "uses AI to extract and distil findings" directly from research — launched in 2022 by programmers Eric Olson and Christian Salem, both in Boston, Massachusetts. Consensus was built for someone who's not an expert in what they're searching for, says Salem. "But we actually have a lot of researchers and scientists using the product," he adds.

Like Elicit, Consensus uses Semantic Scholar's data. "We have a database of 100-million-plus claims that we've extracted

Elicit.org

Scite.ai

Thank
you!

References

OpenAI (2023) GPT-4 Technical Report

UNESCO (2023) ChatGPT and Artificial Intelligence in higher education Quick start guide.
URL: https://www.iesalc.unesco.org/wp-content/uploads/2023/04/ChatGPT-and-Artificial-Intelligence-in-higher-education-Quick-Start-guide_EN_FINAL.pdf

<https://tilburgsciencehub.com/tutorials/more-tutorials/chatgpt-article/chat-gpt-research/>

ChatGPT-4 & ChatGPT-3.5

“As a language model AI developed by OpenAI, I can assist researchers in understanding concepts, finding ways to manage issues, suggesting relevant literature, coding the methods to deal with them, and interpreting the results. Not only that, but I can also help in generating ideas, drafting papers, and proofreading them. I can be a valuable tool for academics.”

Questions for debate

Issues for discussion at a forum about conversational AIs.

- Which research tasks should or should not be outsourced to large language models (LLMs)?
- Which academic skills and characteristics remain essential to researchers?
- What steps in an AI-assisted research process require human verification?
- How should research integrity and other policies be changed to address LLMs?
- How should LLMs be incorporated into the education and training of researchers?
- How can researchers and funders aid the development of independent open-source LLMs and ensure the models represent scientific knowledge accurately?
- What quality standards should be expected of LLMs (for example, transparency, accuracy, bias and source crediting) and which stakeholders are responsible for the standards as well as the LLMs?
- How can researchers ensure that LLMs promote equity in research, and avoid risks of widening inequities?
- How should LLMs be used to enhance principles of open science?
- What legal implications do LLMs have for scientific practice (for example, laws and regulations related to patents, copyright and ownership)?

COMMENT | 03 February 2023

ChatGPT: five priorities for research

Conversational AI is a game-changer for science. Here's how to respond.

[Eva A. M. van Dis](#), [Johan Bollen](#), [Willem Zuidema](#), [Robert van Rooij](#) & [Claudi L. Bockting](#) 

Nature **614**, 224-226 (2023)

doi: <https://doi-org.vu-nl.idm.oclc.org/10.1038/d41586-023-00288-7>

It recognises linearity of differentiation

2. (2 pt) Starting with the likelihood of the binary logistic regression

$$p(\mathbf{t}|\mathbf{w}) = \prod_{n=1}^N y_n^{t_n} \{1 - y_n\}^{1-t_n},$$

show that the negative logarithm of the likelihood gives the *binary cross-entropy* error and that the derivative of the cross entropy error is

$$\nabla E(\mathbf{w}) = \sum_{n=1}^N (y_n - t_n) \phi_n.$$

ANSWER;

$$\begin{aligned} E(\mathbf{w}) &= -\log \prod_{n=1}^N y_n^{t_n} (1 - y_n)^{1-t_n} \\ &= -\sum_{n=1}^N [t_n \log y_n + (1 - t_n) \log(1 - y_n)] \end{aligned}$$

To find the gradient of the error, we first need the derivative of the error with respect to the output y_n :

$$\frac{\partial E}{\partial y_n} = -\frac{t_n}{y_n} + \frac{1 - t_n}{1 - y_n}$$

Now, we need the derivative of the output y_n with respect to the model parameters, \mathbf{w} :

$$\begin{aligned} \frac{\partial y_n}{\partial \mathbf{w}} &= \frac{\partial}{\partial \mathbf{w}} [\sigma(\mathbf{w}^T \phi_n)] \\ &= \sigma(\mathbf{w}^T \phi_n) (1 - \sigma(\mathbf{w}^T \phi_n)) \phi_n \\ &= y_n (1 - y_n) \phi_n \end{aligned}$$

Now, we can find the gradient of the error with respect to the model parameters:

$$\begin{aligned} \nabla E(\mathbf{w}) &= \sum_{n=1}^N \frac{\partial E}{\partial y_n} \frac{\partial y_n}{\partial \mathbf{w}} \\ &= \sum_{n=1}^N \left[-\frac{t_n}{y_n} + \frac{1 - t_n}{1 - y_n} \right] \cdot y_n (1 - y_n) \phi_n \\ &= \sum_{n=1}^N (y_n - t_n) \phi_n \end{aligned}$$

