

# Generative AI Should Not Be Allowed in Psychological Research

John R. Rossiter

Faculty of Business and Law, University of Wollongong, Wollongong, Australia  
Email: jrrossiter14@gmail.com

**How to cite this paper:** Rossiter, J. R. (2025). Generative AI Should Not Be Allowed in Psychological Research. *Psychology*, 16, 521-531.  
<https://doi.org/10.4236/psych.2025.164030>

**Received:** April 1, 2025

**Accepted:** April 24, 2025

**Published:** April 27, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

The author explains in this article why generative artificial intelligence (genAI) should be banned from use in psychological research. It is a form of cheating and therefore unethical because it represents artificial aiding of human effort and draws on the uncredited work of others. Even if this prohibition is ignored, as it has been by most psychology journals to date, I argue that genAI should be banned because it degrades the quality of scientific research and is likely to undermine the learning of good research skills. I will exemplify genAI's inadequacy for research with a critique of two recently published studies investigating its capacity to generate new hypotheses in psychology.

## Keywords

Generative AI, Research Assistance, Writing Assistance, Moral Ethics, Utilitarian Ethics

---

## 1. Introduction

Generative artificial intelligence, genAI, is threatening the very foundations of psychological research. This is due to the ability of genAI computer programs to manufacture fluent text and even whole essays or research reports at the press of a button and accompany it with fake images and graphs and made-up references (Clark, 2025; Spencer, 2022). Generative AI, it must be emphasized, is very different from what can be called robotic AI, as used in medical diagnosis and surgery, in manufacturing and production, and in related technical industries. Here, my focus is on generative AI only, and specifically on the use of text-based, so-called large language AI models (such as Open AI's ChatGPT or Microsoft's Copilot) for assisting researchers to design research studies, but also on language-correcting AI programs (such as the one built into Microsoft Office) for assistance in writing

research reports.

Research-assistance genAI programs and writing-assistance genAI programs raise moral ethical issues, and the research assistance form of genAI raises utilitarian ethical issues. The principle of moral ethical theory is “Do the right thing, regardless of its consequences” whereas the principle of utilitarian ethical theory is “Do the most effective thing, regardless of its morality” (see, e.g., [Blackburn, 1996](#)). These two ethical principles are frequently in conflict, and the best resolution, in my view, is the joint principle proposed by [Ross \(1955\)](#) which is that “The morality of the action comes first, but utility overrides it if the action is likely to result in substantial harm.” As will become evident in this paper, I believe that the use of generative AI in research is both morally wrong and harmful to psychological research.

I will firstly review the moral ethical case for banning genAI from use by university students in research-based assignments, and by researchers in the preparation and writing of research papers. Secondly, I will turn to the utilitarian ethical case for banning genAI, which I will illustrate by exposing the problems in two recent studies on the use of generative AI to produce new hypotheses in psychology. In both cases, moral and utilitarian, I will spell out what I believe should be done.

## 2. GenAI Use by University Students and Researchers

Western universities have basically “rolled over” on the use of genAI by students and by researchers.

### 2.1. Use by University Students

The use of genAI by university students is important to prohibit because if students get into the habit of using it at university, those who go on to pursue a research career may find it much too hard to resist using it and may not have learned the skills to do research on their own.

Universities’ policies on students’ use of genAI differ only in minor details, and universities overall are clearly favorable toward its use. See [Table 1](#) for examples of student policy at four leading western universities (policies current as of March, 2025). Most of them appear to ban its use for tests and exams, but all allow its use for assignments and group projects. [Stanford University \(2023\)](#) and [Harvard University \(n.d.\)](#) in the U.S. seemingly place no restrictions on its use, whereas the [University of Oxford \(2025\)](#) in the U.K. and the [University of Sydney \(2024\)](#) in Australia allow students to use it as a research aid and writing aid provided it is used “ethically” and “appropriately” and the nature of its use is disclosed. But nowhere are “ethical use” and “appropriate use” explained, and disclosure is certainly no guarantee against misuse.

