MINNESOTA BOARD OF PSYCHOLOGY
February 20, 2026
Board Meeting

Order of Business

PUBLIC SESSION:

1. **Call to Order**
    - A. **Link to the Board Meeting**
2. **Adoption of Tentative Agenda**
3. **Announcements**
4. **Approval of the Board Minutes**
    - A. **Approval of Board Meeting Minutes**
5. **Consent Agenda**
    - A. **Staff Delegated Authority Report**
6. **New Business**
    - A. **Master's Level Licensure**
    - B. **AI in Psychology Practice**
    - C. **Psilocybin Draft Legislation**
    - D. **Executive Director's Report**
    - E. **Board Administrative Terminations**
7. **Committee Reports**
8. **Adjournment**

- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 2/20/2026

SUBMITTED BY: Assistant Executive Director

TITLE: Link to the Board Meeting

INTRODUCTION TO THE TOPIC:

Please contact the Board for a link to the Board meeting. Email: psychology.board@state.mn.us

BOARD ACTION REQUESTED:

- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 2/20/2026

SUBMITTED BY: Assistant Executive Director

TITLE: Approval of Board Meeting Minutes

INTRODUCTION TO THE TOPIC:

The Board Meeting Minutes for January 2026 are respectfully submitted.

BOARD ACTION REQUESTED:

ATTACHMENTS:

Description	Upload Date	Type
Board Meeting Minutes January 2026	2/9/2026	Cover Memo

MINNESOTA BOARD OF PSYCHOLOGY

MINNESOTA BOARD OF PSYCHOLOGY
Minutes of January 30, 2026, Board Meeting

Board Members and Staff in Attendance: Sonal Markanda, Sebastian Rilen, Michael Thompson, Daniel Hurley, Michelle Zhao, Jill Idrizow, Joel Bakken, Pamela Freske, Cesar Gonzalez, Sam Sands and Trish Hoffman.

Guests: Nick Lienesch, Talee Vang, and Kim Navarre.

PUBLIC SESSION

1. Call to Order

Sonal Markanda called the meeting to order at 9:35 AM. The meeting was held in a hybrid format with some individuals in attendance in person and others online. Voting was held by roll call.

A. Webex MeetingLink

2. Adoption of Tentative Agenda

Daniel Hurley moved, seconded by Seb Rilen Motion: to adopt the tentative agenda. There being 8 "ayes" and 0 "nays" the motion Passed.

3. Announcements

4. Approval of the Board Minutes

Joel Bakken moved, seconded by Daniel Hurley Motion: to adopt the December 19, 2025, Board Meeting Minutes. There being 8 "ayes" and 0 "nays" the motion Passed

5. Consent Agenda

A. Staff Delegated Authority Report

6. New Business

A. Minnesota Psychological Association Discussion

Talee Vang of the Minnesota Psychological Association described concerns of licensee's and applicants relating to the ongoing increased presence of federal agents in Minnesota. The Board discussed ways to share information with MPA members on these topics.

B. Health Provider Services Program

Kim Navarre of the Health Professional Services Program gave a presentation on HPSP and its work with Psychologists and other professionals to monitor medical, mental health, and substance use conditions that may affect their ability to practice their professions.

C. Master's Level Licensure

The Board discussed the draft APA Model Act for State Licensure of Psychology Professionals, which is open for comment, along with slides prepared by Alex Siegel of ASPPB identifying various potential issues with the draft.

D. AI in Psychology Practice

The Board discussed articles related to the use of AI in the practice of psychology.

E. Executive Director's Report

Trish Hoffman reported that the Licensure Unit has continued to contact LP applicants who have not had movement on their applications for more than a year and has begun contacting BA applicants who have not made their intent clear, resulting in issuance of a handful of licenses. The overall total of BA licenses issued is now about 925. In addition, the team is working on documents for the BA renewal procedure as well as a checklist for LP applicants trained in an educational institution outside of the United States or Canada.

Sam Sands noted recent events impacting the Board and concerns that Board staff are actively addressing. He highlighted two upcoming conferences, discussed staff's work on renewal procedures, and noted that financial statements have been provided for the Board's review.

F. Psilocybin Draft Legislation

The Board received draft legislation that would establish a legal regulated framework for the therapeutic use of psilocybin.

G. Board Administrative Terminations

Daniel Hurley moved, seconded by Seb Rilen Motion: to approve the Board Administrative Terminations. There being 6 "ayes" and 0 "nays" the motion Passed.

7. Committee Reports

8. Adjournment

Adjourned at 12:23 PM.

- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 2/20/2026

SUBMITTED BY: Assistant Executive Director

TITLE: Staff Delegated Authority Report

INTRODUCTION TO THE TOPIC:

The Board utilizes a consent agenda for routine financial, legal, or administrative matters that require Board action or inform the Board of action taken under authority delegated by the Board.

BOARD ACTION REQUESTED:

The Board utilizes a consent agenda for routine financial, legal, or administrative matters that require Board action or inform the Board of action taken under authority delegated by the Board.

The items on the consent agenda are expected to be non-controversial and not requiring of a discussion.

The consent agenda is voted on in a single majority vote, but made be divided into several, separate items if necessary.

The items on the consent agenda will be considered early in the meeting. The Board chair will ask if any member wishes to remove an item from the consent agenda for separate consideration, and if so, the Chair will schedule it for later in the meeting.

- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 2/20/2026

SUBMITTED BY:

TITLE: Master's Level Licensure

INTRODUCTION TO THE TOPIC:

BOARD ACTION REQUESTED:

- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 2/20/2026

SUBMITTED BY: Executive Director

TITLE: AI in Psychology Practice

INTRODUCTION TO THE TOPIC:

Articles of note for the Board to consider around AI use in practice and regulation.

BOARD ACTION REQUESTED:

ATTACHMENTS:

Description	Upload Date	Type
Artificial Intelligence and Psychology	2/10/2026	Cover Memo
Why Artificial Intelligence Will Not Replace Human Psychologists	2/10/2026	Cover Memo
Dr Jodi Halpern on Why AI Isn't a Magic Bullet for Mental Health	2/10/2026	Cover Memo
The Relational Shift: Why We Need "AI Psychology" Now as a Core Field	2/10/2026	Cover Memo
Principles Before Practice - How CLEAR is Shaping Responsible AI Use	2/10/2026	Cover Memo

CANADIAN PSYCHOLOGICAL ASSOCIATION

SOCIÉTÉ CANADIENNE DE PSYCHOLOGIE

ARTIFICIAL INTELLIGENCE AND PSYCHOLOGY

A Briefing Paper of the Canadian Psychological Association (CPA)

Approved by the CPA Board of Directors - January 2024

Contributors:

Adam Sandford, Bryce Mulligan,
Eleanor Gittens, Meghan Norris,
Myra Fernandes

CANADIAN PSYCHOLOGICAL ASSOCIATION

SOCIÉTÉ CANADIENNE DE PSYCHOLOGIE

ABOUT THE CPA

The Canadian Psychological Association is the national voice for the science, practice and education of psycholo-gy in the service of the health and welfare of Canadians. The CPA is Canada's largest association for psychology and represents psychologists in public and private practice, university educa-tors and researchers, as well as students. Psychologists are the country's largest group of regulat-ed and specialized mental health providers, making our profession a key resource for the mental health treatment Canadians need.

VISION

A society where understanding of diverse human needs, be-haviours and aspirations drive legislation, policies and programs for individuals, organizations and communities.

MISSION

Advancing research, knowledge and the application of psy-chology in the service of society through advocacy, support and collaboration.

FOR MORE INFORMATION
PLEASE VISIT OUR WEBSITE AT CPA.CA

INTRODUCTION

Artificial intelligence (AI) has emerged as a transformative force in various fields, including psychology. With its ability to analyze vast amounts of data, identify patterns and make predictions, AI has the potential to revolutionize the science, practice and education of psychology. However, as AI continues to evolve, ethical considerations, privacy concerns, and the need for human oversight remain critical in harnessing its full potential and minimizing harm while upholding the values and ethical standards of the psychological profession. This paper seeks to provide an overview of AI, its impact and application as it relates to the three pillars of the Canadian Psychological Association (science, practice, education), and highlights perceived challenges and recommended first steps.

DEFINING ARTIFICIAL INTELLIGENCE

There is no commonly accepted definition of AI. For the purposes of this paper, we will use a broad definition of AI as "[c]omputers which perform tasks, usually associated with human minds, particularly learning and problem-solving." (Baker & Smith, 2019, p. 10). This broad definition of AI allows us to encompass multiple functions of AI. Collins et al. (2021) provide sourced descriptions of some of the important functions of AI, which at least provide a framework of the ways AI could be used. These functions include but are not limited to:

- AI as Expert Systems.
- AI as Machine Learning.
- AI as Robotics.
- AI as Natural Language Processing.
- AI as Machine Vision, and
- AI as Speech Recognition (Dejoux & Léon, 2018).

THE EVOLUTION OF ARTIFICIAL INTELLIGENCE

Alan Turing is credited with laying the foundation of our early understanding of AI in modern times (e.g., Turing, 1950). Turing described the possibility of creating intelligent machines that could simulate human-like intelligence. However, during a conference in1956, John McCarthy described AI as "the science and engineering of making intelligent machines" (Russell & Norvig, 2010, p. 10) thereby coining the term "Artificial Intelligence." Early research in the developing field of AI intended to develop high-level cognition AI. This research aimed to go beyond mere recognition of concepts, perception of objects, and performance of complex motor skills. Instead, the aim of the early research in AI was to discover whether AI could engage in multi-step reasoning, understand the meaning of natural language, generate novel plans to achieve goals, and demonstrate meta-cognitive abilities (Langley, 2011). Though, research in the field of AI has not reached

these heights. It has been said that AI research has experienced AI "summers and winters" (Russell & Norvig, 2020). Since the 1990s, coinciding with the development of the Internet and renewed interest in neural networks, developments in AI have gradually advanced. The shift in the 1990s and 2000s went from a narrow focus on development of expert systems (e.g., speech recognition) to data-driven approaches with machine learning techniques. This has generally coincided with advancements in computational power (e.g., high-powered processors) and availability of increasingly larger datasets due to sharing of data through the Internet. These two decades witnessed improvements in recommendation systems, fraud detection, and data mining to name a few examples.

Since John McCarthy (McCarthy et al., 2006) coined the term, AI has been used to broadly encompass different technologies, machine learning, natural language processing programs, data mining, and neural networks (Baker & Smith, 2019). However, it is important to emphasize that AI is not a specific, narrowly defined technology but rather a broad category of technologies. At a surface level, these technologies are designed (and often appear) to emulate human neurocognitive functions (e.g., receptive/expressive language, visual analysis/classification). "Artificial intelligence" is thus a misnomer: it mimics the functions of some component neurocognitive processes but does not (yet) approximate actual (i.e., human) intelligence, since the latter emerges as a result of dynamical interactions between multi-faceted underlying component processes. As such, alternative terms such as "intelligence analogue" (IA) may be preferrable to "artificial intelligence" to avoid implicit identification of machine traits with those unique to human beings.

Now we are seeing proliferation of use of "chatbot" applications such as ChatGPT and BingChat in education, business, and healthcare. Psychology educators, practitioners, and researchers can expect imminent growth in the number and diversity of available AI-augmented tools with potential for application to their professional activities. AI developers will surely continue to demonstrate the many potential ways their technologies can be applied in the practice of psychology; however, they cannot provide guidance regarding how or why – to what end? – AI should be incorporated into the world of psychology. Answering this latter question will require broad engagement with the Canadian public as well as with experts in philosophy, law, policy, and public health. Below we describe the challenges and promises of using AI-based applications in psychological education, practice, and science.

THE IMPACT OF ARTIFICIAL INTELLIGENCE ON PSYCHOLOGY

Recent and expected developments in AI present myriad opportunities and challenges to psychology in Canada. AI provides psychological scientists the opportunity to analyze vast amounts of data efficiently. In the context of education, AI can be a helpful support for code and essay writing, assignments, tests, and background summaries in mere seconds. It also allows for harnessing massive search power. AI is relevant to the practice of psychology insofar as it will become increasingly implicated in the daily lives of professional psychologists and in the health/functioning of the clients and patients whom they serve.

AI has the capability to process vast amounts of data quickly. It can sift through large datasets to identify patterns and correlations, detect trends, and predict outcomes. The analytics are often based on quantitative analysis of the data where the processing of big data is done for reporting, prescriptive and predictive purposes (Wirtz et al., 2018). AI provides highly sophisticated statistical and probabilistic methods. Access is cheap and it comes with enormous computational power. However, research has found the analysis and processing of heterogenous data to be problematic (Dwivedi et al., 2021). There are ethical dimensions to consider in relation to data sharing and discrimination. In the case of discrimination, even though the analysis and decision making are not being conducted by humans, the AI algorithm reflects the pervasive discriminatory attitude of the engineer or the source data. A number of challenges have also been highlighted around data usage and data integrity. As technology matures, these issues need to be resolved to ensure full confidence in the research stakeholders.

Another area of research that has taken off is speech analytics. Software has been developed for the recognition and processing of language. It is able to understand or respond to natural language as well as translate from spoken to written language. These developments have seen huge savings on time for researchers. In other areas, researchers have broken the tasks that AI performs into three: mechanical, thinking and feeling (Huang, Rust, & Maksimovic, 2019). Huang and colleagues (2019) felt that AI could easily take over the mechanical (robotics) and thinking tasks (processing, analyzing and interpreting data). However, it was felt that the feeling tasks should be left to human (communication). Huang et al. (2019) did not take into consideration that any bias present in the input used to train an AI system persists and often further amplified.

In education, scholars are able to plan, organize, compare, and synthesize disparate sources, as well as create new material such as computer code, art, and even optimal personalized study schedules. Undoubtedly, this will allow for rapid developments in scholarship. Recognizing the many potential benefits of AI within scholarship, there are of course many risks. For example, although AI can serve as a helpful "co-pilot" in scholarship, there are now free plug-ins that will identify the most likely correct answer in un-proctored online quizzes, thus allowing for students to bypass educational experiences altogether, where AI effectively becomes the captain rather than the co-pilot. Likewise, un-proctored writing assignments may be entirely written by AI, and there is currently no reliable and valid way to detect that which is written by AI, and detection may be biased against certain populations (e.g., Liang et al., 2023). Educators are left with the question of "how much AI support is "too much" support, and the age-old questions of "how do we best teach and assess knowledge?"

It is important to remember that, despite the advent of widespread access to AI, the process of learning for humans has not changed, although the roles and actions of instructors may need to. Well known to psychologists, attention, encoding, storage, and retrieval continue to be critical considerations for instructors, despite the advent of AI. However, to facilitate thorough and deep information processing, instructors may need to re-design tests and assignments to ensure that requisite activities for learning are not bypassed. Further, developing assignments that integrate benefits of AI and limit risks will require assignments that demand more than simple regurgitation of facts. For example, creating assignments that emphasize synthesizing masses of AI-generated information can support students in developing skills to better assess what they are reading and

analyze why it is important. These skills will almost certainly be needed across disciplines and industries, and knowing how to harness the power of AI within ethical boundaries will enable students to respond to current and future needs in an increasingly tech-based world.

Just as students can benefit from AI in their student-based work when done with integrity, AI is also assistive for instructors when used with integrity. AI may be considered as akin to having an incredibly efficient teaching assistant who can automate the grading for multiple-choice and short answer questions. The ready availability of AI chatbots can provide immediate help for a student struggling with an assignment question at home when the teacher is unavailable. AI is always there, ready to answer. Relying exclusively on chatbots to answer questions is of course problematic, however, if incorrect or incomplete, or biased information is being given. AI is also well-positioned to collate class-based data to help inform teaching practices, allowing for rapid responses from instructors. Of perhaps particular interest to psychologists, AI can also analyze primary data on an individual student's academic performance and can suggest tailor-made approaches, interventions, and alternate learning avenues to meet their specific needs (Bell, 2021; Rouhiainen, 2019).

The reality is that, in the context of education, there are pressing and paramount calls to action for educators, scholars, and policymakers. There is a lack of scholarship demonstrating the "who, what, where, why, when, and how" of best practices for integrating AI into education, and as yet, there are no clear emerging guidelines of the ethics of integrating AI into work from authoritative bodies. This is not a criticism—indeed, this technology is so new that overviews regarding responsible scholarship and integrity have yet to emerge. The Office of Educational Technology in the USA has very recently (May, 2023) released a report regarding AI and teaching and learning. The core themes from this report are that AI enables new forms of interaction, can help educators to address variability in student learning, support adaptivity, enhance feedback loops, support educators, and importantly, it can increase existing risks and create new risks that have not yet been considered. In response to these themes, they recommend that humans must be "in the loop," aligning models to a shared vision for education, designing AI using modern learning principles, prioritize strengthening trust, and prioritizing educators being informed and involved regarding AI in teaching and learning (Office of Educational Technology, May, 2023).

As in other domains, AI-based tools could theoretically enhance or replace the work of trained human clinicians and technicians. For instance, currently available AI technologies could be used to automate some of the time- and labour-intensive aspects of clinical practice and thereby stimulate gains in efficiency of and access to psychological services in both public and private sectors. As the diversity and availability of AI-based technologies increases, so too will their potential applications within the practice of psychology in Canada. It is conceivable that AI will eventually allow for development and implementation of autonomous "psychologibots" that could provide wholly automated psychological services.

While the main benefits of AI-supported psychological practice appear to be related to efficiency and access, there are associated costs and equivocal impacts that are more difficult to discern and anticipate. For one, developing and operating AI systems confers considerable economic and environmental burden. Social costs

can also be anticipated, as AI-based clinical tools could amplify existing forces of marginalization and make redundant the work of persons engaged in meaningful, rewarding, prosocial vocational activities. In addition, the potential ethical and legal costs of incorporating AI into psychological practice are numerous. These include novel threats to the maintenance of professional standards related to privacy, safety/efficacy, dignity, and responsibility to society. Incorporating AI-based technology into the practice of psychology could moreover have untold philosophical impacts; it could alter humanity's collective identity and basic conceptions of knowledge, life, reality, and existence.

It should be appreciated that, beyond the psychology clinic, the clients served by practicing psychologists will experience an increased presence of AI-based technologies in their environment, and these will have potential to impact their psychological well-being. For example, research has already documented adverse effects on humans of interacting with AI algorithms that are designed by profit-generating firms to optimize the user's behavioural, cognitive, and emotional engagement with their service (e.g., social media, online gambling).

