

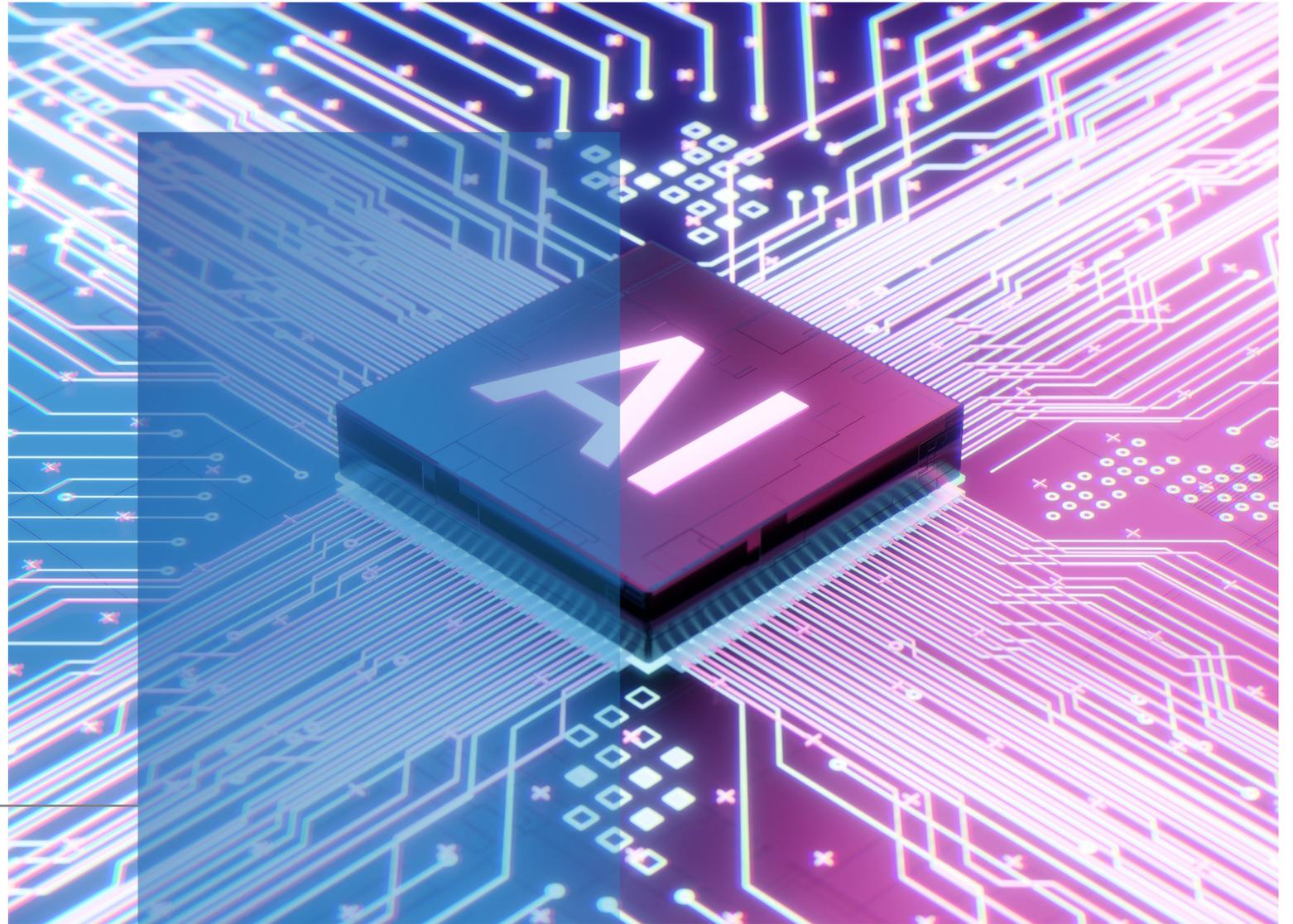


Author Success Webinar 13:

The Impact of AI on Academic Research and Publishing: Opportunities and Challenges

Dr Rachel Baron

17 August 2023



AI Writing and Editing Tools

Grammar Checking



Paraphrasing



Text Generating



What is ChatGPT?