The temptation for students to use genAI is very great. The purveyors of student-serving websites based on genAI have long become adept at turning out essays and reports in which the quality of writing is kept merely at a “good pass”

**Table 1.** Leading universities' policies on student use of Generative AI (websites accessed March, 2025).

Institution	Student Policy	Instructor Policy
Stanford University	GenAI banned for assignments and exams.	Instructor can allow
Harvard University	Banned in some schools or faculties for data security reasons but allowed in others with a disclosure statement	Instructor can allow
University of Oxford	Allowed if used “ethically” and “appropriately.” Students (and teachers) are required to become “AI literate,” and students must differentiate AI-assisted work from their own work. Can use as “research assistant” and “writing assistant.”	Instructor <i>must</i> allow
University of Sydney	Allowed to use only the University’s secure version of Microsoft Copilot for assignments. Can use as “research assistant” and “writing assistant,” but usage must be the disclosed. Banned for tests and exams.	Instructor can allow it for tests and exams

level so as not to arouse too much suspicion about sudden improvements in the student’s performance (Barnes & Filipovic, 2023). It seems that genAI-based translation and writing programs are used in Australia primarily by foreign students who are limited in their reading and speaking of English (Panagopoulos, 2024) and would probably have trouble passing university subjects without it. Many domestic students, however, also use it because, as one student told me confidentially, they fear that they will fall behind their suspected AI-using peers if they do not use it.

The utilitarian justifications put forward by universities for allowing the use of genAI by students mostly fall into two categories—both of them dubious. One frequently heard justification is that genAI will “support their [students’] learning” (the University of Melbourne’s Deputy Vice Chancellor, quoted in Panagopoulos, 2025: p. 12). How genAI could possibly support learning is unclear, and it is far more likely that students use it to *avoid* learning. The other often heard utilitarian justification is that students need to be taught to be “digitally literate” so that they “will have the skills and knowledge to adapt and thrive in a changing world” (University of Sydney, 2024: p. 1) or simply because they “will have to use it in their future jobs” (the rationale given to me at my university, the University of Wollongong, when a couple of years ago I began questioning senior administrators about allowing AI use). But academics know only too well that by the time students enter university they are far ahead of the instructors in terms of digital literacy (e.g., Lee et al., 2024) and will not have any difficulty using genAI in their future jobs if they see an advantage in doing so.

## 2.2. Use by Researchers

Government research funding organizations in western countries allow the use of

genAI for the preparation and writing of research grant applications (Godwin et al., 2024). These include the National Science Foundation (no date) and the National Institutes of Health (2024) in the U.S.; the UK Research and Innovation organization (2024) which is the main research funder in the United Kingdom; and the Australian Research Council (2023) and the National Health and Medical Research Council (2023) in Australia. And, with some largely unenforceable restrictions, all western universities as far as I can tell allow it for the planning and writing of research articles. It is not surprising, therefore, that most researchers consider its usage to be ethically appropriate.

Researchers are allowed to use genAI in *all stages* of research, from planning through to research writing and reporting for publication. Universities Australia, the federal government-associated advisory body for university and university-associated researchers in my country, is typical, and even gives a list of the research stages for which research assistance AI and writing assistance AI is allowed (Universities Australia, 2025). These stages are summarized in Table 2, together with my comments. One of the most pernicious uses of genAI is in the third research stage where researchers can now go to any of the several online services that summarize research papers and articles so that researchers do not have to read or even understand them. Obviously, I am not going to provide any links to such services but I see that one service is claiming it be used by researchers at

**Table 2.** Universities Australia's (2025) advice to researchers on the use of genAI at various stages of research, with my comments and recommendations.

Research Stage	Comment
1. Finding research ideas (including hypothesis generation)	GenAI totally inadequate (see later in this article)
2. Finding background literature and references	ChatGPT, CoPilot and similar programs are too untrustworthy to be used. Search AI programs such as Google Search programs are useful for a wide media search, and Wikipedia is best for locating academic references on a given topic.
3. Evaluating and summarizing the background studies	GenAI programs should not be used. Researcher should do both.
4. Data analysis	Wikipedia is the most convenient source for evaluating alternative methods of data analysis. GenAI data analyses should never be trusted. GenAI is helpful only for preparing data tables and graphs.
5. English language translation assistance	Not permissible because of the risk of substantially changing content. The ethical thing to do is to add a native English-speaker as co-author.
6. Editing and writing assistance	Must be acknowledged and not go beyond spelling and grammar. Recruit a specialist in the field as co-author if needed for content.

leading U.S. universities including Stanford, Harvard, and MIT. Use of genAI at any of the four research planning stages is, in my view, unethical. So too is the undisclosed overuse of writing-assistance AI in the two research reporting stages.