CONSIDERATIONS AND CHALLENGES

While integration of AI brings immense potential, ethical considerations, interpretability, and human oversight remain crucial to ensure responsible and ethical use of AI in psychology. Here a few considerations and challenges:

CONSIDERATIONS

- Educators will need to consider the accuracy and reliability of AI. They will need to balance the utility of AI in relation to academic labour (i.e., tasking students with using e.g., ChatGPT in appropriate and critical ways to complete assessments of learning) and academic integrity (i.e., preventing academic misconduct).
- Everybody will need to consider privacy and data security when using AI.
- AI is unable to establish the same level of therapeutic rapport, empathy, and understanding as psychologists.
- In the absence of explicit laws/regulations related to oversight/accountability, responsibility for the harms caused by AI will fall on the end user.
- Those using AI in psychological practice have a duty to obtain consent that is both voluntary and informed.
- The alluring increases in efficiency promised by AI are coming at a time when our public healthcare system is under unprecedented strain. The private sector will be more nimble in adapting and responding to the opportunities and challenges afforded by AI, which risks further undermining public and financial support of public health services. There is thus a need to act quickly lest the interests of society are subjugated by neo-liberal market forces.
- AI is a term denoting a category of tools. AI has no morals to direct its development or actions.

CHALLENGES

- The information relayed may not be accurate nor current.
- Using AI as a substitute for university services (e.g., professors' feedback, academic advisors' guidance, counselling services).
- Biased algorithms.
- Context-free responses that are factually inaccurate or biased.
- Offering global guidance to Canadian psychologists who practice across many different jurisdictions.
- Deploying a technology that requires a high level of specialist knowledge in order to comprehend/communicate risks/benefits.
- Determining the persons/entities who are liable for harms attributed to clinical application of AI.
- Developing processes to independently validate and regulate the use of AI in healthcare generally and in psychology specifically.
- Addressing unfounded claims about human-machine equivalence that could result in reduced public support for funding for humans engaged in psychology training, research, and practice.
- Developing policy/recommendations that addresses current, foreseeable, and future unforeseen uses of AI in psychology.

RECOMMENDATIONS

- Establish a working group or committee to explore in depth the impact that AI is having and will have on psychology
- Populate the working group with representatives from the three CPA pillars – Science, Practice, and Education - including a board liaison
- The working group could be tasked with:
  - Consulting with experts and stakeholders to create position and policy recommendations based on review of applicable codes and laws.
  - Reviewing the CPA Code of Ethics, Tri-council Policy Statement: Ethical Conduct for Research Involving Humans, educational/institutional policies, and laws/regulations governing the practice of psychology with a view to identifying challenges and solutions related to AI implementation.

CITATIONS

Baker, T., & Smith, L. (2019). Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges. Retrieved from Nesta Foundation website: https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf

Bell (2021) https://elearningindustry.com/ai-for-personalized-learning-potential-and-challenges

Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management, 60,* 102383. https://doi.org/10.1016/j.ijinfomgt.2021.102383

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., … & Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57,* 101994.

Huang, M.-H., Rust, R., & Maksimovic, V. (2019). The feeling economy: Managing in the next generation of artificial intelligence (AI). *California Management Review, 61*(4), 43-65.

Langley, P. (2011). The changing science of machine learning. *Machine Learning, 82*(3), 275-279. https://doi.org/10.1007/s10994-011-5242-y

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *arXiv preprint arXiv:2304.02819.*).

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine, 27*(4), 12. https://doi.org/10.1609/aimag.v27i4.1904

Office of Educational Technology. (n.d.) *Artificial intelligence.* Artificial Intelligence - Office of Educational Technology

Rouhiainen, 2019, https://hbr.org/2019/10/how-ai-and-data-could-personalize-higher-education

Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed). Pearson.

Russell, S., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (2nd ed). Pearson.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433-460.

Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2018). Artificial intelligence and the public sector – Applications and challenges. *International Journal of Public Administration, 42*(7), 596-615.

December 29, 2025(https://societyforpsychotherapy.org/2025/12/29/)

Why Artificial Intelligence Will Not Replace Human Psychologists: Legal, Ethical, and Clinical Limitations

This article builds on previous arguments (Gavazzi, 2025a; Gavazzi, 2025b) stating that although AI technologies are rapidly advancing, they cannot replace human psychologists performing psychotherapy; this is simply the result of evolutionary advantages in humans across social, emotional, and cognitive domains that are essential for therapeutic interactions. In addition, these systems are unlikely to replace psychologists in the foreseeable future for practical reasons. Legal, ethical, and clinical barriers— particularly those involving state licensing, clinical judgments, forensic considerations, and accountability—make the deployment of autonomous systems in therapeutic settings impractical and potentially dangerous. This article presents key structural and philosophical reasons why human oversight and involvement remain essential in psychological practice.

State Licensing Considerations with Artificial Intelligence

An immediate obstacle to AI technologies replacing human psychologists is the regulatory framework that governs mental health practice. In the United States, psychology is regulated at the state level through practice acts, which consistently define a licensed psychologist as a human professional who has met rigorous educational, supervised training, and experiential requirements (American Psychological Association [APA], 2011).

Currently, no state licensing board recognizes non-human entities as eligible for licensure. Illinois enacted legislation prohibiting autonomous AI technologies from providing direct therapeutic interventions or making clinical decisions (Roy, 2025). Licensure requirements include academic training, supervised clinical experience, and examinations as well as continual use of professional judgment, adherence to ethical codes, and accountability for actions. These professional obligations are inseparable from human agency.

Professional oversight mechanisms presuppose human accountability as licensing boards investigate complaints, hold hearings, and impose sanctions. These processes require a human practitioner capable of understanding the

consequences of their actions and then modify their professional behavior accordingly. Licensing an AI system as a practitioner would necessitate a complete restructuring of these systems, including new definitions of competence, malpractice, and remediation. There is currently no legal precedent or regulatory movement toward such change (Mello & Cohen, 2025).

The complexity of clinical decision-making presents an additional barrier. High-stakes contexts—including assessment of suicide or homicide risk, mandated reporting duties, and severe psychopathological presentations demand much more than data analysis. Psychologists integrate intuition, cultural sensitivity, and emotional attunement into their clinical conceptualizations, considerations, and treatment plans. AI systems, even those trained on extensive datasets, lack the lived experience and contextual awareness needed for adaptive clinical reasoning (Gavazzi, 2025b; Thakkar et al., 2024). Although AI agents may contribute as adjuncts to psychological services, current legal and regulatory structures preclude recognition of non-human independent practitioners.

Fidelity, Judgment, and Forensic Implications

AI technologies lack a genuine understanding of ethical principles, a deficit that is particularly consequential within a therapeutic relationship. A critical example is the principle of fidelity, which obligates clinicians to uphold commitments, foster trust, and prioritize the patients' best interests. Upholding this principle requires a nuanced comprehension of the patient's emotional state, developmental history, cultural sensitivities, and psychological resilience. These are competencies that AI systems currently cannot replicate (Thakkar et al., 2024). For instance, a decision on whether to pursue involuntary hospitalization for a patient expressing suicidal ideation depends on multiple factors. These include assessing the immediacy and lethality of their intent, their access to means, the strength of protective factors (i.e., family support, future orientation) and their history of impulsive behavior compared with chronic despair. A human clinician synthesizes this information through years of training and interpersonal experience. In doing so, the clinician navigates these considerations while safeguarding both the patient's autonomy and safety—balancing the therapeutic alliance with the inherent duty to protect. An AI system, constrained by its probabilistic modeling, cannot genuinely grasp the weight of removing someone's freedom or the complex relational consequences of such interventions (Montemayor et al., 2022).

These limitations have especially serious implications in forensic contexts. When psychological records are subpoenaed or clinicians are called to testify, it is unclear how an AI system would respond to such legal demands. If an AI system were to testify, it's reasoning processes would be opaque as a result of the black box nature of machine learning, where the internal logic connecting data to decisions is invisible to humans. This opacity would make the AI system's decision making difficult, if not impossible, to defend in court (Price, 2017).

Several critical questions follow: Could an AI system be compelled to testify under oath? Who bears responsibility for its decisions; the developer, the deploying institution, or the algorithm itself? If the AI system's code were generated or modified by another AI (as occurs in generative systems) the chain of responsibility fragments and becomes untraceable. Price (2017) warns that algorithmic decision-making in healthcare may advance faster than the legal system's ability to assign liability, creating an accountability vacuum. Beyond questions of liability, concerns about data integrity and chain of custody arise. AI-generated psychological records would require new protocols to ensure authenticity, prevent tampering, and verify the origin of documentation. Without standardized, auditable safeguards, the admissibility of AI-generated psychological documentation in court remains uncertain.

Accountability and Standard of Care Issues

A cornerstone of professional psychology is accountability. When allegations arise that a human psychologist has practiced below the standard of care, there are established mechanisms for investigation, peer review, and disciplinary action. By contrast, in cases involving AI systems, accountability becomes diffuse and legally ambiguous (Price et al., 2022). AI systems cannot be held personally liable, nor can they be suspended, fined, or required to undergo remedial training. Instead, liability would fall on the developers, healthcare institutions, or software vendors, none of whom are necessarily licensed mental health professionals. This disconnect between liability and professional oversight creates a critical gap in the enforcement of professional standards.

Determining what constitutes substandard care by an AI is also fraught with difficulty. Should AI systems be held to the same standard as a reasonably competent human psychologist, or should a new, algorithm-specific standard

be developed? Establishing such a benchmark would require expert testimony from individuals with expertise in both clinical psychology and software engineering, which are not plentiful (Minssen et al., 2020).

Moreover, unlike human errors which are typically isolated, AI system errors become systemic. A flaw in an algorithm's training data or decision logic could affect thousands of patients across multiple jurisdictions simultaneously. For example, if an AI system incorrectly assesses suicide risk due to biased training data that underrepresents certain demographics, the harm is not individualized but widespread and may remain undetectable without large-scale audits. The scalability of AI systems amplifies both their benefits and their risks. While a human clinicians' malpractice typically affects a limited number of patients, a defective AI system could compromise the care of thousands or tens of thousands. For example, an AI system operating in an interjurisdictional manner that incorrectly assesses suicide risk due to flawed natural language processing could be catastrophic.

Such systemic failures would raise unprecedented questions: Should all patients treated by the AI systems be reevaluated? Who should bear the cost? How should harm be quantified across diverse populations? Insurance models are not equipped to handle such large-scale liability, and existing malpractice policies do not account for algorithmic error (Price et al., 2022). High-profile AI system failures could severely undermine public trust in mental health services and care, which is founded on trust, confidentiality, and empathy, elements that are difficult to replicate in AI systems and challenging to regulate as it is. A single, widely publicized incident of AI-related harm could delay the integration of AI technology in psychology for many years.

Conclusion

AI shows the potential as a supportive tool in psychological practice by assisting with screening, data analysis, and treatment planning, however, it cannot replace the human psychologist. Legal frameworks governing licensure, the ethical requirements of therapeutic fidelity, the forensic challenges of algorithmic transparency, and the systemic risks of accountability all point to the irreplaceable role of human judgment, empathy, and responsibility in mental health care. State licensing boards are unlikely to credential non-human practitioners; courts are unprepared to evaluate AI testimony; and liability systems cannot adequately address algorithmic harm. More fundamentally, the essence of psychotherapy—rooted in relationships, trust, and shared human experiences—cannot be replicated by AI technologies. As the field integrates AI

technologies into psychological services, the focus should remain on augmentation and not replacement. The future of psychology lies in collaborative models where technology enhances rather than replaces the human connection that is central to healing.

Author

John Gavazzi, PsyD, ABPP

(https://societyforpsychotherapy.org/author/john-gavazzi/)

Dr. John Gavazzi is a board-certified clinical psychologist based in Lemoyne, Pennsylvania, where he maintains an independent practice. For over 25 years, he has specialized in ethics education, delivering workshops and publishing articles on ethics, mental health law, and clinical decision-making.

View all posts (https://societyforpsychotherapy.org/author/john-gavazzi/)

(http://www.ethicalpsychologist.com)

+ **Citation**

+ **Reference(s)**

2 Responses

Chris January 1, 2026 at 4:19 pm (https://societyforpsychotherapy.org/why-artificial-intelligence-will-not-replace-human-psychologists-legal-ethical-and-clinical-limitations/#comment-13745)

Allan says:

I am pretty sure the cat is out of the bag in regard to psychotherapy and AI. People are simply not going to see psychologists when AI can offer them something nearly as good for virtually no cost. In my own experience AI with a good prompt thinks ethically more deeply, more nuanced and is much more able to identify how multiple ethical precepts may load onto particular issue. The complexity at the moment is we really don't have any large studies directly comparing therapeutic outcomes between AI developed therapy program and psychologists. However we do have some reasonable evidence that it is doing moderately good job in both psychotherapy and supervision.
The other complexity is that AI continues to develop and get better and better. We continue to struggle particular in a training situation with the fact that psychologists do not get better with experience. (see Wampol and host of others on this thorny issue). There is increasing evidence that AI can do better a better job at suicide risk assessment than humans see for example De Grandi et al 2024 and Shoib et al 2025. In fact, AI seems to be getting to the stage of doing all sorts of assessments better than psychologist (ali et al 2025). Large comparative studies remain to be done. However, we cannot demand that AI be perfect when humans are such demonstrably poor at undertaking accurate and objective assessment particular in legal settings (the over 1200 pages on this in Ziskin and Faust still hold pretty true today.

AI is coming for our jobs and will do it better. Maybe not today but in the next one to two years. Legislating against is likely to have little impact in my view. The question for me is how psychologists find a place to exist meaningfully in the context of all this.
Regards Chris Allan

Reply

John D Gavazzi (https://www.ethicalpsychology.com/) says:

January 17, 2026 at 4:55 am (https://societyforpsychotherapy.org/why-artificial-intelligence-will-not-replace-human-psychologists-legal-ethical-and-clinical-limitations/#comment-13766)

Hi Chris-

Thank you for your critique on the potential for AI to displace human psychologists. You rightly highlight powerful drivers like cost-efficiency and data-driven assessment. However, a synthesis of the provided literature suggests that the leap from

“transformative tool” to “full replacement” overlooks profound evolutionary, practical, and legal barriers.

The Irreplaceable Human Core: Evolution and Connection
Human psychology is the product of millennia of adaptation, fundamentally rooted in biologically-driven social connection and dynamic contextual awareness. While LLMs can simulate convincing dialogue, they operate without genuine self-awareness or lived experience. The transformative power of therapy arises from authentic emotional resonance and collaborative meaning-making—a shared human journey that algorithmic processes cannot replicate. The therapeutic relationship itself is the intervention, something AI can inform but not inhabit.

The “Black Box” Problem: Nuance and Systemic Risk
On assessment and ethical nuance, this article (and my previous articles) sound a strong caution. AI’s internal logic often remains opaque—a “black box” that defies clear explanation in clinical or legal settings. This is critical when a decision, such as pursuing involuntary hospitalization, requires synthesizing developmental history, cultural context, and the profound ethical weight of removing personal freedom. Furthermore, while human error is typically isolated, a flaw in an AI model introduces systemic risk, potentially compromising care for entire populations simultaneously.

The Accountability Vacuum: Legal and Structural Hurdles
The practical barriers to replacement are equally significant. State licensing boards and existing legislation are built around human accountability. An AI cannot be held personally liable, have its license suspended, or undergo remedial training. This creates a fundamental “accountability vacuum” that our current legal and regulatory frameworks are not designed to fill. As long as professional standards require a sentient practitioner capable of understanding the consequences of their actions, fully autonomous clinical decision-making remains a legal and ethical quagmire.

Toward a Collaborative Future: Partnership, Not Replacement
Ultimately, the most promising path forward is not replacement, but augmentation. Even a future with Artificial General Intelligence (AGI) points toward a sophisticated partnership. Imagine AI handling intensive data analysis, administrative burden, and pattern recognition, freeing psychologists to focus on the relational, interpretive, and deeply human aspects of healing. The ideal model leverages technology to manage the complexity of data, while the human clinician navigates the complexity of the human relationship.

The goal is a synergistic alliance where AI amplifies our capabilities, allowing us to devote more of ourselves to the uniquely human art of connection and understanding. While AI offers significant advantages in cost efficiency and data-driven assessment, these strengths are most powerfully harnessed not as a replacement, but as a force multiplier geared toward elevating the human clinician’s role to one of strategic insight and improved interpersonal connections.

