Editorial

Artificial intelligence and psychology: balancing innovation and ethics

Inteligencia artificial y psicología: equilibrando innovación y ética

Sarah Frances Gordon
Universidad Iberoamericana, Mexico City, México
sarah.gordon@ibero.mx
https://orcid.org/0000-0001-5131-8519

Psicología Iberoamericana vol. 33 núm. 1
e331859 2025

Universidad Iberoamericana, Ciudad de México
México

Introduction

Artificial intelligence (AI) has transformed psychology by changing how mental health services are provided and researched. AI uses algorithms and data to simulate human intelligence, offering tools for data analysis, process automation, and personalised treatments (Benítez Rojas, 2024). This article examines AI's transformative role in psychology, focusing on its history, applications, and ethical considerations, specifically, how generative AI puts research integrity at risk.

The History of Artificial Intelligence

AI technology enables computers to perform tasks typically requiring human intelligence, achieved through advanced algorithms and extensive datasets. In 1950, Alan Turing introduced the Turing Test and famously asked, "Can machines think?" (Turing, 1950, p. 433). This question marked the beginning of the AI era. At the Dartmouth Conference in 1956, John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon coined "artificial intelligence" and proposed a summer research project focused on AI (Benítez Rojas, 2024). Notable early advancements also included the development of ELIZA in 1965, the first chatbot, and IBM's Deep Blue defeating Garry Kasparov at chess in 1997 (Benítez Rojas, 2024; Thompson, 2022). The field evolved significantly, starting in 2016 with the founding of OpenAI. AlphaGo's win over Lee Sedol in 2017

highlighted deep reinforcement learning. The introduction of generative models like GPT-1 in 2018 and ChatGPT (GPT-3.5) in 2022 transformed conversational AI, and by 2023, developments continued with ChatGPT and other large language models (LLMs).

The Relationship between Artificial Intelligence and Psychology

As AI becomes increasingly important, we must recognise its potential to transform the field of psychology. In August 2024, the American Psychological Association released a policy statement acknowledging that while AI can revolutionise assessment, intervention, and research, its development must be guided by ethical principles rooted in human rights and rigorous scientific standards (APA, 2024). Leading scholars suggest that "psychology and AI" are emerging as a new sub-discipline, building on three decades of cyberpsychology (Krägeloh & Medvedev, 2025). This development reflects our field's growing responsibility to promote ethical and practical uses of AI in clinical practice and research.

The Uses of Artificial Intelligence in Psychology

Psychology is crucial in developing AI. By utilising theoretical frameworks, psychologists can validate diagnostic algorithms, identify hidden biases in language models, and evaluate digital tools' ethical and social implications. This collaboration enhances diagnostic accuracy and personalisation (Lee et al., 2021). There are also various ways psychologists can apply AI in their field, which is discussed in detail below.

Digital Mental Health Applications

Digital mental health AI-driven mobile and web applications offer scalable, on-demand psychological support—yet few have undergone rigorous clinical trials (Casu et al., 2024). For example, Woebot (launched in 2017) delivers Cognitive-Behavioural Therapy, Interpersonal Psychotherapy and Dialectical Behaviour Therapy techniques via a rule-based conversational engine but will cease operation on 30 June 2025; users may download their anonymised conversation history until that date (Lovett, 2025). Wysa, by contrast, uses an AI companion to monitor emotional states and guide users through CBT exercises, guided meditation and motivational interviewing, illustrating how psychology informs design and efficacy assessment (Wysa, 2025). AI chatbots are effective for mental well-being and addressing mental illness, but challenges related to engagement and integration with healthcare remain. Large randomised controlled trials are needed to investigate the integration

of human and AI support in treating mental health problems (Casu et al., 2024).

Machine Learning and Diagnostic Precision

Machine-learning algorithms can help psychologists spot complex behavioural and neuropsychological data patterns and improve psychiatric diagnoses and treatment predictions (Lee et al., 2021).

NLP in Psychological Research and Clinical Practice

Natural language processing (NLP) is a branch of AI that analyses clinical notes, patient interviews, narrative writing, and social media posts to identify linguistic markers of mental distress (Lee et al., 2021; Malgaroli et al., 2023; Zhang et al., 2022). This technology can provide early warning signals for conditions such as depression and psychosis, as it leverages NLP techniques to analyse texts to find early indicators of mental illness (Zhang et al., 2022). However, NLP models may carry cultural or linguistic biases, potentially misclassifying some non-standard dialects or uncommon phrases (Laricheva et al., 2024).

NLP in qualitative research provides a range of AI-driven techniques that simplify every stage of text analysis, from data preparation to reporting. Specifically, AI can automatically extract keywords and key phrases using statistical and machine learning methods, highlighting essential terms and themes across large text sets.