The use of genAI in research is clearly unethical. It has led to a rash of plagiarism charges and an unprecedented number of research article retractions, even in the top scientific journals such as *Nature* and *Science* (Subbaraman, 2025). Most journals have neglected to publish an unambiguous policy concerning the use of AI, and those that have are issuing ineffective warnings such as that researchers must ensure the accuracy of information and disclose the use of AI wherever it has been employed. Going further, the publisher of many of our leading journals in psychology, the *American Psychological Association* (2025), offers for-credit continuing education programs on the use of genAI, arguing that psychology researchers must learn to master this rapidly evolving technology so that they can provide human oversight over its use. This is a naive position to adopt, in my opinion, because the training of researchers in the use of genAI will surely encourage its unethical use.

GenAI use for assistance with research and writing also will cause massive utilitarian harm. We are already seeing a lowering of trust in psychological research findings (e.g., Wu et al., 2023) and this can only get worse with the widespread use of genAI. Trust is reducing because genAI programs have been shown to have extremely high error rates when used for literature searches and research summaries (Nolan, 2025). ChatGPT Search, for example, is estimated to provide false information 67% of the time, and Elon Musk's Grok 3 an almost incredible 97% of the time. GenAI programs also fabricate references, according to Nolan, with well over half the citations on Google's Gemini search program and even more on Grok 3 leading to false sources or broken and unusable links.

A final utilitarian problem with genAI is that it is incapable of *evaluating* research. This is because all it can do is pull together evaluations made by other researchers, evaluations that are anonymous so you cannot check them and are of varying and questionable quality. Furthermore, researchers who rely on genAI to conduct evaluations will never learn how to evaluate research for themselves.

Researchers' poor ability to evaluate research is nowhere better illustrated than in recent studies using genAI to produce new hypotheses in psychology.

### 3. GenAI Use for Hypothesis Generation in Psychology

In this section, I will provide a critique of the only two published studies that I could find on AI use for hypothesis generation in psychology, as of March, 2025. Both studies, it should be noted, were reported on favorably by the British Psychological Society (Young, 2024) and the second one, by Banker et al. (2024), appeared as the lead article in the *American Psychologist*, the major journal of the American Psychological Association, and was praised in editorially invited commentaries in the same issue.

### 3.1. The Study by Tong et al. Published in *Humanities & Social Sciences Communications*

Tong, Mao, Huang, Zhao, and Peng (2024) used genAI to produce hypotheses about the causes of psychological well-being. Here is what Tong et al. did. They fed into ChatGPT-4 the text from over 43,000 open access English-language research articles on well-being research in psychology, instructing the program to look for keywords such as “psychol” and “clin psychol” in the title or abstract of the article. They then instructed GPT-4 to search the full text of each article for concepts related to the focal concept, psychological well-being, and to identify and list pairs of cause-and-effect concepts in those articles. This step produced the incredibly large number of 197,000 concepts and 235 apparently causal relationships between pairs of the concepts. The researchers then selected the 20 most frequently occurring concepts for their study along with their most frequent cause-and-effect pairings. These concepts pairs were converted by the researchers into “causal graphs,” or network diagrams, that began with psychological well-being and linked it to other concepts as presumed direct causes, which in turn were pairwise linked to further concepts as causes of the direct causes (see Figure 1 on p. 3 of their article). As the final step, the researchers instructed GPT-4 to convert the most frequently found linkage paths into fully worded hypotheses (see Table 3 on p. 5 of their article). A typical hypothesis and its origin in GPT-4 is shown in **Table 3**, together with my analysis of what this hypothesis means when broken down into its causal components.

**Table 3.** Example of the derivation of a GenAI-produced hypothesis in the “psychological well-being” study by Tong et al. (2024).

The Process	Example
1. GenAI identifies pairs of concepts to form a causal network graph	Pandemic → Online social connectivity Pandemic → Access to well-being resources Online social connectivity → Resilience Access to well-being resources → Resilience Resilience → Psychological well-being
2. The hypothesis is worded by GenAI	“Online social connectivity and access to well-being resources can build ‘virtual resilience’ and enhance psychological well-being during stressful events like pandemics”
3. What the hypothesis means in terms of its causal components (Tong et al. failed to do this)	Conditional variable: Stressful event, such as a pandemic Causal variables: Online social connectivity and Access to well-being resources Mediating variable: Resilience Dependent variable: Psychological well-being

As can be seen in the second panel of the table, the hypothesis is complex but is made to look less so by the smooth continuous wording that GPT-4 came up with. Its complexities are exposed in the bottom panel where I show that there is a con-

ditional variable, two jointly necessary causal variables, and one mediating variable before we get to the dependent variable of psychological well-being. The hypothesis, moreover, would be very difficult to test. First, researchers would have to agree on the most appropriate definition of each of the five variables as well as agree on the best measure of each, and then place these into a longitudinal time-series panel survey with a lagged correlational analysis to infer the hypothesized causality and estimate the forms, linear or otherwise, and magnitudes of the relationships. Lastly, the hypothesis is far from being insightfully new and is not much more than common observation would suggest.