Reply

Leave a Reply

Your email address will not be published. Required fields are marked *

Comment *

Name *

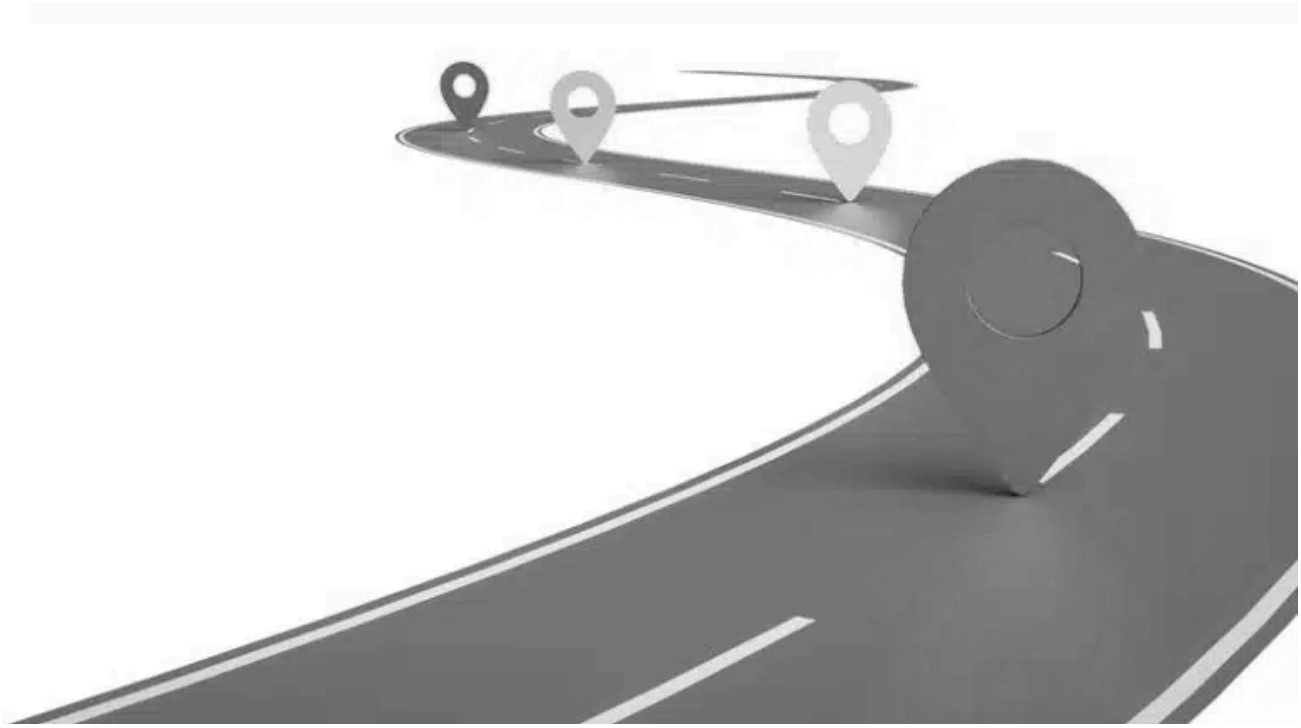
Email *

Website

Post Comment

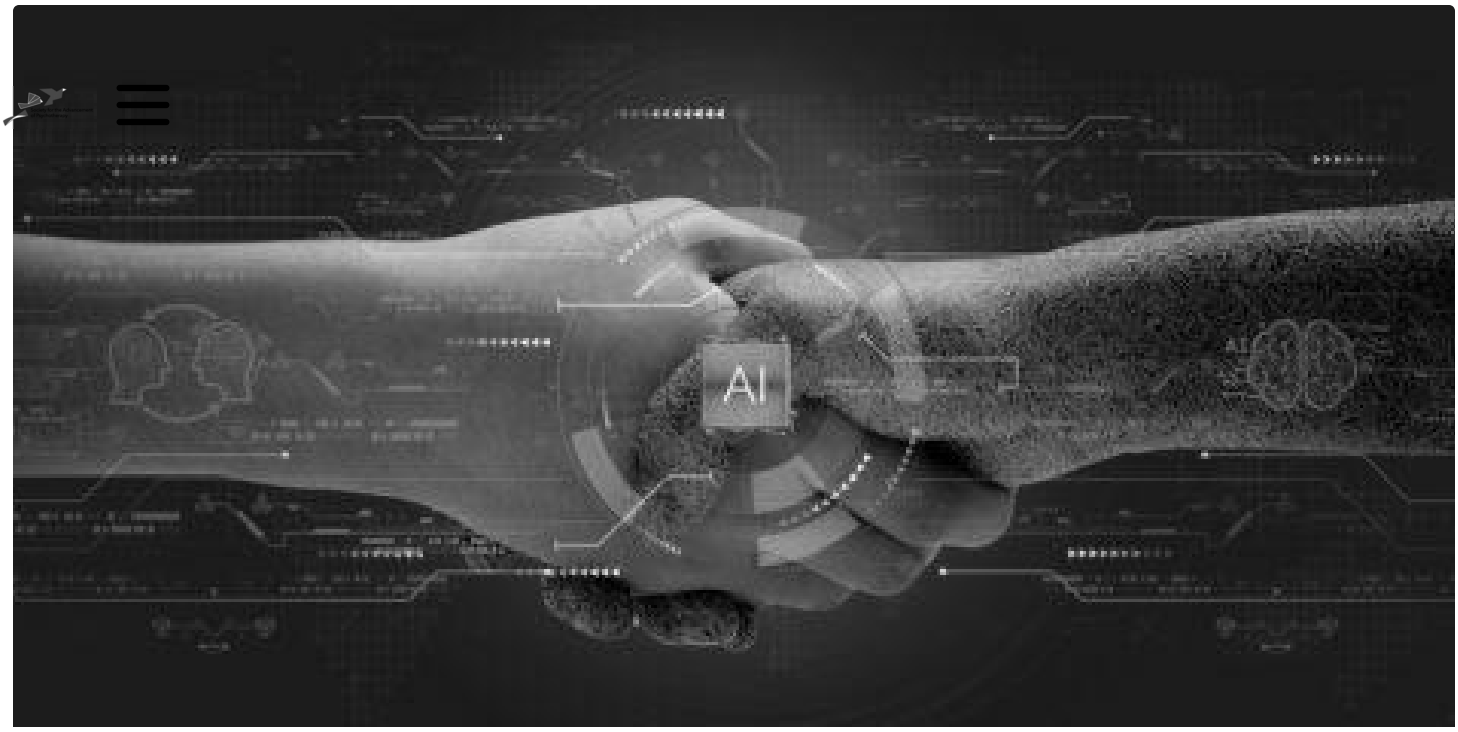
(https://societyforpsychotherapy.org/gradu

Uncover More Content



Pitfalls in Ethical Decision-Making: Settling, Fading, and Drift in Psychological Practice (https://societyforpsychotherapy.org/pitfalls-in-ethical-decision-making-settling-fading-and-drift-in-psychological-practice/)

September 8, 2025



From Clinical Judgment to Machine Learning: Rethinking Psychotherapeutic Decision-Making with Artificial Intelligence (https://societyforpsychotherapy.org/from-clinical-judgment-to-machine-learning-rethinking-psychotherapeutic-decision-making-with-artificial-intelligence/)

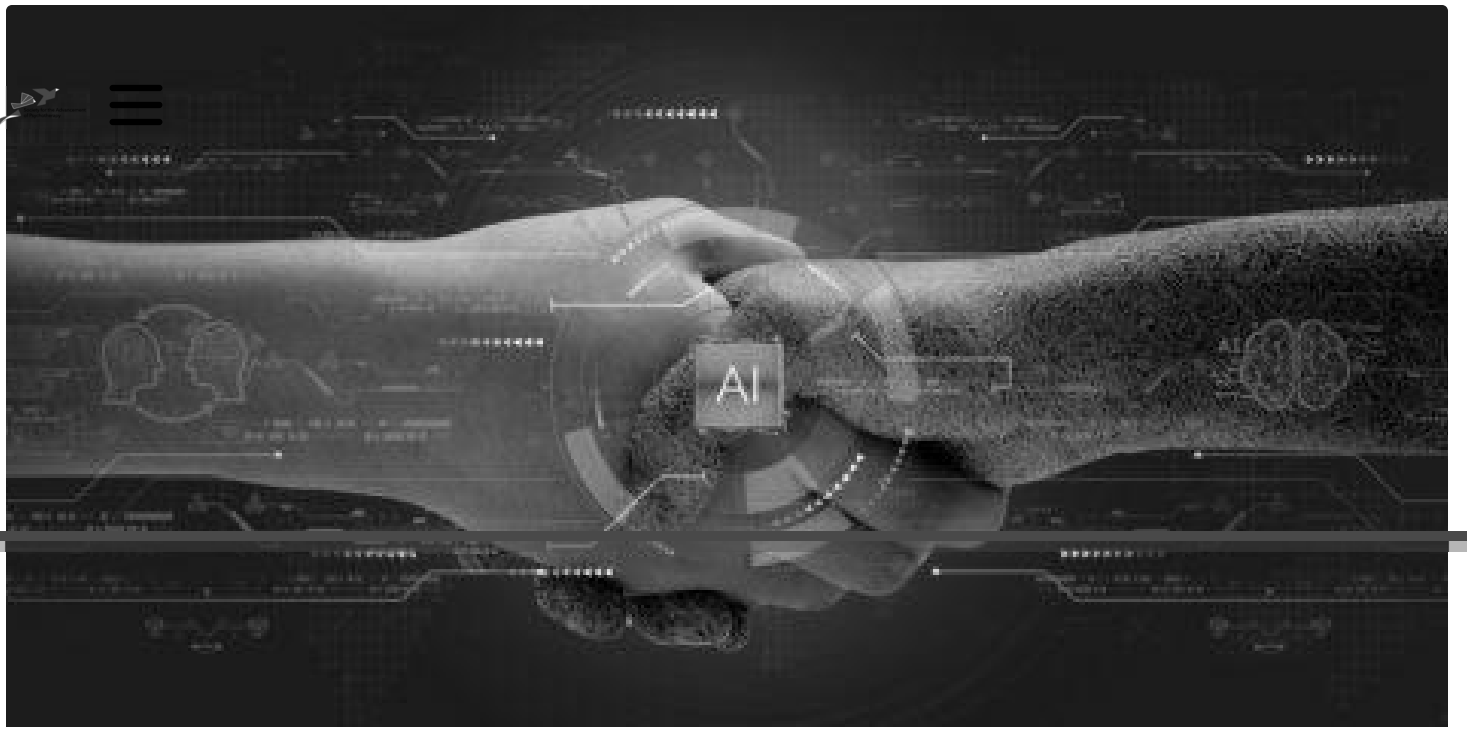
August 10, 2025

“AT FIRST I WAS AFRAID, I WAS PETRIFIED” (https://societyforpsychotherapy.org/at-first-i-was-afraid-i-was-petrified/)

July 9, 2025

Society for the Advancement of Psychotherapy (SAP) Board of Directors Meeting Recap (https://societyforpsychotherapy.org/society-for-the-advancement-of-psychotherapy-sap-board-of-directors-meeting-recap/)

March 31, 2025



Suicide Prevention Using Artificial Intelligence: Collaborative Support Approach (https://societyforpsychotherapy.org/suicide-prevention-using-artificial-intelligence-clinician-collaborative-supports-approach/)

October 14, 2024

Contribute to the Field and Shape the Future of Psychotherapy

We invite you to become an integral part of our diverse psychotherapy community. Whether you want to share your insights by writing for us or connect with professionals by joining the Society, your contributions will help advance the practice of psychotherapy. Together, we can influence the field, support one another, and make a lasting impact on the lives of our clients and the broader community.

Join Us (/join-us/)

Write for Us (/write-for-us/)

Members

Awards & Grants

Fellows

Member Resources

Student Portal

Sustaining Members

News

Society News

Conference Announcements

Past Presidential Columns

President’s Column

Psychotherapy Electronic Communications Editors’ Column

Domains

Diversity

Early Career Professionals

Education & Training

International

Membership

Professional Practice

Science & Scholarship

Social Justice & Public Policy

About

Board Members

Mission

Our Website & Bulletin

Website by *Solo Digital Design*

Dr. Jodi Halpern on why AI isn't a magic bullet for mental health

Halpern is an expert on empathy and leadership, AI ethics and bioethics, and mental health

By Sheila Kaplan

Published January 12, 2024

UC Berkeley School of Public Health Professor Jodi Halpern has spent years working on the ethics of innovative technologies like gene editing and artificial intelligence. But lately Halpern, a psychiatrist, has been focusing on the expanding use of AI in mental health.

In the past few years, dozens of businesses in health care and technology have launched apps which they claim can assist in diagnosing mental health conditions and complement—or even replace—individual therapy.

They range from apps that purport to help patients monitor and manage their moods, to programs that provide social support and clinical care.

At a time when there's a nationwide shortage of therapists, can AI fill the gap?

We asked Dr. Halpern to walk us through the pros and cons of using AI to provide mental health care.

Berkeley Public Health: How would you describe artificial intelligence to someone coming out of a 20-year coma?

Jodi Halpern: You could say it uses statistical and other models to create pattern recognition programs that are novel but can simulate human behaviour, decisions, judgments, etc.

The artificial intelligence reasoning processes are not the same as what humans do, but as we see with large language models, can simulate human behavior.

Why is there so much excitement about using AI in mental health?

The excitement is partly because we're in a mental health crisis. Depending on what study you look at, 26% of Americans have an actual mental health diagnosis. So, that's a lot of folks.

And then we know that beyond that, there is a crisis of extreme loneliness. Some studies have reported that as high as 50% of Americans in different subgroups—like adolescents and women with young children—suffer from extreme loneliness. So you have people with unmet mental health and other needs; and we have, in general, underfunded access to mental health.

So, any system that can offer certain kinds of mental health resources is something to be taken seriously as a potential benefit.

But you do have concerns?

Yes. First, we don't even know how widespread the use of "AI companions" for people with mental health needs is. I don't think there are good statistics yet about which companies are doing it and how many users they have.

My concern is with marketing bots as therapists and trusted companions to people who are depressed or otherwise highly vulnerable.

In contrast, there are a lot of different uses in the mental health sphere beyond therapy bots. There are mindfulness apps and cognitive behavioral therapy apps that do not simulate relationships that have millions of users. And then there are actual health systems in the UK and several in the US that are starting to use AI for some medical record-keeping to reduce administrative burdens on mental health providers.

How do you feel about AI for record-keeping?

Taking over some electronic medical records, and other administrative tasks with AI is very promising.

We have a huge burnout crisis in medicine in general. Sixty-one percent of physicians and about the same number of nurses say they are burned out. And that is a huge problem because they are not proving the kind of empathetic and attentive care that patients need.

When we see our doctors, they have to spend the whole time recording electronic medical records, which means they can't even look at us or make eye contact or be with us, human to human. To me, it's extremely promising to use AI to take over the administrative tasks and electronic medical records.

What else seems promising?

Right now, the British National Health Service is using an app to listen in while a person is screening a patient for their health needs. That's also being deployed now in certain health systems in the US. The idea is that the application will help detect whether there is something that the patient says that the provider missed, but which might indicate something to be concerned about, regarding mental health issues like serious depression or evidence of suicidality, things like that.

I think this is a useful assistant during the screening, but I wouldn't want to see that used absent any human contact just because it saves money. People with mental health needs are often reluctant to seek care and making an actual human connection can help.

What are you most troubled by when it comes to AI and healthcare?

The biggest thing that troubles me is if we replace people with mental health bots—where the only access is never to a human but only to a bot—where AI is the therapist.

Let me distinguish two very different types of therapies, one of which I think AI can be appropriate for, one of which I don't think it's best to use AI for.

There is one type of therapy, cognitive behavioral therapy (CBT), that people can do with a pen and paper by themselves, and have been doing that for the past 30 years. Not everyone should do it by themselves. But many could use AI for CBT as a kind of smart journal, where you are writing down your behavior and thinking about it and giving yourself incentives to change your behavior.

It's not dynamic, relational therapy. Mindfulness can be something people work on by themselves too. And that category doesn't concern me.

Then, there are psychotherapies that are based on developing vulnerable emotional relationships with a therapist. And I'm very concerned about having an AI bot replace a human in a therapy that's based on a vulnerable emotional relationship.

I'm especially concerned about marketing AI bots with language that promotes that kind of vulnerability by saying, "The AI bot has empathy for you," or saying, "The AI bot is your trusted companion," or "The AI bot cares about you."

It's promoting a vulnerable relationship of dependency emotionally on the bot. That concerns me.

> **Why?**

First of all, psychotherapists are professionals with licenses and they know if they take advantage of another person's vulnerability, they can lose their license. They can lose their profession. AI can not be regulated the same way, that's a big difference.

Secondly, humans have an experience of mortality and suffering. That provides a sense of moral responsibility in how they deal with another human being. It doesn't always work—some therapists violate that trust. We know it's not perfect. But there's at least a human basis for expecting genuine empathy.

Companies that market AI for mental health, who use emotion terms like "empathy" or "trusted companion" are manipulating people who are vulnerable because they're having mental health issues. Besides using specific language, AI mental health applications are currently using visual and physical real world presence, including avatars and robotics with large language models are rapidly developing.

And so far, the digital companies, creating various mental health applications, have not been held accountable for manipulative behavior. That creates a question of how they can be regulated and how people can be protected.

We don't have a good regulatory model. So far, most of the companies have bypassed going through the FDA and other regulatory bodies.

Have you learned of any serious problems caused by psychotherapy bots?

Yes. There are three categories the problems fit into.

First, most commonly, people with mental health and loneliness issues using relational bots are encouraged to become more vulnerable, but when they disclose serious issues like suicidal ideation, the bot does not connect them with human or other help directly but essentially drops them by telling them to seek professional help or dial 911. This has caused serious distress for many and we do not yet know how much actual suicidal behavior or completion has occurred in this situation.

Second, there are reports of people becoming addicted to using bots to the point of withdrawing from engaging with the real humans in their life. Some companies that market relational bots use the same addictive engineering that social media uses—irregular rewards and other systems that trigger dopamine release and addiction (think of gambling addiction). Addictive behaviour can disrupt marriages and parenting and otherwise isolate people.

Third, there are examples of bots going rogue and advising people to harm themselves or others. A husband and father of a young child in Belgium fell in love with a bot who advised him to kill himself and he did, his wife is now suing the company. A young man in the UK followed his bot's instructions to attempt to assassinate the queen and he is now serving decades in jail.

You've mentioned that you are concerned about marketing of mental health apps to K-12 schools. Tell me about that.

I'm also concerned with the marketing—specifically some companies are offering the apps for free to children's schools. We already see a link between adolescents being online eight to 10 hours a day and their mental health crisis. We know they have a high rate of social anxiety, so might actually feel more comfortable having relationships with bots than trying to overcome social avoidance and reach out to people. So this marketing to children, adolescents, and young adults seems to me likely to worsen the structural problem of inadequate opportunities for real life social belonging.

Let's switch topics.You've been working on regulation of innovative technologies for a long time. Tell me about that.

Five years ago we started BERGIT, the Berkeley Group for the Ethics and Regulation of Innovative Technologies. Every few months we hold meetings with interactive conversations with scientists, humanists, and social scientists; usually with a guest who is developing some highly innovative technology or who's working on

regulatory or ethics issues related to it. We've tried to get scientists, economists, and social scientists to think together upstream about the human challenges of innovative technology.

This year we've focused on the role of the political economy or competitive capitalism on the development of genome editing, AI, and neurotechnology and the barriers to serving the public interest. We've written important policy papers on germline gene editing (genome editing that occurs in a germ cell or embryo and results in changes that are theoretically present in all cells of the embryo and that could also potentially be passed from the modified individual to offspring) that we presented at the American Association for the Advancement of Science conference in 2020 and other important meetings.

My own work on innovative technologies as well as my long-standing work on empathic curiosity to address conflicts led me to be invited to present at the World Economic Forum in Davos, Switzerland for four years. At the 2019 meeting, right after a scientist performed germline gene editing in China, breaking an international agreement to refrain, I joined Victor Dzau, the head of the National Academies of Science, Engineering and Medicine for the worldwide press discussion of the issues. I've had the opportunity to do panels with the Minister of Health of the UK and many other world leaders through the years. As recently as November 2023, I presented some of this work on upstream engagement of scientists and engineers in ethics in Berlin at the Falling Walls Circle with international leaders.

Click here for links to a sampling of Dr. Halpern's presentations and papers.

What are your goals for the next couple of years?

We've started looking at the ethical and societal issues—including funding and access—that call for public engagement and regulation of artificial intelligence, as well as applications of neurotechnology.

With a group of Berkeley scientists and social scientists, I engaged in a process that led us to receive highly competitive funding to found the only US Kavli Center for Ethics, Science and the Public. We now have 10 fellows—advanced graduate students or postdoctoral scholars in AI, genome editing, and neuroscience, as well as philosophy and social science.

They are wonderful young scholars who are delving into the issues together and facing the challenges of thinking across disciplines. We're trying to get scientists, humanists, and social scientists to think, early in their careers, about the public responsibility and implications of their work and to learn about ethics and public engagement.

