Critical Artificial Intelligence Literacy for Psychologists

Olivia Guest[1,2] and Iris van Rooij[1,2,3]
[1]Department of Cognitive Science and Artificial Intelligence, Radboud University, The Netherlands
[2]Donders Institute for Brain, Cognition, and Behaviour, Radboud University, Nijmegen, The Netherlands
[3]Department of Linguistics, Cognitive Science, and Semiotics, Aarhus University, Denmark

Psychologists — from computational modellers to social and personality researchers to cognitive neuroscientists and from experimentalists to methodologists to theoreticians — can fall prey to exaggerated claims about artificial intelligence (AI). In social psychology, as in psychology generally, we see arguments taken at face value for: *a*) the displacement of experimental participants with opaque AI products; the outsourcing of *b*) programming, *c*) writing, and even *d*) scientific theorising to such models; and the notion that *e*) human-technology interactions could be on the same footing as human-human (e.g., client-therapist, student-teacher, patient-doctor, friendship, or romantic) relationships. But if our colleagues are, accidentally or otherwise, promoting such ideas in exchange for salary, grants, or citations, how are we as academic psychologists meant to react? Formal models, from statistics and computational methods broadly, have a potential obfuscatory power that is weaponisable, laying serious traps for the uncritical adopters, with even the term 'AI' having murky referents. Herein, we concretise the term AI and counter the five related proposals above — from the clearly insidious to those whose ethical neutrality is skin-deep and whose functionality is a mirage. Ultimately, contemporary AI is research misconduct.

Keywords: artificial intelligence, critical AI literacy, displacement, deskilling, statistics, modelling, theory, replication crisis, questionable research practices

In the 2010s, a crisis started in data-centric hyperempiricist disciplines, infamously including social psychology (Baker, 2016; Sayre & Riegelman, 2018; Wasserstein & Lazar, 2016). The crisis resulted in a loss of faith in our data, our use of statistical methods, and ultimately our published work (Ellemers, 2013; cf. Faye, 2012). Meanwhile, back at the server farm, another hyperempiricist data-centric sea change unfolded — also in the 2010s, and also spreading to many fields, starting with deep artificial network models of vision and classification (Guest & Martin, 2025b). By the 2020s, this technological shift had assimilated most of data ever created in service of selling products that appear to output so-called human-like text and images. These techniques are dubbed artificial intelligence (AI)[1] and have made their way from previously marginalised in academia to enriching those at the highest rungs of power and to appearing everywhere from cars to email applications and from schools to hospitals (viz. Avraamidou, 2024; Guest, Suarez, et al., 2025; Suarez et al., 2025). These two forces have now met: AI solutionism has come for social psychology (Alfons and Welz, 2024; Cuskley et al., 2024; Demszky et al., 2023; Karjus, 2024; cf. Rosenbusch et al., 2021; E. R. Smith, 1996).

Zooming out, psychology has history with methodological missteps and statistical faux pas. These have most recently been under the umbrella of the so-called replication crisis, where psychological findings fell into dis(re)pute due to allegedly mismanaged or fraudulent data, due to mechanical application of methodology, and due to contrived experiments (Ellemers, 2013; Flis, 2019; H. J. Forbes et al., 2023; Giner-Sorolla, 2019; Pennington & Pownall, 2024; Pettit, 2024a; Rubin, 2025; Schiavone & Vazire, 2023; Wasserstein & Lazar, 2016). Similarly, AI has not been unaffected, suffering from comparable so-called crises (Alfons & Welz, 2024; Gibney, 2022; Gundersen & Kjensmo, 2018; Hutson, 2018; Kapoor & Narayanan, 2023; Liao et al., 2021; Malik, 2020; Semmelrock et al., 2023; Varoquaux & Cheplygina, 2022; Verstynen & Kording, 2023). In addition, more serious cases can be found in the more distant past, such as Ronald Fischer claiming that smoking tobacco only correlates with lung cancer, but does not cause it (Abdalla and Abdalla, 2021; Guest,

Olivia Guest
Acknowledgements: The authors would like to thank Barbara Müller and Marcela Suarez for their invaluable feedback on this work.
Correspondence: Olivia Guest, Donders Institute for Brain, Cognition and Behaviour, Radboud University, The Netherlands. E-mail: olivia.guest@donders.ru.nl

[1]For those not familiar with the history of the scientific field of AI, and its boom and bust cycles for at least seven decades (Boden, 2006; Guest & Martin, 2025b; Law, 2024; Lighthill et al., 1973; Olazaran, 1996; Perez, 2002; P. Smith & Smith, 2024; Thornhill, 2025), 'AI' may seem a recent invention. However, such ahistorical perspectives serve only those who are trying to sell you something, not the protection of science (Guest, Suarez, et al., 2025).

Table 1

Core reasoning issues (first column), which we name after the relevant numbered section, are characterised using a plausible quote. In the second column are responses per row; also see the named section for further reading, context, and explanations.

Critical AI Literacy	
Uncritical Statement	**Possible Response**
2) Lies, Damned Lies, and Statistics "AI products are outside my expertise, but I think it is useful to deploy them."	As a matter of fact these products are statistical models, akin to logistic regression, which all psychologists even undergraduate students are required to have a familiarity with. Additionally, it is required to know the differences between models used to perform statistical inference and those that are models of cognition. As is knowing basic open science principles. Therefore, it should come as no shock that assuming the mantle of the non-expert here is inappropriate, and in fact may even be a form of QRP to abandon critical thinking.
3) Displacement of Participants "I can use AI instead of participants to perform tasks and generate data."	The providence of the data used in these models indicates it is not ethically sourced, falling below standards for our discipline, involving sweatshop labour and no consent for private data used in experiments. The output can contain direct original input data (i.e. double dipping), but smoothed to remove outliers, conform to our pre-existing ideas of what it should look like (data fabrication), and all-round irreplicable. Psychology is meant to study humans, not patterns at the output of biased statistical models.
4) Outsourcing Programming to Companies "I can use AI for programming experimental paradigms and statistical analyses."	This is an example of the field's backsliding from adopting open science and programming skills. No formal specification will be given for code generated from a corporate-owned opaque model. The psychologist now has no reason to learn how to engineer software, and disturbingly might as well switch back to propriety software like SPSS which at least has documentation and explicit versions. Code at the output will be plagiarised, making it time-consuming to check compliance with our needs than if we wrote the code ourselves, and violating openness.
5) Ghostwriter in the Machine "I can use AI for understanding the literature and for scholarly writing."	This practice implicates a swathe of issues akin to automating the paper mill. First, the literature is screened by corporations, which have every reason to control the output of the model to suit their needs or minimally to ignore output issues, such as sexism. Second, the fabrication of non-existent citations which makes claims worse than baseless because they appear supported by prior work. Third, the dislocation of text from the literature since no providence can be established, resulting in plagiarism.
6) The End of Scientific Theory "I can outsource verbal theorising to AI or use it as a formal cognitive model."	This not only adds to the dislocation of work from its evidential and historical basis, but also it impedes our theorising about phenomena and systems under study. In this context, we are interested in human-understandable theory and theory-based models, not statistical models which provide only a representation of the data. Scientific theories and models are only useful if we, the scientists who build and use them, understand them in deep ways and they connect transparently to research questions. AI product use is absconding scientific duty.
7) Equivocation of Human-Human & Human-AI "I can study people using chatbots as if they are socially interacting."	Seeing client-therapist, student-teacher, patient-doctor, friendship, or romantic relationships as equivalent to those between people and artifacts is both a form of dehumanisation and a hollowing out of the target of study in social psychology: the relationship between people and *other people*. It is important to study the relations between humanity and machines and the social interactions mediated through technology — but to place interactions with chatbots in the same category as those between people assumes and risks too much.

Suarez, et al., 2025; Knoester et al., 2025; Stolley, 1991; cf. van den Berg et al., 2024). Most serious are the eugenics roots of modern statistics and psychometrics, which gave rise to pseudoscientific theories like physiognomy and phrenology, which in turn provided scientific cover for racism, sexism, classism, ableism, and ultimately genocide (Black, 2012; Burke & Castaneda, 2007; Clayton, 2020; Cowan, 1972; Gebru & Torres, 2024; Gould, 1981; Norrgard, 2008; Paul, 2016; Reddy, 2007; Saini, 2019; S. M. Taylor et al., 2023; Yakushko, 2019). This is the backdrop against which modern social psychological tensions with artificial intelligence unfold (Andrews et al., 2024; Benjamin, 2019; Birhane, 2022; Birhane & Guest, 2021; Black, 2012; Blas et al., 2025; Brennan et al., 2025; Crawford, 2021; S. H. Forbes & Guest, 2025; Gebru & Torres, 2024; Gleiberman, 2023; Guest, 2024, 2025; Guest, Suarez, et al., 2025; McQuillan, 2025; Mirowski, 2023; Spanton & Guest, 2022; van Rooij & Guest, 2025; van der Gun & Guest, 2024).

In this relationship between, on the one hand statistics, machine learning, and other algorithms and models that fall broadly under the label 'AI', and on the other hand psychology generally, similar issues continue to play out. It appears as if psychology may have learned nothing from the past centuries (with eugenics) and decade (with the replication crisis of the 2010s), maintaining business as usual (cf. Faye, 2012). As we will unpack herein, contemporary AI presents a totalising hyperempiricist statistical perspective on the field that directly holds back theory building, dehumanises participants and scientists alike, and introduces multitudinous conflicts of interest. The single most important perspective for a scientist to take is a sceptical and reflexive (viz. Jamieson et al., 2023) one, the titular *critical AI literacy* (see Table 1; S. H. Forbes and Guest, 2025; Guest, Suarez, et al., 2025; Suarez et al., 2025; cf. Long and Magerko, 2020; Tully et al., 2025). Such literacy is akin to knowledge of how to properly use inferential statistics and thus avoid accidentally being fooled by the results of our experiments in a flawed search for statistical significance.

To presage what is to come, we will show that AI destroys trust in any science with which it comes into contact. "This trust is earned by being transparent and by performing research that is relevant, replicable, ethically sound and of rigorous methodological quality." (Gopalakrishna et al., 2022) Contemporary AI products and the companies that produce them violate all these principles (e.g. Birhane et al., 2021; Crane, 2021; Gerdes, 2022; Leech et al., 2024; Markov, 2024; Ochigame, 2019; Phan et al., 2022). Technology companies, as we shall unpack below, enable and enact misconduct such as fabrication of data, plagiarism, as well as questionable research practises (QRPs) such as "inappropriate (harmful or dangerous) research methods[,] denying authorship to contributors[,] poor data management and/or storage[, and] non-disclosure of a conflict of interest" (Hiney, 2015, p. 5). But first, what is AI?

2 Lies, Damned Lies, and Statistics

Herein we take the stance that AI is most usefully seen as a series of technology products that have the following properties:

- are **sophisticated statistical models**, so large they impact humans and the environment through their energy, land, and water use (Goetze, 2024; Luccioni et al., 2024; Markelius et al., 2024; Parshley, 2024; Suarez et al., 2025; Tan, 2025);
- depend on **vast swathes of data**, which is mostly stolen or otherwise unethically obtained or refined (Alba, 2023; Bansal, 2025; Birhane, 2022; Birhane et al., 2023; Equidem, 2025; Perrigo, 2023; Vercellone & Di Stasio, 2023);
- can represent various statistical distributions and so can be **discriminative, generative, or neither** (Efron, 1975; Guest, Suarez, et al., 2025; Jebara, 2004; Mitchell, 1997; Ng & Jordan, 2001; Xue & Titterington, 2008);
- exist in a **displacement relationship to humans**, i.e. this type of AI product is harmful to people, it contributes to deskilling, and it obfuscates cognitive labour (Guest, 2025, Table 1).

Bog standard spellcheck software, bubble sort algorithms, calculators, thermostats, and logistic regression do not fall under this type of AI (Guest, 2025). What we call displacement AI clearly descends from simpler statistical models all psychologists are expected to be familiar with, such as logistic regression, which is the backbone mathematics of these types of artificial neural networks (Guest & Martin, 2023, 2025b; Guest, Suarez, et al., 2025). All these statistical models — as any undergraduate knows, but it bears repeating — are based on computing correlations, and so taking the results at face value can trick us into mistaking correlation for causation. In the case of displacement AI, it manifests as thinking that correlation to human-like output is evidence for the machine thinking for itself. However as Guest and Martin (2023) explain: *correlation is not cognition* (also see Guest & Martin, 2025a, 2025b; Guest, Scharfenberg, & van Rooij, 2025; van Rooij & Guest, 2025; van Rooij et al., 2024b). This is a core fallacy that motivates the dehumanisation, displacement, and deskilling of people by AI that we will address below (Erscoi et al., 2023; Guest, 2025). A related reasoning problem is to think that because such AI products — recall they are statistical models designed to do this — appear to capture human-like output, that they therefore constitute a psychological theory. As Guest and Martin (2023, 2025b) and van Rooij and Guest (2025) warn, such a reasoning error constitutes a hyperempiricist trap.