Large language model
developed by OpenAI



Trained on large corpus of text
from the Internet



Learns to predict next word in
a sentence



Chat:
Focuses on generating
conversational responses



GP:
Generative Pre-training



Transformer:
the type of machine learning
model it's based on

How Does ChatGPT Work?

Training

Pre-training learns grammar and facts about the world. Predicts next word, e.g.: “The sky is...blue”



Fine tuning: further learning using a narrower dataset, with input from reviewers



However, it doesn't “understand” text, only patterns and relationships between textual elements

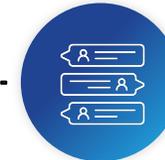


Generating

To use ChatGPT, open an account: most features available for free



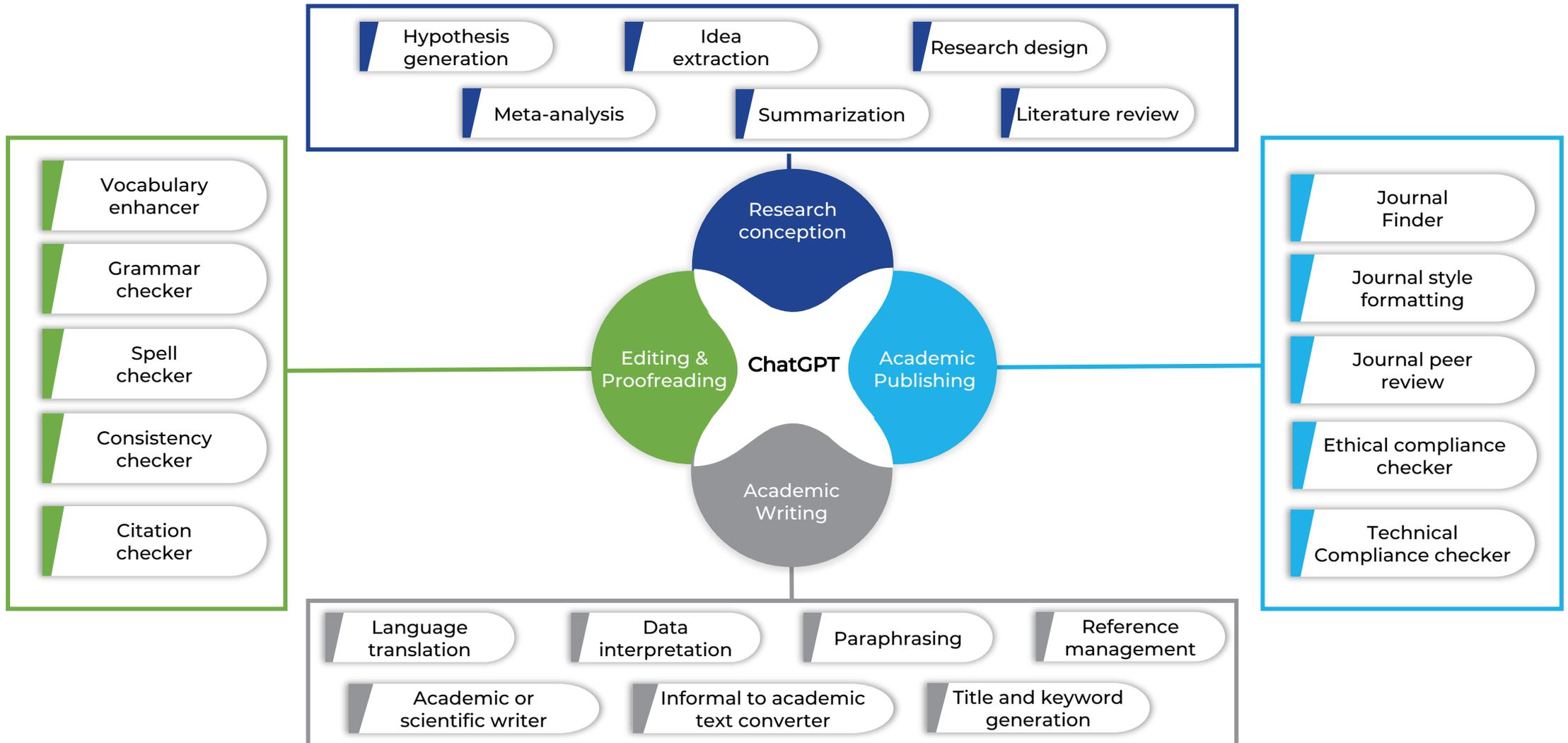
Type your question (called a prompt) into the chat box, e.g., “summarise the content of the following article...”