Whereas I have focused on the scientific utility of Tong et al.'s study, I also question it on moral ethical grounds. It is hard to believe that the authors—who, as far as I can tell, are Chinese native speakers—could fully understand and evaluate GPT-4's summaries of the English language journal articles, and the poor and often quaint spelling and often awkward grammar throughout their article suggest that they used, unacknowledged, an AI translation program in the writing of it.

### 3.2. The study by Banker et al. published in the *American Psychologist*

Banker, Chatterjee, Mishra, and Mishra (2024) employed a large-language AI program to generate new hypotheses in social psychology that were then compared with social psychological hypotheses designed by humans. Their method was as follows. They firstly trained ChatGPT-3 on more than 100,000 social psychology research abstracts and asked it to generate 300 hypotheses from the abstracts and 300 new hypotheses. They then selected independent sets of 30 hypotheses containing 15 of each type, listed the hypotheses in random order, and presented one set each to 47 presumed social psychology research experts, 21 of whom had PhDs and 26 of whom were PhD students, asking them to rate, completely blind as to where the hypothesis came from, the 30 hypotheses on the dimensions of clarity, originality, and impact, after first being given four practice hypotheses drawn from the larger pool to rate in order to familiarize themselves with the rating procedure.

As is becoming all too common with research articles these days, Banker et al. placed vital details in an online appendix, where only a few diligent researchers would bother to locate and check them. In the appendix, Banker et al. gave numerous examples of the hypotheses but did not give any examples in the published paper. However, their online appendix (p. 11) does reveal the four *practice* hypotheses, two of which presumably are AI-generated “new” hypotheses and two are human-derived existing hypotheses. These four practice hypotheses are shown in **Table 4**, with my comments alongside.

The vague nature of the test hypotheses and the confusing rating instructions combined to make Banker et al.'s comparison of the AI and human hypotheses meaningless. The rating instructions asked the raters “to focus on the proposed variable relationship being described” (online appendix, p. 11) but in each

**Table 4.** The practice examples used in Banker et al.'s (2024) study, with my comments. (The slash divisions are mine and were not in the original.) It was not revealed to the raters (or to the reader) whether the hypothesis was AI-generated or a human-generated existing hypothesis, but presumably there are two of each.

	Hypothesis	Comments
H1	That people who are high in both conscientiousness and agreeableness/ are sensitive to the social climate/ and react more positively to favorable than unfavorable social interaction	The instructions said to focus on “the proposed variable relationship” but there are four variables and six relationships: the additive relationship between the two independent variables, each of their relationships with the mediating variable, its relationship to the dependent variable, and the two direct relationships between the independent variables and the dependent variable. An impossibly complex hypothesis.
H2.	That the gender differences in aggression/, both physical and verbal/, are present from the youngest of children to the oldest of adults	Firstly, <i>are</i> there gender differences, and if so, are males favored or are females favored? And what if the gender differences differ for physical and verbal aggression? An inadequately worded hypothesis.
H3.	That a person’s perception of the self/ is less socially defined/ when he or she holds a complex versus a simple schema of others	What does “socially defined” mean, and what does a “schema of others” mean? A hopelessly vague hypothesis.
H4.	That feelings of autonomy/promote future self-control/but that feelings of relatedness/promote future self-control to a greater extent	This hypothesis involves variables that are far too general and hard to measure.

hypothesis there are multiple variables and therefore more than one relationship, rendering it completely arbitrary for the rater to decide which is supposed to be the main one. Furthermore, as in Tong et al.'s study, the hypotheses are complexly worded—try reading them without the slashes that I inserted—and remember that there were 30 of these to be read and evaluated by each rater in addition to the four practice hypotheses. And the hypotheses are all too hopelessly vague for it to be worthwhile comparing them as Banker et al. did.

In summary, AI-produced hypotheses are useless. In the next section, I will argue that we do not need new hypotheses in psychology anyway.