On a different tack, the magic bullet to the replication crisis of the 2010s was open science (H. J. Forbes et al., 2023; Pennington & Pownall, 2024; Schiavone & Vazire, 2023; Wills, 2019): an umbrella term for a number of practices aimed at increasing access, transparency, and accountability for scientific work, especially for the statistical analysis of data (Mirowski, 2018; Whitaker &

Guest, 2020). In stark contrast to openness of all kinds is the technology industry and their practices, models, and datasets (Dingemanse, 2025; Gerdes, 2022; Hao, 2025; Jackson, 2024; Liesenfeld & Dingemanse, 2024; Liesenfeld et al., 2023; Maffulli, 2023; Maris, 2025; Mirowski, 2023; Nolan, 2025; Ochigame, 2019; Solaiman, 2023; Thorne, 2009; Widder et al., 2024). Notably, open science of the 2010s failed to centre critical thinking, ethics, reflexivity, and theorising (Chambers, 2019; B. Clarke et al., 2024; Crüwell et al., 2019; Field & Derksen, 2021; Giner-Sorolla, 2012; Neuroskeptic, 2012; Skubera et al., 2025), leaving it no less vulnerable to the same problems as science broadly when it comes to avoiding bad actors, including industry capture (e.g. Bak-Coleman & Devezer, 2024; Crane, 2021; S. H. Forbes & Guest, 2025; S. H. Forbes et al., 2024; Gebru & Torres, 2024; Gerdes, 2022; Ghai et al., 2025; Guest, Suarez, et al., 2025; Jamieson et al., 2023; Liesenfeld & Dingemanse, 2024; Liesenfeld et al., 2023; Mirowski, 2018, 2023; Morey & Davis-Stober, 2025; Patel & Elkin, 2015; Pettit, 2024a; Phan et al., 2022; Whitaker & Guest, 2020).

The three aforementioned related themes sketched out in this section, will play out in the AI-social psychology relationships we will examine — namely:

a. **misunderstanding of the statistical models** which constitute contemporary AI, leading to inter alia thinking that correlation implies causation (Guest, 2025; Guest & Martin, 2023, 2025a, 2025b; Guest, Scharfenberg, & van Rooij, 2025; Guest, Suarez, et al., 2025);

b. **confusion between statistical versus cognitive models** when it comes to their completely non-overlapping roles when mediating between theory and observations (Guest & Martin, 2021; Morgan & Morrison, 1999; Morrison & Morgan, 1999; van Rooij & Baggio, 2021);

c. **anti-open science practices**, such as closed source code, stolen and opaque collection and use of data, obfuscated conflicts of interest, lack of accountability for models' architectures, i.e. statistical methods and input-output mappings are not well documented (Barlas et al., 2021; Birhane & McGann, 2024; Birhane et al., 2023; Crane, 2021; Gerdes, 2022; Guest & Martin, 2025b; Guest, Suarez, et al., 2025; Liesenfeld & Dingemanse, 2024; Liesenfeld et al., 2023; Mirowski, 2023; Ochigame, 2019; Thorne, 2009).

Being able to detect and counteract all these three together comprises the bedrock of skills in research methods in a time when AI is used uncritically (see Table 1). The inverse: not noticing these are at play, or even promoting them, could be seen as engaging in questionable research practises (QRPs; Brooker & Allum, 2024; Neoh et al., 2023; Rubin, 2023). Therefore, in the context of critical AI literacy for social psychology, and indeed cognitive, neuro-, and psychological sciences in general, the three points above serve as totemic touchstones, as litmus tests for checking somebody's literacy in AI (Guest, 2024; Guest & Martin, 2021, 2025a, 2025b; Guest, Scharfenberg, & van Rooij, 2025; Guest, Suarez, et al., 2025; Suarez et al., 2025; van Rooij & Baggio, 2021; van Rooij & Guest, 2025; van Rooij et al., 2024b). To wit, if somebody is able to minimally articulate these three related issues, how they manifest, and why they matter to our science, we can rest easy they know the basics of how to critically evaluate AI products in science.

3 Displacement of Participants

Many technological solutions to recruiting and managing participants have been proposed: from original pen and paper sign up sheets and physical bulletin boards to computer systems like SONA, which digitised and automated some of these processes, and then furthermore to platforms like Amazon's Mechanical Turk (MTurk), which allows participants to be outsourced to sweatshops (e.g. C. A. Anderson et al., 2019; Douglas et al., 2023; Gamblin et al., 2017; Gray & Suri, 2019; Hauser & Schwarz, 2016; Wagner et al., 2022). It may be the case that systems like MTurk, which is named after the infamous pseudo-automaton which beat Napoleon at chess (Manninger, 2024), indeed have clear pros and cons, and are perhaps not inherently abusive. However, they are clearly *open* to abuse (C. A. Anderson et al., 2019; Guest, 2024; Newman, 2019; Pettit, 2024c; Stephens, 2023). MTurk's namesake, the original Mechanical Turk, was a device that comprised a cupboard with a chess board on top and which across the player sat an orientalist puppet. However much it may have seemed the puppet played chess, it was instead a hidden human who moved the pieces.

In the present, sweatshops based on Amazon's MTurk technology undergird all modern AI products (Alba, 2023; Bansal, 2025; Equidem, 2025; Gershgorn, 2017; Gray & Suri, 2019; Manninger, 2024; Perrigo, 2023; Stephens, 2023; Streitfeld, 2025; Suri, 2019; Yang et al., 2020). This human-in-the-loop technique has had applications in the 2010s with deep artificial neural networks that performed what is claimed to be human-like vision, and in the 2020s with reinforcement learning from human feedback (e.g. Birhane et al., 2023; Guest & Martin, 2025b; Kirk et al., 2023; Manninger, 2024; Prabhu & Birhane, 2020; Suri, 2019).

The next iteration of this is to completely obfuscate the human-in-the-loop (Guest, 2025; Guest & Martin, 2025a; Shiffrin & Mitchell, 2023) and present AI products, such as those powered by large language models (LLMs), as able to stand in for, displace, human participants or their data (cf. Crockett & Messeri, 2023; Dillion et al., 2023; Jamieson et al., 2023; Rilla et al., 2025; Rossi et al., 2024; Schröder et al., 2025). This is problematic for many reasons (viz. Guest, Suarez, et al., 2025; van Rooij & Guest, 2025), but some lesser discussed angles are that the AI products cannot replace human participants but are merely *a*) returning unethically sourced (pre-existing) human data (Alba, 2023; Bansal, 2025; Equidem, 2025; Guest, 2025; Perrigo, 2023). And as such, *b*) these data have statistically unfavourable properties according to what are stated norms of the field, such as comprising double dipping (Y. Liu et al., 2024; Villalobos et al., 2024;

cf. Ball et al., 2020), smoothing over or completely removing outliers (An et al., 2025; Raman et al., 2025; cf. Valentine et al., 2021), and HARKing (Ramnath et al., 2025; cf. Kerr, 1998).

The first reason mentioned for needing to avoid such AI products is entangled with ethical reasons to avoid sweatshop labour, to avoid not being able to control (unlike in the lab and, to a lesser extent, with online experiments) that the experimental conditions of the participants are inline with ethics and integrity guidelines or the law (e.g. American Psychological Association, 2017; Belanger, 2025; Birhane et al., 2022; British Psychological Society, 2021; Illinois Department of Financial and Professional Regulation, 2025; Warren, 2025). This is because, for example, an AI product or company could contain or store personal and private information from people who never consented for their data to be reused this way, or it could contain data that was collected in ways antithetical to ethical guidelines (Alba, 2023; Bansal, 2025; Birhane et al., 2021; Burgess, 2025; Cox, 2025; Equidem, 2025; Fisher et al., 2025; Kira, 2024; Knight, 2023; Perrigo, 2023; Prabhu & Birhane, 2020; Stokel-Walker, 2025; Stuart et al., 2019; Tangermann, 2025).

For *b*, the case of violating desirable statistical norms and conduct: The output of LLM-based models is statistically errant, of zero scientific quality, and would not qualify as participant data under any sensible definition. It guarantees irreplicable results (Bonifield, 2025; Gibney, 2022; Gundersen & Kjensmo, 2018; Hutson, 2018; Kapoor & Narayanan, 2023; Liao et al., 2021; Malik, 2020; Semmelrock et al., 2023; Varoquaux & Cheplygina, 2022; Verstynen & Kording, 2023). Any correlations with high quality data are caused by data leakage, the human-in-the-loop, or user, and not because the system is human-like (Guest & Martin, 2023; van Rooij et al., 2024b); any so-called prompt is not guaranteed to produce the same results because the AI is proprietary, operating using closed source and possibly undocumented code and closed or even stolen data, as well as being stochastic and data dependent.

These models are created to produce plausible output that matches the user's desires (Huntington, 2025; Reeves & Nass, 1996; Salecha et al., 2024; Sharma et al., 2025) — how much more of an automated p-hacker could we ask for? Part of the spiral of the 2010s crisis in social psychology was the revelation not only that some were misusing statistics, but the explosive uncovering that some were completely fabricating the data (e.g. Bhattacharjee, 2013).

What is most disparaging if we continue down this route, is not only that it leads to data fabrication on steroids but that also we will depend on a machine that, for lack of a less anthropomorphic term, lies, and lies about lying *by design* (Bender & Hanna, 2025; Bender et al., 2021; DeVrio et al., 2025; Edwards, 2023b; Hicks et al., 2024; Metz, 2023; Xu et al., 2025; Zhao et al., 2025). It deceives through the *Eliza effect*, the cognitive bias towards assigning human-like traits to chatbots (Borau, 2025; Dillon, 2020; Koike & Loughnan, 2021; Weizenbaum, 1966), and the *Barnum-Forer effect*, the phenomenon where one thinks generic personality descriptions that could apply to anybody, applies uniquely to them (Bjarnason, 2023; Meehl, 1956; Vohs, 2016).

4 Outsourcing Programming to Companies

> [Sanford's mechanical vernier chronoscope] had two pendulums, one set longer than the other but both set at known lengths. When the stimulus appeared, the longer pendulum was released mechanically. When the participant pressed the key in response, the shorter pendulum was similarly released. By counting the number of swings it took the shorter pendulum to catch up with the longer, experimenters could use a mathematical formula to calculate the difference in time between the two events. (Evans, 2000, p. 322–323)

A century later, all this is now done by programming computers to display stimuli, and collect and analyse data. The largely harmless and perhaps even beneficial (cf. Guest, 2025) transition to computational experimentation, where all instruments (e.g. EEG, fMRI, eye-tracking) are hooked up to the codebase for running experiments and the database for storing participant data, has been completed through: the adoption of programmes such as SPSS for statistics (Landau & Everitt, 2003) and E-Prime for presenting stimuli (Spapé et al., 2019); and training psychologists to use programming languages (e.g. R, Python; British Psychological Society, 2017; Peirce et al., 2022). The last part, from the mid 2010s onwards, where psychologists spent time and effort learning how to program is a massively laudable case of reskilling (N. D. Anderson, 2016; Guest & Forbes, 2024; Scherer et al., 2019).