Writing good prompts is the **key** to good output



How Can It Be Used In Research And Writing?



Medhat Zohery. (2023). Artificial Intelligence in Academia, Research and Science: ChatGPT as a Case Study. (pp.10-61). Chapter: 2. Achatgo Publishing



Research Conception

Can you summarise the literature on xxx and identify some research gaps?

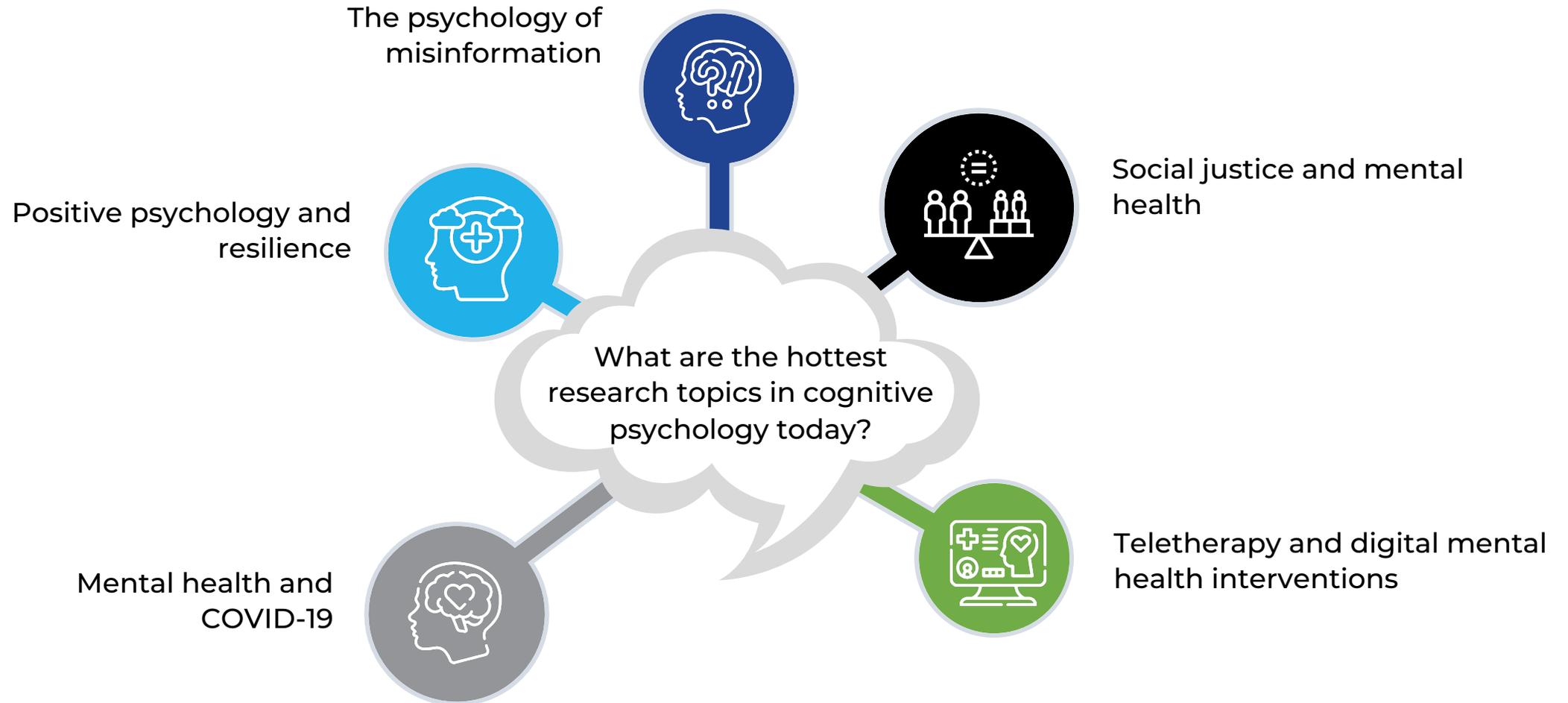
I am interested in studying xxx. Can you suggest an appropriate, novel hypothesis?

What are the hottest research topics in xxx currently?