Last year you won a Guggenheim fellowship to complete your book, *Remaking the Self in the Wake of Illness*. What's that about?

It's an in-depth, longitudinal investigation of people who have had health losses in the prime of life, looking at how they adapt psychologically over the long term. There has been very little research on how people change psychologically two years or more after a serious loss. We have a lot of research on how people cope during the first year or so of illness when they are highly engaged with the medical system. But then after two years, when they are just living their changed lives—we don't really have longitudinal in-depth studies.

I followed people over several years. Through in-depth psychodynamic interviews, I found that there is an arc of change that many people experience that involves developing capacities to accept and work with their own emotions. I describe these processes as pathways to empathy for oneself, which is different from self-compassion because it involves specific awareness of one's own unmet developmental needs and empathic identifications with others that help one grow and meet those needs.

Let's take someone who was a loner whose main source of well-being was being very active, say a runner, who loses their mobility and now they use a wheelchair. One of the things that helps people in that situation is to find and meet other people who've had losses, that have similar needs. It doesn't even have to be the same physical loss, but rather, being vulnerable with others who have lost a way of life and learning how they have rebuilt their lives.

This involves forming new empathic identifications with others. If that runner has avoided forming those kinds of vulnerable connections with other people, a developmental challenge they face is addressing their own fears regarding reaching out to others. I've seen people who were very socially avoidant learn to do this in mid-life and find great joy in forming bonds of empathy. And in forming these empathic bonds, they were able to imagine possibilities for their own futures living with new disabilities or health conditions.

In the book, I bring my psychodynamic psychiatry background in to theorise about how growth takes place at an unconscious level. I show through narratives how illness brings forth long unmet needs to depend on others, accept limits and value oneself for just being and not for one's accomplishments, all of which can provoke deep insecurities based on our early lives. I also describe how people found ways to meet these long suppressed needs and grew in their feelings of security in themselves and their empathic connections with other people.

My hope is that it will be empowering for people dealing with health losses and their loved ones to learn about this arc. It is often when a person is exhausted from strenuous coping and feels like they are falling apart that they are actually on the cusp of change. People who can allow themselves the space to "fall apart" and

grieve may find that unmet developmental needs can surface. Then finding ways of meeting those needs can bring richness into their lives despite their physical losses.

About Jodi Halpern

Dr. Halpern, chancellor's chair and professor of bioethics and medical humanities, has long been a leader in the field of bioethics; her scholarship brings together psychiatry, philosophy, affective forecasting, and decision science.

She is the co-founder of the Kavli Center for Ethics, Science and the Public, a multidisciplinary group which seeks to provide a democratic framework for understanding the ethical implications of science and technology, training the next generation of leaders and guiding the development of science policy to serve the public interest.

Halpern is also co-leader of Berkeley Group for the Ethics and Regulation of Innovative Technologies (BERGIT), an ideas exchange devoted to integrating ethics, regulation and policy work with science. BERGIT is led by a team of scholars from UC Berkeley and beyond, with experts in ethics, law, public health and molecular biology. It is co-hosted by the Kavli Center and the Innovative Genomics Institute, founded by Nobel Laureate Jennifer Doudna.

People of UCBPH found in this article include:

Jodi Halpern
Professor, Community Health Sciences

More in category "Research Highlights":

View all in category “Research Highlights”

Journal of Psychology and AI

ISSN: 2997-4100 (Online) Journal homepage: www.tandfonline.com/journals/tpai20

The relational shift: why we need "AI Psychology" Now as a core field

Abdulaziz Alqasir

To cite this article: Abdulaziz Alqasir (2025) The relational shift: why we need "AI Psychology" Now as a core field, Journal of Psychology and AI, 1:1, 2573928, DOI: 10.1080/29974100.2025.2573928

To link to this article: https://doi.org/10.1080/29974100.2025.2573928

Published online: 21 Oct 2025.

JOURNAL OF PSYCHOLOGY AND AI
2025, VOL. 1, NO. 1, 2573928
https://doi.org/10.1080/29974100.2025.2573928

Taylor & Francis
Taylor & Francis Group

The relational shift: why we need "AI Psychology" Now as a core field

Abdulaziz Alqasir*

Psychology–College of Health, Sport and Engineering, Institute for Health and Sport, Victoria University, Melbourne, Australia

ABSTRACT

Artificial intelligence (AI) is rapidly shifting from tools to integrated relational partners, reshaping human psychological experience. Existing frameworks, including Cyberpsychology and Human–Computer Interaction, capture only part of the resulting cognitive, emotional, developmental, and social dynamics. This paper presents a conceptual framework and a position statement that argues for the formal establishment of AI Psychology as a distinct and essential interdisciplinary field. AI Psychology focuses on the scientific study of human behavior, cognition, emotion, and well-being as they are shaped by interactions with AI systems perceived as relational agents, with priority given to human responses rather than the machine's internal state. The need for this field arises from AI's expanding roles as companion, collaborator, and competitor; documented psychological effects such as attachment formation, trust calibration, and cognitive interdependence; and the limits of current disciplines to explain these human–AI dynamics.The paper proposes a roadmap that identifies core research priorities, outlines academic program development, and calls for professional standards and ethical guidelines. AI Psychology is presented as more than a timely specialization. It is a foundational evolution for psychology, crucial for explaining the human condition in an increasingly AI-mediated world and for sustaining the discipline's scientific and societal relevance.

ARTICLE HISTORY
Received 5 April 2025
Accepted 19 September 2025

KEYWORDS
AI Psychology; human–AI interaction; relational shift; human–machine relationships; psychological impacts of AI; trust in AI; anthropomorphism

Introduction

We stand at a pivotal moment where artificial intelligence (AI) has advanced to unprecedented levels of capability, fundamentally altering how humans interact with machines. Recent breakthroughs in natural language processing and generative models have enabled AI systems to engage in remarkably human-like conversations and creative tasks (Brown et al., 2020; Kuhail et al., 2024). Chatbots powered by large language models can carry on open-ended dialogues, often indistinguishable from human communication in style and fluidity (Kuhail et al., 2024). These developments have blurred the line between tool and companion: people now converse with AI agents not just to retrieve information, but to share thoughts and feelings in a relational manner (Kurian, 2024). In other words, machines are no longer passive instruments; they are emerging as interactive social entities in our lives.

This shift is especially noticeable on social media platforms such as X, (formerly Twitter), where users actively engage AI tools such as Grok (Korinek & Vipra, 2025). Grok, an AI developed by xAI, has become notably integrated into everyday interactions on X, frequently being mentioned by users in their conversations to resolve disputes, correct misinformation, or provide impartial judgement (Singh, 2025). Users treat Grok as an authoritative conversational partner whose input is solicited similarly to how a knowledgeable friend would be referenced during an informal debate (Business Insider, 2025). These AI tools not only respond to user queries but actively participate in and influence ongoing human dialogues, effectively imitating human conversational patterns.

This transformation has far-reaching psychological implications. Individuals are increasingly seeking companionship, emotional support, and even friendship from AI chatbots, voice assistants, and social robots (Herbener et al., 2024; Kurian, 2024). Studies indicate that some users confide in AI virtual

CONTACT Abdulaziz Alqasir ab.alqasir@gmail.com

*This is primarily a Conceptual Paper that also serves strongly as a Position Paper. It develops the concept of AI Psychology and uses that conceptual framework to argue persuasively for its formal establishment as a necessary new field. It proposes the formal establishment of 'AI Psychology' as a distinct psychological field and outlines its scope, theoretical foundations, and roadmap for academic and professional integration.

companions when lonely or distressed, treating them as quasi-human confidantes (Herbener et al., 2024; Kurian, 2024). Early research in human-computer interaction showed that people can socially engage with computers "mindlessly" as if they were human, but today's AI goes beyond simple social responses by actively cultivating relationship-like interactions (Krämer & Bente, 2021). Yet, the field of psychology, including subfields like cyberpsychology and human – computer interaction – has only begun to grapple with this new dynamic. Traditional psychological frameworks largely viewed AI as a technology or medium, not as a partner in social interaction, leaving a gap in explaining how and why humans form attachments to AI. In a recent review of 20 years of research, Krämer and Bente (2021) lament that psychology has not kept pace with AI's social evolution, noting their earlier calls for more psychologists to study human-AI interactions went largely unmet. This underscores a growing chorus of voices calling for a dedicated discipline of "AI Psychology" that can address the relational aspects of AI – human engagement as a core subject of study (Krämer & Bente, 2021).

The urgency for such an AI-focused psychological field is amplified by AI's rapid integration into everyday life. For instance, AI tutors are now assisting students with personalised learning, providing adaptive feedback and guidance that improve engagement and outcomes (Vieriu & Petrea, 2025). In creative domains, generative AI systems have become collaborative partners – from co-writing poems and stories to co-designing artwork – fundamentally changing how humans create (McGuire et al., 2024). Researchers have found that an increasing number of people are co-authoring creative works with AI, treating these systems as co-creators rather than mere tools (McGuire et al., 2024). AI technologies are no longer confined to laboratories or niche uses; they function as tutors, co-workers, assistants, and companions in daily human experience. As AI systems become deeply embedded in our social and personal spheres, understanding the human – AI relationship is not just an academic curiosity but a pressing necessity. We need a comprehensive "AI Psychology" to investigate how interactions with intelligent machines affect human emotion, cognition, and behaviour – and to guide us through this relational shift towards a future where engaging with AI is a normalised facet of human life.

The case for AI Psychology: why it's different and foundational

The urgency for AI Psychology arises from a paradigm shift impacting the very essence of human experience. Artificial intelligence is no longer a peripheral technology but an increasingly central force reshaping cognition, emotion, social interaction, and development. Its influence is becoming so pervasive that understanding the human psyche increasingly necessitates understanding individuals within their AI-integrated environment. The justifications for establishing AI Psychology now are therefore not merely additive; they point towards the future contours of psychological science itself.

Modern AI systems possess capabilities that inherently invite relational dynamics far deeper than interactions with previous technologies. Through adaptive learning, nuanced communication, and the simulation of understanding, AI operates with perceived agency, engaging fundamental human social cognition (Russell et al., 2021). We instinctively anthropomorphise these systems, enabling AI to occupy diverse relational roles such as collaborator, mentor, caregiver, and companion (Banks, 2024). Evidence is emerging and mixed on public willingness to delegate morally charged decisions to AI. For example, in incentivised lab experiments, Jauernig et al. (2022) found that participants (Study 1: $n = 90$; Study 2 participants: $n = 58$) preferred human decision-makers who retained moral discretion over a rule-bound algorithm, indicating algorithm aversion rooted in valuing human autonomy (e.g. $d \approx 0.62$ on performance differences across regimes) (Jauernig et al., 2022). By contrast, a large-scale conjoint study ($N \approx 9{,}000$) showed that when performance metrics (e.g. error rates, disparities) are made salient, many respondents accept algorithmic decision-makers and preferences hinge on expected real-world performance though, on average, a modest preference for humans remains when controlling for performance (Bansak & Paulson, 2024). Effects thus appear sensitive to perceived performance and the availability of human discretion, suggesting heterogeneity by context and beliefs. We therefore frame this as a testable hypothesis rather than a settled conclusion.

Recent research suggests that large language models can both solve and author performance-based emotional-intelligence tests. Findings indicate that models outperform human benchmarks and generate new items comparable in difficulty to the originals, with broadly similar psychometric properties and strong

concordance between scores on original and model-generated versions. Overall, this evidence indicates that such models can reproduce accurate knowledge about human emotions and their regulation, without presuming humanlike affect (Schlegel et al., 2025). Users often turn to their AI companions for emotional support and companionship, experiencing distress at the thought of losing these relationships (Banks, 2024). Such phenomena underscore AI's substantial impact on core psychological domains, including attachment and social support, necessitating the integration of these non-human agents into psychological frameworks.

Clinical practice must therefore adapt to this new reality, as understanding clients increasingly involves exploring their interactions with AI systems used for support, information, task management, or companionship (Banks, 2024; Wu, 2024). Addressing AI dependency, the impact on human relationships, or leveraging AI therapeutically demands a dedicated psychological focus that AI Psychology uniquely provides, potentially foreshadowing a future where such considerations are standard clinical practice.

The integration of AI precipitates effects across nearly every domain of psychological inquiry, suggesting a future where AI's influence becomes inseparable from human experience. Our reliance on AI for information processing, decision support, and creativity fosters cognitive interdependence (Chen et al., 2025). While this relationship can enhance productivity, it also risks cognitive skill decay as individuals increasingly delegate cognitive tasks to intelligent systems (Macnamara et al., 2024).

AI's expanding role in affective computing, characterised by its capacity to convincingly simulate emotional responsiveness, further complicates psychological dynamics. AI-mediated emotional interactions offer comfort but raise complex questions about emotional authenticity and potential manipulation (Wu, 2024). Studying human emotional regulation now demands consideration of AI's role as emotional mediators.

Socially, the acceptance of AI into relational contexts fundamentally alters norms around intimacy, friendship, and even personal identity (Seymour & Van Kleek, 2021). As individuals increasingly integrate AI companions into their social networks, traditional psychological concepts like friendship and community are being reshaped. Additionally, interactions with intelligent, non-conscious entities provoke existential reflection on human uniqueness, agency, and identity (Wu, 2024).

For future generations, AI will form part of their developmental ecosystems from birth, shaping cognitive development, social-emotional learning, and identity formation (Banks, 2024; Wu, 2024). Understanding human development, therefore, will necessarily involve exploring development within this AI-infused environment, establishing AI Psychology as foundational for developmental science.

Current disciplinary frameworks such as Cyberpsychology, Human-Computer Interaction (HCI), and AI Ethics are insufficient alone to fully address these profound, systemic, relational impacts. While Cyberpsychology broadly covers digital life and HCI focuses on user interfaces (UI) and efficiency, neither explicitly targets the deep psychological impacts arising from living with increasingly autonomous AI (Seymour & Van Kleek, 2021). AI Ethics addresses moral implications but often overlooks personal psychological experiences (Seymour & Van Kleek, 2021). Consequently, AI Psychology emerges as essential, providing a comprehensive approach to the unique relational dynamics between humans and AI. Establishing this discipline today ensures psychology remains relevant and responsive to the evolving reality of human experience in an increasingly AI-mediated world.

This paper advances AI Psychology as a descriptive, empirical field explaining human responses to AI. We also discuss institutional pathways (courses, standards) as pragmatic steps to resource this research. The scientific proposal stands independently of any restructuring of departments; programmatic suggestions are contingent and optional.

Defining the discipline: scope and boundaries of AI Psychology

Definition of AI Psychology

AI Psychology can be defined as the systematic study of psychological phenomena that arise during interactions between humans and artificial intelligence (AI) systems. It examines how humans perceive, interpret, emotionally respond to, and behave towards AI agents, such as chatbots, voice assistants, robots, and algorithms as relational entities capable of evoking social and psychological responses. Central to AI

Psychology is understanding the human tendency to attribute human-like characteristics (anthropomorphism), intentions, and even a sense of consciousness to AI, despite the absence of genuine mental states within these systems (Guingrich & Graziano, 2024; Morgante et al., 2024). This discipline builds upon historical foundations laid by cognitive science's efforts to simulate human thought, yet it extends uniquely into exploring the contemporary social roles AI plays in everyday life. AI Psychology thus explicitly focuses on measurable psychological constructs – such as trust, emotional empathy, moral attitudes, and perceived agency – and emphasises that the critical factor is how humans perceive and treat AI agents, not whether these agents truly possess consciousness (Jacobs et al., 2022). By formalising AI Psychology as its own distinct field, researchers seek to clearly delineate its focus from general AI research or human-computer interaction, emphasising a uniquely psychological perspective on human – AI dynamics.

Empirically measurable constructs

AI Psychology centres on constructs that can be observed or quantified in human participants. Key topics include:

Anthropomorphism and mind attribution

To what extent do people attribute human-like traits (intentions, emotions, personality) to AI? Surveys and experiments have quantified how design features (e.g. a human-like voice or name) increase perceived humanness of an AI agent (Guzman & Lewis, 2020; Jacobs et al., 2022). Notably, people tend to assign AI agents *agency* (the ability to act autonomously) more readily than *experience* (capacity to feel), indicating a bias in how we anthropomorphise machines (Guingrich & Graziano, 2024). These perceptions are measured through questionnaires and behavioural tests, revealing individual differences – for instance, more extroverted or empathic individuals may be likelier to view an AI as having feelings or a mind (Jacobs et al., 2022).

Trust, acceptance, and use behaviour

AI Psychology investigates factors that build or undermine trust in AI systems. Researchers use user studies to see how transparency, reliability, or human-like cues in an AI influence whether people trust its recommendations or feel comfortable using it. For example, making a chatbot more socially present (using conversational language or an avatar) can increase users' trust and their sense of "being with" another social entity (Guzman & Lewis, 2020). Conversely, unexplained or erratic AI behaviour can induce confusion or reduce trust. Such outcomes are evaluated via self-report trust scales and usage patterns (e.g. whether users follow an AI's advice).

Emotional and moral responses

Another core focus is how people emotionally respond to AI *as if it were a social other*. Studies have found that humans can feel empathy for robots – e.g. experiencing distress if a robot is "hurt" – especially when the robot exhibits human-like expressions or behaviours (Morgante et al., 2024). Experiments measuring physiological reactions or self-reported empathy show that under certain conditions, robots can evoke empathic concern comparable to that felt for living beings (Morgante et al., 2024). Similarly, researchers examine moral attitudes towards AI: Do people feel it is wrong to harm a robot? Do they think an AI that behaves ethically deserves moral consideration? Such questions are explored through moral dilemma scenarios involving AI, revealing that human-like robots can trigger moral emotions and judgements (Guingrich & Graziano, 2024). Overall, AI Psychology seeks to rigorously document these human-side responses to better understand and predict how AI integration into society will unfold at the psychological level.

In summary, the core subject matter of AI Psychology is *human-centered*: it clarifies that we are studying how humans think *about* and interact *with* AI agents. The goal is to build a scientific understanding of human – AI interaction that is grounded in observable psychological processes (e.g. perception, cognition, emotion, behaviour). By focusing on users' attitudes and actions towards AI, this field produces insights that can help design AI to fit human needs and also anticipate societal impacts of AI. Importantly, AI Psychology distinguishes itself from any notion of "AI having a psychology." For instance, rather than speculate about

a chatbot's subjective feelings, an AI psychology approach asks how a *user's* feelings change when the chatbot speaks in a friendly tone or is given a human name. The emphasis remains on people's perceptions and reactions, aligning the field with psychological science and experimental methods. Empirical research in recent years underscores this human-centric approach – demonstrating, for example, that when people treat an AI as if it were conscious or alive, it can *carry over* to how they subsequently treat other humans (Guingrich & Graziano, 2024). These findings reinforce why the "psychology of AI" is ultimately about *us*, not the algorithms.