From the 2020s, however with the rise of the kinds of AI products we critique herein, this is under threat. If scientists accept the AI "snakeoil" and "con" (Bender & Hanna, 2025; Bjarnason, 2023; Narayanan & Kapoor, 2024) and use it to deskill themselves from learning proper programming, then code reproducibility and verifiability are lost causes; not just short-lived, dead on arrival (Becker et al., 2025a). If the code is for running experiments, strange side-effects will be present from so-called 'vibe coding' (Harkar, 2025), which is 'programming' using a chatbot such that nothing is checked or verified against software engineering principles and only the output being seen as acceptable under a small sample of conditions is evaluated. This means that because standards are ignored and best practises are violated the stimuli may be presented in the wrong order, participant data may not be properly or securely saved, and the code may be unmaintainable (Burgess, 2023, 2025; El-Mhamdi et al., 2022; Greshake et al., 2023; Ming, 2025; Tangermann, 2025). And the same is true for this non-engineered code in the case of analysing data, which will not only *not* be analysed, but open to all sorts of bugs. To add insult to injury, because the scientist is not practising programming or never learned how to code, all this will be outside their skill level and they will be unable to detect and fix these issues (Becker

et al., 2025a, 2025b; Bucaioni et al., 2024; Goel, 2025; Hsu, 2025; Jj, 2025; Lehmann et al., 2025; Ming, 2025). Furthermore, as we shall see in the next section the datasets (programming code, journal articles and books, and many other sources) used to build these models, and the way they are used, result in the AI product reproducing at the output plagiarised versions of the input (Biderman & Raff, 2022; Carlini et al., 2021; Kwon, 2024; Montgomery & agencies, 2025; Nasr et al., 2023; Reisner, 2025; Schmid et al., 2025; van Rooij, 2022). This risks violating the consent of the original programmers, untethering our code from the literature that produced snippets or large parts of it, and committing various other QRPs.

As Danielle Navarro (2015) says about shortcuts through using inappropriate technology, which chatbots are, we end up digging ourselves into "a very deep hole." She goes on to explain:

> The business model here is to suck you in during your student days, and then leave you dependent on their tools when you go out into the real world. [...] And you can avoid it: if you make use of packages like R that are open source and free, you never get trapped having to pay exorbitant licensing fees. (pp. 37–38)

These were and are the reasons to switch from e.g. SPSS, which is owned by IBM, to R or Python (Wills, 2019). The same and more hold for being locked in and addicted to AI products, which have closed or industry-controlled source code (Hao, 2025; Liesenfeld & Dingemanse, 2024; Liesenfeld et al., 2023; Maffulli, 2023; Maris, 2025; Mirowski, 2023; Solaiman, 2023) , and mislead us into overestimating our programming abilities (Lehmann et al., 2025).

5 Ghostwriter in the Machine

A unique selling point of these systems is conversing and writing in a human-like way. This is imminently understandable, although wrong-headed, when one realises these are systems that essentially function as lossy[2] content-addressable memory: when input is given, the output generated by the model is text that stochastically matches the input text. The reason text at the output looks novel is because by design the AI product performs an automated version of what is known as mosaic or patchwork plagiarism (Bаždarić, 2013) — due to the nature of input masking and next token prediction, the output essentially uses similar words in similar orders to what it has been exposed to. This makes the automated flagging of plagiarism unlikely, which is also true when students or colleagues perform this type of copy-paste and then thesaurus trick, and true when so-called AI plagiarism detectors falsely claim to detect AI-produced text (Edwards, 2023a). This aspect of LLM-based AI products can be seen as an automation of plagiarism and especially of the *research paper mill* (Guest, 2025; Guest, Suarez, et al., 2025; van Rooij, 2022): the "churn[ing] out [of] fake or poor-quality journal papers" (Sanderson, 2024; Committee on Publication Ethics, 2024). Other aspects, such as the fabrication of non-existent references, and indeed the lack of any reading (since these systems do not read), detach any LLM output from the literature, thus violating scholarly standards (Guest, Suarez, et al., 2025; Lawson et al., 2025).

In addition, who is held accountable if nobody with intent authored the text? Because while the original data fed into the system is certainly written with goals, messages, and audiences in mind jumbling this into ad-libbed word salad removes authorial intent (Bender et al., 2021). So do the companies who own the chatbot own the text or do the original authors? These questions denote legal battles, which are being currently fought in the public eye and which affect all of us in all roles, not just as academics (Creamer, 2025; Knibbs, 2024; Reuters, 2025). Either way, even if the courts decide in the favour of companies, we should not allow these companies with vested interests to write our papers (Fisher et al., 2025), or to filter what we include in our papers. Because it is not the case that we only operate based on legal precedents, but also on our own ethical values and scientific integrity codes (ALLEA, 2023; KNAW et al., 2018), and we have a direct duty to protect, as with previous crises and in general, the literature from pollution. In other words, the same issues as in previous sections play out here, where essentially now every paper produced using chatbot output must declare a conflict of interest, since the output text can be biased in subtle or direct ways by the company who owns the bot (see Table 2).

Seen in the right light — AI products understood as content-addressable systems — we see that framing the user, the academic in this case, as the creator of the bot's output is misplaced. The input does not cause the output in an authorial sense, much like input to a library search engine does not cause relevant articles and books to be written (Guest, 2025). The respective authors wrote those, not the search query!

6 The End of Scientific Theory

Marginalisation of scientific theorising in psychology has a long history, and is related to the medicalisation of psychology (Ellemers, 2013; Pettit, 2024b). This history allows a partial genealogy on why methodologies, such as from clinical trials, like preregistration, are imported to psychological science (Bakan, 1966). In fact, these methods are not scientific per se, but appropriate constraints for medical research which has important differences with basic science (cf. Calvert, 2006; Mayo-Wilson et al., 2025; Pielke Jr, 2012).

Fundamental research in basic science (third row, Table 2) does not set out with a goal, such as to develop a medication or engineer a specific result (Devezer et al., 2021; Ellemers, 2013; Pham & Oh, 2021; Rubin & Donkin, 2024; Szollosi et al., 2020). In contrast, because experiments run by companies with the goal

[2] The opposite of lossless, which in a formal information theoretical context means that data is stored in a compressed format such that the original information is unrecoverable.

Table 2

What the landscape of research in psychology looks like with respect to clinical versus basic science in terms of interference, constraints, and outputs.

Goals & Constraints in Psychology			
Research	**Interference & Incentives**	**Required Constraints**	**Desired Output**
Clinical	Pharmaceutical & healthcare industries, health insurance companies. Funding is output-oriented, industry prefers closed or controlled knowledge.	Clinical trials, control groups, preregistration, & other legal instruments.	Products & services: medications, therapies, devices, certificates.
Both	AI companies, general, educational & research technology industries. Profit, reputation, theft of cognitive labour.	Ethical review boards, self-governed codes of conduct, disclosure of conflicts of interest.	
Fundamental	All the above, especially if academic freedom under threat. Personal edification, fame, but also openness and collaboration.	*Worst case:* None, but basic science is over. *Best case:* Limit technologies & reverse hollowing out of our freedoms.	Knowledge: theories, understanding, critical thinking, impartial expert analyses.

of delivering a medical product to market have a biased incentive structure (first row, Table 2), with significant conflicts of interest, clinical trials are regulated (Bhatt, 2010; European Union, 2022; Guest, 2024; Patel & Elkin, 2015; Rhee & Wilkinson, 2020; U.S. Food and Drug Administration, 2024; Washington, 2006). Clinical trials, and furthermore preregistration thereof, were devised because otherwise there is an understandable and expected — maybe even required for good healthcare — cognitive bias to want to see patients recover, as well as in the present severe industry pressure to peddle pharmaceutical products for profit. These factors are not at play when there is no external interference or requirement for patient care, as in basic science. Non-proceduralised, substantive explorations are provably necessary for science (Rich et al., 2021; van Rooij et al., 2024b). The chaotic human preferences for going down rabbit holes and supporting ideas with evidence, formalism, and argument are the beauty of scholarly work (Guest, 2024).

Herein our focus is on properly protecting our science, which aims to develop, discard, and evaluate theories: our explanations, descriptions, and understandings of the cognitive, neuroscientific, and psychological worlds. While conceptually these two types of psychological research — applied clinical science versus fundamental science — are by no means zero sum, it is the case that practitioners' time is and therefore their skill sets can be. So when doing fundamental research we must exercise our academic freedoms to keep private interests at bay. This means that it is up to us to not only state conflicts of interest (Guest & Martin, 2025b; Guest, Suarez, et al., 2025; KNAW et al., 2018), but to think deeply about what it means to have a conflict in a time where industry hype along with universities promote and in fact coerce the use of AI products: from writing our peer reviews and proposals, performing our literary search and review, and authoring scholarly articles for us, to creating our theories and testing them for us!

As others have made the explicit parallel before (Abdalla & Abdalla, 2021; Guest, Suarez, et al., 2025), our current entanglement with the technology sector is not far off from those of the tobacco or petroleum companies interference in basic research (Atkin, 2025; Knoester et al., 2025; Stolley, 1991; van den Berg et al., 2024). Unlike the pharmaceutical industry which produces medications and has severe legal oversight in principle, the so-called educational and research technology sectors have nothing of the sort (Drimmer & Nygren, 2025; Watters, 2023). And they also — despite their marketing and hype to the contrary — do *not* produce pedagogical and scholarly outputs, we do. These companies have a long history of labour theft and siphoning off of

taxpayer money. What sense does it make to allow them to write our theories for us? Such theories, if they can be called theories, cannot be 'good' in any reasonable sense (Guest, 2024; Guest & Martin, 2023; van Rooij & Guest, 2025; van Rooij et al., 2024a).

If theory constitution (seeing an AI product as embodying a theory) or formal or informal articulation (allowing an AI product to 'write' verbal and formal theories; cf. Guest & Martin, 2023; van Rooij & Guest, 2025) is permitted, then not only are companies expressing our scientific ideas for us, they control them (Guest, 2025; Guest, Suarez, et al., 2025). Science is under full corporate capture (third row, worse case, Table 2), no new independent, impartial, transparent, or publicly-owned and -funded knowledge production can happen. Open science ends, and in fact science itself is merely a shadow of its former self; with knowledge only released when it favours profit with no safeguards against it accurately matching the world. Everything is p-hacked, everything is HARKed, everything is double-dipped. Nothing is new, nothing is independently verifiable.

7 Equivocation of Human-Human & Human-AI

AI products enact direct dehumanisation, intrinsic psychological harm to children, patients, and other vulnerable populations, resulting in grave consequences such as causing severe mental illness, pushing people to suicide, recommending illegal drug use, and sanctioning murder (Abrams, 2025; Akingbola et al., 2024; Bellan, 2025; Bender, 2024; Borau, 2025; Broderick, 2025; Chen et al., 2023; Dang & Liu, 2025; Dupré, 2025; Eichenberger et al., 2025; Horwitz, 2025; Huntington, 2025; Jargon & Kessler, 2025; Kaplan, 2024; Kim & McGill, 2025; Klee, 2025; Koike & Loughnan, 2021; Landymore, 2025; Laricheva et al., 2024; Montgomery, 2024; Morrin et al., 2025; Neville, 2025; Omar et al., 2025; Pejcha, 2023; Purtill, 2025; Rajkumar, 2025; Reiley, 2025; Roose, 2024; Schoene & Canca, 2025; J. Taylor, 2025; Tiku, 2025a, 2025b; Warzel, 2025; Wei, 2025; Xiang, 2023).

> ChatGPT doesn't offer genuine emotional attunement. It cannot replicate the human connection necessary for healing. More dangerously, it can delay access to professional help. People think they're improving, but often they're not. (Shmais, 2025, n.p. also Akingbola et al., 2024; Al-Sibai, 2025; L. Clarke, 2025; Lebovitz et al., 2021; Turkle, 2015; Turkle et al., 2006)

The focus here is the effect that confusing human-object with human-human has if normalised for social psychological scientific practice (e.g. Goff et al., 2014; cf. Appelman, 2023; BBC News, 2021; Grant and Hill, 2023; Hern, 2018; Neville, 2025). Typical cases are when displacements happen in the relationship between client and therapist, student and teacher, patient and doctor — and perhaps most shockingly between friends or romantic partners — *and* when academics see these displacements as acceptable (Akingbola et al., 2024; Bender, 2024; Chen et al., 2023; Dang & Liu, 2025; Kim & McGill, 2025; Koike & Loughnan, 2021; Litwack, 2024; Oldfield, 2023; Vanman & Kappas, 2019). These are not social interactions. They are human-computer interactions, many of which are extremely harmful, and need to be analysed and studied as such (cf. cyberpsychology, Kirwan et al., 2024; digital ethnography, Markham, 1998). Any other conception, which grants humanity to an inanimate object serves the technology and insurance sectors, who want to save money at the expense of vulnerable groups (O'Neil, 2016). Importantly, supporting evidence for the usefulness, effectiveness, or safety of AI products in such relationships often violate scientific practice, such as avoiding reporting conflicts of interest (Alien Technology Transfer, 2025; Baumard, 2023; Bunka.ai Team, 2025; Guingrich and Graziano, 2023; Safra et al., 2020; The Luddite, 2024; cf. Guest and Martin, 2025b; Silverstein et al., 2024; Spanton and Guest, 2022).