I have the following idea for a study.... Could you suggest a suitable research design, specifying the type of study, the sample size, and the analysis methods?

Can you suggest some ideas for novel research in the area of xxx?

Example Outputs



Example: Generating A Research Question

Prompt:

“Suggest a novel research question on the effects of digital overload on cognitive functions and mental health”

Suggestion:

"What are the long-term effects of sustained digital overload, characterized by excessive screen time, information overload, and constant multitasking, on cognitive functions and mental health outcomes in different age groups?"

Explanation:

This research question aims to explore the potential consequences of prolonged exposure to digital devices and online activities on cognitive abilities (such as attention, memory, and problem-solving) and mental well-being (including symptoms of anxiety, depression, and stress). By considering different age groups, the study can examine how digital overload affects individuals across various developmental stages, such as children, adolescents, and adults. Additionally, it acknowledges that the impact of digital overload may vary depending on factors like neuroplasticity, coping mechanisms, and susceptibility to technology-induced stressors.



Example: Generating a Research Design

Study Design: Longitudinal Cohort Study

Participants:

- Sample: A diverse sample of individuals from different age groups (e.g., children, adolescents, adults)
- Sample Size: Determine an appropriate sample size based on power analysis and the specific research objectives

Data Collection:

- Baseline Assessment: Gather demographic information, digital device usage patterns, and self-reported cognitive function and mental health measures.
- Digital Overload Measurement: Use objective measures (e.g., screen time tracking apps, activity logs) to quantify the extent of digital overload experienced by participants.
- Cognitive Function Assessment: Administer validated cognitive assessment tools (e.g., attention tasks, memory tests, executive function measures) to evaluate participants' cognitive abilities.
- Mental Health Assessment: Administer standardized self-report measures (e.g., anxiety, depression, stress scales) to assess participants' mental health status.
- Follow-up Assessments: Conduct regular follow-up assessments at predetermined intervals (e.g., every six months or annually) to track changes in cognitive functions and mental health over time.

Data Analysis:

- Quantitative Analysis: Utilize statistical methods (e.g., regression analysis, correlation analysis) to examine the relationship between digital overload, cognitive functions, and mental health outcomes. Control for potential confounding variables (e.g., socioeconomic status, pre-existing mental health conditions).
- Subgroup Analysis: Analyze data separately for different age groups to investigate potential age-specific effects of digital overload.
- Longitudinal Analysis: Analyze changes in cognitive functions and mental health measures over time using appropriate longitudinal data analysis techniques (e.g., mixed-effects models).

Ethical Considerations:

- Obtain informed consent from participants and ensure confidentiality of their data.
- Comply with ethical guidelines and regulations for research involving human subjects.

Limitations:

- Self-report measures may be subject to bias and may not capture the full extent of digital overload.
- It may be challenging to establish causality due to the complex nature of digital overload and potential confounding factors.
- Attrition and participant dropouts over the course of the longitudinal study may affect the generalizability of the findings.



How Good Are Its Suggestions?



- Highly useful for new students and early career researchers in developing a basic understanding of complex topics. Also helpful for overcoming language barriers and making research more accessible
- Can provide inspiration and ideas - prompts you to think outside the box
- Can help to cross-link between disciplines - you could ask it to suggest how developments in one field relate to findings in another



- Level of specificity and novelty rather low
- Lahat et al. (2023) investigated the use of ChatGPT for generating research ideas in four gastroenterology topics
 - Topics reviewed by experts – found them to be excellent in terms of clarity and relevance, but not particularly original or specific
 - Newer versions of CGPT might perform better

Plug-Ins Can Improve Quality of Output



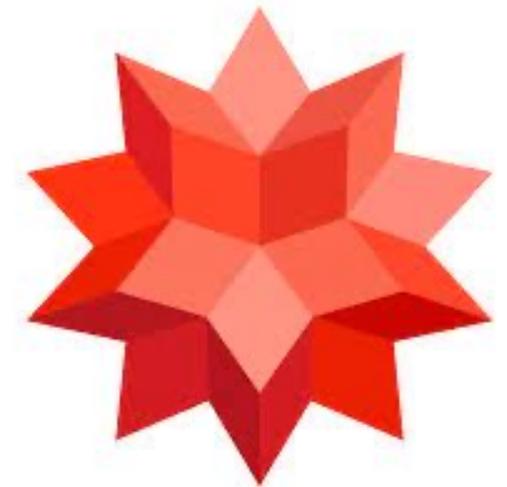
ScholarAI



BrowserPilot



AiPDF



Wolfram

Plug-ins Can Improve Quality of Output



ScholarAI prompt: Based on the recent psychology research literature, what would be an interesting research question for a cognitive psychologist interested in the effects of digital overload on cognitive functions and mental health?