Using AI in Qualitative Psychological Research

There are various ways to use AI in qualitative research. Qualitative researchers can efficiently convert interviews, focus groups, and audio recordings into transcripts using AI-powered transcription services. Advanced analysis software like NVivo and MAXQDA employs machine-learning algorithms to identify themes, patterns, and keywords in the text. These AI applications offer automatic coding suggestions and seamlessly integrate with larger workflows. Reducing the manual effort in transcription and coding enhances analytical rigour and AI-powered coding tools, like ATLAS.ti's AI Coding Beta suggest codes by identifying data patterns. Visualisation tools create word clouds and thematic clusters, simplifying complex qualitative data through clear graphics.For multilingual datasets, translation tools like ChatGPT and other LLMs allow researchers to read and analyse responses in various languages accurately (Chan & Tang, 2024; Lee, 2024; Linlin, 2024). The quality of a translation depends on the effectiveness of the AI prompt and the researcher's expertise in prompt engineering. Including details like the translation's purpose,

target audience, and contextual information is essential for achieving high-quality results (Chan & Tang, 2024).

Ethics and Privacy

While AI can automate intake assessments and monitor progress—freeing psychologists to tackle complex clinical work—it also introduces serious ethical and privacy risks around informed consent, data security, research integrity, and the therapeutic alliance. Psychologists play a vital role in identifying ethical violations in research and practice (Evans, 2024). They apply principles like beneficence, justice, and individual respect concerning new technologies. Human oversight is vital for validating AI outputs and fighting misinformation. AI can streamline administrative tasks and improve accessibility for non-native English researchers, but strict ethical oversight and ongoing human evaluation are essential to prevent harm and maintain public trust (Abrams, 2025; Eacersall et al., 2024).

For example, as a journal editor, I've witnessed a surge of manuscripts containing AI-generated citations. This is a common feature of using AI tools like ChatGPT to write your manuscript. One type of AI-generated reference is the chimeric reference, "in which elements of one reference are combined with other elements from an unrelated one" (Dunford et al., 2024, p. 151). Chimeric references and other AI-generated references can be detected by searching for the digital object identifier (DOI) or by searching directly in the journal volume and number in the reference list. Fabricated citations undermine research integrity and represent an attack on science (Emsley, 2023). Often, we forget that ChatGPT is fundamentally a language-processing tool. It is not an information retrieval system; therefore, it does not concern itself with what is truthful or accurate, as we should be scientists (Walters & Wilder, 2023). Generative AI can also be used for image manipulation and data fabrication. Paper mills also use generative AI to operate on a much larger scale. An easy way to detect whether a paper was produced using generative AI is to analyse the bibliography or reference list, as this analysis will offer clues regarding a paper's potential research integrity issues (Dunford et al., 2024).

But how do we address the challenges generative AI pose to our field? Regarding clinical practice, we must advocate for ethical AI practices in psychology. Practitioners should prioritise transparent informed consent, which involves disclosing what data is collected, how it is used, and with whom it may be shared. This information should be presented through easy-to-understand and interactive consent forms. Practitioners must practice data minimisation by collecting only the necessary information for their work. They should also conduct regular security audits and publish summary reports to

ensure accountability and transparency. Incorporating human oversight by requiring clinicians to review all high-stakes decisions (e.g., diagnosis, risk prediction) is also an important factor.

In the research context, it is crucial to provide more training and guidelines on the appropriate uses of generative AI and the risks associated with its unethical use. Journal editors should carefully review every reference list or bibliography in the manuscripts submitted to their journals to identify potential threats to research integrity. As the trend of generative AI grows, we must stay alert to its possible risks to research integrity.

Conclusion

The face of psychology is starting to change as AI infiltrates research and clinical practice. Researchers can use machine learning to uncover human behaviour, and clinicians may employ conversational agents for support in their administrative tasks. As AI becomes increasingly important, we must contemplate Alan Turing's seminal question, "Can machines think?" which compels us to distinguish between computational mimicry and genuine understanding. However, one thing we have as humans that AI models still lack is our ability to think critically and make choices based on principles of fairness and social justice. At this stage, we must contemplate AI's capabilities and moral and epistemic boundaries- protecting ourselves against an overreliance on its black-box models and trying to preserve the unique human dimensions of trust, empathy, ethics and original thought- all vital to our work as psychologists and researchers.

References

Abrams, Z. (2025, January 1). Artificial intelligence is impacting the field. As AI transforms our world, psychologists are working to channel its power and limit its harm. *Monitor on Psychology, 56*(1), 46. https://www.apa.org/monitor/2025/01/trends-harnessing-power-of-artificial-intelligence?utm_source=chatgpt.com

American Psychological Association (APA). (2024). *Artificial intelligence and the field of psychology*.https://www.apa.org/about/policy/statement-artificial-intelligence.pdf

Benítez Rojas, R. V. (2024). Artificial intelligence: Genesis, development, and future. In R. V. & Benítez Rojas & F. J. Martínez-Cano (Eds.), *Revolutionizing communication* (pp. 1-16). CRC Press.