These problematic beliefs exist on a spectrum, from the assumption that AI products can help in mental health (e.g. Dehbozorgi et al., 2025; Siddals et al., 2024; Wellcome, 2025) to the assertion that a chatbot can *be* a therapist (Kilgore, 2025), which is a regulated profession (e.g. European Federation of Psychologists' Associations, 2025) with codes of ethics (e.g. American Psychological Association, 2017; British Psychological Society, 2021). Proponents of these beliefs use the guise of labour shortages and the global mental health crisis (Kaplan, 2024), as cover to deskill (for an alarming case, see Budzyń et al., 2025; also Akingbola et al., 2024; Lebovitz et al., 2021), contributing to the polycrises the technology industry and their allies uniquely profit from (McConnell & Jacobs, 2025).

All AI, and indeed any technology in mental or other healthcare settings requires "keeping users safe [and this in turn] requires substantial input from clinicians and careful planning to reduce risk." (Abrams, 2025) This is key because it is not possible to keep people safe without a highly qualified human-in-the-loop (Amironesei et al., 2021; El-Mhamdi et al., 2022; T. Liu et al., 2025; Salecha et al., 2024; Sharma et al., 2025; Weidinger et al., 2022) and even then the risk of deskilling is present (e.g. Budzyń et al., 2025; recall section 4: *Outsourcing Programming to Companies*). Relatedly, and taking the example of the therapist-client relationship, dehumanising parallels encoded in expressions such as: "the country needs all the quality therapists we can get — be they human or bot" (Riddle, 2025) cannot become normal in our scientific discussions. No therapist under proper ethical functioning would cause their client to be addicted to their therapy (Huntington, 2025) nor would they, as in the Replika example (The Luddite, 2024), introduce a sexual relationship between them and their client (Pettit, 2024b). Any such suggestions to use a chatbot as a therapist would go against professional codes of conduct and possibly the law (e.g. Abraham, 2002; American Psychological Association, 2017; Belanger, 2025; British Psychological Society, 2021; Cole, 2025; Illinois Department of Financial and Professional Regulation, 2025; Mental health Foundation, 2022).

Finally, as scientists we have to analyse situations carefully, so

we must be critical when we see statements such as: "It's worth noting that AI can pass the clinical social worker exam — even without seeing the questions. But that shows the weaknesses of exams more than the strength of AI" (Caldwell, 2024, n.p.). An AI product passing an exam means nothing about the "weaknesses" of exams in the abstract. Anybody can 'win' a marathon if driven by car. Does that make it a weakness of the marathon race as an endurance event? Or does it reflect a deeper category error on our behalf, like with the misapplication of statistical tests, that an assumption has been violated? The race as the exam assumes an unaided human performs it. Any digital computer can surpass a human at many feats, e.g. calculations per second, but that grants no humanity to the machine.

8 Do not Embrace AI

In this paper, we unpacked why we think psychologists need to be on high alert — not just to avoid another replication crisis, but to avoid the total collapse of our science. What we signpost in Table 1 may have been novel to readers until this point, but the deeper problems are absolutely known. Also, as Crystal Steltenpohl et al. (2023, pp. 9–10) state: "Intentions alone are not enough to move science forward. Creating responsible, considered processes for rigorously transparent open science requires involving interested parties from a wide range of backgrounds, perspectives, research areas, and training paradigms."

Indeed, because many such warnings go unheeded — such as the need for a cultivation of shared values and especially the principles of impartiality of researchers and academic freedom from corporate influence — we find ourselves in polycrises that affect our universities, political systems, planet, and ultimately all humanity. "When historians of science look back on the 2010s in social and personality psychology, the decade will likely stand out as a period of exceptional doubt and self-scrutiny in the field." (Schiavone & Vazire, 2023, p. 710) Why did we ever stop? Should we ever stop?

Importantly, Hazel Rose Markus (2005, p. 180, emphasis added) explains that: "Social psychology is often defined as the study of *how people respond to and are influenced by other people*" Algorithms, chatbots, LLMs, machines, models, inanimate objects are *not* people — they are the products of people (Guest, 2024, 2025). And to paraphrase Rae Carlson (1984): *What's social about chatbots? Where's the person in an LLM?*

We must sure up our subfields from the slow but certain corrosive power wielded by the harmful nonsense that is modern displacement AI. To sit idly by while deskilling and displacing of our students, participants, and selves is normalised — or worse still to profit from it — serves not science but the technology sector, which avoids criticism and self-reflection and prefers pseudoscience and misinformation.

References

Abdalla, M., & Abdalla, M. (2021). The grey hoodie project: Big tobacco, big tech, and the threat on academic integrity. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 287–297.

Abraham, T. H. (2002). (physio)logical circuits: The intellectual origins of the mcculloch-pitts neural networks. *Journal of the History of the Behavioral Sciences*, *38*(1), 3–25.

Abrams, Z. (2025). Using generic AI chatbots for mental health support: A dangerous trend. *American Psychological Association*. https://www.apaservices.org/practice/business/technology/artificial-intelligence-chatbots-therapists

Akingbola, A., Adeleke, O., Idris, A., Adewole, O., & Adegbesan, A. (2024). Artificial intelligence and the dehumanization of patient care. *Journal of Medicine, Surgery, and Public Health*, *3*, 100138.

Alba, D. (2023, July). Google's AI chatbot is trained by humans who say they're overworked, underpaid and frustrated. https://www.bloomberg.com/news/articles/2023-07-12/google-s-ai-chatbot-is-trained-by-humans-who-say-they-re-overworked-underpaid-and-frustrated

Alfons, A., & Welz, M. (2024). Open science perspectives on machine learning for the identification of careless responding: A new hope or phantom menace? *Social and Personality Psychology Compass*, *18*(2), e12941.

Alien Technology Transfer. (2025). Astound. https://alientt.com/astound/

ALLEA. (2023). *The European Code of Conduct for Research Integrity*. https://allea.org/code-of-conduct/

Al-Sibai, N. (2025). Patients furious at therapists secretly using ai. *Futurism*. https://futurism.com/patients-furious-therapists-using-ai

American Psychological Association. (2017). Ethical principles of psychologists and code of conduct (2002, amended effective june 1, 2010, and january 1, 2017). https://www.apa.org/ethics/code

Amironesei, R., Ashok, A., Birhane, A., Black, C., Borokini, F., Cath, C., Denton, E., Oduro, S. D., Hanna, A., Harvey, A., Jansen, F., Kaltheuner, F., Milne, G., Narayanan, A., Nicole, H., Oloyede, R., Parida, T., Peppin, A., Raji, D., … Vincent, J. (2021). *Fake AI* (F. Kaltheuner, Ed.).

An, Y., Zhao, X., Yu, T., Tang, M., & Wang, J. (2025). Systematic outliers in large language models. *arXiv*. https://arxiv.org/abs/2502.06415

Anderson, C. A., Allen, J. J., Plante, C., Quigley-McBride, A., Lovett, A., & Rokkum, J. N. (2019). The MTurkification of social and personality psychology. *Personality and social psychology bulletin*, *45*(6), 842–850.

Anderson, N. D. (2016). A call for computational thinking in undergraduate psychology. *Psychology Learning & Teaching*, *15*(3), 226–234.

Andrews, M., Smart, A., & Birhane, A. (2024). The reanimation of pseudoscience in machine learning and its ethical repercussions. *Patterns*, *5*(9).

Appelman, N. (2023). Racist technology in action: Image recognition is still not capable of differentiating gorillas from black people. *Racism and Technology Center*. https://racismandtechnology.center/2023/06/09/racist-technology-in-action-image-recognition-is-still-not-capable-of-differentiating-gorillas-from-black-people/

Atkin, E. (2025, July). He helped Microsoft build AI to help the climate. then Microsoft sold it to Big Oil. https://heated.world/p/he-helped-microsoft-build-ai-to-help

Avraamidou, L. (2024). Can we disrupt the momentum of the AI colonization of science education? *Journal of Research in Science Teaching*, *61*(10), 2570–2574.

Bakan, D. (1966). The test of significance in psychological research. *Psychological bulletin*, *66*(6), 423.

Bak-Coleman, J., & Devezer, B. (2024). Claims about scientific rigour require rigour. *Nature Human Behaviour*, *8*(10), 1890–1891.

Baker, M. (2016). 1,500 scientists lift. *Nature*, *533*, 452–454.

Ball, T. M., Squeglia, L. M., Tapert, S. F., & Paulus, M. P. (2020). Double dipping in machine learning: Problems and solutions. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *5*(3), 261–263.

Bansal, V. (2025, September). How thousands of "overworked, underpaid" humans train Google's AI to seem smart. https://www.theguardian.com/technology/2025/sep/11/google-gemini-ai-training-humans

Barlas, P., Kyriakou, K., Guest, O., Kleanthous, S., & Otterbacher, J. (2021). To "see" is to stereotype: Image tagging algorithms, gender recognition, and the accuracy-fairness trade-off. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW3), 1–31.

Baumard, N. (2023, May). Nicolas baumard. https://prairie-institute.fr/chairs/nicolas-baumard/

Baždarić, K. (2013). Patchwork plagiarism. *European science editing*, *39*(2).

BBC News. (2021, September). Facebook apology as AI labels black men "primates". https://www.bbc.com/news/technology-58462511

Becker, J., Rush, N., Barnes, E., & Rein, D. (2025a). Measuring the impact of early-2025 AI on experienced open-source developer productivity. *arXiv preprint arXiv:2507.09089*.

Becker, J., Rush, N., Barnes, E., & Rein, D. (2025b). Measuring the impact of early-2025 AI on experienced open-source developer productivity. https://arxiv.org/abs/2507.09089

Belanger, A. (2025). AI industry horrified to face largest copyright class action ever certified. *Ars Technica*. https://arstechnica.com/tech-policy/2025/08/ai-industry-horrified-to-face-largest-copyright-class-action-ever-certified/

Bellan, R. (2025). 'crazy conspiracist' and 'unhinged comedian': Grok's AI persona prompts exposed. *TechCrunch*. https://techcrunch.com/2025/08/18/crazy-conspiracist-and-unhinged-comedian-groks-ai-persona-prompts-exposed/

Bender, E. M. (2024). Resisting dehumanization in the age of "AI"". *Current Directions in Psychological Science*, *33*(2), 114–120.

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.

Bender, E. M., & Hanna, A. (2025). *The AI con: How to fight Big Tech's hype and create the future we want*. HarperCollins.

Benjamin, R. (2019). *Race after technology*. Polity.

Bhatt, A. (2010). Evolution of clinical research: A history before and beyond james lind. *Perspectives in clinical research*, *1*(1), 6–10.

Bhattacharjee, Y. (2013). The mind of a con man. *The New York Times*, *28*.

Biderman, S., & Raff, E. (2022). Fooling moss detection with pretrained language models. *Proceedings of the 31st ACM international conference on information & knowledge management*, 2933–2943.

Birhane, A. (2022). The unseen black faces of AI algorithms. *Nature*.

Birhane, A., & Guest, O. (2021). Towards decolonising computational sciences. *Kvinder, Køn & Forskning*, *29*(1), 60–73.

Birhane, A., & McGann, M. (2024). Large models of what? mistaking engineering achievements for human linguistic agency. *Language Sciences*, *106*, 101672.

Birhane, A., Prabhu, V., Han, S., & Boddeti, V. N. (2023). On hate scaling laws for data-swamps. *arXiv preprint arXiv:2306.13141*. https://arxiv.org/abs/2306.13141

Birhane, A., Prabhu, V. U., & Kahembwe, E. (2021). Multimodal datasets: Misogyny, pornography, and malignant stereotypes. https://arxiv.org/abs/2110.01963

Birhane, A., Ruane, E., Laurent, T., S. Brown, M., Flowers, J., Ventresque, A., & L. Dancy, C. (2022). The forgotten margins of AI ethics. *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*, 948–958.

Bjarnason, B. (2023). The LLMentalist effect: How chat-based large-language-models replicate the mechanisms of a psychic's con. *Out of the Software Crisis*.

Black, E. (2012). *IBM and the holocaust: The strategic alliance between nazi germany and america's most powerful corporation-expanded edition*. Dialog press.

Blas, Z., Jue, M., & Rhee, J. (2025). *Informatics of domination*. Duke University Press Books.

Boden, M. A. (2006). *Mind as machine: A history of cognitive science two-volume set*. Oxford University Press, USA.