1. "Brief Digital Interventions: An Implementation-Sensitive Approach to Addressing School Mental Health Needs of Youth with Mild and Emerging Mental Health Difficulties" by Eli Cwinn et al. [Full Text](#) - This study discusses the use of brief digital interventions (BDIs) to address mental health needs of students. A potential research question could be: "How can BDIs be optimized to mitigate the effects of digital overload on cognitive functions and mental health in students?"
2. "Usefulness of Implementation Outcome Scales for Digital Mental Health (iOSDMH): Experiences from Six Randomized Controlled Trials" by E. Obikane et al. [Full Text](#) - This study discusses the use of the Implementation Outcome Scale of Digital Mental Health (iOSDMH) in evaluating digital mental health interventions. A potential research question could be: "How does the implementation of digital mental health interventions impact cognitive functions and mental health in the context of digital overload?"

RA

can you summarise the following article: <https://www.nature.com/articles/s41746-023-00819-6#citeas>



Used Link Reader UNVERIFIED



The article titled "Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers" published in npj Digital Medicine discusses an experiment conducted to assess the ability of OpenAI's ChatGPT to generate convincing scientific abstracts. Here's a summary of the key points:

1. **Background:** The authors note that large language models (LLMs) like ChatGPT, which is based on GPT-3, can generate coherent and fluent text that can be hard to distinguish from human-written text. Given the increasing use of AI in various fields, including medical technologies and scientific writing, they highlight the need to determine if ChatGPT can write convincing medical research abstracts.
2. **Methodology:** The authors collected 50 abstracts from five high-impact journals as a control group of well-written abstracts. They then asked ChatGPT to generate 50 scientific abstracts based on the titles and specific journals from this list.
3. **Findings:** The AI output detector they used found a high probability of AI-generated output in the generated abstracts. Blinded human reviewers were able to correctly identify 68% of generated abstracts as being generated, and correctly identified 86% of original articles as being original. However, they incorrectly identified 32% of generated abstracts as being real, and 14% of original abstracts as being generated.

Regenerate

Send a message



ChatGPT may produce inaccurate information about people, places, or facts. [ChatGPT August 3 Version](#)

4. **Limitations:** The study's limitations include its small sample size and few reviewers. The authors also note that ChatGPT is sensitive to small changes in prompts, and they did not exhaust different prompt options or deviate from their prescribed prompt.
5. **Implications:** The authors suggest that this technology could be used both ethically and unethically. For instance, it could be used by organizations such as paper mills to entirely falsify research. On the other hand, it could be used in conjunction with a researcher's own scientific knowledge as a tool to decrease the burden of writing and formatting.
6. **Recommendations:** The authors recommend clear disclosure when a manuscript is written with assistance from large language models such as ChatGPT. They also suggest that AI output detectors may be one tool to include in the research editorial process, depending on the publisher or conference's guidelines.
7. **Conclusion:** The authors conclude that while ChatGPT is a powerful tool to create readable scientific abstracts, it requires careful review for factual accuracy. The optimal use and ethical boundaries of AI-generated writing remain to be determined as discussion within the scientific community evolves.



Writing a Literature Review



Potentially very useful –

- extensive knowledge base
- highly effective in synthesising complex research findings
- can evaluate the quality of studies, including design, methodology, sample size and statistical analyses



But ChatGPT doesn't search the Internet

- developments after 2021 will not be included



Example Prompts for Literature Review

Summarise this paper in about 500 words

What are the strengths and weaknesses of this study?

What contribution does this study make to the field of xxxx?

What are the main findings of this study?

How do the findings of these two studies differ?

What are the implications of this study for government policy?

In what way can these findings be applied to education?

Example Prompts for Literature Review

Summarize the main methodologies used in studies on xxx

Summarise the timeline of developments in the field of xxx

What theoretical frameworks have been developed in the field of xxx?

How has theory xxx been extended over the past decade?