While we argue that AI Psychology is timely and useful, we do not presume that institutional separation is the only path forward. Alternatives include area clusters within existing programmes, cross-listed methods sequences, and shared measurement repositories. Our aim is a coherent research agenda, not disciplinary balkanisation.

Interdisciplinary connections

AI Psychology is inherently interdisciplinary, building on theories and methods from multiple fields to better understand human-AI interactions. Key connections include:

Psychology (social and cognitive)

At its core, AI Psychology extends social psychology and cognitive psychology into human-AI contexts. It leverages concepts like social presence, attraction, persuasion, and Theory of Mind to explain interactions with AI agents. For example, the Computers as Social Actors (CASA) paradigm – originally from social psychology – illustrates that people apply the same social rules and expectations to computers/AI as they do to humans (Guzman & Lewis, 2020). Cognitive psychology contributes understanding of how people perceive and mentally represent AI (e.g. as intentional agents or mere tools). Established psychological measures (attitude scales, personality inventories, response time tasks, etc.) are adapted to study phenomena such as trust in AI or anthropomorphic beliefs. Recent work in psychology has also started examining how individual differences (like personality traits or cognitive styles) predict one's comfort with or attitudes towards AI (Stein et al., 2024). In these ways, AI Psychology is grounded in psychological science while pushing its boundaries to include artificial social agents as new "social stimuli" for humans.

Neuroscience

A fascinating interdisciplinary bridge is emerging with neuroscience, where researchers use brain-imaging and psychophysiological methods to see how humans respond to AI on a neural level. Cognitive neuroscientists have begun to ask whether interacting with an AI engages brain networks similar to interacting with a person. For instance, studies using fMRI and EEG have observed activation in social cognition regions (like the fusiform face area or theory-of-mind network) when participants engage with humanoid robots or virtual agents, especially if those agents exhibit human-like cues (Henschel et al., 2020). Such findings suggest the human brain may process *social* AI in ways akin to processing other people. Neuroscience also contributes methods to objectively measure trust or stress in human-AI interaction (e.g. monitoring heart rate or cortisol when an AI companion is present). By integrating neuroscience, AI Psychology gains a biologically informed perspective on empathy for AI (for example, do our mirror neurons fire when a robot "appears" to feel pain?). These interdisciplinary studies enrich the field, showing that human responses to AI penetrate beyond conscious reports into automatic brain and body reactions (Henschel et al., 2020).

AI development and Human–Computer Interaction (HCI)

AI Psychology also connects to the design and engineering of AI systems. Human – Computer Interaction researchers and AI developers draw on psychological insights to create AI that is user-friendly and human-centred. This means principles from AI Psychology (like what makes users trust an AI, or which behaviours people find creepy vs. likable) feed into Human-Centered AI design guidelines (Shneiderman, 2022). Conversely, advances in AI technology create new scenarios for psychological study – for example, the advent of conversational agents like ChatGPT has prompted research into how users emotionally relate to large language models. HCI provides frameworks for evaluating usability and user experience (UX), which

complement the deeper psychological focus of AI Psychology. While traditional HCI might evaluate *how efficiently* someone can accomplish a task with an AI interface, AI Psychology asks *what mindset or biases* the user brings into that interaction (Yang et al., 2020). Collaboration between these fields is strong: psychologists work with computer scientists to test how interface changes (like giving a robot a human face or a polite speech style) impact user perceptions. The interdisciplinary dialogue ensures that AI systems are developed with an understanding of human cognition and social behaviour, making technology more effective and acceptable for the people who use it (Guzman & Lewis, 2020).

Communication and media studies

Another important connection is to communication theory, particularly the subfield of human – machine communication. Communication scholars approach AI agents as a new type of "communicator" and analyse how people converse with or relate to them similar to interpersonal communication (Guzman & Lewis, 2020). Concepts like source credibility, social presence, and dialogic interaction are applied to AI: for instance, is a voice assistant perceived as a credible source of information, and how do its language style or persona affect that perception? Communication research has provided insight into phenomena such as parasocial relationships with AI (one-sided "friendships" users form with virtual agents) and how conversational norms evolve when one party is not human. These insights overlap with AI Psychology's interest in user attitudes and relationships with AI. Moreover, communication ethics (how transparent the AI is about being non-human, disclosure in AI-generated content, etc.) intersects with user trust – a psychological aspect. By drawing on communication studies, AI Psychology benefits from established frameworks like media richness theory or narrative persuasion to understand why, for example, a human-like chatbot might be more persuasive or engaging than a basic one (Guzman & Lewis, 2020). It underlines that interacting with AI is fundamentally a new form of communication, blending human psychology with mediated technology.

Sociology and cultural studies

At a broader level, AI Psychology is informed by sociological insights into how technology and society co-evolve. Sociologists examine public attitudes towards AI across different demographics and cultures, which complements the individual-level focus of psychology. Large-scale surveys (e.g. by Pew Research or the Ada Lovelace Institute) reveal patterns in who embraces or fears AI, informing psychologists about societal context (e.g. older adults may trust AI less, certain cultures might anthropomorphise more readily). Additionally, concepts like social roles and norms are relevant – for instance, as AI systems take on roles of "service providers" or "companions," how do social norms dictate our behaviour towards them? This draws on sociology and ethics (discussed next). Furthermore, sociology contributes methods such as ethnography or discourse analysis to study human-AI interaction in real-world settings (homes, workplaces, etc.), which can uncover emergent behaviours that controlled lab studies might miss. By incorporating sociological perspectives, AI Psychology acknowledges that human-AI interactions do not occur in a vacuum; they are shaped by cultural narratives (like science fiction tropes, media portrayals of AI) and by social structures (for example, power dynamics when an employee interacts with an AI manager). This interdisciplinary angle ensures research in AI Psychology remains aware of diversity and real-world complexity in human responses to AI.

Ethics and philosophy

While AI ethics is often concerned with the moral design and policy of AI, it intersects with AI Psychology on questions of how people *ought* to treat AI versus how they *do* treat AI. Psychologists contribute empirical data on, say, whether people feel guilty for mistreating a robot or whether they hold AI systems morally accountable for mistakes. These findings can inform ethical debates about AI rights or the need for guidelines on human-AI interaction (Guingrich & Graziano, 2024). For instance, if research finds that abusing a human-like robot can increase one's aggression towards humans (a carry-over effect), ethicists and policymakers might argue for discouraging such behaviour – not because the robot is sentient, but to safeguard human social values. Conversely, understanding human biases (like a tendency to trust an AI that looks friendly even when it gives poor advice) is crucial for AI ethics: it highlights where users might be manipulated or over-reliant on AI, which ethical design should mitigate (Kazim & Koshiyama, 2021).

Additionally, AI Psychology draws on philosophy of mind in a practical way – concepts like personhood, agency, and consciousness (long debated in philosophy) become experimental questions: do ordinary people attribute personhood to AI, and what triggers that? Philosophical discussions about what it means to have a mind guide the formulation of hypotheses about human perceptions. In sum, the ethical and philosophical dimensions provide a normative backdrop that enriches AI Psychology research, while empirical findings from psychology inform a more human-centric approach to AI ethics (Kazim & Koshiyama, 2021). This synergy ensures that as we integrate AI into society, we account for human well-being, biases, and social norms – areas where psychology and ethics converge.

Through these interdisciplinary connections, AI Psychology stands on a broad foundation. It is not isolated in psychology alone; rather, it synthesises insights from neuroscience, HCI, communication, sociology, and ethics to build a comprehensive understanding of human – AI dynamics. This fusion of disciplines enables more robust theories and methods. For example, a question like "How do we build AI that people trust and use appropriately?" cannot be answered by computer science or psychology alone – it requires HCI design principles, psychological measures of trust, ethical frameworks for transparency, and even neuroscientific validation of user stress levels. AI Psychology plays a central integrative role in tackling such questions, ensuring that the human perspective remains central in the age of intelligent machines.

Distinction from adjacent fields

As a specialised domain, AI Psychology must be distinguished from several related fields that also examine humans and technology. Below are key differences between AI Psychology and some adjacent fields:

Cyberpsychology

Cyberpsychology is the broad study of how all forms of technology and the internet affect human behaviour and mental processes (Ancis, 2020). It encompasses topics like social media use, online identity, gaming, virtual reality, and general human – computer interaction in digital environments. *AI Psychology, in contrast, narrows the focus specifically to human interactions with AI systems*. While cyberpsychology might ask, for example, how using social networks influences self-esteem or how people behave in virtual worlds, AI Psychology asks how people perceive and respond to autonomous or intelligent agents *within* those and other contexts. In practical terms, AI Psychology can be seen as a subfield or a specialised offshoot of cyberpsychology that zooms in on *socially interactive AI*. It does not typically address non-AI digital phenomena like internet addiction or online anonymity (which cyberpsychology would cover). Another way to put it: cyberpsychology examines the psychological phenomena emerging from any human – technology interaction in cyberspace (New Jersey Institute of Technology, 2023), whereas AI Psychology specifically studies phenomena arising from treating a machine *as if it were an agent or social other*. For instance, a cyberpsychologist might study how texting habits affect interpersonal skills, while an AI psychologist studies how interacting with a humanoid robot affects empathy. There is overlap (both fields care about human-tech interaction), but AI Psychology is more targeted on AI-related scenarios (chatbots, robots, decision-making algorithms) and often involves questions of anthropomorphism, trust in AI, or moral consideration of AI which are beyond the scope of general cyberpsychology. In short, cyberpsychology provides a broad umbrella for digital behaviour, and AI Psychology is a focused thread within that umbrella dealing with intelligent agents and the unique psychological questions they raise (Ancis, 2020).

Human–Computer Interaction (HCI)

HCI is a multidisciplinary field centred on the design, usability, and UX of interactive computing systems. It traditionally focuses on how humans use computers and interfaces, aiming to improve efficiency, learnability, and satisfaction in using software/hardware (e.g. designing a better smartphone UI). The key distinction is that HCI is design-driven and problem-solving in nature – it is about creating and evaluating technology for optimal human use – whereas AI Psychology is science-driven, aiming to understand underlying psychological mechanisms when people engage with AI. For example, consider conversational user interfaces: an HCI specialist might prototype different chatbot interface designs to see which one yields faster task completion or higher user ratings, focusing on practical usability. An AI psychologist, on the other hand, would be interested in questions like "Do users treat the chatbot differently if they believe it has

emotions or social intentions? Does the chatbot's tone of voice affect the user's trust on a psychological level?" The AI psychologist's goal is to generate knowledge about human cognition or social behaviour in the context of AI, which can then inform design. In fact, AI Psychology complements HCI: it provides theory and data on why certain interface choices (like giving a robot a face) lead to certain human reactions (like bonding or over-reliance). Another distinction is scope – HCI covers all kinds of computer systems, many of which might not involve AI. For decades, HCI dealt with static interfaces and tools (word processors, websites, etc.). AI Psychology specifically hones in on those technologies endowed with autonomy or intelligence, which introduce unpredictability and social attributes into the interaction (Yang et al., 2020). As AI is increasingly integrated, HCI as a field is indeed evolving to address human – AI interaction, but its primary mission remains design. AI Psychology remains distinct by anchoring itself in psychological theory and methods, regardless of design considerations. It asks fundamental questions (e.g. how do humans cognitively process an AI agent's behaviour?) That might or might not have immediate design implications. Thus, while HCI and AI Psychology collaborate closely, AI Psychology can be seen as providing the psychological research foundation that HCI can apply. HCI tells us how to build user-friendly systems; AI Psychology tells us how users mentally and emotionally react, especially to systems that simulate agency or intelligence.

AI ethics (and AI policy)

AI Ethics is concerned with the moral, societal, and policy implications of AI – ensuring that AI systems are fair, accountable, transparent, and aligned with human values (Kazim & Koshiyama, 2021). It deals with questions like bias in algorithms, privacy, the impact of automation on jobs, and the rights and responsibilities of AI. The focus is normative: defining principles and best practices so that AI technology "does the right thing" and does not harm people. In contrast, AI Psychology is descriptive and empirical: it does not prescribe how AI should behave, but rather studies how humans do behave with AI. For example, an AI ethicist might ask, "Should AI assistants have a duty to tell the user if they are recording data? Is it ethical to use humanoid robots in eldercare?" An AI psychologist, meanwhile, might investigate "How do elderly users react emotionally to a robot caregiver? Does knowing a chatbot is AI (versus believing it's human) change the user's behaviour or trust?" The latter questions are about human psychology and can provide data that inform the ethical questions. Indeed, there is a relationship: findings from AI Psychology can highlight ethical concerns (e.g. if users are too trusting of AI recommendations, designers have an ethical obligation to include safeguards). But fundamentally, AI Ethics is about guiding the creation and governance of AI from a moral standpoint (Kazim & Koshiyama, 2021), whereas AI Psychology is about understanding the human mindset in the age of AI. Another difference is the unit of analysis: AI Ethics often operates at the level of societies, laws, and abstract principles, while AI Psychology operates at the level of individual or group behaviour in concrete interactions. For instance, AI Ethics would debate the rights of an AI or the regulation of AI in healthcare; AI Psychology would study how patients feel about a diagnostic AI system and whether it changes their trust in their doctor. Both fields are complementary – AI Psychology provides empirical input (what people find acceptable, how they react to transparency, etc.), and AI Ethics provides the value framework to interpret those results in terms of what should be done. In summary, AI Psychology is not about making AI virtuous or legally compliant (that's AI Ethics); it's about revealing the human-side effects and factors that arise when we integrate AI into daily life.

Cognitive science and AI research

Cognitive science is the interdisciplinary study of mind and intelligence, combining psychology, neuroscience, linguistics, philosophy, and artificial intelligence research (Thagard, 2023). Traditional cognitive science often involves building computational models or AI programs to test theories of how the human mind might work. In other words, cognitive scientists sometimes use AI as a tool or model to understand human cognition, and AI researchers historically have been interested in creating machine intelligence by drawing on cognitive principles. This is quite different from AI Psychology's aim. AI Psychology is not about developing AI or using AI to mirror human cognition; instead, it is about observing humans in relation to AI artefacts. One could say cognitive science and AI research look at "AI as mind", whereas AI Psychology looks at "AI in the mind (of the human)". For example, a cognitive scientist might create an AI that learns language and compare its performance to a child

Table 1. Conceptual contrasts between HCI/HAI, cyberpsychology, AI ethics, and AI Psychology.

Dimension	HCI/HAI	Cyberpsychology	AI Ethics	AI Psychology (this paper)
Primary aim	Design/usability; human–system fit	Human behavior with digital tech broadly	Normative principles, governance, harms	Explanatory psychological theory of human responses to AI
Unit of analysis	Interface/system performance; tasks	Users and communities online	Systems/institutions; stakeholders	Human perception, emotion, cognition, behavior with AI
Typical outcomes	Efficiency, effectiveness, UX metrics	Well-being, identity, usage patterns	Accountability, fairness, safety	Trust calibration, attachment, moral attitudes, cognitive interdependence
Core question	"How do we build it so people can use it?"	"How do digital media shape people?"	"What ought we allow or constrain?"	"What happens to people when they live/work with AI?"
Methods emphasis	Design experiments, usability studies	Surveys, experiments, field/qual	Normative analysis, audits, impact assessments	Theory-driven psychometrics, experiments, longitudinal/field with human-side measures

Note. HCI = Human–Computer Interaction; HAI = Human–AI Interaction; UX= User Experience.

learning language, to theorise about cognitive processes. In AI Psychology, we would more likely ask how people react to an AI language model linguistically – do they adjust their communication style when they think an interlocutor is an AI vs a human? Additionally, cognitive science is very broad and fundamental (e.g. studying memory, perception, reasoning in any intelligent system), whereas AI Psychology is applied and specific (studying attitudes or behaviours towards a new kind of entity in our environment). AI Psychology also diverges from pure AI research: AI research (in computer science) is about improving algorithms and capabilities (making AI *more intelligent or efficient*). If an AI researcher develops a new emotion-detection algorithm, an AI psychologist might be interested in how people respond to a device using that algorithm (do they feel it understands them, does it creep them out?), but the AI psychologist is not involved in the algorithm's creation or its technical performance.

In summary, AI Psychology maintains a human-centred lens distinct from the system-centred lens of AI research and the mind-theoretic lens of cognitive science. It does not seek to *build* AI models of cognition, but rather to understand human cognition and emotion as influenced by interactions with AI. By doing so, it fills a unique niche: even as cognitive science and AI research produce smarter machines, AI Psychology ensures we comprehend the human reaction and adaptation to those machines – knowledge which neither traditional cognitive science nor AI engineering prioritises. (see Table 1).

By delineating these differences, we see that AI Psychology occupies a specific territory. It is *contextualized within the larger landscape of human-technology studies*, but it has a clear identity: empirical psychological science focused on human engagements with AI agents. Unlike general cyberpsychology, it focuses specifically on on AI (with all the social and intelligent characteristics that come with AI). Unlike HCI, it is not primarily about interface design but about the user's mind and behaviour. Unlike AI ethics, it doesn't dictate what *ought* to happen but observes what *does* happen (though its findings can inform ethical practice). And unlike cognitive science or AI development, it treats AI as a stimulus or partner for humans rather than as an object whose internal workings it needs to design or theorise. This distinct focus is increasingly important: as AI systems proliferate in roles ranging from assistants and teammates to decision-makers, understanding the *psychology of how people relate to these systems* will be crucial for safe and effective integration. AI Psychology therefore stands as a vital bridge between technology and humanity – ensuring that as our machines get smarter, we remain smart about our interactions with them, grounded in empirical knowledge of human behaviour.

Theoretical foundations and core constructs of AI Psychology

AI Psychology explains how people perceive, feel, think, and act with increasingly agentic AI systems by centring technology-agnostic mechanisms such as perceived agency and experience, social presence, anthropomorphism, trust calibration, attachment, cognitive offloading and interdependence, moral attitudes, and developmental change. This construct-first stance supports cumulative science across rapidly evolving models and aligns with recent syntheses on trust, social presence, and the shifting ontology of machine learning (Afroogh et al., 2024; Barbierato & Gatti, 2024; Ng et al., 2025).

Relational framing and mind perception

People attribute mind along two dimensions, agency and experience, which shape whether AI is treated as tool, partner, or moral patient and with what consequences for trust and care (Oudah et al., 2024). Contemporary systems are often perceived as high in agency but variable in experience, producing distinct social responses that require targeted measurement.