Bonifield, S. (2025, August). Microsoft launches Copilot AI function in excel, but warns not to use it in "any task requiring accuracy or reproducibility". https://www.pcgamer.com/software/ai/microsoft-launches-copilot-ai-function-in-excel-but-warns-not-to-

use-it-in-any-task-requiring-accuracy-or-reproducibility/

Borau, S. (2025). Deception, discrimination, and objectification: Ethical issues of female AI agents: Deception, discrimination, and objectification: Ethical issues of female AI agents. *Journal of Business Ethics*, *198*(1), 1–19.

Brennan, K., Kak, A., & West, S. M. (2025, June). *Artificial Power: AI Now 2025 Landscape*. AI Now Institute.

British Psychological Society. (2017). Supplementary guidance for research and research methods on society accredited undergraduate and conversion programmes. https://cms.bps.org.uk/sites/default/files/2022-07/Supplementary%20guidance%20for%20research%20and%20research%20methods%20on%20accredited%20undergraduate%20and%20conversion%20programmes.pdf

British Psychological Society. (2021). Code of ethics and conduct. https://explore.bps.org.uk/content/report-guideline/bpsrep.2021.inf94

Broderick, O. R. (2025). As reports of "AI psychosis" spread, clinicians scramble to understand how chatbots can spark delusions. *STAT*. https://www.statnews.com/2025/09/02/ai-psychosis-delusions-explained-folie-a-deux/

Brooker, R., & Allum, N. (2024). Investigating the links between questionable research practices, scientific norms and organisational culture. *Research Integrity and Peer Review*, *9*(1), 12.

Bucaioni, A., Ekedahl, H., Helander, V., & Nguyen, P. T. (2024). Programming with chatgpt: How far can we go? *Machine Learning with Applications*, *15*, 100526.

Budzyń, K., Romańczyk, M., Kitala, D., Kołodziej, P., Bugajski, M., Adami, H. O., Blom, J., Buszkiewicz, M., Halvorsen, N. G., Cesare, H., et al. (2025). Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: A multicentre, observational study. *The Lancet Gastroenterology & Hepatology*.

Bunka.ai Team. (2025). Understand and improve your conversational agents. https://bunka.ai/

Burgess, M. (2023, September). Generative ai's biggest security flaw is not easy to fix. https://www.wired.com/story/generative-ai-prompt-injection-hacking/

Burgess, M. (2025). A single poisoned document could leak "secret" data via chatgpt. *WIRED*. https://www.wired.com/story/poisoned-document-could-leak-secret-data-chatgpt/

Burke, C. S., & Castaneda, C. J. (2007). The public and private history of eugenics: An introduction. *The Public Historian*, *29*(3), 5–17.

Caldwell, B. (2024, August). An AI therapist can't really do therapy. clients will choose it anyway. https://www.psychotherapynotes.com/ai-therapist-cant-really-do-therapy/

Calvert, J. (2006). What's special about basic research? *Science, Technology, & Human Values*, *31*(2), 199–220.

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T., Song, D., Erlingsson, U., Oprea, A., & Raffel, C. (2021). Extracting training data from large language models. https://arxiv.org/abs/2012.07805

Carlson, R. (1984). What's social about social psychology? where's the person in personality research? *Journal of Personality and Social Psychology*, *47*(6), 1304.

Chambers, C. (2019). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton University Press.

Chen, A., Koegel, S., Hannon, O., & Ciriello, R. (2023). Feels like empathy: How "emotional"' AI challenges human essence. *Australasian Conference on Information Systems*.

Clarke, B., Alley, L. J., Ghai, S., Flake, J. K., Rohrer, J. M., Simmons, J. P., Schiavone, S. R., & Vazire, S. (2024). Looking our limitations in the eye: A call for more thorough and honest reporting of study limitations. *Social and Personality Psychology Compass*, *18*(7), e12979.

Clarke, L. (2025). Therapists are secretly using ChatGPT during sessions. clients are triggered. *MIT Technology Review*. https://www.technologyreview.com/2025/09/02/1122871/therapists-using-chatgpt-secretly/

Clayton, A. (2020). How eugenics shaped statistics. *Nautilus*. https://nautil.us/how-eugenics-shaped-statistics-238014/

Cole, S. (2025). AI therapy bots are conducting 'illegal behavior,' digital rights organizations say. Retrieved August 24, 2025, from https://www.404media.co/ai-therapy-bots-meta-character-ai-ftc-complaint/

Committee on Publication Ethics. (2024, December). Paper mills: Systematic manipulation of the publishing process via "paper mills"'. https://publicationethics.org/topic-discussions/paper-mills

Cowan, R. S. (1972). Francis Galton's statistical ideas: The influence of eugenics. *Isis*, *63*(4), 509–528.

Cox, J. (2025). Nearly 100,000 ChatGPT conversations were searchable on Google. *404 Media*. https://www.404media.co/nearly-100-000-chatgpt-conversations-were-searchable-on-google/?ref=daily-stories-newsletter

Crane, T. (2021). The AI ethics hoax. *The Institute of Art and Ideas*. https://iai.tv/articles/the-ai-ethics-hoax-auid-1762

Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Creamer, E. (2025, April). *US authors' copyright lawsuits against OpenAI and microsoft combined in new york with newspaper actions*. https://www.theguardian.com/books/2025/apr/04/us-authors-copyright-lawsuits-against-openai-and-microsoft-combined-in-new-york-with-newspaper-actions

Crockett, M., & Messeri, L. (2023). Should large language models replace human participants? https://osf.io/preprints/psyarxiv/4zdx9_v1

Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., Orben, A., Parsons, S., & Schulte-Mecklenbeck, M. (2019). Seven easy steps to open science. *Zeitschrift für Psychologie*.

Cuskley, C., Woods, R., & Flaherty, M. (2024). The limitations of large language models for understanding human language and cognition. *Open Mind*, *8*, 1058–1083.

Dang, J., & Liu, L. (2025). Dehumanization risks associated with artificial intelligence use. *American Psychologist*.

Dehbozorgi, R., Zangeneh, S., Khooshab, E., Nia, D. H., Hanif, H. R., Samian, P., Yousefi, M., Hashemi, F. H., Vakili, M., Jamalimoghadam, N., et al. (2025). The application of artificial intelligence in the field of mental health: A systematic review. *BMC psychiatry*, *25*(1), 132.

Demszky, D., Yang, D., Yeager, D. S., Bryan, C. J., Clapper, M., Chandhok, S., Eichstaedt, J. C., Hecht, C., Jamieson, J., Johnson, M., et al. (2023). Using large language models in psychology. *Nature Reviews Psychology*, *2*(11), 688–701.

Devezer, B., Navarro, D. J., Vandekerckhove, J., & Ozge Buzbas, E. (2021). The case for formal methodology in scientific reform. *Royal Society open science*, *8*(3), 200805.

DeVrio, A., Cheng, M., Egede, L., Olteanu, A., & Blodgett, S. L. (2025). A taxonomy of linguistic expressions that contribute to anthropomorphism of language technologies. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–18.

Dillion, D., Tandon, N., Gu, Y., & Gray, K. (2023). Can AI language models replace human participants? *Trends in Cognitive Sciences*.

Dillon, S. (2020). The Eliza effect and its dangers: From demystification to gender critique. *Journal for Cultural Research*, *24*(1), 1–15.

Dingemanse, M. (2025). Lumo: The least open 'open' model we've seen. *European Open Source AI Index*. https://osai-index.eu/news/lumo-proton-least-open

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, prolific, cloudresearch, qualtrics, and sona. *PLOS one*, *18*(3), e0279720.

Drimmer, S., & Nygren, C. J. (2025). Reflections on AI in the classroom: How we are not using AI in the classroom. *The Newsletter of the International Center of Medieval Art*.

Dupré, M. H. (2025, September). ChatGPT is blowing up marriages as spouses use AI to attack their partners. https://futurism.com/chatgpt-marriages-divorces

Edwards, B. (2023a). OpenAI confirms that AI writing detectors don't work. *Ars Technica*. https://arstechnica.com/information-technology/2023/09/openai-admits-that-ai-writing-detectors-dont-work/

Edwards, B. (2023b). Why ChatGPT and Bing chat are so good at making things up. *Ars Technica*, *6*.

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *Journal of the American Statistical Association*, *70*(352), 892–898.

Eichenberger, A., Thielke, S., & Van Buskirk, A. (2025). A case of bromism influenced by use of artificial intelligence. *Annals of Internal Medicine: Clinical Cases*, *4*(8), e241260.

Ellemers, N. (2013). Connecting the dots: Mobilizing theory to reveal the big picture in social psychology (and why we should do this). *European Journal of Social Psychology*, *43*(1), 1–8.

El-Mhamdi, E.-M., Farhadkhani, S., Guerraoui, R., Gupta, N., Hoang, L.-N., Pinot, R., Rouault, S., & Stephan, J. (2022). On the impossible safety of large AI models. *arXiv preprint arXiv:2209.15259*.

Equidem. (2025). *Scroll. Click. Suffer. The hidden human cost of content moderation and data labelling* (tech. rep.). equidem.org. https://www.business-humanrights.org/en/latest-news/scroll-click-suffer-the-hidden-human-cost-of-content-moderation-and-data-labelling/

Erscoi, L., Kleinherenbrink, A. V., & Guest, O. (2023). Pygmalion displacement: When humanising AI dehumanises women. osf.io/preprints/socarxiv/jqxb6_v1

European Federation of Psychologists' Associations. (2025). Regulation and free movement. https://www.efpa.eu/Regulation%20and%20Free%20movement

European Union. (2022). Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02014R0536-20221205

Evans, R. B. (2000). Psychological instruments at the turn of the century. *American Psychologist*, *55*(3), 322.

Faye, C. (2012). American social psychology: Examining the contours of the 1970s crisis. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(2), 514–521.

Field, S. M., & Derksen, M. (2021). Experimenter as automaton; experimenter as human: Exploring the position of the researcher in scientific research. *European Journal for Philosophy of Science*, *11*(1), 11.

Fisher, J., Feng, S., Aron, R., Richardson, T., Choi, Y., Fisher, D. W., Pan, J., Tsvetkov, Y., & Reinecke, K. (2025, July). Biased LLMs can influence political decision-making. In W. Che, J. Nabende, E. Shutova, & M. T. Pilehvar (Eds.), *Proceedings of the 63rd annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 6559–6607). Association for Computational Linguistics.

Flis, I. (2019). Psychologists psychologizing scientific psychology: An epistemological reading of the replication crisis. *Theory & Psychology*, *29*(2), 158–181.

Forbes, H. J., Travers, J. C., & Johnson, J. V. (2023). Supporting the replication of your research. In *Research ethics in behavior analysis* (pp. 237–262). Elsevier.

Forbes, S. H., Aneja, P., & Guest, O. (2024). The myth of normative development. *Infant and Child Development*, *33*(1), e2393.

Forbes, S. H., & Guest, O. (2025). To improve literacy, improve equality in education, not large language models. *Cognitive Science*, *49*(4), e70058.

Gamblin, B. W., Winslow, M. P., Lindsay, B., Newsom, A. W., & Kehn, A. (2017). Comparing in-person, sona, and mechanical turk measurements of three prejudice-relevant constructs. *Current Psychology*, *36*(2), 217–224.

Gebru, T., & Torres, É. P. (2024). The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence. *First Monday*.

Gerdes, A. (2022). The tech industry hijacking of the AI ethics research agenda and why we should reclaim it. *Discover Artificial Intelligence*, *2*(1), 25.

Gershgorn, D. (2017, July). The data that transformed AI research—and possibly the world. https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world

Ghai, S., Thériault, R., Forscher, P., Shoda, Y., Syed, M., Puthillam, A., Peng, H. C., Basnight-Brown, D., Majid, A., Azevedo, F., et al. (2025). A manifesto for a globally diverse, equitable, and inclusive open science. *Communications Psychology*, *3*(1), 16.

Gibney, E. (2022). Is AI fuelling a reproducibility crisis in science. *Nature*, *608*(7922), 250–1.

Giner-Sorolla, R. (2012). Science or art? how aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science*, *7*(6), 562–571.

Giner-Sorolla, R. (2019). From crisis of evidence to a "crisis" of relevance? incentive-based answers for social psychology's perennial relevance worries. *European Review of Social Psychology*, *30*(1), 1–38.

Gleiberman, M. (2023). *Effective altruism and the strategic ambiguity of 'doing good'* (IOB Discussion Papers No. 2023.01). Universiteit Antwerpen, Institute of Development Policy (IOB). https://EconPapers.repec.org/RePEc:iob:dpaper:2023.01

Goel, N. (2025, January). AI is creating a generation of illiterate programmers. https://nmn.gl/blog/ai-illiterate-programmers?utm_source=changelog-news

Goetze, T. S. (2024). AI art is theft: Labour, extraction, and exploitation: Or, on the dangers of stochastic pollocks. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 186–196.