#### 4. No Need for New Hypotheses

Hypotheses are of little scientific value unless they follow from, or result in, a comprehensive theory, and theories are totally beyond the capability of AI to formulate and to evaluate. We have been devoid of new theories in psychology (and in psychiatry, for that matter) for more than 40 years. In the field of learned behavior,

for example, the major theories of Pavlov on classical conditioning and Thorndike and Skinner on operant learning were first formulated in 1906 and 1938 respectively (see the excellent introductory psychology textbook by [Westen & Kowalski, 2005](#)) and the only theoretical advances since then arguably have been the Rescorla-Wagner 1972 theory of compound conditioning and Premack's 1965 operant learning theory which radically proposed that desirable behavior can be used to reinforce less desirable behavior. This neglect of learning theory is unforgivable because if the behavior is not biologically caused, then it must be learned (see [Rossiter, 2022](#)). The situation is similar in the field of social psychology, the focus of Banker et al.'s study. The main social psychology theories listed in Westen and Kowalski's textbook are Allport's 1937 theory of personality, Asch's 1946 theory of conformity, Festinger's 1957 theory of cognitive dissonance, and Milgram's 1965 theory of obedience to authority, and I maintain that the only substantial social psychological theories developed since then are Fishbein and Ajzen's 1975 theory of reasoned action and Cialdini's 1984 theory of social influence (see [Rossiter et al., 2018](#)). Psychologists these days do not seem willing to expend the enormous amount of effort needed to develop and test new theories. Instead, they have shifted their attention to easier and quicker to do micro-studies to build up their publication output as rapidly as possible, or they have fallen back on meta-analyses that do no more than combine the results of previous studies. Psychology (and psychiatry I should add) will not advance unless we can come up with new theoretical explanations for the many big questions that remain unanswered.

Finally, many accepted hypotheses are supported only by single or limited studies often using very questionable measures, which makes nonsense of popular utterances such as "evidence-based" or "the science tells us." Not only do we need new hypotheses, we researchers should be doing much more to try to disprove *existing* hypotheses because, as the philosopher of science Karl Popper said long ago (e.g., [Popper, 1983](#)), this is the way that science advances. And we need intelligent human brains, not computer brains, to do this.

## 5. Conclusion

At no stage of the research process is generative AI ethical or useful, and it should be banned. Universities should not be teaching it and encouraging students to use it because not only is it an invitation to cheat, it is also likely to undermine actual learning and be carried forward into students' professional careers. Likewise, genAI should be banned from use by researchers, and I would like to see our journals require a statement by the authors in every article that it was not used at any stage.

Even if this ethical prohibition is ignored, as seems to be the trend among government funding organizations and research journals at present, there is still the utilitarian argument for banning generative AI in psychological research. It is inaccurate and cannot be trusted. Neither is genAI in any way useful for researchers because they will learn nothing from it. This was illustrated in the present article

by the worthless studies that used genAI for hypothesis generation.

Even if genAI is banned for research assistance, there remains the question of whether it should be allowed for writing assistance. My view is that should not be allowed because there is far too much risk that AI will be used to modify content rather than merely improve spelling and grammar. If it is used, I believe the authors should include a statement that it was used to improve structure only. Alternatively, if substantial improvements in the content need to be made, the best policy is to add an English native-speaking researcher as a co-author.

## Disclosures

I would like to disclose that an earlier version of this paper criticizing the Banker et al. study and critical of the Editor of the *American Psychologist* for accepting it was rejected by that same editor as “not acceptable for publication”.

I would also like to disclose, of course, that generative AI was not used anywhere in present paper.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- American Psychological Association (2025). APA Guide for Assessing AI Tools. *Monitor on Psychology*.  
<https://apa.org/topics/artificial-intelligence-machine-learning/ai-research-writing>
- Australian Research Council (2023). *Policy on Use of Generative Artificial Intelligence in The ARC's Grants Programs*.
- Banker, S., Chatterjee, P., Mishra, H., & Mishra, A. (2024). Machine-Assisted Social Psychology Hypothesis Generation. *American Psychologist*, 79, 789-797.  
<https://doi.org/10.1037/amp0001222>
- Barnes, L., & Filipovic, M. (2023). Robot Overlords Are Here but Haven't Conquered Human Expertise Yet. *The Australian*.
- Blackburn, S. (1996). *Oxford Dictionary of Philosophy*. Oxford University Press.
- Clark, J. (2025). *Artificial Intelligence in Research*. Universities Australia.  
<https://universitiesaustralia.edu.au/policy-submissions/research-innovations/artificial-intelligence-in-research/>
- Godwin, R. C., DeBerry, J. J., Wagener, B. M., Berkowitz, D. E., & Melvin, R. L. (2024). Grant Drafting Support with Guided Generative AI Software. *SoftwareX*, 27, Article 101784. <https://doi.org/10.1016/j.softx.2024.101784>
- Harvard University (n.d.). *Initial Guidelines for the Use of Generative AI Tools at Harvard*.  
<https://www.huit.harvard.edu/ai/guidelines>
- Lee, V. R., Pope, D., Miles, S., & Zárte, R. C. (2024). Cheating in the Age of Generative AI: A High School Survey Study of Cheating Behaviors before and after the Release of ChatGPT. *Computers and Education: Artificial Intelligence*, 7, Article 100253.  
<https://doi.org/10.1016/j.caeai.2024.100253>
- National Health and Medical Research Council (2023). *Policy on Use of Generative Artificial Intelligence in Grant Applications and Peer Review*.