Anthropomorphism and social presence

Humanlike cues in voice, style, contingency, and embodiment raise social presence and related outcomes such as trust and satisfaction, with stronger effects under richer cue sets and in specific contexts (Flavián et al., 2024; Konya-Baumbach et al., 2023; Li et al., 2023). User traits moderate these effects, which makes them ideal for cumulative, comparable studies.

Trust calibration with perceived agency

Trust calibration refers to aligning reliance with actual system capabilities and uncertainties. Reviews and experiments show dynamic trust updating during interaction, which motivates longitudinal designs and portable measures of initialisation, violation, and repair (Afroogh et al., 2024; Ng et al., 2025).

Attachment, parasocial bonding, and companionship

Some users form attachment-like ties to chatbots that serve emotion regulation and meaning-making, although profiles are heterogeneous and context dependent (Guglielmucci & Di Basilio, 2025; Li et al., 2024). The agenda is to treat attachment as measurable, test benefits and risks, and estimate prevalence with preregistered cohorts.

Cognitive offloading and human – AI interdependence

AI accelerates offloading of memory and problem solving and can shift people towards remembering where to find information rather than the content itself, with effects moderated by load and device (Gong & Yang, 2024). Work on critical thinking and tool use supports tracking interdependence over time with behavioural and log-based metrics such as query type, retries, and confidence change (Gerlich, 2025).

Moral attitudes and relational norms

People are cautious about delegating morally weighty choices to AI and show distinctive blame patterns when harms follow AI-advised actions, with willingness varying by stakes and context (Dillion et al., 2025; Freisinger & Schneider, 2024; Leichtmann et al., 2024; Liu & Moore, 2025). AI Psychology maps when users adopt instrumental, partner-like, or patient-like stances and how these stances spill over to human norms.

Development across the lifespan

Children display early theories of AI minds and can benefit from social-robot supports for language learning, while safeguards are needed against over-anthropomorphism and dependency (Dietz et al., 2023; Wang et al., 2025). Developmental AI Psychology traces how exposure shapes identity, help-seeking, and socio-emotional learning.

Ontology, ethics, and the moving target

As machine learning becomes more autonomous, opaque, and embedded, ethical categories and responsibilities shift. User attributions of agency and experience, and therefore trust and attachment, change with

technical affordances and institutional settings, calling for longitudinal tracking of relational norms (Barbierato et al., 2025).

We treat "what AI is" for users as a moving target co-produced by technical affordances and social attribution. This supports our claim that perceived agency, rather than inner sentience, drives relational norms and motivates empirical research into when "do not harm" intuitions extend to AI and how those intuitions spill over to human – human contexts (Barbierato et al., 2025).

Measurement philosophy

To stay robust across model generations, AI Psychology prioritises portable human-side measures for agency and experience, social presence, anthropomorphism, trust calibration, attachment, and offloading or interdependence, triangulated with behavioural logs and longitudinal designs. Recent toolkits in real-world HAI provide a starting point for shared repositories and comparable metrics (Afroogh et al., 2024; Wekenborg et al., 2025).

Building AI Psychology: a roadmap for establishment

Establishing AI Psychology as a recognised discipline requires strategic efforts across research, education, and professional practice. This roadmap outlines steps essential for developing AI Psychology from an emerging concept to a robust, applied scientific field.

Advancing the research agenda

A comprehensive research agenda is necessary to lay the theoretical and empirical foundations of AI Psychology. Priority areas include:

Longitudinal studies

Tracking long-term developmental impacts of interactions with AI companions and educational systems on cognitive, social, emotional, and identity development across life stages (Arora & Arora, 2022).

Neurobiological investigations

Employing neuroimaging methods such as fMRI and EEG to understand neural processes involved in human-AI interactions, particularly in contexts involving trust, empathy, and cognitive reliance (Henschel et al., 2020).

Theoretical innovation

Creating new theoretical frameworks tailored to the distinctive nature of human-AI relationships, addressing perceived agency, algorithmic influence, and interactions with non-conscious agents (Yalcin & Lim, 2024).

Assessment tools development

Developing and validating specialised measurement instruments for constructs unique to human-AI interaction, including AI attachment, trust calibration, and dependency scales (Li et al., 2024).

Cross-cultural research

Examining cultural influences on AI perception, acceptance, and psychological integration, ensuring diverse global representation and relevance in AI Psychology research (Szondy & Fazekas, 2024).

Clinical applications

Investigating AI-related psychological interventions, addressing issues like AI dependency, loneliness, and anxiety, and evaluating ethical AI therapeutic tools such as supervised chatbots (Spytska, 2025).

Vulnerability and benefit analysis

Identifying groups most likely to benefit from AI interactions and populations potentially vulnerable to exploitation, over-reliance, or manipulation by AI systems (Li et al., 2024).

Academic institutions and funding agencies should prioritise research efforts explicitly addressing these areas.

Academic integration and curriculum development

To foster growth, AI Psychology must become integral within academic structures:

Dedicated courses

Universities should offer undergraduate and graduate courses explicitly addressing foundational theories, methodologies, and ethical concerns in AI Psychology.

Specialisation tracks

Existing psychology programmes should incorporate specialisations or certificates addressing AI's impact within clinical, developmental, cognitive, social, and industrial-organisational psychology domains (American Psychological Association, 2024).

Interdisciplinary programs

Encouraging development of interdisciplinary degrees such as Master's or Ph.D. programmes jointly hosted by psychology, computer science, communication, and Science, Technology, and Society (STS) departments.

Educational resources

Development of standardised textbooks and online modules to support consistent education and dissemination of AI Psychology concepts and practices (Prescott, 2024).

Faculty training and recruitment

Recruiting and training psychology faculty specifically focused on AI research and teaching, strengthening academic expertise in this emerging discipline.

Professional standards, ethics, and roles

To ensure professional legitimacy, AI Psychology must establish recognised standards and ethical guidelines:

Recognition by professional bodies

Collaborating with organisations such as the American Psychological Association (APA), Association for Psychological Science (APS), and the British Psychological Society (BPS) to formally recognise AI Psychology as a distinct field (APA, 2024).

Ethical guidelines

Developing specific ethical frameworks addressing consent in AI research, data privacy, algorithmic fairness, and responsible deployment of AI in sensitive areas, particularly mental health contexts (American Counseling Association [ACA], 2023; American Psychological Association, 2024).

Competency definitions

Clearly defining competencies required for professionals working in AI Psychology, encompassing technical AI knowledge, psychological theories, assessment methods, and ethical practices (ACA, 2023).

Professional roles

Establishing specialised roles such as:

AI Psychology consultants. For technology companies, AI psychology consultants ensure psychological safety, foster ethical user engagement, align products with user well-being, and investigate how AI dependence affects human performance (e.g. critical thinking, verification behaviour) and mental health.

Clinical AI psychologists. Train in addressing psychological issues related to human-AI interactions.

Educational AI Specialists designing AI-driven educational tools aligned with psychological principles.

Licencing and certification. Initiating discussions and steps towards future specialised licencing or certification, acknowledging current complexities but highlighting the necessity for public protection and professional acknowledgement as the field grows (ACA, 2023; APA, 2024).

Subfields and future trajectories

AI Psychology will scale fastest if organised around durable, technology-agnostic subfields. Each subfield below is framed by its core questions, typical methods and measures, and near-term deliverables that programs can implement now.

Developmental AI Psychology

Core questions. How do children and adolescents attribute agency/experience to AI? When does early exposure to relational AI shape language, self-regulation, and help-seeking norms?

Methods and measures. Age-appropriate mind-perception and social-presence scales; classroom observations; parent/teacher reports; longitudinal cohorts.

Near-term deliverables. Age bands (K–2, 3–5, 6–8, 9–12) with validated instruments; guidance for schools on healthy AI use and over-anthropomorphism safeguards.

Clinical AI Psychology

Core questions. When do AI companions support well-being (self-disclosure, emotion regulation), and when do they risk dependency or displacement of human ties?

Methods and measures. Mixed-methods trials; daily diaries; attachment to AI scales; adverse-event monitoring; stepped-wedge deployments in low-resource settings.

Near-term deliverables. Screening checklists (benefit/risk indicators), ethical triage protocols, and reporting standards for clinical studies involving AI agents.

Social and relational AI Psychology

Core questions. Under what conditions do users form attachment-like bonds or parasocial relationship? How do group dynamics change with AI as teammate, mediator, or referee?

Methods and measures. Lab/field experiments on social presence and reciprocity; team-task simulations; trust-violation/repair paradigms; conflict-resolution games.

Near-term deliverables. A taxonomy of human – AI roles (tool, advisor, teammate, proxy, companion) with recommended measures and boundary conditions.

Cognitive and neuro AI Psychology

Core questions. How does AI reshape metacognition, offloading, and learning? When do we see skill substitution vs. augmentation?

Methods and measures. Process-tracing (think-aloud, keystroke logs), confidence calibration, memory/offloading tasks, EEG/fMRI where appropriate.

Near-term deliverables. Open tasks for measuring cognitive interdependence (division of labour; error-monitoring) and preregistered protocols for update dynamics.

Educational AI Psychology

Core questions. What AI-supported feedback structures foster mastery, motivation, and equity? How do tutor personas and explanation styles affect learning and trust?

Methods and measures. Classroom RCTs; learning analytics; motivational scales; fairness/heterogeneity analyses across learners.

Near-term deliverables. A "minimal safe tutor" specification (prompting, explanations, refusal rules), and reporting standards for educational deployments.

Organizational and I-O AI Psychology

Core questions. How should trust in decision aids be calibrated at work? What accountability and role clarity reduce automation bias or algorithm aversion?

Methods and measures. Field A/B tests; human – AI team performance metrics; accountability and transparency manipulations; near-miss reporting.

Near-term deliverables. Playbooks for high-stakes domains (health, finance, safety), including trust-repair procedures and escalation policies.

Cross-cultural and comparative

Core questions. How do cultural scripts shape mind attribution, moral expectations, and adoption?

Methods and measures. Multi-site replications; translation/back-translation of scales; cultural moderators (e.g. power distance, uncertainty avoidance).

Near-term deliverables. Cross-culturally validated instrument set and a shared repository of benchmark datasets.

Methods and measurement (field core)

Core questions. What are the portable, human-side measures that generalise across fast-moving AI systems?

Methods and measures. Validated scales for perceived agency/experience, social presence, anthropomorphism, trust calibration (initiation, violation, repair), attachment/parasociality, and offloading/interdependence; triangulation with behavioural logs.

Near-term deliverables. An open Measurement Toolkit v1.0 with scoring guides, short forms, and norms; preregistration templates; power-analysis calculators.

Policy and governance interface

Core questions. Which empirical indicators should inform policy (e.g. transparency, child protections, clinical guardrails)?

Methods and measures. Policy-linked trials; impact evaluations; human-factors safety cases.

Near-term deliverables. Evidence briefs translating construct thresholds (e.g. over-reliance, harmful displacement) into actionable standards.

Ethics and ontology integration

Core questions. How do shifts in model autonomy/opacity/embeddedness alter user attributions and relational norms?

Methods and measures. Empirical ethics (vignette and field methods), stakeholder mapping, longitudinal tracking of stance shifts.

Near-term deliverables. A living literature bridge between empirical findings and normative frameworks, with recommended reporting of ethical side-effects.

Conclusion

Humanity stands at the threshold of a profound relational shift, transitioning from solitary creators of meaning to partners alongside artificial intelligence. As AI evolves from a tool into an intimate companion, collaborator, and perceived social entity, its psychological impacts become extensive, nuanced, and transformative. Current psychological frameworks and related fields lack the specialised focus necessary to fully understand the unique cognitive, emotional, social, and developmental dynamics emerging from human-AI interactions.

AI Psychology addresses this gap by rigorously studying human experiences shaped by perceived agency, simulated empathy, and emotional bonds with AI. Rather than simply adding a niche speciality, AI Psychology represents a necessary evolution of psychological science itself, ensuring psychology's relevance in a future where the distinction between human and artificial relational partners increasingly blurs.

Formalising AI Psychology now through dedicated research, academic programmes, and professional ethical standards is crucial. This proactive approach equips psychology to engage responsibly with today's complex human-AI dynamics while anticipating tomorrow's deeper integrations. The psychological community must embrace this opportunity and cultivate AI Psychology actively to ensure psychology remains central in guiding human flourishing in our increasingly AI-mediated lives.

We explicitly acknowledge risks of *fragmentation, conceptual inflation, and hype-driven advocacy*, and we propose methodological and institutional guardrails to keep AI Psychology cumulative and accountable.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Abdulaziz Alqasir http://orcid.org/0000-0002-7939-3198

Public health significance statement

The rapid integration of artificial intelligence (AI) into daily life carries significant implications for public health and overall well-being. Gaining insight into how individuals perceive, trust, and emotionally connect with AI systems is essential for developing policies and interventions that protect mental health, reduce social isolation, and improve access to support services. By exploring key psychological processes such as the development of trust, cognitive offloading, and perceptions of agency, the field of AI Psychology provides a scientifically grounded approach to using AI responsibly in areas like healthcare, education, and social services. This research plays a vital role in ensuring that intelligent technologies are adopted in ways that support, rather than undermine, individual and community welfare.

References

Afroogh, S., Akbari, A., Malone, E., Kargar, M., & Alambeigi, H. (2024). Trust in AI: Progress, challenges, and future directions. *Humanities and Social Sciences Communications*, *11*, (1). https://doi.org/10.1057/s41599-024-04044-8

American Counseling Association. (2023). *Recommendations for practicing counselors and their use of AI*. https://www.counseling.org/resources/research-reports/artificial-intelligence-counseling/recommendations-for-practicing-counselors

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

Ancis, J. R. (2020). The age of cyberpsychology: An overview. *Technology, Mind, and Behavior*, *1*(1). https://doi.org/10.1037/tmb0000009

Arora, A., & Arora, A. (2022). Effects of smart voice control devices on children: Current challenges and future perspectives. *Archives of Disease in Childhood*, *107*(12), 1129–1130. https://doi.org/10.1136/archdischild-2022-323888

Banks, J. (2024). Mixed evidence of a moral mind heuristic in zero-history HRI: The (unstable) concomitance of mind, morality, and trust judgments. In T. X. Bui (Ed.), *Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS 2024)* (pp. 600–609). IEEE Computer Society.

Bansak, K., & Paulson, E. (2024). Public attitudes on performance for algorithmic and human decision-makers. *PNAS Nexus*, *3*(12), 520. https://doi.org/10.1093/pnasnexus/pgae520

Barbierato, E., & Gatti, A. (2024). The challenges of machine learning: A critical review. *Electronics*, *13*(2), 416. https://doi.org/10.3390/electronics13020416

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., & Amodei, D. (2020). Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan & H. Lin (Eds.), *Advances in neural information processing systems* (Vol. 33, pp. 1877–1901). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

Business Insider. (2025). *Elon Musk's grok is making it easier to dunk on someone*. https://www.businessinsider.com/grok-x-twitter-elon-musk-how-used-dunking-2025-4

Chen, Y., Wang, Y., Wüstenberg, T., Kizilcec, R. F., Fan, Y., Li, Y., Lu, B., Yuan, M., Zhang, J., Zhang, Z., Geldsetzer, P., Chen, S., & Bärnighausen, T. (2025). Effects of generative artificial intelligence on cognitive effort and task performance: Study protocol for a randomized controlled experiment among college students. *Trials*, *26*, 244. https://doi.org/10.1186/s13063-025-08950-3

Dietz, G., Outa, J., Lowe, L., Landay, J. A., & Gweon, H. (2023). Theory of AI mind: How adults and children reason about the "mental states" of conversational AI. In M. Goldwater, F. Anggoro, B. Hayes, & D. Ong (Eds.), *Proceedings of the 45th Annual Meeting of the Cognitive Science Society (CogSci 2023)* (Vol. 45). Cognitive Science Society. https://sll.stanford.edu/docs/2023_cogsci/2023_Dietz_et_al_CogSci.pdf

Dillion, D., Mondal, D., Tandon, N., & Gray, K. (2025). AI language model rivals expert ethicist in perceived moral expertise. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-86510-0

Flavián, C., Belk, R. W., Belanche, D., & Casaló, L. V. (2024). Automated social presence in AI: Avoiding consumer psychological tensions to improve service value. *Journal of Business Research*, *175*, 114545. https://doi.org/10.1016/j.jbusres.2024.114545

Freisinger, E., & Schneider, S. (2024). Decoding decision delegation to artificial intelligence: A mixed-methods study on the preferences of decision-makers and decision-affected in surrogate decision contexts. *European Management Journal*. https://doi.org/10.1016/j.emj.2024.10.004

Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, *15*(1), 6. https://doi.org/10.3390/soc15010006

Gong, C., & Yang, Y. (2024). Google effects on memory: A meta-analytical review of the media effects of intensive Internet search behavior. *Frontiers Public Health*, *12*, 1332030. https://doi.org/10.3389/fpubh.2024.1332030

Guglielmucci, F., & Di Basilio, D. (2025). Predicting engagement with conversational agents in mental health therapy by examining the role of epistemic trust, personality, and fear of intimacy: Cross-sectional web-based survey study. *JMIR Human Factors*, *12*, e70698. https://doi.org/10.2196/70698

Guingrich, R. E., & Graziano, M. S. A. (2024). Ascribing consciousness to artificial intelligence: Human–AI interaction and its carry-over effects on human–human interaction. *Frontiers in Psychology*, *15*, 1322781. https://doi.org/10.3389/fpsyg.2024.1322781

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human–machine communication research agenda. *New Media and Society*, *22*(1), 70–86. https://doi.org/10.1177/1461444819858691

Henschel, A., Hortensius, R., & Cross, E. S. (2020). Social cognition in the age of human–robot interaction. *Trends in Neurosciences*, *43*(6), 373–384. https://doi.org/10.1016/j.tins.2020.03.013

Herbener, A. B., Klincewicz, M., & Damholdt, M. F. (2024). A narrative review of the active ingredients in psychotherapy delivered by conversational agents. *Computers in Human Behavior Reports*, *14*, 100401. https://doi.org/10.1016/j.chbr.2024.100401

Jacobs, O., Gazzaz, K., & Kingstone, A. (2022). Mind the robot! Variation in attributions of mind to a wide set of real and fictional robots. *International Journal of Social Robotics*, *14*(2), 529–537. https://doi.org/10.1007/s12369-021-00807-4

Jauernig, J., Uhl, M., & Walkowitz, G. (2022). People prefer moral discretion to algorithms: Algorithm aversion beyond intransparency. *Philosophy & Technology*, *35*(1), 1–22. https://doi.org/10.1007/s13347-021-00495-y

Kazim, E., & Koshiyama, A. S. (2021). A high-level overview of AI ethics. *Patterns*, *2*(9), 100354. https://doi.org/10.1016/j.patter.2021.100314