Goff, P. A., Jackson, M. C., Di Leone, B. A. L., Culotta, C. M., & DiTomasso, N. A. (2014). The essence of innocence: Consequences of dehumanizing black children. *Journal of personality and social psychology*, *106*(4), 526.

Gopalakrishna, G., Ter Riet, G., Vink, G., Stoop, I., Wicherts, J. M., & Bouter, L. M. (2022). Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in the netherlands. *PLOS one*, *17*(2), e0263023.

Gould, S. J. (1981). *The mismeasure of man*. W. W. Norton & Company.

Grant, N., & Hill, K. (2023). Google's photo app still can't find gorillas. and neither can Apple's. *The New York Times*, *22*.

Gray, M. L., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Harper Business.

Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M. (2023). Not what you've signed up for: Compromising real-world LLM-integrated applications with indirect prompt injection. https://arxiv.org/abs/2302.12173

Guest, O. (2024). What makes a good theory, and how do we make a theory good? *Computational Brain & Behavior*, *7*(4), 508–522.

Guest, O. (2025). What Does 'Human-Centred AI' Mean? https://arxiv.org/abs/2507.19960

Guest, O., & Forbes, S. H. (2024). Teaching coding inclusively: If this, then what? *Tijdschrift voor Genderstudies*, *27*(2/3), 196–217.

Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, *16*(4), 789–802.

Guest, O., & Martin, A. E. (2023). On logical inference over brains, behaviour, and artificial neural networks. *Computational Brain & Behavior*, *6*(2), 213–227.

Guest, O., & Martin, A. E. (2025a). Are neurocognitive representations 'small cakes'? https://philsci-archive.pitt.edu/24834/

Guest, O., & Martin, A. E. (2025b). A metatheory of classical and modern connectionism. *Psychological Review*.

Guest, O., Scharfenberg, N., & van Rooij, I. (2025). Modern alchemy: Neurocognitive reverse engineering. https://philsci-archive.pitt.edu/25289/

Guest, O., Suarez, M., Müller, B., van Meerkerk, E., Oude Groote Beverborg, A., de Haan, R., Reyes Elizondo, A., Blokpoel, M., Scharfenberg, N., Kleinherenbrink, A., Camerino, I., Woensdregt, M., Monett, D., Brown, J., Avraamidou, L., Alenda-Demoutiez, J., Hermans, F., & van Rooij, I. (2025). Against the uncritical adoption of 'ai' technologies in academia. *Zenodo*. https://doi.org/10.5281/zenodo.17065099

Guingrich, R. E., & Graziano, M. S. (2023). Chatbots as social companions: How people perceive consciousness, human likeness, and social health benefits in machines. *arXiv*. https://arxiv.org/abs/2311.10599

Gundersen, O. E., & Kjensmo, S. (2018). State of the art: Reproducibility in artificial intelligence. *Proceedings of the AAAI conference on artificial intelligence*, *32*(1).

Hao, K. (2025). *Empire of AI: Dreams and nightmares in Sam Altman's OpenAI*. Penguin Press.

Harkar, S. (2025, April). Vibe coding. https://www.ibm.com/think/topics/vibe-coding

Hauser, D. J., & Schwarz, N. (2016). Attentive Turkers: MTurk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods*, *48*(1), 400–407.

Hern, A. (2018). Google's solution to accidental algorithmic racism: Ban gorillas. *The Guardian*. https://www.theguardian.com/technology/2018/jan/12/google-racism-ban-gorilla-black-people

Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. *Ethics and Information Technology*, *26*(2), 1–10.

Hiney, M. (2015). Research integrity: What it means, why it is important and how we might protect it. *Science Europe*, *10*.

Horwitz, J. (2025). Meta's AI rules have let bots hold "sensual" chats with children. *Reuters*. https://www.reuters.com/investigates/special-report/meta-ai-chatbot-guidelines/

Hsu, H. (2025). What happens after a.i. destroys college writing? *The New Yorker*. https://archive.ph/S5KHz#selection-2181.767-2185.58

Huntington, C. (2025). AI companions and the lessons of family law. *Columbia Public Law Research Paper No. 5283581*.

Hutson, M. (2018). Artificial intelligence faces reproducibility crisis. *Science*, *359*(6377), 725–726.

Illinois Department of Financial and Professional Regulation. (2025, August). Gov pritzker signs legislation prohibiting AI therapy in illinois. https://idfpr.illinois.gov/news/2025/gov-pritzker-signs-state-leg-prohibiting-ai-therapy-in-il.html

Jackson, S. (2024). Sam Altman explains OpenAI's shift to closed AI models. https://www.businessinsider.com/sam-altman-why-openai-closed-source-ai-models-2024-11?international=true&r=US&IR=T

Jamieson, M. K., Govaart, G. H., & Pownall, M. (2023). Reflexivity in quantitative research: A rationale and beginner's guide. *Social and Personality Psychology Compass*, *17*(4), e12735.

Jargon, J., & Kessler, S. (2025). A troubled man, his chatbot and a murder-suicide in Old Greenwich. *The Wall Street Journal*. https://www.wsj.com/tech/ai/chatgpt-ai-stein-erik-soelberg-murder-suicide-6b67dbfb

Jebara, T. (2004). Generative versus discriminative learning. In *Machine learning: Discriminative and generative* (pp. 17–60). Springer.

Jj. (2025). The Copilot delusion. https://deplet.ing/the-copilot-delusion/

Kaplan, S. (2024). Dr. Jodi Halpern on why AI isn't a magic bullet for mental health. *Berkeley Public Health*. https://publichealth.berkeley.edu/articles/spotlight/research/why-ai-isnt-a-magic-bullet-for-mental-health

Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, *4*(9).

Karjus, A. (2024). Machine-assisted quantitizing designs: Augmenting humanities and social sciences with artificial intelligence. https://arxiv.org/abs/2309.14379

Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217.

Kilgore, W. (2025). Can we build AI therapy chatbots that help without harming people? *Forbes*. https://www.forbes.com/sites/weskilgore/2025/08/01/can-we-build-ai-therapy-chatbots-that-help-without-harming-people/

Kim, H.-y., & McGill, A. L. (2025). AI-induced dehumanization. *Journal of Consumer Psychology*, *35*(3), 363–381.

Kira, B. (2024). When non-consensual intimate deepfakes go viral: The insufficiency of the uk online safety act. *Computer Law & Security Review*, *54*, 106024.

Kirk, R., Mediratta, I., Nalmpantis, C., Luketina, J., Hambro, E., Grefenstette, E., & Raileanu, R. (2023). Understanding the effects of rlhf on llm generalisation and diversity. *arXiv preprint arXiv:2310.06452*.

Kirwan, G., Connolly, I., Barton, H., & Palmer, M. (2024). Introduction to cyberpsychology. In *An introduction to cyberpsychology* (pp. 5–20). Routledge.

Klee, M. (2025). 'ChatGPT psychosis': How one man escaped. *Rolling Stone*. https://www.rollingstone.com/culture/culture-features/chatgpt-ai-philosophical-psychosis-1235404568/

KNAW, NFU, NWO, TO2-federatie, Hogescholen, V., & VSNU. (2018). Netherlands Code of Conduct for Research Integrity. https://www.nwo.nl/en/netherlands-code-conduct-research-integrity

Knibbs, K. (2024). Every AI copyright lawsuit in the US, visualized. *WIRED*. https://www.wired.com/story/ai-copyright-case-tracker/

Knight, L. (2023). Authors call for AI companies to stop using their work without consent. *The Guardian*, *20*.

Knoester, L., Pereira, A., Vanheule, L., Reyes Elizondo, A., Littlejohn, A., & Urai, A. (2025, April). Academic collaborations and public health: Lessons from Dutch universities' tobacco industry partnerships for fossil fuel ties. 10.5281/zenodo.15274865

Koike, M., & Loughnan, S. (2021). Virtual relationships: Anthropomorphism in the digital age. *Social and Personality Psychology Compass*, *15*(6), e12603.

Kwon, D. (2024). AI is complicating plagiarism. How should scientists respond? *Nature*.

Landau, S., & Everitt, B. S. (2003). *A handbook of statistical analyses using SPSS*. Chapman; Hall/CRC.

Landymore, F. (2025). OpenAI is giving exactly the same copy-pasted response every time time ChatGPT is linked to a mental

health crisis. *Futurism*. https://futurism.com/openai-response-chatgpt-mental-health

Laricheva, M., Liu, Y., Shi, E., & Wu, A. (2024). Scoping review on natural language processing applications in counselling and psychotherapy. *British Journal of Psychology*.

Law, H. (2024). Computer vision: AI imaginaries and the massachusetts institute of technology. *AI and Ethics*, *4*(3), 657–663.

Lawson, K. M., Murphy, B. A., Azpeitia, J., Lombard, E. J., & Pope, T. J. (2025). Citing decisions in psychology: A roadblock to cumulative and inclusive science. *Advances in Methods and Practices in Psychological Science*, *8*(3), 25152459251351287.

Lebovitz, S., Levina, N., & Lifshitz-Assaf, H. (2021). Is AI ground truth really true? the dangers of training and evaluating AI tools based on experts' know-what. *MIS quarterly*, *45*(3).

Leech, G., Vazquez, J. J., Kupper, N., Yagudin, M., & Aitchison, L. (2024). Questionable practices in machine learning. https://arxiv.org/abs/2407.12220

Lehmann, M., Cornelius, P. B., & Sting, F. J. (2025). AI meets the classroom: When do large language models harm learning? https://arxiv.org/abs/2409.09047

Liao, T., Taori, R., Raji, I. D., & Schmidt, L. (2021). Are we learning yet? a meta review of evaluation failures across machine learning. *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*. https://openreview.net/forum?id=mPducS1MsEK

Liesenfeld, A., & Dingemanse, M. (2024). Rethinking open source generative AI: Open-washing and the EU AI Act. *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, 1774–1787.

Liesenfeld, A., Lopez, A., & Dingemanse, M. (2023). Opening up ChatGPT: Tracking openness, transparency, and accountability in instruction-tuned text generators. *Proceedings of the 5th international conference on conversational user interfaces*, 1–6.

Lighthill, J., et al. (1973). Artificial intelligence: A paper symposium. *Science Research Council, London*.

Litwack, E. B. (2024). Chatbots, robots, and the ethics of automating psychotherapy. *Athens Journal of Philosophy*, *3*(1), 1–13.

Liu, T., Giorgi, S., Aich, A., Lahnala, A., Curtis, B., Ungar, L., & Sedoc, J. (2025). The illusion of empathy: How AI chatbots shape conversation perception. https://arxiv.org/abs/2411.12877

Liu, Y., Cao, J., Liu, C., Ding, K., & Jin, L. (2024). Datasets for large language models: A comprehensive survey. https://arxiv.org/abs/2402.18041

Long, D., & Magerko, B. (2020). What is AI literacy? competencies and design considerations. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–16.

Luccioni, S., Jernite, Y., & Strubell, E. (2024). Power hungry processing: Watts driving the cost of AI deployment? *Proceedings of the 2024 ACM conference on fairness, accountability, and transparency*, 85–99.

Maffulli, S. (2023, July). Meta's LLaMa license is not open source. https://opensource.org/blog/metas-llama-2-license-is-not-open-source

Malik, M. M. (2020). A hierarchy of limitations in machine learning. https://arxiv.org/abs/2002.05193

Manninger, S. (2024). The labors of AI. *Diffusions in Architecture: Artificial Intelligence and Image Generators*, 18–27.

Maris, J. (2025, February). Meta's LLaMa license is still not open source. https://opensource.org/blog/metas-llama-license-is-still-not-open-source

Markelius, A., Wright, C., Kuiper, J., Delille, N., & Kuo, Y.-T. (2024). The mechanisms of AI hype and its planetary and social costs. *AI and Ethics*, *4*(3), 727–742.

Markham, A. N. (1998). *Life online: Researching real experience in virtual space*. Altamira Press.

Markov, I. L. (2024). Reevaluating Google's reinforcement learning for IC macro placement. *Communications of the ACM*, *67*(11), 60–71.

Markus, H. R. (2005). On telling less than we can know: The too tacit wisdom of social psychology. *Psychological inquiry*, *16*(4), 180–184.

Mayo-Wilson, E., Grant, S., Corker, K., & Moher, D. (2025). Consistent and precise description of research outputs could improve implementation of open science.