Discuss "author's" contribution to the field of xxx

Summarise the strengths and weaknesses of xxx's theoretical model

How do the findings on xxx differ between Western and Eastern context?

Potential Problems



Hallucinations: content may be inaccurate and needs to be fact-checked



Misses nuances and complexities: review may seem superficial



Can reproduce prejudices and biases from original dataset

Example Of A Hallucination



“We queried ChatGPT about engorged female hard ticks and where they might choose to lay their eggs. We were particularly interested in how far they travel while looking for an oviposition site. This was the generated response:

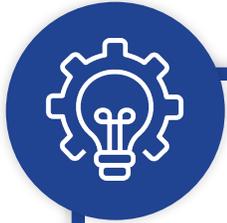
In one study, researchers found that engorged female *Amblyomma americanum* ticks can travel up to several hundred meters before selecting a suitable site for egg laying. However, the majority of the ticks tended to stay within a few meters of the location where they fed. The study also found that the distance traveled by the ticks was influenced by their physiological state, with ticks in better physiological condition being more likely to travel greater distances.”

When asked to provide a reference for this misleading statement, it made one up:

Burtis, J.C., Sullivan, P., Levi, T., Oggenfuss, K., Fahey, A., Gaff, H.D., Falco, R.C., Fish, D. 2019. Off-host physiological state of an ixodid tick species impacts the distance at which it engages in host-seeking and likelihood of subsequently attaching. *Parasites and Vectors* 12, 326. <https://doi.org/10.1186/s13071-019-3565-1>

Goddard (2023). Hallucinations in ChatGPT: A cautionary tale for biomedical researchers. <https://doi.org/10.1016/j.amimed.2023.06.012>

Potential Ethical Problems



Inadvertent plagiarism
Production of fictitious
references

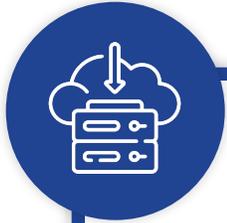


Ethical concerns over
authorship
ChatGPT can't be listed
as author



AI use needs to be
acknowledged
Potential questions
about academic
seriousness and integrity

Potential Data Security and Privacy Problems



Data input as prompts may be subject to leakages and breaches of privacy



Cyber attacks: Potential for AI systems to be hacked or manipulated



Potential loss of intellectual property rights

Can AI-Written Text Be Detected?



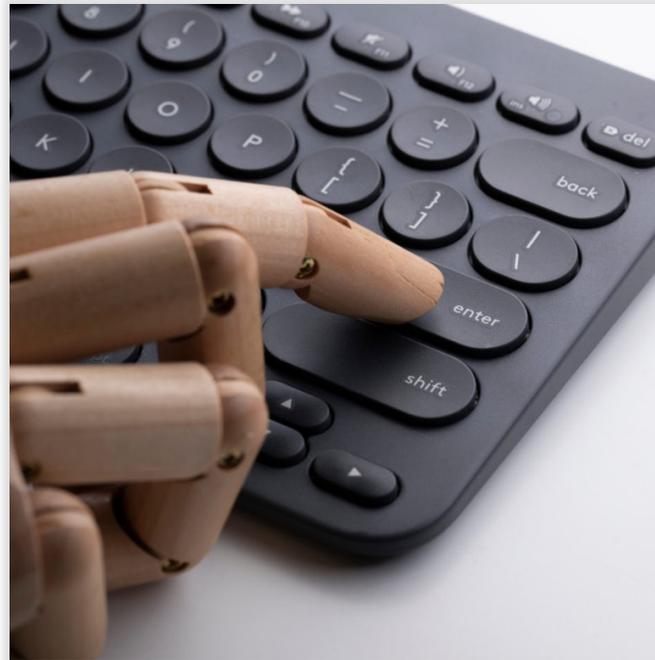
Gao, C.A., Howard, F.M., Markov, N.S. *et al.* Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *npj Digit. Med.* **6**, 75 (2023). <https://doi.org/10.1038/s41746-023-00819-6>

Using AI For Editing And Proofreading

AI-edited text
reads very well



- Corrects almost all grammar
- Sophisticated vocabulary
- Improves syntax



Many challenges
to overcome



- Requires sophisticated prompts
- Introduces as many problems as it solves
- Doesn't yet save time or effort

Editing and Proofreading

Original



The link between student engagement and student learning is underexplored and deserving of greater attention. In surveying the range of studies of real-time response systems, the principal conclusion is that student engagement is the most apparent benefit (Wood and Shirazi 2020: 13). Student engagement via SRS improves the classroom dynamic, but does it actually spur learning? There is a debate on this question and cautionary voices abound. Dangel and Wang (2008) point out how difficult it is to improve critical thinking skills when using SRS. Graham et al. (2007: 236-237) see little crossover between student engagement and student achievement. Newland and Black's (2020: 160) reading of the literature is more ambivalent, suggesting that the evidence is mixed on whether SRS promote significant learning.