Casu, M., Triscari, S., Battiato, S., Guarnera, L., & Caponnetto, P. (2024). AI chatbots for mental health: A scoping review of effectiveness, feasibility, and applications. *Applied Sciences, 14*, 5889. https://doi.org/10.3390/app14135889

Chan, V., & Tang, W. K. W. (2024). GPT for translation: A systematic literature review. *SN Computer Science, 5*(8), 1-9. https://doi.org/10.1007/s42979-024-03340-z

Dunford, R., Rosenblum, B., & Hunter, S. I. (2024). Using automated analysis of the bibliography to detect potential research integrity issues. *Learned Publishing, 37*, 147–153. https://doi.org/10.1002/leap.1600

Eacersall, D., Pretorius, L., Smirnov, I., Spray, E., Illingworth, S., Chugh, R., Strydom, S., Stratton-Maher, D., Simmons, J., Jennings, I., Roux, R., Kamrowski, R., Downie, A., Thong, C., & Howell, K. A. (2024). *Navigating ethical challenges in generative AI-enhanced research: The ethical framework for responsible generative AI use.* ArXiv https://doi.org/10.48550/arXiv.2501.09021

Emsley, R. (2023). ChatGPT: These are not hallucinations - they're fabrications and falsifications. *Schizophrenia, 9*(1), 52. https://doi.org/10.1038/s41537-023-00379-4

Evans, A. C. Jr. (2024, October 1). Psychology's role in tackling artificial intelligence, misinformation, loneliness, and immigration. *Monitor on Psychology, 55*(7), p. 10. https://www.apa.org/monitor/2024/10/psychologists-impact-on-world?utm_source=chatgpt.com

Krägeloh, C. U., & Medvedev, O. N. (2025). A new journal for a new era. *Journal of Psychology and AI*, *1*(1), 2450105. https://doi.org/10.1080/29974100.2025.2450105

Laricheva, M., Liu, Y., Shi, E., & Wu, A. (2024). Scoping review on natural language processing applications in counselling and psychotherapy. *British Journal of Psychology, 00*, 1-25. https://doi.org/10.1111/bjop.12721

Lee, E. E., Torous, J., De Choudhury, M., Depp, C. A., Graham, S. A., Kim, H. C., Paulus, M. P., Krystal, J. H., & Jeste, D. V. (2021). Artificial intelligence for mental health care: Clinical applications, barriers, facilitators, and artificial wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 6*(9), 856-864. https://doi.org/10.1016/j.bpsc.2021.02.001

Lee, T. K. (2024). Artificial intelligence and posthumanist translation: ChatGPT versus the translator. *Applied Linguistics Review, 15*(6), 2351-2372. https://doi.org/10.1515/applirev-2023-0122

Linlin, L. (2024). Artificial intelligence translator Deepl translation quality control. *Procedia Computer Science, 247*, 710-717. https://doi.org/10.1016/j.procs.2024.10.086

Lovett, L. (2023). Woe is me: Woebot says farewell to signature app. *Behavioral Health Business.*https://bhbusiness.com/2025/04/23/woe-is-me-woebot-says-farewell-to-signature-app/

Malgaroli, M., Hull, T. D., Zech, J. M., & Althoff, T. (2023). Natural language processing for mental health interventions: A systematic review and research framework. *Translational Psychiatry*, *13*(1), 309. https://doi.org/10.1038/s41398-023-02592-2

Thompson, C. (2022, February 18). Artificial intelligence. What the history of AI tells us about its future. *MIT Technology Review.*https://www.technologyreview.com/2022/02/18/1044709/ibm-deep-blue-ai-history/

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59,* 433–460. https://doi.org/10.1093/mind/LIX.236.433

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports, 13*, 14045. https://doi.org/10.1038/s41598-023-41032-5

Wysa. (2025). *Mental health, redefined.*https://www.wysa.com/

Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: A narrative review. *NPJ Digital Medicine*, *5*(1), 46. https://doi.org/10.1038/s41746-022-00589-7

Información adicional

redalyc-journal-id: 1339

Sarah Frances Gordon

Artificial intelligence and psychology: balancing innovation and ethics
Inteligencia artificial y psicología: equilibrando innovación y ética

Psicología Iberoamericana
vol. 33, núm. 1, e331859, 2025
Universidad Iberoamericana, Ciudad de México, México
revista.psicologia@ibero.mx

ISSN: 1405-0943

DOI: https://doi.org/10.48102/pi.v33i1.859