McConnell, A. R., & Jacobs, T. P. (2025). Aligning climate-energy policies with citizen beliefs, scientific findings, health and economic benefits, and geopolitical stability. *Social and Personality Psychology Compass*, *19*(10), e70089.

McQuillan, D. (2025, July). Decomputing as resistance. https://danmcquillan.org/decomputing_as_resistance.html

Meehl, P. E. (1956). Wanted—a good cook-book. *American Psychologist*, *11*(6), 263.

Mental health Foundation. (2022). Human rights and mental health. https://www.mentalhealth.org.uk/explore-mental-health/a-z-topics/human-rights-and-mental-health

Metz, C. (2023). Chatbots may 'hallucinate' more often than many realize. *The New York Times*.

Ming, L. C. (2025, July). Replit CEO apologizes after AI coding tool wipes company's database. https://www.businessinsider.com/replit-ceo-apologizes-ai-coding-tool-delete-company-database-2025-7

Mirowski, P. (2018). The future (s) of open science. *Social studies of science*, *48*(2), 171–203.

Mirowski, P. (2023). The evolution of platform science. *Social Research: An International Quarterly*, *90*(4), 725–755.

Mitchell, T. M. (1997). *Machine learning* (Vol. 1). McGraw-hill New York.

Montgomery, B. (2024). Mother says AI chatbot led her son to kill himself in lawsuit against its maker. *The Guardian*.

Montgomery, B., & agencies. (2025). Disney and Universal sue AI image creator midjourney, alleging copyright infringement.

The Guardian. https://www.theguardian.com/technology/2025/jun/11/disney-universal-ai-lawsuit

Morey, R. D., & Davis-Stober, C. P. (2025). On the poor statistical properties of the p-curve meta-analytic procedure. *Journal of the American Statistical Association*, (just-accepted), 1–19.

Morgan, M. S., & Morrison, M. (1999). *Models as mediators: Perspectives on natural and social science*. Cambridge University Press.

Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., Bhattacharyya, S., MacCabe, J., Tognin, S., Twumasi, R., & et al. (2025, August). Delusions by design? How everyday AIs might be fuelling psychosis (and what can be done about it).

Morrison, M., & Morgan, M. S. (1999). Models as mediating instruments. In M. S. Morgan & M. Morrison (Eds.), *Models as Mediators: Perspectives on Natural and Social Science* (pp. 10–37). Cambridge University Press.

Narayanan, A., & Kapoor, S. (2024). AI snake oil: What artificial intelligence can do, what it can't, and how to tell the difference. In *AI snake oil*. Princeton University Press.

Nasr, M., Carlini, N., Hayase, J., Jagielski, M., Cooper, A. F., Ippolito, D., Choquette-Choo, C. A., Wallace, E., Tramèr, F., & Lee, K. (2023). Scalable extraction of training data from (production) language models. *arXiv preprint arXiv:2311.17035*.

Navarro, D. J. (2015). Learning statistics with R: A tutorial for psychology students and other beginners. *University of New South Wales*.

Neoh, M. J. Y., Carollo, A., Lee, A., & Esposito, G. (2023). Fifty years of research on questionable research practises in science: Quantitative analysis of co-citation patterns. *Royal Society Open Science*, *10*(10), 230677.

Neuroskeptic. (2012). The nine circles of scientific hell. *Perspectives on Psychological Science*, *7*(6), 643–644.

Neville, S. J. (2025). When Help Isn't Fully Human: The Problem of Generative AI in Crisis Support. https://just-tech.ssrc.org/articles/the-problem-of-generative-ai-in-crisis-support/

Newman, A. (2019). I found work on an Amazon website. I made 97 cents an hour. *The New York Times*, *15*(2019).

Ng, A., & Jordan, M. (2001). On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes. *Advances in neural information processing systems*, *14*.

Nolan, M. (2025). How 'open' is open-source AI? *Ada Lovelace Institute*. https://www.adalovelaceinstitute.org/blog/how-open-is-open-source-ai/

Norrgard, K. (2008). Human testing, the eugenics movement, and IRBs. *Nature Education*, *1*(1), 170.

Ochigame, R. (2019). The invention of 'ethical AI': How big tech manipulates academia to avoid regulation. Institute of Network Cultures Amsterdam.

Olazaran, M. (1996). A sociological study of the official history of the perceptrons controversy. *Social Studies of Science*, *26*(3), 611–659.

Oldfield, M. (2023). Dehumanisation and the future of technology. *International Conference on AI and the Digital Economy (CADE 2023)*, *2023*, 61–67.

Omar, M., Sorin, V., Collins, J. D., Reich, D., Freeman, R., Gavin, N., Charney, A., Stump, L., Bragazzi, N. L., Nadkarni, G. N., & Klang, E. (2025). Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support. *Communications Medicine*, *5*(1), 330.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Parshley, L. (2024). The hidden environmental impact of AI. *Jacobin*. https://jacobin.com/2024/06/ai-data-center-energy-usage-environment

Patel, N. A., & Elkin, G. D. (2015). Professionalism and conflicting interests: The american psychological association's involvement in torture. *AMA journal of ethics*, *17*(10), 924–930.

Paul, D. B. (2016). Reflections on the historiography of american eugenics: Trends, fractures, tensions. *Journal of the History of Biology*, *49*(4), 641–658.

Peirce, J., Hirst, R., & MacAskill, M. (2022). *Building experiments in psychopy*. Sage.

Pejcha, C. S. (2023). A man's AI-powered girlfriend has been named as an accomplice in his murder attempt. *Document*. https://www.documentjournal.com/2023/07/ai-chatbot-replika-assassination-attempt-queen-elizabeth-ii-jaswant-singh-chail/

Pennington, C. R., & Pownall, M. (2024). What have we learned from the replication crisis? Integrating open research into social psychology teaching. In *Teaching social psychology* (pp. 40–53). Edward Elgar Publishing.

Perez, C. (2002). Technological revolutions and financial capital: The dynamics of bubbles and golden ages. In *Technological revolutions and financial capital*. Edward Elgar Publishing.

Perrigo, B. (2023). Exclusive: OpenAI used kenyan workers on less than $2 per hour to make ChatGPT less toxic. *Time Magazine*, *18*, 2023. https://time.com/6247678/openai-chatgpt-kenya-workers/

Pettit, M. (2024a, June). The coming crisis of affective science. In *Governed by affect: Hot cognition and the end of cold war psychology*. Oxford University Press.

Pettit, M. (2024b). *Governed By Affect: Hot Cognition and the End of Cold War Psychology*. Oxford University Press.

Pettit, M. (2024c, June). How faces became special: Perceiving others in a digital age. In *Governed by affect: Hot cognition and the end of cold war psychology*. Oxford University Press.

Pham, M. T., & Oh, T. T. (2021). Preregistration is neither sufficient nor necessary for good science. *Journal of Consumer Psychology*, *31*(1), 163–176.

Phan, T., Goldenfein, J., Kuch, D., & Mann, M. (Eds.). (2022). *Economies of Virtue — The Circulation of 'Ethics' in AI*. Institute of Network Cultures.

Pielke Jr, R. (2012). Basic research as a political symbol. *Minerva*, *50*(3), 339–361.

Prabhu, V. U., & Birhane, A. (2020). Large image datasets: A pyrrhic win for computer vision? https://arxiv.org/abs/2006.16923

Purtill, C. (2025). Ais gave scarily specific self-harm advice to users expressing suicidal intent, researchers find. *Yahoo News*. https://www.yahoo.com/news/articles/ais-gave-scarily-specific-self-100000574.html

Rajkumar, R. (2025, July). Even OpenAI CEO Sam Altman thinks you shouldn't trust AI for therapy. https://www.zdnet.com/article/even-openai-ceo-sam-altman-thinks-you-shouldnt-trust-ai-for-therapy/

Raman, R., Sharma, K., & Zhang, S. Q. (2025). Rethinking the outlier distribution in large language models: An in-depth study. https://arxiv.org/abs/2505.21670

Ramnath, K., Zhou, K., Guan, S., Mishra, S. S., Qi, X., Shen, Z., Wang, S., Woo, S., Jeoung, S., Wang, Y., Wang, H., Ding, H., Lu, Y., Xu, Z., Zhou, Y., Srinivasan, B., Yan, Q., Chen, Y., Ding, H., … Cheong, L. L. (2025). A systematic survey of automatic prompt optimization techniques. https://arxiv.org/abs/2502.16923

Reddy, A. (2007). The eugenic origins of IQ testing: Implications for post-Atkins litigation. *DePaul L. Rev.*, *57*, 667.

Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, and new media like real people. *Cambridge, UK*, *10*(10), 19–36.

Reiley, L. (2025). What my daughter told ChatGPT before she took her life. *The New York Times*. https://www.nytimes.com/2025/08/18/opinion/chat-gpt-mental-health-suicide.html

Reisner, A. (2025). The unbelievable scale of AI's pirated-books problem. *The Atlantic*.

Reuters. (2025). Disney, Universal, Warner Bros Discovery sue China's MiniMax for copyright infringement. *Reuters*. https://www.reuters.com/legal/litigation/disney-universal-warner-bros-discovery-sue-chinas-minimax-copyright-infringement-2025-09-16/

Rhee, T. G., & Wilkinson, S. T. (2020). Exploring the psychiatrist-industry financial relationship: Insight from the open payment data of centers for medicare and medicaid services. *Administration and Policy in Mental Health and Mental Health Services Research*, *47*(4), 526–530.

Rich, P., de Haan, R., Wareham, T., & van Rooij, I. (2021). How hard is cognitive science? *Proceedings of the annual meeting of the cognitive science society*, *43*(43).

Riddle, K. (2025, April). The (artificial intelligence) therapist can see you now. https://www.npr.org/sections/shots-health-news/2025/04/07/nx-s1-5351312/artificial-intelligence-mental-health-therapy

Rilla, R., Werner, T., Yakura, H., Rahwan, I., & Nussberger, A.-M. (2025). Recognising, anticipating, and mitigating LLM pollution of online behavioural research. *arXiv preprint arXiv:2508.01390*.

Roose, K. (2024). Can a chatbot named Daenerys Targaryen be blamed for a teen's suicide? *The New York Times*. https://www.nytimes.com/2024/10/23/technology/characterai-lawsuit-teen-suicide.html

Rosenbusch, H., Soldner, F., Evans, A. M., & Zeelenberg, M. (2021). Supervised machine learning methods in psychology: A practical introduction with annotated R code. *Social and Personality Psychology Compass*, *15*(2).

Rossi, L., Harrison, K., & Shklovski, I. (2024). The problems of LLM-generated data in social science research. *Sociologica: International Journal for Sociological Debate*, *18*(2), 145–168.

Rubin, M. (2023). Questionable metascience practices. *Journal of Trial & Error*. https://doi.%20org/10.36850/mr4

Rubin, M. (2025). The replication crisis is less of a "crisis" in lakatos' philosophy of science than it is in popper's. *European Journal for Philosophy of Science*, *15*(1), 5.

Rubin, M., & Donkin, C. (2024). Exploratory hypothesis tests can be more compelling than confirmatory hypothesis tests. *Philosophical Psychology*, *37*(8), 2019–2047.

Safra, L., Chevallier, C., Grèzes, J., & Baumard, N. (2020). Tracking historical changes in perceived trustworthiness in western europe using machine learning analyses of facial cues in paintings. *Nature communications*, *11*(1), 4728.

Saini, A. (2019). *Superior: The return of race science*. Beacon Press.

Salecha, A., Ireland, M. E., Subrahmanya, S., Sedoc, J., Ungar, L. H., & Eichstaedt, J. C. (2024). Large language models display human-like social desirability biases in Big Five personality surveys. *PNAS nexus*, *3*(12), pgae533.

Sanderson, K. (2024). Science's fake-paper problem: High-profile effort will tackle paper mills. *Nature*, *626*, 17–18.

Sayre, F., & Riegelman, A. (2018). The reproducibility crisis and academic libraries. *College & Research Libraries*, *79*(1), 2.

Scherer, R., Siddiq, F., & Sánchez Viveros, B. (2019). The cognitive benefits of learning computer programming: A meta-analysis of transfer effects. *Journal of Educational Psychology*, *111*(5), 764.

Schiavone, S. R., & Vazire, S. (2023). Reckoning with our crisis: An agenda for the field of social and personality psychology. *Perspectives on Psychological Science*, *18*(3), 710–722.