Edited



The connection between student engagement and student learning deserves greater attention. Although student engagement is the most apparent benefit of SRS, its impact on learning remains a subject of debate. Some cautionary voices, such as Dangel and Wang (2008) and Graham et al. (2007: 236-237), suggest limited crossover between engagement and achievement. Newland and Black (2020: 160) present a more ambivalent reading of the literature, indicating mixed evidence regarding the promotion of significant learning through SRS.



Problems With Editing Text



- Wholesale deletion of text
 - sometimes an entire lit review replaced with a brief summary
- Tendency to over-edit in general
 - many edits subjective and unnecessary
 - too many synonyms changed for the sake of it
 - details removed, e.g., sample size, details of experimental design
- Sometimes changes the meaning, especially when original is poorly written or ambiguous
- Changes the author's style
 - e.g., introducing first person (we)

Problems With Editing Text

- Changes in verb tense, often introducing and/or not correcting errors and inconsistencies
- Tendency to remove citations
- Paraphrasing of direct quotes and citations deleted (plagiarism!)
- Repeats or introduces errors that our editors would correct
 - use of “on the other hand” without “on the one hand”
 - poor use of definite articles
- Does not query things that a human editor would
 - e.g., ambiguities in the text
 - factual queries (e.g., dates)
 - missing information
 - missing references
- Doesn't correct inconsistencies as well as a human editor



The sample size

~~A total of 70 chronic patients will be recruited for the proposed project. The sample size is determined by G*power (G*Power v3.1.2; Universität Düsseldorf, Germany) based on one previous study for acupressure on sleep quality (Zheng et al., 2014). The mean difference of sleep quality (i.e., Pittsburgh Sleep Quality Index; PSQI) between groups is 2.44, and the pooled SD is 7.74 with the effect size $d = 0.63$. To detect results with calculation, based on a previous study on acupressure for sleep quality, determined that 35 participants are required per group to achieve 80% power with a one-tailed probability level of 0.05, and assuming a 10% dropout rate, 35. Eligible participants for each group are required.~~

~~Participants are eligible if they are: (1) must be aged between 55 and 85 yrs; (2) years, diagnosed with stroke by magnetic resonance imaging or computed tomographic scan longer than one year; (3) able to walk 6 m independently; (4) scored ≥ 18 but less or equal than 27 (Tombaugh & McIntyre, 1992; Zhu et al., 2020) in mini-mental state examination (MMSE); (5) self-reported poor sleep quality (PSQI, scores ≥ 6) in the past four weeks (Tsai et al., 2005).~~

~~Participants will be excluded if they: (1) have a able to walk independently for 6 meters, have a Mini-Mental State Examination (MMSE) score between 18 and 27, and self-report poor sleep quality in the past four weeks (Pittsburgh Sleep Quality Index score ≥ 6). Participants with specific exclusions, such as cardiac pacemaker; (2) have a use, severe disease that precludes the receipt of Acu-TENS; (3) are taking diseases, medication that may affect measured outcomes; (4) have interference, skin lesions, infection, or inflammation near selected acupoints; (5) are participating, or participation in other drug/treatment programs. will be excluded from the study.~~

Supporting information
about sample size
deleted

Details about eligibility
deleted

Some detail reinstated,
but without references

In Summary: Why We're Not Yet Converts!

Quality Issues

- No time saving due to excessive amount of checking needed
- Possibility of introducing plagiarism
- High probability of introducing factual errors or distortions
- Tendency to over-edit
- Makes too many assumptions and doesn't query text

Practical Issues

- Technical problems such as word length limitation
- Tracked changes not shown so need to perform comparison in Word
- Data leakage concerns
- Ethical concerns and transparency

Thank You!

Any Questions?