- <https://www.nhmrc.gov.au/about-us/resources/policy-use-generative-artificial-intelligence>
- National Institutes of Health (2024). *NIH Issues New Centralized Resource of Policies and Practices for Promoting Responsible Use of Artificial Intelligence*.  
<https://dcp.usc.edu/2024/08/07/nih-issues-new-centralized-resource-of-policies-and-practices-for-promoting-responsible-use-of-artificial-intelligence/>
- National Science Foundation (n.d.). *Artificial Intelligence*.  
<https://www.nsf.gov/focus-areas/artificial-intelligence>
- Nolan, L. (2025). Study: AI Search Engines Cite Incorrect Sources at a 60% Rate. *Breitbart*.  
<https://www.breitbart.com/techt/2025/03/15/study-ai-search-engines-cite-incorrect-sources-at-a-60-rate/>
- Panagopoulos, J. (2024). Foreign Students Top Cheat Inquiries. *The Australian*.
- Panagopoulos, J. (2025). Foreign Student Academic Misconduct Cases on Rise. *The Australian*.
- Popper, K. (1983). *Realism and the Aim of Science*. Routledge.
- Ross, D. (1955). *The Right and the Good*. Oxford University Press.
- Rossiter, J. R. (2022). Learning Theory Principles for Psychotherapy. *Psychology*, 13, 743-766. <https://doi.org/10.4236/psych.2022.135049>
- Rossiter, J. R., Percy, L., & Bergkvist, L. (2018). *Marketing Communications: Objectives, Strategy, Tactics*. Sage.
- Spencer, B. (2022). Robots Make It So Easy to Cheat. *The Australian*.
- Stanford University (2023). *Generative AI Policy Guidance*.  
<https://communitystandards.stanford.edu/generative-ai-policy-guidance>
- Subbaraman, N. (2025). Scientists Take on Scholarly Journals with Walkouts, Scathing Letters and Delistings. *The Wall Street Journal*.  
[https://www.wsj.com/business/media/scientific-journals-quality-publishers-6399fc95?mod=hp\\_lead\\_pos11](https://www.wsj.com/business/media/scientific-journals-quality-publishers-6399fc95?mod=hp_lead_pos11)
- Tong, S., Mao, K., Huang, Z., Zhao, Y., & Peng, K. (2024). Automating Psychological Hypothesis Generation with AI: When Large Language Models Meet Causal Graph. *Humanities and Social Sciences Communications*, 11, Article No. 896.  
<https://doi.org/10.1057/s41599-024-03407-5>
- UK Research and Innovation (2024). *£12 Million for UK Projects to Address Rapid AI Advances*.  
<https://www.ukri.org/news/12-million-for-uk-projects-to-address-rapid-ai-advances>
- Universities Australia (2025). *Use of Artificial Intelligence (AI) Tools in Research*.
- University of Oxford (2025). *AI in Teaching and Assessment*.  
<https://academic.admin.ox.ac.uk/AI-in-teaching-and-assessment>
- University of Sydney (2024). *Academic Integrity*.  
<https://www.sydney.edu.au/students/academic-integrity/artificial-intelligence.html>
- Westen, D., & Kowalski, R. (2005). *Psychology* (4th Ed.). Wiley.
- Wu, Y. Y., Yang, Y., & Uzzi, B. (2023). A Discipline-Wide Investigation of the Replicability of Psychology Papers over the Past Two Decades. *Proceedings of the National Academy of Sciences*, 120, e2208863120. <https://doi.org/10.1073/pnas.2208863120>
- Young, E. (2024). *Should Psychologists Embrace AI-Powered Hypothesis Generation?* British Psychological Society.  
<https://www.bps.org.uk/psychologist/should-psychologists-embrace-ai-powered-hypothesis-generation>