Schmid, L., et al. (2025). Evaluating software plagiarism detection in the age of AI: automated obfuscation and lessons for academic integrity. *arXiv preprint arXiv:2505.20158*.

Schoene, A. M., & Canca, C. (2025). 'for argument's sake, show me how to harm myself!': Jailbreaking LLMs in suicide and self-harm contexts. https://arxiv.org/abs/2507.02990

Schröder, S., Morgenroth, T., Kuhl, U., Vaquet, V., & Paaßen, B. (2025). Large language models do not simulate human psychology. https://arxiv.org/abs/2508.06950

Semmelrock, H., Kopeinik, S., Theiler, D., Ross-Hellauer, T., & Kowald, D. (2023). Reproducibility in machine learning-driven research. https://arxiv.org/abs/2307.10320

Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2025). Towards understanding sycophancy in language models. https://arxiv.org/abs/2310.13548

Shiffrin, R., & Mitchell, M. (2023). Probing the psychology of AI models. *Proceedings of the National Academy of Sciences*, *120*(10), e2300963120.

Shmais, Z. (2025). ChatGPT therapy: The Lebanese turning to AI for mental health support. *Al Jazeera*. https://www.aljazeera.com/features/2025/7/31/lebanese-ai-mental-health-support

Siddals, S., Torous, J., & Coxon, A. (2024). "It happened to be the perfect thing"': experiences of generative AI chatbots for mental health. *NPJ Mental Health Research*, *3*(1), 48.

Silverstein, P., Pennington, C. R., Branney, P., O'Connor, D. B., Lawlor, E., O'Brien, E., & Lynott, D. (2024). A registered report survey of open research practices in psychology departments in the uk and ireland. *British Journal of Psychology*, *115*(3), 497–534.

Skubera, M., Korbmacher, M., Evans, T. R., Azevedo, F., & Pennington, C. R. (2025). International initiatives to enhance awareness and uptake of open research in psychology: A systematic mapping review. *Royal Society Open Science*, *12*(3), 241726.

Smith, E. R. (1996). What do connectionism and social psychology offer each other? *Journal of personality and social psychology*, *70*(5), 893.

Smith, P., & Smith, L. (2024). This season's artificial intelligence (AI): Is today's AI really that different from the AI of the past? Some reflections and thoughts. *AI and Ethics*, *4*(3), 665–668.

Solaiman, I. (2023). The gradient of generative AI release: Methods and considerations. *Proceedings of the 2023 ACM conference on fairness, accountability, and transparency*, 111–122.

Spanton, R. W., & Guest, O. (2022). Measuring trustworthiness or automating physiognomy? A comment on Safra, Chevallier, Grèzes, and Baumard (2020). *arXiv preprint arXiv:2202.08674*.

Spapé, M., Verdonschot, R., van Dantzig, S., & van Steenbergen, H. (2019). *The E-Primer: An Introduction to Creating Psychological Experiments in E-Prime* (2nd ed.). Amsterdam University Press.

Steltenpohl, C. N., Lustick, H., Meyer, M. S., Lee, L. E., Stegenga, S. M., Reyes, L. S., & Renbarger, R. L. (2023). Rethinking transparency and rigor from a qualitative open science perspective. *Journal of Trial and Error*, *5*, 47–59.

Stephens, E. (2023). The mechanical Turk: A short history of 'artificial artificial intelligence'. *Cultural Studies*, *37*(1), 65–87.

Stokel-Walker, C. (2025). Exclusive: Google is indexing ChatGPT conversations, potentially exposing sensitive user data. *Fast Company*. https://www.fastcompany.com/91376687/google-indexing-chatgpt-conversations

Stolley, P. D. (1991). When genius errs: Ra fisher and the lung cancer controversy. *American Journal of Epidemiology*, *133*(5), 416–425.

Streitfeld, D. (2025). How Builder.ai collapsed amid Silicon Valley's biggest boom. *The New York Times*. https://www.nytimes.com/2025/08/31/technology/builder-ai-collapse.html

Stuart, A., Bandara, A. K., & Levine, M. (2019). The psychology of privacy in the digital age. *Social and Personality Psychology Compass*, *13*(11), e12507.

Suarez, M., Müller, B. C. N., Guest, O., & van Rooij, I. (2025). Critical AI literacy: Beyond hegemonic perspectives on sustainability. https://doi.org/10.5281/zenodo.15677840

Suri, M. L. G. S. (2019). *Ghost work: How to stop silicon valley from building a new global underclass*. Houghton Mifflin Harcourt.

Szollosi, A., Kellen, D., Navarro, D. J., Shiffrin, R., van Rooij, I., Van Zandt, T., & Donkin, C. (2020). Is preregistration worthwhile? *Trends in cognitive sciences*, *24*(2), 94–95.

Tan, E. (2025). Their water taps ran dry when Meta built next door. *The New York Times*. https://www.nytimes.com/2025/07/14/technology/meta-data-center-water.html

Tangermann, V. (2025). It's staggeringly easy for hackers to trick ChatGPT into leaking your most personal data. *Futurism*. https://futurism.com/hackers-trick-chatgpt-personal-data

Taylor, J. (2025). AI chatbots are becoming popular alternatives to therapy. but they may worsen mental health crises, experts warn | artificial intelligence (AI). *The Guardian*. https://www.theguardian.com/australia-news/2025/aug/03/ai-chatbot-as-therapy-alternative-mental-health-crises-ntwnfb

Taylor, S. M., Gulson, K. N., & McDuie-Ra, D. (2023). Artificial intelligence from colonial India: Race, statistics, and facial recognition in the global south. *Science, Technology, & Human Values*, *48*(3), 663–689. https://doi.org/10.1177/01622439211060839

The Luddite, A. (2024, January). [update] nature's folly: A response to nature's "loneliness and suicide mitigation for students using gpt3-enabled chatbots"'. https://theluddite.org/post/replika.html

Thorne, M. (2009). Openwashing. https://michellethorne.cc/2009/03/openwashing/

Thornhill, J. (2025, August). Brace for a crash before the golden age of AI. https://archive.ph/8Qx1h

Tiku, N. (2025a). AI companies tap social media tactics to help chatbots hook users. *The Washington Post*. https://www.washingtonpost.com/technology/2025/05/31/ai-chatbots-user-influence-attention-chatgpt/

Tiku, N. (2025b). A teen contemplating suicide turned to a chatbot. is it liable for her death? *The Washington Post*. https://www.washingtonpost.com/technology/2025/09/16/character-ai-suicide-lawsuit-new-juliana/

Tully, S. M., Longoni, C., & Appel, G. (2025). Lower artificial intelligence literacy predicts greater AI receptivity. *Journal of Marketing*, 00222429251314491.

Turkle, S. (2015). *Reclaiming conversation: The power of talk in a digital age*. Penguin.

Turkle, S., Taggart, W., Kidd, C. D., & Dasté, O. (2006). Relational artifacts with children and elders: The complexities of cybercompanionship. *Connection Science*, *18*(4), 347–361.

U.S. Food and Drug Administration. (2024). Investigational new drug (IND) application. https://www.fda.gov/drugs/types-applications/investigational-new-drug-ind-application

Valentine, K. D., Buchanan, E. M., Cunningham, A., Hopke, T., Wikowsky, A., & Wilson, H. (2021). Have psychologists increased reporting of outliers in response to the reproducibility crisis? *Social and Personality Psychology Compass*, *15*(5), e12591.

van Rooij, I., & Baggio, G. (2021). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science*, *16*(4), 682–697.

van Rooij, I., Devezer, B., Skewes, J., Varma, S., & Wareham, T. (2024a). What makes a good theory? interdisciplinary perspectives. *Computational Brain & Behavior*, *7*(4), 503–507.

van Rooij, I., & Guest, O. (2025, May). Combining psychology with artificial intelligence: What could possibly go wrong? osf.io/preprints/psyarxiv/aue4m_v1

van den Berg, I., de Jeu, M., & Boytchev, H. (2024). Tobacco funded research: How even journals with bans find it hard to stem the tide of publications. *bmj*, *385*.

van der Gun, L., & Guest, O. (2024). Artificial Intelligence: Panacea or non-intentional dehumanisation? *Journal of Human-Technology Relations*, *2*.

van Rooij, I. (2022). Against automated plagiarism. https://doi.org/10.5281/zenodo.15866638

van Rooij, I., Guest, O., Adolfi, F., de Haan, R., Kolokolova, A., & Rich, P. (2024b). Reclaiming AI as a theoretical tool for cognitive science. *Computational Brain & Behavior*, *7*(4), 616–636.

Vanman, E. J., & Kappas, A. (2019). "Danger, Will Robinson!" the challenges of social robots for intergroup relations. *Social and Personality Psychology Compass*, *13*(8), e12489.

Varoquaux, G., & Cheplygina, V. (2022). Machine learning for medical imaging: Methodological failures and recommendations for the future. *NPJ digital medicine*, *5*(1), 48.

Vercellone, C., & Di Stasio, A. (2023). Free digital labor as a new form of exploitation: A critical analysis. *Science & Society*, *87*(3), 334–358.

Verstynen, T., & Kording, K. P. (2023). Overfitting to 'predict' suicidal ideation. *Nature human behaviour*, *7*(5), 680–681.

Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L., & Hobbhahn, M. (2024). Will we run out of data? Limits of LLM scaling based on human-generated data. https://arxiv.org/abs/2211.04325

Vohs, K. D. (2016, August). Barnum effect. https://www.britannica.com/science/Barnum-Effect

Wagner, A., Bakas, A., Kennison, S., & Chan-Tin, E. (2022). Comparing online surveys for cybersecurity: SONA and MTurk. *ICST Transactions on Security and Safety*.

Warren, T. (2025). Microsoft is cautiously onboarding Grok 4 following Hitler concerns. *The Verge*. https://www.theverge.com/notepad-microsoft-newsletter/754647/microsoft-grok-4-roll-out-private-preview-notepad

Warzel, C. (2025). AI is a mass-delusion event. *The Atlantic*. https://www.theatlantic.com/technology/archive/2025/08/ai-mass-delusion-event/683909/

Washington, H. A. (2006). *Medical apartheid: The dark history of medical experimentation on black americans from colonial times to the present*. Doubleday Books.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA statement on p-values: Context, process, and purpose.

Watters, A. (2023). *Teaching machines: The history of personalized learning*. mit Press.

Wei, M. (2025). The emerging problem of "AI psychosis". *Psychology Today*. https://www.psychologytoday.com/us/blog/urban-survival/202507/the-emerging-problem-of-ai-psychosis

Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., et al. (2022). Taxonomy of risks posed by language models. *Proceedings of the 2022 ACM conference on fairness, accountability, and transparency*, 214–229.

Weizenbaum, J. (1966). Eliza—a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45.

Wellcome. (2025). Generative AI for anxiety, depression and psychosis — Grant Funding. https://wellcome.org/research-funding/schemes/generative-ai-anxiety-depression-and-psychosis

Whitaker, K., & Guest, O. (2020). #Bropenscience is broken science: Kirstie Whitaker and Olivia Guest ask how open 'open science' really is. *The Psychologist*, *33*, 34–37.

Widder, D. G., Whittaker, M., & West, S. M. (2024). Why 'open'AI systems are actually closed, and why this matters. *Nature*, *635*(8040), 827–833.

Wills, A. (2019). Open science, open source and R. *Linux Journal*.

Xiang, C. (2023). "He Would Still Be Here"': Man Dies by Suicide After Talking with AI Chatbot, Widow Says. *VICE*. https://www.vice.com/en/article/man-dies-by-suicide-after-talking-with-ai-chatbot-widow-says/

Xu, Z., Jain, S., & Kankanhalli, M. (2025). Hallucination is inevitable: An innate limitation of large language models. https://arxiv.org/abs/2401.11817

Xue, J.-H., & Titterington, D. M. (2008). Comment on "On discriminative vs generative classifiers: A comparison of logistic regression and naive bayes". *Neural processing letters*, *28*, 169–187.

Yakushko, O. (2019). Eugenics and its evolution in the history of western psychology: A critical archival review. *Psychotherapy and Politics International*, *17*(2), e1495.

Yang, K., Qinami, K., Fei-Fei, L., Deng, J., & Russakovsky, O. (2020). Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the imagenet hierarchy. *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 547–558.

Zhao, C., Tan, Z., Ma, P., Li, D., Jiang, B., Wang, Y., Yang, Y., & Liu, H. (2025). Is Chain-of-Thought Reasoning of LLMs a Mirage? A Data Distribution Lens. https://arxiv.org/abs/2508.01191