It knows the
classic
+1 -1 trick
In mathematics

(1 pt) Verify that the derivative of the sigmoid

$$\sigma(a) = \frac{1}{1 + \exp(-a)},$$

is

$$\frac{d\sigma}{da} = \sigma(1 - \sigma).$$

ANSWER;

$$\begin{aligned} \frac{d\sigma}{da} &= \frac{d}{da} \left(\frac{1}{1 + \exp(-a)} \right) \\ &= \frac{\exp(-a)}{(1 + \exp(-a))^2} \\ &= \frac{1}{1 + \exp(-a)} \cdot \frac{\exp(-a)}{1 + \exp(-a)} \\ &= \sigma(a) \cdot \frac{\exp(-a)}{1 + \exp(-a)} \\ &= \sigma(a) \cdot \frac{\exp(-a) + 1 - 1}{1 + \exp(-a)} \\ &= \sigma(a) \cdot \left(\frac{1 + \exp(-a)}{1 + \exp(-a)} + \frac{-1}{1 + \exp(-a)} \right) \\ &= \sigma(a) \cdot \left(1 - \frac{1}{1 + \exp(-a)} \right) \\ &= \sigma(a)(1 - \sigma(a)) \end{aligned}$$

It's good at rewriting text

Model: GPT-4

BS

Act as: Editor

Degree of revision: Substantial Revision

Type of edit: Enhance clarity and consistency

Change style to: Academic, PhD Work

Change tone to: Analytical



Change reader comprehension level to: advanced, assume extensive prior knowledge

Change length to: 1000 Words

My Text:

Brav.. hav u chaked the gpt-4 thing?? U can writ any branles drunken drivel in it and it changes it to any level even PhD hyper brain style.. trulli insanse.. you, got to check it bro! Its a tru gift for all of us iliterate assholes around the the global flat plane! Now we can have universit jobs.. hhah crazi stuff...



Title: A Comprehensive Analysis of the GPT-4 Language Model and its Potential Impact on Scholarly Communication  

Abstract: This article explores the functionality and capabilities of OpenAI's GPT-4, a sophisticated language model capable of generating coherent and semantically rich text. We examine the model's potential to transform written communication, particularly for those with limited literacy skills, by analyzing the model's capacity for tailoring text to various academic and linguistic requirements. Furthermore, we discuss the implications of the model's wide-ranging applications, including its potential to democratize access to higher education and professional opportunities.

Introduction

I want you to be an academic in a STEM field. I want to start work on writing a paper. Can you provide me with steps I need to follow to get through the research process? Please also indicate how you can help me with each step.

Identify a Research Topic

Before you can begin writing, you need to identify a topic that is both interesting and appropriate for your field. This should be something you are passionate about, as you'll be spending a lot of time on it. I can help by suggesting potential topics based on the latest trends in your field, or by helping you refine your ideas into a specific, manageable research question.

Conduct a Literature Review

The next step is to see what has already been published about your topic. This will help you understand the current state of research, identify gaps that your work could fill, and find sources to cite in your own paper. I can assist by guiding you through databases and search engines, helping you evaluate the quality and relevance of sources, and summarizing key findings.

Develop a Hypothesis or Research Question

Based on your literature review, you should be able to develop a clear research question or hypothesis. This will guide your data collection and analysis. I can help you articulate this in a way that is both specific and researchable, and suggest ways to operationalize your variables if necessary.

Design the Research Methodology

Depending on your field, you may be conducting experiments, collecting survey data, performing simulations, or using other research methods. I can provide guidance on designing a robust methodology, including suggestions for data collection and analysis techniques.

Collect and Analyze Data

Once your methodology is in place, it's time to start gathering and analyzing data. I can provide tips for data collection, help you troubleshoot any issues that arise, and guide you through the process of analyzing your data, whether that involves statistical tests, coding qualitative data, or interpreting simulation results.

Write the Paper

After your data analysis is complete, you can start writing your paper. This typically includes sections like the introduction, literature review, methodology, results, discussion, and conclusion. I can help you structure your paper, write clearly and persuasively, and properly cite your sources. I can also help with the revision process, providing feedback on drafts and suggesting improvements.

Peer Review and Revision

Before a paper is published, it often goes through a peer review process. Reviewers may provide feedback and suggest revisions. I can assist you in understanding and responding to reviewer comments, and in revising your paper to address any issues they raise.

Publication

Finally, you'll need to select a journal to submit your paper to, and possibly go through several rounds of revisions before it is accepted. I can provide advice on choosing a suitable journal and help you navigate the submission process.