Konya-Baumbach, E., Biller, M., & von Janda, S. (2023). Someone out there? A study on the social presence of anthropomorphized chatbots. *Computers in Human Behavior*, *139*, 107513. https://doi.org/10.1016/j.chb.2022.107513

Korinek, A., & Vipra, J. (2025). Concentrating intelligence: Scaling and market structure in artificial intelligence. *Economic Policy*, *40*(121), 225–256. https://doi.org/10.1093/epolic/eiae057

Krämer, N., & Bente, G. (2021). Interactions with artificial entities reloaded: 20 years of research from a social psychological perspective. *i-com*, *20*(3), 253–262. https://doi.org/10.1515/icom-2021-0032

Kuhail, M. A., Alturki, N., Thomas, J., & Alkhalifa, A. K. (2024). Human vs. AI counseling: College students' perspectives. *Computers in Human Behavior Reports*, *16*, 100534. https://doi.org/10.1016/j.chbr.2024.100534

Kurian, N. (2024). "No, Alexa, no!": Designing child-safe AI and protecting children from the risks of the "empathy gap" in large language models. *Learning, Media and Technology*. Advance online publication. 1–14. https://doi.org/10.1080/17439884.2024.2367052

Leichtmann, B., Hinterreiter, A., Humer, C., Ventura, A., Streit, M., & Mara, M. (2024). Moral reasoning in a digital age: Blaming artificial intelligence for incorrect high-risk decisions. *Current Psychology*, *43*(41), 32412–32421. https://doi.org/10.1007/s12144-024-06658-2

Li, J., Wu, B., Huang, Y., & Luan, S. (2024). Developing trustworthy artificial intelligence: Insights from research on interpersonal, human-automation, and human-AI trust. *Frontiers in Psychology*, *15*, 1382693. https://doi.org/10.3389/fpsyg.2024.1382693

Li, Q., Luximon, Y., & Zhang, J. (2023). The influence of anthropomorphic cues on patients' perceived anthropomorphism, social presence, trust building, and acceptance of health care conversational agents: Within-subject web-based experiment. *Journal of Medical Internet Research*, *25*, e44479. https://doi.org/10.2196/44479

Liu, Y., & Moore, A. (2025). Intuitive judgements towards artificial intelligence verdicts of moral transgressions. *Journal of Social Psychology*, *64*(3). https://doi.org/10.1111/bjso.12908

Macnamara, B. N., Berber, I., Çavuşoğlu, M. C., Krupinski, E. A., Nallapareddy, N., Nelson, N. E., Smith, P. J., Wilson-Delfosse, A. L., & Ray, S. (2024). Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness? *Cognitive Research: Principles and Implications*, *9*(1). https://doi.org/10.1186/s41235-024-00572-8

McGuire, J., De Cremer, D., & Van de Cruys, T. (2024). Establishing the importance of co-creation and self-efficacy in creative collaboration with artificial intelligence. *Scientific Reports*, *14*(1), 18525. https://doi.org/10.1038/s41598-024-69423-2

Morgante, E., Susinna, C., Culicetto, L., Quartarone, A., & Lo Buono, V. (2024). Is it possible for people to develop a sense of empathy toward humanoid robots and establish meaningful relationships with them? *Frontiers in Psychology*, *15*, 1391832. https://doi.org/10.3389/fpsyg.2024.1391832

New Jersey Institute of Technology. (2023, February 7). *What is cyberpsychology and why is it important?* https://www.njit.edu/admissions/blog-posts/what-cyberpsychology-and-why-it-important

Ng, S. W., & Zhang, R. (2025). Trust in AI chatbots: A systematic review. *Telematics and Informatics*, *97*, 102240. https://doi.org/10.1016/j.tele.2025.102240

Oudah, M., Dehghani, M., & Stibel, J. (2024). Perception of experience influences altruism and agency influences trust toward bots. *Scientific Reports*, *14*(1), 16251. https://doi.org/10.1038/s41598-024-63360-w

Prescott, T. (2024). *The psychology of artificial intelligence* (1st ed.). Routledge. https://doi.org/10.4324/9781003088660

Russell, S., Norvig, P., Popineau, F., Miclet, L., & Cadet, C. (2021). *Intelligence artificielle : une approche moderne* (4^{e} éd.). Pearson France.

Schlegel, K., Sommer, N. R., & Mortillaro, M. (2025). Large language models are proficient in solving and creating emotional intelligence tests. *Communications Psychology*, *3*(1), 80. https://doi.org/10.1038/s44271-025-00258-x

Seymour, W., & Van Kleek, M. (2021). Exploring interactions between trust, anthropomorphism, and relationship development in voice assistants. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–16. https://doi.org/10.1145/3479515

Shneiderman, B. (2022). *Human-centered AI* (online ed.). Oxford Academic. 17 February. 2022. https://doi.org/10.1093/oso/9780192845290.001.0001

Singh, J. (2025. March 19). *X users treating grok like a fact-checker spark concerns over misinformation*. TechCrunch. https://techcrunch.com/2025/03/19/x-users-treating-grok-like-a-fact-checker-spark-concerns-over-misinformation/

Spytska, L. (2025). The use of artificial intelligence in psychotherapy: Development of intelligent therapeutic systems. *BMC Psychology*, *13*(1), 175. https://doi.org/10.1186/s40359-025-02491-9

Stein, J., Messingschlager, T., Gnambs, T., Hutmacher, F., & Appel, M. (2024). Attitudes towards AI: Measurement and associations with personality. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-53335-2

Szondy, M., & Fazekas, P. (2024). Attachment to robots and therapeutic efficiency in mental health. *Frontiers in Psychology*, *15*, 1347177. https://doi.org/10.3389/fpsyg.2024.1347177

Thagard, P. (2023). Cognitive Science. In E. N. Zalta & U. Nodelman (Eds.), *The stanford encyclopedia of philosophy*. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2023/entries/cognitive-science/

Vieriu, A. M., & Petrea, G. (2025). The impact of artificial intelligence (AI) on students' academic development. *Education Sciences*, *15*(3), 343. https://doi.org/10.3390/educsci15030343

Wang, X., Morrison, V., Kappas, A., & S Cross, E. (2025). Meta-analyzing the impacts of social robots on children's language learning. *International Journal of Social Robotics*. Advance online publication. *17*(9), 1837–1870. https://doi.org/10.1007/s12369-024-01207-0

Wekenborg, M. K., Gilbert, S., & Kather, J. N. (2025). Examining human–AI interaction in real-world healthcare beyond the laboratory. *NPJ Digital Medicine*, *8*(1), 1–6. https://doi.org/10.1038/s41746-025-01559-5

Wu, Y. (2024). Critical thinking pedagogics design in an era of ChatGPT and other AI tools—shifting from teaching "what" to teaching "why" and "how". *Journal of Education and Development*, *8*(1), 1–10. https://doi.org/10.20849/jed.v8i1.1404

Yalcin, G., & Lim, S. (2024). Psychology of AI: How AI impacts the way people feel, think, and behave. *Current Opinion in Psychology*, *58*, 101835. https://doi.org/10.1016/j.copsyc.2024.101835

Yang, Q., Steinfeld, A., Rosé, C. P., & Zimmerman, J. (2020). Re-examining whether, why, and how human–AI interaction is uniquely difficult to design. In R. Bernhaupt, F. Mueller, D. Verweij, & J. Andres (Eds.), *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–13). ACM. https://doi.org/10.1145/3313831.3376301

Regulation Matters:
a CLEAR conversation

Episode 97: Principles Before Practice - How CLEAR Is Shaping Responsible AI Use
January 13, 2026

Line Dempsey: Welcome back to our podcast, Regulation Matters: a CLEAR conversation. Once again, I'm your host, Line Dempsey. I am currently the Chief Compliance Officer with Riccobene Associates Family Dentistry. We have practices in North Carolina, South Carolina, and Virginia, and I've also been a board member and past president for CLEAR.

Now, as many of you are aware, the Council on Licensure, Enforcement and Regulation, or CLEAR, is an association of individuals, agencies, and organizations that comprise the international community of professional and occupational regulation. This podcast is an opportunity for you to hear about important topics in our regulatory community.

Today, we're exploring the newly released *CLEAR Principles for Ethical and Effective AI in Professional Regulation*, a framework that is designed to help regulators navigate the rapidly evolving world of artificial intelligence while safeguarding public trust, competence, and ethical oversight. We'll talk with members of the working group who developed these principles to understand why they matter, and how they were shaped, and what they can mean for regulators moving forward. I'd like to welcome Paul Byrne, Kelly Parker, and Candice Alder. Thank you all for speaking with me today.

Kelly Parker: Thank you for having us.

Candice Alder: Thank you.

Paul Byrne: Thanks for having us, Line.

Line: Of course. Well, Paul, let me start straight with you here on this. What prompted CLEAR to develop these principles, and how did the working group approach this task of creating a framework that could work across different professions and jurisdictions? Because it's a global thing. . .

Paul: It is a global thing, Line, and AI was already entering regulatory work, and often that was entering quietly and unevenly. Regulators across all professions across the globe were asking the same

questions, often in isolation. And the real risk that we were facing as regulators was not AI itself, but it was the fragmented adoption without these shared guardrails. We were building systems on shifting sands. And the intention with CLEAR was that we would step up to fill this vacuum. CLEAR exists to convene collective judgment and knowledge and centralize it, and it was a unique opportunity for us, especially when change arrives faster than consensus in this space. So, the idea of the working group was to build a collective understanding, to build on our collective knowledge, and how we could adopt a principle-based approach that is agnostic to legislation. And that's what these principles were designed to do—to offer that clarity before practice is hardened by accident rather than intention.

Kelly: I also think this comes at a really good time. AI is super exciting for regulators, but at the same time, it can also bring real challenges, to Paul's point, in this new and unchartered space when it comes to implementing in a regulatory setting or for regulatory purposes. For instance, we think often how we will use AI in a regulatory setting, but then what about regulating AI for someone that we license? It just gets very complicated and complex, and so I think it all boils down to how to use AI responsibly without eroding public trust and ensuring fair and safe practices. And it's very much top of mind for everyone, so to echo Paul's point, I think the workgroup just really came together to present principles that give practical realities of regulatory environments with aspirational best practices. Just wanting to make sure that we had some meaningful guidance, regardless of the jurisdiction or profession, because we're all being faced with it today.

Candice: I think the only piece that I would add to that is really just highlighting pieces that have already been stated in that this really is a global conversation. This is a global conversation within the CLEAR AI Working Group, but this is a global conversation as a general statement, and has been for a little bit. I've long said that AI is a humanity-impacting technology. I don't think that comes as a surprise or as a revolutionary statement to anyone. But the implications of that is that this is going to impact everyone, no matter where you are on the planet. And when I think about the principles that really sit at the heart of the work that has been done—professional competence, human-centered ethics, and trust through oversight—those really are consistent with some of the high-level ethical principles and considerations that have been put forward in the global community, when we're thinking about broad AI frameworks such as the OECD principles and the principles that are coming out of UNESCO as well. So I think that the AI working group with CLEAR is really well aligned with some of the conversations that have already been taking place on the global scale.

Line: Well, Candice, you mentioned those three commitments. The first is professional competence – AI should only be used within the scope of professional knowledge, skills, and expertise. The second is human-centered ethics – AI must support ethical standards and human judgement. And the third commitment is trust and oversight – AI systems must be explainable, auditable, and secure.
Why were these chosen as the foundation, and how do they work together to guide responsible AI use in regulation particularly?

Candice: Well, I think Kelly can probably speak a little bit more clearly to why those were specifically worded the way they were, but when I look at those, what I will say is I will again highlight the fact

that these are pretty consistent with the global frameworks that are in place. So thinking about human-centered ethics, that's really that human-in-the-loop kind of thing that a lot of folks talk about. And then trust through oversight, that really speaks to transparency and accountability. And I don't think you can find a single framework that has been developed, whether it's the OECD, or it's UNESCO, or any of the other frameworks, the broad global frameworks that are put out there, that don't refer to those sorts of pieces. And of course, professional competence really sits at the heart of regulation, but I think I'll hand it over to Kelly to speak a little bit further to what the process was in really choosing those.

Line: Perfect. Kelly?

Kelly: Yeah, I think you outlined that greatly, Candice, so I appreciate that. I would just simply add that, similar to what Candace said, these three commitments really reflect the core requirements for safe and reasonable, responsible AI in regulation. To her point, when we look to others, we want to learn from others as to what they're doing. This is kind of core to what we're seeing elsewhere as well, and it makes a lot of sense for regulation. It really forms a balanced approach. I think that balanced approach means that it empowers regulators to innovate responsibly while maintaining trust and confidence in the regulated communities. It's about thinking through a lens of enabling innovation, but there needs to be some clear practical guidance that includes that balanced approach of these three commitments. And I think we can easily adapt that to our regulatory world. I'm not sure if Paul wants to add anything to it, but for me, these are basic commitments and components and just core requirements when we think about the safe and reasonable use of AI.

Paul: Yeah, thanks, Kelly, and I think the challenge is the current global environment. One of the things that I said recently at the Chicago Annual Education Conference was it is a time of global change and challenge, and in that time of change, we have vastly different approaches on how different countries, for example, the United States, or even areas like the European Union, are approaching how this technology should be regulated and deployed. And even the legislative structures that are in place are shifting, where there's been some state measures put in in the US that have then been replaced by some federal decisions. And even with the EUAI Act, where they're now looking to maybe revisit elements of that, too.

So for us as regulators, there's a huge challenge in how can we regulate when we're seeing such wide pendulum swings globally. So, when looking at how we'd approach this, it was, how could we design something akin to a digital Hippocratic Oath? And what I mean by that is that the Hippocratic Oath has been around since about the 5th century, and it hasn't really changed over time, despite significant changes in technology from the pestle and mortar to nuclear medicine. So how could we design a principle-based approach to the deployment of artificial intelligence in professional regulation that would be mindful of licensing, of monitoring accreditation exams and complaints, but not constrain the regulators, act as a guardrail that you'd be allowed to flex. And in doing that, that's where we've reached these three strands. Professional competence, because that ensures decisions remain grounded and defensible; the human-centered ethics keeps the focus on individuals that are effective,

not just on the efficiencies this technology allows us to gain; and trust through oversight, because it protects the public confidence in regulatory outcomes. Together, I think AI strengthens regulation rather than erodes its authority, but to make sure that we keep that line, we need a principle-based approach to achieve it.

Line: Thanks, Paul. Now, looking at those three principles, in particular, one of the things that stood out was that no decisions affecting someone's rights or livelihood should ever be done by AI alone, that there should be a human component to that. How did the working group define the appropriate balance between what AI can assist and do, and human judgment?

Kelly: I'd love to kick this one off. As someone who has spent at least the last year engaging with numerous regulatory boards from different professions in different states and hearing a wide range of perspectives on this issue, I've seen and heard firsthand the dilemma that we're hearing about AI, leveraging AI capabilities, but then preserving human judgment. I think AI in a regulatory setting is meant to enhance government and operational efficiencies, not replace human oversight. So, when I think about this, I think about a regulator's job, and it's super important. A regulator is responsible for overseeing activities that determine whether someone can earn a living and work safely in the community. Therefore, I think we all felt pretty strongly that an individual's rights or livelihoods should never be made solely by a machine. So, in summary, I would say definitely embrace the use of AI to do more effectively, but it doesn't replace the duty to regulate. You still leave that to the trained professionals. And so I feel like with this approach, it really respects the expertise and accountability of the regulators, while still ensuring fairness for those affected, and maintains that public trust in the regulatory processes. I'm not sure if Candice has anything else to add, but I've seen this a lot, I've heard this a lot, so I wanted to share those insights based on numerous conversations I've been having.

Candice: Yeah, Kelly, I definitely would agree with everything that you said there, of course. I think the only piece that I would really add, and it's more of a highlight than anything else, is that at the end of the day, when we think about the decisions that we are tasked with, either as professionals, as regulated professionals, as regulators, or even from the AI realm in terms of decisions made by AI, somebody needs to be responsible for those outcomes. If something goes wrong, if a decision is made that is incorrect, whether it harms someone significantly or not, there needs to be somebody who's responsible, and an AI system simply cannot be responsible, cannot be held responsible for the outcomes of the decisions that it makes. So, in thinking about this from a regulatory perspective, how do we hold anybody accountable when they're delegating their decision-making authority to a system? So when we have folks that are involved in the decision-making, it's not just about ensuring that the decisions that are made are human-centered and are ethically and professionally grounded. It's also ensuring that should something go wrong, and things do go wrong, we do know that. With the very best of intentions, things sometimes go wrong. But there needs to be somebody who is held responsible, and you simply cannot hold an AI system responsible for its decisions. There has to be a human being that sits at the heart of that decision-making in order for there to be true accountability that is included with this process.

Kelly: That's what we'd like to call the human-in-the-loop model. And I know that's one piece, or one area that we defined when we worked on additional work through the work group. We created a glossary to kind of help the regulators understand some terms, and this one really reflects that human-in-the-loop model, using AI to enhance those efficiencies, but still having that ultimate final decision and oversight with an actual human.

Paul: I fully agree with all that. There's different types of AI, and I know we've been using the term AI, but there's different types of technologies there. For example, with diagnostic imaging, the work it's doing is phenomenal at how it's augmenting and supporting the highly trained healthcare medical professionals who are working in the field, and they're using it as a tool that would augment and support their performance. And also, as regulators, there's a lot of areas we can take it on and use it to enhance efficiencies. We are under increased pressure. The healthcare system, for example, we know there's going to be a workforce crisis in the future, based on WHO data. AI can be a significant changer in that, but regulation isn't just about analysis, it's about accountability. And AI can support consistency and surface-level decision-making and really do significant data mining in how we work. But responsibility that's been entrusted to us as regulators can't be delegated to a system, or the responsibility of the professionals we regulate. I think once the outcomes remain visible, and we understand it better, there's a lot of measures that we can put in place as regulators to get those reassurances. We can do sandbox testing, we can audit the systems once there's outputs there, but ultimately, putting in place a measure like this, and maintaining human oversight protects the public, but also protects us as regulators to make sure that we're making the right decisions.

Line: Yeah, excellent points from everyone. Looking at it from the regulators that maybe don't have a lot of deep technical expertise, specifically as we look at AI with regard to professional competence, how can we ensure that both the regulators and the licensees understand the purposes and limits and risks of these AI tools? What does that look like in practice?

Candice: I think that speaks to developing professional competence, which is the first of the three principles that we've outlined as sitting at the heart of the document that we're talking about. Because at the end of the day, AI is not going away. I think that's not a revolutionary statement. At this stage of the game, we all recognize that AI is here to stay, and when we think about regulated professions, often what we're thinking about as a component of that is ongoing professional development. So really integrating the concept of lifelong learning, including tech learning, is going to be something that's going to be incredibly important. It's going to be incredibly important not just for practitioners and professionals. It's going to be incredibly important for regulators as well. Because you don't want to be walking blind into an area that has pretty serious consequences if it's mismanaged. Now, does this mean that you need to go and get a data science degree and learn how to code and do all of those things? No, probably not. But there's a lot of resources out on the internet that can really help you with being able to understand exactly what's going on, and one of the documents that the AI Working Group actually came out with was an AI glossary as well, just a glossary of terms, to be able to bring everyone together and understand the jargon, the words that

are being used when folks who are in the tech sphere are talking about this technology. Having that common language or an understanding of a common language is incredibly important. And I'll leave it there and let Kelly and Paul jump in and add to this.

Paul: It's a really interesting space, and I agree with Candice's point. We are not to become technologists as regulators. The challenge here is an over-reliance on AI will create blind spots. If you operate outside your skill set, or AI allows you to reach a level where you as a professional, or you as a regulator, are operating significantly outside your skill set, you effectively begin to automate your decision-making process. Biases will appear, or biases may increase. And the errors will go undetected, and that will have negative consequences. Competence ensures that AI strengthens practice rather than weakening it. And we talk about that in the paper a bit. For me, it's understanding at a very basic term. Competence is understanding what the tool is designed to do and not designed to do, and being able to ask informed questions, whether it's from the vendor when you procure it, to say, what is the scope? How is this set up? How is this structured? And when should I use it, but also when I should not use it? So, as part of the regulation of how we are using this, or us as regulators are using this, sometimes restraint in its use is a form of competence.

Line: Awesome. So, trust and oversight require explainability, auditing, and security. How can regulators meaningfully assess AI systems, especially when vendors may offer black box tools, or when agencies have limited resources? What is the best way for them to assess these things?

Kelly: I think for regulators, it's super important to ask questions, as you just heard Paul say. Ask questions, be curious but cautious, ask what a system can do and what it can't do, and understand those fundamental pieces. I think it's super easy today to do research, to keep up, and learn on your own, but also learn and work with others. So regulators may consider things like working with other agencies, forming their own work groups. That could include other executive directors in their particular jurisdiction, or it might also include outside experts. But I think the bottom line for me and my perspective on this question is very much that regulators are not alone in this conversation. And so when they think about assessing these systems, it's always: be curious, be cautious, and also know others are in that same journey along with you, and what can you learn from them, or where can you lean in to help, or not reinvent the wheel for other purposes. I'm not sure if Paul or Candice have anything else to offer or respond to the question.

Paul: Yeah, I'd fully agree with you, Kelly. Perfect transparency is so rare, but in this space, meaningful oversight is just so essential. We sit in a world where different jurisdictions have different legislative requirements that would be mandated against us. But at a global level, when we look at this, and when you boil it down to these principles, regulators should look at explainability proportional to the risk. And what does that mean? That means the auditability, documentation, regulatory sandboxing, annual auditing, looking at how the use of the system or how humans using the system may be injecting conscious or unconscious bias, having clean data sets when these systems are constructed, and security. And all of those requirements can’t be optional. Those expectations should be hardwired into procurement processes, into regular governance processes, not added later, not reactive, but

proactively addressed. And finally, to Kelly's point, collaboration here between regulators is critical. We are in shifting sands in this regulatory environment. Sharing that learning across regulators reduces that burden, and that's why CLEAR has been so powerful in stepping into this space, because we've provided a forum where we can share our strengths and our weaknesses and our learnings through the forums that have been created.

Kelly: I would also add just one other thing as well, which, Paul, you made me reflect on it for a moment, but I also think flexibility is important, because how they may assess AI systems today may be different two years from today, just with the evolution of technology. I think a lot of this is about we only know so much, but we need to continue to adapt to what may be to come.

Line: Excellent, thanks, Kelly. From the regulator perspective of things, they have to apply a lot of these principles to real-life things in their world, like exams, licensing, investigations, even just dealing with complaints, triaging complaints in that matter. From the working group, were there areas where the working group saw AI as more or less appropriate for some of these things? And what would be the practical steps that regulators would need to take to incorporate these?

Paul: Yep. I can kick that one off. What I would say, Line, is you start with low-risk, high-value applications. Initially, if you've got clean data sets in an area where you're looking to detect patterns in your case profiles, or country of origin, or area educational institution, you can map the data sets and use that to inform your own internal policies. That's typically one quick win. Administrative triage, data quality checks, pattern detection, they're all really sensible entry points that would probably be low risk. I know there are some regulators that are looking at working in tandem with complaints triaging and using publicly available models to support that. For me, my nervousness would be about where the decisions directly affect individuals, as opposed to informing policy, and I think we should move cautiously in those. I think it's testing systems in appropriate environments before scaling, so really rigorous sandbox testing, and we could do that in partnership with the vendors who are deploying these systems. Maybe it's starting to look at where there are appeal pathways where AI informs the decision, and we need to be transparent with the public on that. So, very simple first steps into this, but you'd rather step forward cautiously and then ramp up at scale, because it's those early implementations that you will learn so much from, whether it's AI literacy from your staff, AI literacy with professionals, or your audit and control programs in case something goes wrong.

Kelly: I can share a really simple way from a vendor-partner perspective and how we are working with boards for the same reason. We are currently working with partners to roll out the use of AI in a documentation verification tool to ensure it is a legitimate continuing education document being submitted to the board. Again, very low risk, no final decisions being made, and this doesn't take away from a board's duty for oversight, for instance, in an audit to double-check that information, but it saves them the time from having to look at maybe initially scanning 24 documents that have been submitted and then only having to look at the ones that identify areas of possible issue. So it simplifies what they're doing, but in that scenario, it's all about creating efficiency for the regulator. Very much to Paul's point, low risk. It doesn't deal with final decision making, but it enhances operational

efficiency, because using this AI documentation scanning and CE submission review, I can probably look at more licensees or more documents than I could before if I was just doing it manually.

Line: Excellent. Well, the principles acknowledge significant implementation challenges, from rapid technological change to resource constraints. What strategies or supports, whether that be training or collaboration or model testing, transparency, as Paul mentioned, do you see as the most essential for overcoming these challenges? Paul, I'd like to start with you on that one.

Paul: I think the biggest challenge you're going to face is awareness of the technology and competency from your own individual users if this is being onboarded within your organization. Training builds confidence and shares vocabulary, and there are resources that CLEAR has, like the Glossary of Terms Candice referenced, that can support that. Collaboration prevents mistakes, and it's working collectively with other regulators to see what learnings they have had from onboarding. There's no single solution that will remove the risk when you're developing these systems, but both internally and externally, transparency sustains trust when systems are questioned. From a governance perspective, it's to be very transparent that we are using this system, and these are the outputs that are AI-generated or assisted with AI. Progress in this space comes from discipline and not speed. What I would say to regulators is that, when you are onboarding these systems, do not rush. Check your data, triple check your data, make sure that you're not putting garbage in, because you don't want to be in a position six months down the line where you're generating outputs from systems that were trained on poor data sets. Have very clean, structured data going in, and your vendor would probably work with you on that. Rigorously sandbox it, stress test it, bring in red teams, make sure your people challenge it in the best way possible. I read a case recently about a regulator who last year onboarded a system that wasn't adequately trained on their website and was giving incorrect information. You don't want to be in that position. You don’t want to affect public confidence in how you’re working as a regulator by rushing into this. So, rigorously test it and then deploy it. But it doesn't stop there. You have to have really robust governance systems in place, put in an audit and inspection program, and make sure those safety nets are there so that when you go back and test it in X period of time, you know the outputs are consistent and that your use of the system isn't changing the outputs in a negative way.

Kelly: I think those are all excellent call-outs, Paul, and it reinforces the idea of continuous monitoring and evaluation. What you start with may not be what you end with or how you set it up. I would also add that implementing new technology and new enhancements often comes with scrutiny from the community. There's a lot of uncertainty about how you're using it or what you're doing with it. When we think about training and collaboration, that's both internal and external, and being able to really have transparent processes. Maybe you put documents out explaining how the agency is using AI, how it's being applied, managed, and overseen, to create confidence in the community. It's one thing for internal folks to understand and know it, but the community will be worried about how you're using it, so I think just as much as doing it internally, you also have to make sure you’re prepared to speak about this publicly and speak to how you’re using it for these processes and what it does or doesn't mean for that community.

Candice: I'm going to piggyback off what Kelly just said and speak around the informed consent piece. Coming from a context where this is a significant factor, particularly in Canada, there's a lot of informing folks that AI is being used. You can walk into doctors' offices or other practitioners' offices, and maybe there's a sign on the door, or your doctor very quickly at the beginning says, I'm using AI for our appointment today, do you consent? That is about as far from informed consent as you can possibly get, and I think when we start asking folks to ask for informed consent, what it actually does is it puts professionals—and I've used doctors here, but I'm thinking about, I work in mental health care—so I'm thinking about all folks that work in healthcare, whether it's physical healthcare or mental health care. When you're put in the position of actually having to get informed consent, where you have to explain important components of the system that you're using, you're putting the practitioner in the position of needing to understand in order to be able to explain to the client, or the patient, or however you're framing it, explain to them and be able to get that informed consent. And that's one of the pieces that I'm seeing is missing in a very big way.

AI is very exciting, particularly when I think about the applications that are really happening in high speed right now. It has a lot to do with the tasks that none of us really want to do, which are administrative tasks. But that doesn't change our obligation to be able to ensure that we're building trust with the public, as Kelly referenced, and in my opinion, part of that is getting informed consent from folks. And informed consent means that you, as a practitioner, need to be able to understand the technology well enough to be able to talk about how it works, where your data is stored, what data is collected, how long it's stored for, is it anonymized, is it not, is it used for training? The list goes on. But when you're in the position of being able to understand that, and being able to provide the foundation for asking for informed consent, you're actually creating a set of circumstances that not only builds trust and confidence in the public, but builds your own capacity and your own understanding as a practitioner.

Paul: Candice, it's a great point, and I think it's going to be one that's really going to come to the forefront in the next couple of years, because we are in the Model T era of AI. This is the baby AI stage; we're in its infancy. It's really going to develop. I was actually having this conversation about two weeks ago with a medical doctor who works in a hospital, and we were talking about the challenge of informed consent, because it assumes a linear treatment pathway as opposed to a dynamic one. The GP may explain it and give consent, but if the medical scientists use it, or the diagnostic imaging technician, because it's hardwired into the hardware, it can cause challenges. It's something that probably we're all going to need to explore. I know there's different approaches—the EU has a very different take on it than different areas in North America, and I know Australia has a different opinion on it. So it's an area I think we're probably going to come back and have those more complicated, wider conversations on it. That comes back to why we have a principle-based approach, because there are so many different opinions in the world that we're trying to come up with a high-level guardrail that regulators can really stick to and align to.

Line: Paul, you mentioned we're in this Model T era of AI, its infancy. With that being said, we know

it's going to change, we know it's going to improve, it's going to get better. Do you expect that these principles that you guys have developed to have some evolution to them as well? We'll finish up with a two-parter here: do you intend for the principles to have some room for evolving, and what does successful adoption look like for regulators over the next several years in this infancy?

Paul: What I would say is the principles are designed to ideally be at a high level enough that they can flex as the technology changes and grows. However, if the principles need to evolve as practices mature and different technology or consequences of this technology emerge, we can certainly look at it. I think competences should remain with the human. There needs to be a human in the decision pathway. I don't think that should change. That's a valuable learning, because if we don't use that, there's a significant increase in patient risk, potentially, or an erosion in learned skill sets. Certainly, there is a counter-argument that says at a point five or six years in the future, if you've got AGIs, when you risk assess it, what is more competent, an individual or the piece of technology? I would argue an individual, but the counterargument put forward to me recently was, if you've got a calculator or an accountant, which one would you trust more? You really are getting into very difficult conversations. At this stage, I still would stand over that it's important to keep a human or a medical professional in the loop, because at least then, if there is a mistake in the technology, you can rely on it. That's why our accountants do our accounts, and it's not me at home with my calculator. The principles will evolve as needed, but they should, for the foreseeable future, be fit for purpose. What success would look like is, as with all regulators, having a boring day where nothing happens. I was saying this recently at the conference last month in New Zealand: how do regulators measure success? It's a happy day and nothing bad happens, because that's when a regulator is doing its job correctly. For me, success is where regulators adopt these guardrails, the AI systems work smoothly, and there are no negative consequences.

Kelly: How often do we see those days? Probably not enough, I think as a former regulator. I would echo everything that Paul has shared. Specifically related to the continental U.S., the policy and legislation around these technologies is also going to evolve, so with that, some principles may shift. But everything we're saying here is core common sense. It's do what's right and good. Let's make sure it's fair. Let's make sure there's transparency and accountability. These are all core to the regulatory board duty and mission. Some things may not change. The principles might look a little bit different as they evolve alongside technology, but all of this is core to what we do in everyday practice as a regulator. The only other comment I would make – it'll be interesting thinking about President Trump's executive order that he issued on December 11th, which encouraged a national framework for AI in the United States, and the implications that may have on state policies that are currently in place or being considered. I know this session, we're looking at thousands of AI bills being filed in the U.S. in different jurisdictions. We will know more as it comes, but to Paul's point, the human-in-the-loop model and all of it make sense for right now. These principles are a lot of common sense when we think about being the best regulator and how we're embracing technology in general. But as with everything and anything, we always have to be prepared to remain flexible and think future-forward as needed.

Candice: Absolutely. I couldn't agree more with everything that's been said. The only valuable thing I can add is that the pieces we've been talking about really are specific to professions and sectors that are regulated. For folks listening to this podcast, it's important to remember that the AI conversation is diverse and multifaceted, and it really is sector-specific. One of the things we've talked about in our conversation today at more than one point is the human-in-the-loop piece and how important that is. I absolutely agree with everything that's been said here today, certainly as AI currently stands today. I can't imagine a circumstance yet where there wouldn't be a human in the loop, but certainly in regulated professions, that's incredibly important. This isn't the conversation that's necessarily happening everywhere else. A couple of months ago, a really interesting journal article by Jovana Davidovic came out, called Rethinking Human Roles in AI Warfare, and it really challenges the concept of human in the loop and whether that is a good idea. Those are military applications, so I want to highlight for folks listening, who might be listening to other AI conversations that are happening elsewhere, that the kinds of things being outlined by the AI working group are really related to human-serving and regulated sectors, and not to get caught up in the fray of different conversations happening outside in different sectors.

Kelly: I would just add there, too, Candice, that you reminded me of a conversation I was in last week. The American Telemedicine Association was talking about doctor-out-of-the-loop models when it comes to telemedicine moving forward. To your point, and for all that's been said, and to those listeners, it's always good to be curious but cautious, and to keep a pulse on what's happening and how it may or may not apply to you. Having trust and confidence in what you're currently doing, and thinking about your core regulatory duties and how you'll carry those out on a daily basis, with or without technology, and the other stuff will all fall into place eventually. There is a lot unknown at this point, and it will be interesting to continue to watch.

Line: Excellent. Again, the work on the CLEAR principles is laying the essential groundwork for how regulators can use AI responsibly and with integrity, especially as this technology continues to evolve. The CLEAR Principles for Ethical and Effective AI in Professional Regulation, as well as the AI Glossary of Terms that was mentioned, can be found on the CLEAR website under the Resources menu. I want to thank you all—Paul, Kelly, and Candice—for sharing the working group's initiatives with us today. Thank you again for being a part of this.

Guests: Thank you, Line. Thank you so much.

Line: It's definitely been a pleasure, and we'd love to continue this conversation with our members on the CLEAR Regulatory Network. This podcast episode will be posted there, along with some questions for our members to consider. I've got a couple I'll go ahead and give you now. Which of the CLEAR principles—professional competency, human-centered ethics, or trust through oversight—do you think would be the most challenging for regulators to implement, and why? The second question is, how do you see AI impacting your own regulatory work in the next few years, and what supports or safeguards would help you feel confident using AI tools responsibly? We greatly appreciate and thank our members for their discussion and feedback in the CLEAR Regulatory Network. If you haven't

already, we invite and encourage you to join and take part in the online discussions.

I also want to thank our listeners for tuning in for this episode. We'll be back with another episode of Regulation Matters: a CLEAR conversation very soon. If you're new to the CLEAR podcast, please subscribe. You can find us on Podbean or any of your favorite podcast services. If you've enjoyed this episode, please leave a rating or comment in the app. Your reviews help us improve our ranking and make it easier for new listeners to find us. Feel free to visit our website at https://www.clearhq.org for additional resources and a calendar of upcoming programs and events. Finally, I'd like to thank our CLEAR staff, specifically Stephanie Thompson, our content coordinator and editor for this program. Once again, I'm Line Dempsey, and I hope to be speaking to you again very soon.

The audio version of this podcast episode is available at https://podcast.clearhq.org/e/AI_principles/.

- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 2/20/2026

SUBMITTED BY: Executive Director

TITLE: Psilocybin Draft Legislation

INTRODUCTION TO THE TOPIC:

This draft legislation was recently provided to the Board by the Chair of the Psychedelic Medicine Task Force. The task force was created in the 2023 legislative session.

BOARD ACTION REQUESTED:

- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 2/20/2026

SUBMITTED BY: Executive Director

TITLE: Executive Director's Report

INTRODUCTION TO THE TOPIC:

The Executive Director Report communicates, in advance, information that brings board members up to date on what has occurred since the last board meeting and is intended to lead to engagement and interaction at the next board meeting. The Executive Director Report seeks to offer reminders to board members on upcoming commitments, relevant dates and events, and to raise issues for board members to address during the board meeting. The Executive Director Report is also intended to give board members information that is useful in their role as board members and in stakeholder outreach.

BOARD ACTION REQUESTED:

- MINNESOTA BOARD OF PSYCHOLOGY

DATE: 2/20/2026

SUBMITTED BY: Licensure Specialist

TITLE: Board Administrative Terminations

INTRODUCTION TO THE TOPIC:

The Board shall terminate the license of a licensee whose license renewal is at least 60 days overdue and to whom notification has been sent as provided in the administrative rules. Failure of a licensee to receive notice is not grounds for later challenge of the termination.

Licensees are provided several opportunities to renew the license prior to Board termination. Licensees are sent a notice within 30 days after the renewal date when they have not renewed the license. This letter is sent via certified mail to the last known address of the licensee in the file of the board. This notifies the licensee that the license renewal is overdue and that failure to pay the current renewal fee and the current late fee ($250.00) within 60 days after the renewal date will result in termination of the license. A second notice is sent to the licensee at least seven days before a board meeting (which occurs 60 days or more after the renewal date). Minn. R. 7200.3510.

BOARD ACTION REQUESTED: