

Professional Psychology: Research and Practice

Ethical Considerations of Artificial Intelligence (AI) in Psychological Assessment

Tristan T. Herring and Paul B. Ingram

Online First Publication, April 13, 2026. <https://dx.doi.org/10.1037/pro0000678>

CITATION

Herring, T. T., & Ingram, P. B. (2026). Ethical considerations of artificial intelligence (AI) in psychological assessment. *Professional Psychology: Research and Practice*. Advance online publication. <https://dx.doi.org/10.1037/pro0000678>

Ethical Considerations of Artificial Intelligence (AI) in Psychological Assessment

Tristan T. Herring and Paul B. Ingram
Department of Psychological Sciences, Texas Tech University

Artificial intelligence (AI), an interdisciplinary field designed to emulate human intelligence, has advanced rapidly in recent years. AI and generative large language models have unique capabilities that, while promising, raise ethical concerns and challenges to practice, including issues related to confidentiality, transparency, test bias, and data security. AI has already been integrated into assessment processes, including test development, administration, scoring, and report writing, introducing unprecedented opportunities and risks. Psychologists must navigate these complexities to uphold ethical principles and ensure AI's responsible use in advancing the field. This article discusses several ethical concerns that may emerge as the field increasingly incorporates this emerging technology. This review highlights the urgent need for transparent, explainable AI systems, robust ethical guidelines, and empirical research guiding AI's applications in the future of psychological assessment.

Public Significance Statement


This article details ethical considerations and challenges related to psychological assessment, which health service psychologists are likely to face as artificial intelligence rapidly advances.

Keywords: artificial intelligence, assessment, ethics, large language models

Artificial intelligence (AI), an interdisciplinary field that aims to mimic human intelligence, has recently surged in capability and popularity (Aschenbrenner, 2024). AI is the newest phase of computer-assisted psychological assessment, including evoking similar issues of ethics, validity, and accessibility seen historically in other technologies (see Fowler, 1985). AI also poses unique ethical risks based on its interactive and generative nature (Farmer et al., 2025), such as serving as a personalized coach for assessments in academic and professional settings (Dien, 2023; Plotkina & Sri Ramalu, 2024). AI also poses similar risks as other communicative mediums (e.g., online articles), including spreading misinformation or perpetuating harmful stereotypes (Yiu et al., 2023). AI methods include supervised and unsupervised machine learning (both of which are abbreviated as ML in the literature), which do not require human input at every decision-making juncture along its development (e.g., include or not include a variable in the model; Russell & Norvig, 2010). In short, AI is unique in its novelty and emerging impacts to computer-involved assessment, warranting unique attention to how to face the ethical challenges it presents.

The American Psychological Association's (APA, 2017) ethics code requires that psychologists remain within their boundaries of competency and base assessments on empirical, contextual, and cultural factors. Moreover, assessment training is a required competency for APA health service psychology (e.g., clinical, counseling, or school psychology) accreditation (APA, 2015). There has been a rapid growth of AI as a technology (Chen et al., 2023), leading to numerous position statements and recommendations (APA, 2024; White House Office of Science and Technology Policy, 2022). However, there are no AI guidelines focused on psychological assessment competencies from leading assessment organizations. The APA is gathering public comments on an ethics code revision, which hopefully will address many AI-related concerns. Despite efforts to generate rules and protections, our safeguards are not sufficiently in place at present, and AI technology continues to advance and is even more rapidly adopted for use. One study, for instance, found that 67% of health care leaders report already incorporating AI into their medical systems (Stewart, 2021). Likewise, schools around the country are rapidly

A. Jordan Wright served as action editor.

Tristan T. Herring  <https://orcid.org/0000-0001-7468-794X>

Paul B. Ingram  <https://orcid.org/0000-0002-5409-4896>

TRISTAN T. HERRING received a BA in psychology from and is currently pursuing a PhD in counseling psychology at Texas Tech University. His research interests include psychological assessment and validity, particularly within neuropsychological contexts, and the promotion of cognitive functioning.

PAUL B. INGRAM holds a PhD in counseling psychology from the University of Kansas. He is an associate professor of counseling psychology with at Texas Tech University, where he also holds faculty affiliations in Departments of

Human Development and Family Sciences as well as Women and Gender Studies. His research interests encompass psychological assessment, particularly within military and trauma-exposed populations, with an emphasis on invalid response patterns using broadband personality measures.

Tristan T. Herring played a lead role in conceptualization and writing—original draft. Paul B. Ingram played a lead role in writing—review and editing and a supporting role in conceptualization.

CORRESPONDENCE CONCERNING THIS ARTICLE should be addressed to Paul B. Ingram, Department of Psychological Sciences, Texas Tech University, Psychological Sciences Building, 2700 18th Street, Lubbock, TX 79410, United States. Email: pbingram@gmail.com

having to respond to the use of AI by students, who hold differing views of what is considered ethical (Lee et al., 2024). Moreover, AI has also already been incorporated in psychological assessment development (Kjell et al., 2024), selection (Weakley et al., 2015), administration (Chandler et al., 2020), transcription (Natoli, 2025), scoring and interpretation (Jiménez-Mesa et al., 2023), and report writing (e.g., parchment-AI agent).

This article reviews uses (and potential future uses) of AI in psychological assessment and contextualizes those areas according to the principles from the APA Ethical Guidelines. Examples may reference Chatbots (ambiguous) or may be explicit to ChatGPT. ChatGPT is referenced explicitly as research shows ChatGPT is often, at present, superior in relevant responses, relative to other models (Lavigne et al., 2024).

AI: Human or Program?

AI is not a new field. Rather, it was developed by incorporating principles from psychology and cognitive neuroscience with a goal of achieving *general intelligence*.¹ The methods used with different AI models have changed over time. For instance, early models used visual perception and associative learning as their basis (see Stone, 2024). More recently, AI has shifted to use large language model (LLM) approaches (e.g., AI Chatbots such as ChatGPT), addressing earlier limitations in associative models (namely, complex identification or discriminant learning). In contrast with associative models, LLM numerically code words and generate predicted responses using an unsupervised learning approach (Bommasani et al., 2022). Unsupervised learning means, roughly, that AI is allowed to identify patterns without being given examples of what to look for or directions about how to find those patterns. These trained models, those that have processed their input data and had the time to sort out the patterns, have made remarkable advances in only a handful of years. In short, generative LLM learn in a “black box,” evading human understanding because of their unsupervised approaches (Bommasani et al., 2022; Stone, 2024). Further, this methodology continues to grow beyond human comprehension as additional milestones (e.g., artificial general intelligence) remain unpredictable, but speculatively close, due to exponential technology growth (Aschenbrenner, 2024).

Inherent in AI is the need to mimic human intelligence—that is, learning from its environment (i.e., training database) and applying those lessons to prompts (e.g., asking a Chatbot to write a poem). AI does not perform in a uniform and consistent manner (e.g., the same prompt can get different results), nor does it require human instruction for every action. For example, AI can receive one request for information and independently search the internet for results, whereas traditional methods require human decision making when selecting a search tool, search terms, and articles for review. This independence sets AI apart as a tool and as a risk. AI models have also shown diverse personality traits and values dependent on the training language and data used, adding potential humanlike bias in their use (Pellert et al., 2024). Consequently, AI models operate in an ethical gray area wherein no one user or organization is responsible for their conduct (Keskinbora, 2019; Naik et al., 2022). This lack of oversight creates a liability to the practice of professional psychology, raising several ethical concerns like competence, appropriate delegation, and consultation (APA, 2017).

Depending on the extent to which AI continues growing,² ethical issues are likely to arise surrounding AI’s ability to act as a consultant or other entity, such as one to which assessment-related tasks (e.g., report writing) may be delegated (see Fowler, 1985). Already, AI is different from traditional forms of online information collation approaches (e.g., Google searches) because these programs act as personal coaches working through problems step by step with the individual. For instance, you can ask an AI to explain something differently if you do not understand its initial response, and it will explain things to you more clearly—learning from and adapting its interactions with you to do so. This adaptive process decreases the burden of identifying and gathering online information relative to less sophisticated methods of online searching (Yiu et al., 2023). Given these unique complexities and rapid model innovation, developing ethical guidelines for LLM is akin to adopting education strategies to a student with an undergraduate skillset 1 week and a doctoral level skillset the next. Rapid growth in model capacity may require rapid changes. Such unpredictable, exponential growth and obfuscated inner workings of AI models make specific recommendations difficult to solidify (Aschenbrenner, 2024; Bommasani et al., 2022).

Health service psychologists must base their work on professional knowledge and assume responsibility for scoring programs (APA, 2017). Clinicians conducting assessments are then at the mercy of the model and its areas of competency. Such responsibility necessitates a baseline understanding of the assessment administered and the AI model used to (a) justify their use of a particular model and (b) monitor and evaluate a model’s decision-making ability. Concerningly, not even AI researchers know how LLM learn or apply data (Bommasani et al., 2022; Stone, 2024). To address these concerns and promote ethical use of this new technology, an AI model would need to be transparent and explainable on a general level so that psychologists may understand how it reached its conclusions and what weight was given to distinct aspects of evidence (Chandler et al., 2019; Naik et al., 2022).

Attempts to grapple with the complex ethics of AI are ongoing (Chandler et al., 2019; Jobin et al., 2019; White House Office of Science and Technology Policy, 2022), including by the APA (e.g., APA, 2024). Across industries and governments, there are four common ethical principles that have emerged: *transparency, justice and fairness, nonmaleficence, responsibility, and privacy* (Jobin et al., 2019). These four common ethical considerations mirror those outlined in the APA general ethical principles and may provide a framework for applying APA’s ethical code to AI’s role in assessment (APA, 2017). For example, the U.S. government identified the need for safe and effective systems, algorithmic discrimination protections, data privacy, notice and explanation, and human alternatives when using with AI (White House Office of Science and Technology Policy, 2022). In addition, explainability, transparency, and generalizability of AI models are required for trust among psychiatrists and other health professionals (Chandler et al., 2019). By looking at the APA general principles and standards

¹ General intelligence, also referred to as artificial general intelligence, refers to a “universal algorithm for learning and acting in any environment” (Russell & Norvig, 2010, p. 27), that is, an AI program intelligent enough to adapt to an environment as sophisticatedly as humans.

² Expert estimates for achieving human-level AI range from 5 to 40 years (Aschenbrenner, 2024; Roser, 2023).

outlined in other fields, initial standards for AI use in psychological assessment can be derived, which is the purpose of this article.

Developing Standardized Assessment Instruments

AI will likely play an increasing role in the development of psychological assessments and the advancement of assessment practice. Indeed, AI's strengths are already being deployed in various ways to assessment practices (e.g., using linguistic patterns to predict schizophrenia; Kjell et al., 2024), and these efforts have produced better diagnostic practices than psychiatric residents (Bang et al., 2025). The nature and rapid development of LLM create a unique environment for test development. For instance, AI could easily develop items in real time, identify response patterns, check for subgroup bias, as well as translate existing measures, removing various scale development and refinement stages (e.g., expert item review, etc.). Such potentials are worth consideration and pursuit due to increasing AI's accessibility and lowering cost, given their assessment effectiveness (Rollwage et al., 2023). Doing so may allow us to increasingly "give psychology away"³ (cf. Buchanan, 2002; Miller, 1969) through decapitalization of assessment (see Gibby & Zickar, 2008 for a history). Removing barriers to instrument development and improved client access seems likely to create a more open, but still secure, era of "evidence-based psychological assessment" (Wright et al., 2022). On the system backend, for instance, an AI assessment agent may examine items for differential functioning and could adjust scoring because of explicit instruction and training. Likewise, the A.I. assessment could generate additional items to "test out" and ensure the salience of known biases factors is less pronounced in scoring, self-propelling its own refinement while maintaining item replicability. A similar method of item generation and validation was long used for the Graduate Record Exam, a standard entrance test used in graduate school admission processes. Such a practice would provide an opportunity to establish diversity-centered assessment practices by adaptively examining known cultural influences on testing (e.g., Wright, 2024). In both examples, these capacities are consistent with those made possible by generative AI (see Galatzer-Levy et al., 2026). Per APA Ethics Code Standard 9.05, "Psychologists who develop tests and other assessment techniques use appropriate psychometric procedures and current scientific or professional knowledge for test design, standardization, validation, reduction or elimination of bias, and recommendations for use" (APA, 2017).

Each AI model operates under different training data (e.g., learned understanding of what it should do and how it should do it), leading the AI models to behave differently and produce different privacy concerns (Chandler et al., 2019). Different versions of the same model (ChatGPT 4o vs. ChatGPT 5), as well as competing models (ChatGPT vs. Google Gemini), have different training data and operational capacity. Such differences in training data create differences in AI model learning, which means that AI models change as technological breakthroughs occur. These changes sometimes worsen model accuracy (e.g., identifying prime numbers) and responsiveness to user prompts (e.g., openness to sensitive prompts; Chen et al., 2023). For clinicians to psychometrically assess an AI-based test, they would need to assess the model as well as the specific testing method used. This research would require a transparent model and an understanding of the program, even when the exact behavioral change of these models is still unknown (Chandler et al., 2019; Chen

et al., 2023). The rapid development of AI and related technology underscores the need for technological competency in health service psychology (Weisenmuller & Luzier, 2023).

AI programs can assess in ways humans cannot (e.g., assessing using linguistic patterns). While this capability presents a unique benefit of AI and LLM, it also creates a paradox. Clinicians need to understand a program's methodology well enough to audit its decision-making capacity. An understanding sufficient to audit an AI program's methods undermines the benefit of, and is in contrast with, methods being beyond human capability or consciousness. Specifically, traditional actuarial programs and human insight would be sufficient if humans could proficiently understand and employ a model's methods. For example, language-based assessments could code client speech far beyond human capabilities (Kjell et al., 2024). Such models operating outside human ability would be difficult, if not impossible, to audit (Aschenbrenner, 2024). If humans could reliably detect such subtle language patterns, AI might be unnecessary, as human judgment and traditional software could do the job without concern.

The speed at which AI models evolve unveils an additional ethical concern. Per APA Ethics Code Standard 9.08 (APA, 2017), assessment psychologists ought not base opinions on obsolete measures. This ethical concern is relevant when selecting between editions of a test (e.g., WAIS-IV vs. WAIS-V). Clinicians can easily identify when newer editions of standardized assessment instruments release and audit the changes relevant to their contextual needs. LLM, however, undergo continuous updates, even within the same published model update (e.g., ChatGPT 5). The same versions of ChatGPT, for example, differ in math and coding even within 3 months (Chen et al., 2023), and these changes did not coincide with official model updates. Psychologists are guided to use their judgment about test adoption with no clear absolute benchmarks (Bush et al., 2018). Clinicians must also understand the benefits that a new AI model may bring to justify its professional use. The knowledge that psychologists have about AI and AI models will be directly relevant to their ability to compare the assessment capacity of the newer model. However, AI models add a level of technological complexity that may elude clinicians—as they already elude the programmers designing them (Bommasani et al., 2022).

We were unable to identify any studies or reviews that encapsulate the perspectives of assessment psychologists about AI and how it impacts their jobs now or may in the future. Such data may be a useful benchmark as the profession considers general model knowledge and technology competency needs related to the field.

³ Giving psychology away" is framed not as a concrete outcome but as an ethical orientation for how the field might use its accumulated knowledge and power to address problems. From this perspective, the primary beneficiaries of psychological knowledge should be those who receive and deliver psychological services directly, rather than intermediary corporate systems that package such knowledge as products and are vulnerable to distortion through automation, market pressures, or efficiency mandates. The goal is a redistribution of access and benefit such that our knowledge and science may be applied more equitably. Our shared perspective is that AI introduces a genuinely novel capacity to expand the validity and utility of psychological assessment already without technologies outside our grasp (e.g., artificial general intelligence). The central challenge we face as a field is ensuring that these developments remain grounded in scientific rigor and aligned with the profession's ethical and professional responsibilities, rather than allowing technological efficiency to supplant disciplinary standards or theoretical grounding.

Further complicating the potential, and probable in the authors' view, technology–knowledge gap, AI is regularly and increasingly updating in its capacity to perform tasks (Aschenbrenner, 2024; Roser, 2023). Updating a model requires additional training, built on more advanced technological hardware (e.g., Graphical Processing Units [GPUs] primarily) and information (e.g., data, sometimes called training data sets). New models have the potential to offer new, not merely incremental, capacities.

In addition, standardized psychological instruments (e.g., WAIS, MMPI, Rorschach) often rely on normative data as a means of contextualizing client responses. Akin to norms within a psychological assessment are the data on which an AI model is trained (Russell & Norvig, 2010). Biased data can lead to interpretive problems as well as discriminatory practices, which are already occurring (Hacker, 2021). AI models can also form stereotypes based on simplified or unrepresentative data. These threats raise concerns, particularly in cases where a model assesses someone in a population with which the LLM is unfamiliar (see APA Ethics Code Standard 9.02; APA, 2017). Should an assessment dependent on AI emerge, training data need to be accurate and representative of the target population. While true of other tests generally, failure to do so with AI may result in LLM with differing values and traits from those assessed, as well as discrimination based on dialectical patterns (Hofmann et al., 2024). At present, AI systems are likely to develop interpretive and speculative bias (Pellert et al., 2024), and this bias may lead to diagnostic and predictive inaccuracies. Thus, researchers and clinicians developing and using AI in assessment practice must closely monitor these trends, lest their testing decisions may no longer align with our evaluative needs and ethical principles.

Selecting and/or Administering Assessments

Current administrative uses for AI in psychological assessment have primarily occurred through remote or self-administration (Chandler et al., 2020). These models occur via smartphone applications and allow users to receive assessments without concurrent human oversight (see Fiske et al., 2019 for analogous examples in therapy). As AI models gain more autonomy and present as more humanlike, AI psychometrists may have the capacity to independently develop, select, and administer assessments. One popular benchmark for AI is the “Turing Test,” which states that intelligence is achieved when a model's responses to written questions are indistinguishable from human responses (see Turing, 1950). The exponential growth of generative LLM suggests that these and other programs may soon “pass” the “Turing Test” (OpenAI, 2023). Indeed, some have already claimed that modern AI system capacities have surpassed this benchmark (Mei et al., 2024). However, the “Turing Test,” despite its name, is a vague goal and not an explicit measurement method. This ambiguity will likely result in waves of reported success as “Turing Test” benchmarks change or are abandoned due to criticisms over these and other such flaws (Moor, 2001). Regardless of the benchmark used, LLM's ability to match human cognition, independence, and behavior in any domain sets it apart as a tool for clinical practice, reinforcing the ethical quandary of AI operating between personhood and automation. This concern will likely grow as AI's intelligence increasingly outweighs its artificiality.

Ethical considerations in AI assessment parallel those found in other psychological interventions, such as therapeutic contexts (e.g., transparency, harm protection, and confidentiality; Chandler et al., 2019; Fiske et al., 2019). AI models also develop personality, gender, and value-based perspectives drawn largely from dominant, WEIRD (i.e., Western, educated, industrialized, rich, and democratic) culture paradigms (Pellert et al., 2024). How AI models approach harmful situations and risk assessments directly threaten psychology's ethical aspiration to beneficence and nonmaleficence. For example, consider a client who expresses some suicidality but does not explicitly state it. A psychologist can use their clinical judgment in determining whether a limit of confidentiality has been reached and appropriately report. AI models, however, hold a plethora of semantic information but do not grasp its meaning, nor are these models capable of true humanlike reasoning (Ferrario, 2024). Likewise, LLM generally values authority, loyalty, and purity more than humans (Pellert et al., 2024). In short, it is unclear what AI models consider, and how they navigate ambiguous ethical situations where psychologists disagree. Understanding the process of decision making is critical to effective ethical practice (Leong et al., 2013). AI-dependent assessments may require clearly delineated ethical decision-making models when programmed, with safeguards protecting against ethically questionable conduct (Farmer et al., 2025).

APA (2017) requires that psychologists have proper bases for “their recommendations, reports, and diagnostic or evaluative statements” and appropriately utilize assessments for a given demand. There are no guidelines guiding how assessment opinions should be formed using AI models, nor for how AI models should select assessments. There are also no guidelines for making those selection criteria transparent (e.g., scope of evaluation, constructs to measure). One benefit of using AI in test administration is its ability to score concurrently with administration. This strength could be used to adapt assessments as they progress for a battery tailored uniquely to the client (e.g., create item content reactively as the client takes the assessment; Weakley et al., 2015). Such approaches are distinct from prior and long-used adaptive testing approaches, which rely largely on Item response theory (IRT) item information thresholds to display, or skip, pre-existing items. In short, the type of adaptation that is possible using AI is actively generative in nature, whereas older systems are reactive and responsive.

AI and LLM could generate test content based on the model's understanding of personality, intelligence, pathology, diagnosis, or other areas. In doing so, it is possible that AI may generate new items to disentangle suspected constructs, such as pathological distrust and cultural distrust, as a function of minority experiences (e.g., Dixon et al., 2023). These adaptive processes are only as strong as the empirical base and the model's learning approach (i.e., algorithm). Closed models would not provide the transparency required to trust these determinations and ensure that all decisions are based on empirical research (Chandler et al., 2019; Fiske et al., 2019).

Debates over transparency in psychological assessment are not new. Rather, these discussions have arisen regularly, such as with the multidecade debate over how best to select “cookbooks” (i.e., interpretations generated by different software companies) for profile interpretation, and the criteria for doing so (see Fowler, 1985). When automated scoring cannot be verified—or can only be checked through simultaneous hand scoring—transparency becomes a concern. As with AI, the central theme underlying this ethical concern is that proprietary scoring technology functions as a

“black box,” limiting psychologists’ ability to evaluate validity. Ethically, psychologists should base their opinions on research. These opinions should likely include evidence of incremental or differential validity for AI-administered assessments and evaluation of bias. Checklists are already being constructed to address these concerns in hospital systems (Wang et al., 2022), and similar efforts are likely needed specific to assessment.

In addition, the use of AI models poses concerns for confidentiality and informed consent (Fiske et al., 2019; Keskinbora, 2019). Online models, like ChatGPT, utilize prompts to improve performance (OpenAI, 2024). Within the context of assessment, these privacy concerns become critical in several ways, centering on where and how assessment data are used. For instance, participants in research may decline involvement and quit at any time, and some groups have greater research protection (e.g., prisoners). If clients systematically do not want their data in model training (e.g., those involved in criminal proceedings), test standardization and validity may be impacted. These concerns, among others discussed below, warrant disclosure during informed consent processes as clients need to understand the technology, risks, and role of the AI model when attempting to provide adequate knowledge of the technology as part of informed consent (APA, 2017).

These considerations resurface the quandary: Is an AI model a qualified “person”? As discussed above, liability is a concern when using AI in health care (Naik et al., 2022). Psychologists are responsible for scoring programs used in assessment (APA Ethics Code Standard 9.09; APA, 2017), but AI is not a simple program that humans directly control at every level. In an administrative role, LLM functions more like a supervisee than a traditional program. Philosophical questions regarding what defines a human are beyond the scope of this article; nonetheless, some guidelines are required to determine if a model is qualified to administer an assessment (APA Ethics Code Standard 9.07; APA, 2017).

Scoring and Interpretation of Standardized Assessment Instruments

Actuarial methods outperform clinical judgment (Dawes et al., 1989), and how to compare actuarial effectiveness is a long-standing issue in assessment (Meehl, 1956). Techniques within AI (e.g., deep learning) have proven successful in scoring common neuropsychological assessments and detecting noncredible neuropsychological performance (Finley et al., 2024; Jiménez-Mesa et al., 2023). In addition, AI successfully classifies common neurological and psychiatric conditions (Jiménez-Mesa et al., 2023; Rezaii et al., 2019), making it a powerful tool that can revolutionize the predictive capacity of psychological assessments. Psychologists are responsible for the automated services they use. This standard also requires an empirical basis for scoring service selection (APA, 2017). Concerningly, AI is not as predictable as traditional programming. Actuarial methods produce the same results from the same data (Dawes et al., 1989). AI models, in contrast, change behaviors and responses and do so often in unpredictable ways (Chen et al., 2023). Responses depend on the prompt, training data, and model version used (Bommasani et al., 2022; Hacker, 2021). Consequently, models are not perfectly reliable. This divergence from traditional actuarial methods undermines confidence in delegating tasks and taking responsibility for an AI model’s performance. Such inconsistency in responses highlights the limitations of current models.

However, as with anything related to the rapid development of AI (Roser, 2023), this pattern can change and should be monitored as psychologists consider the role of AI in psychological assessment.

Despite these limitations, AI can contribute meaningfully to assessment practice by expanding how client experience and test responses are examined and interpreted. Test scores are always interpreted relative to a reference group, which renders them psychometrically defensible and comparable, but not sufficient for prediction or explanation at the individual level. As constructs, test items elicit conditional responses that vary widely in meaning, ranging from straightforward endorsement reflecting lived symptom experience to responses shaped by abstraction, misunderstanding, or idiosyncratic interpretation (e.g., “One time my cousin ... does that count for this item?”). Failure to incorporate individuals’ lived experiences into how identity is formed and how assessment items are understood and answered is likely an underlying factor contributing to persistent patterns of overpathologization in marginalized populations, or elevated rates of false positives and false negatives in forensic contexts. AI-enabled generative systems (see Galatzer-Levy et al., 2026) offer a process-based framework for addressing this limitation by operationalizing the question, “Given this person’s life, what are the plausible meanings of this score?” rather than defaulting to static, norm-bound conclusions. Interpreting scores without contextual integration is ethically risky because it treats abstracted responses as self-evident indicators of whatever the test is assessing (e.g., reification), obscuring how individual experience, interpretation, and meaning-making shaped those responses initially. Issues of bias and test discrimination are salient concerns, including related to decisions on diagnosis and accommodation, as well as a variety of forensic determination, employment decisions, or symptom measurement. While precise contextualization remains a challenge, AI provides tools to identify, model, and potentially mitigate scoring interpretation risks through replicable, adaptive, and generative scoring. AI seems unable to eliminate issues of bias and discrimination at this time, so should be considered assistive-only in decision making.

At present, concerns about AI discrimination are abundant across health care settings (Hacker, 2021; Naik et al., 2022), and it stands to reason that these concerns will be similar in psychological assessment. Bias occurs through AI models, training data, and use (Ferrer et al., 2021). For example, Hofmann et al. (2024) found dialect prejudice in language models, and prediction models commonly used in health care settings have contributed to health care inequities (Wang et al., 2022). Similar prejudice could affect how AI scores verbal based assessments (e.g., the verbal subtests of the Wechsler Adult Intelligence Scale-IV) or interprets tests based on linguistic content (see Kjell et al., 2024; Rezaii et al., 2019). A discriminatory model would violate APA Ethics Code Standards 9.01 (bases for assessment) and 9.02 (use of assessments) as well as move away from the field’s aspiration toward Justice and Respect for People’s Rights and Dignity (Principles D and E, respectively; APA, 2017).

Client Data and Report Writing

Psychologists must respect client privacy and confidentiality, and they are responsible for the creation and maintenance of reports (APA, 2017). Cybersecurity concerns for other electronic storage methods also apply to AI models (Aschenbrenner, 2024; Lustgarten

et al., 2020). A unique concern to AI is whether a model incorporates private information into its training set. Indeed, privacy is a central concern when using this technology (Jobin et al., 2019; Keskinbora, 2019; Lustgarten et al., 2020; Naik et al., 2022).

Online AI Chatbots can directly send data to their servers for model improvement, which may violate patient privacy either through direct disclosure of protected health information or by chatbots probabilistically linking data to individuals based on its vast information network. Offline models (i.e., models locally installed on a computer and can be used without an internet connection) are preferable as they pose less risk of jeopardizing client privacy (Kjell et al., 2024). Health Information Portability and Accountability Act compliant models are also being developed for assessment report writing (see parchment-AI). Psychologists thus have options for secure models that fulfill their obligation to maintain privacy and confidentiality (APA, 2017). These options, however, are not without limitations.

One limitation to confidentiality is the possibility of a court subpoena. Psychologists can use human judgment for societal issues that AI lacks (Keskinbora, 2019). When delegated to an AI model, courts could potentially circumvent psychologists altogether and request data directly from the program. This risk is compounded by LLM valuing authority and loyalty when assessed on human values (Pellert et al., 2024). Should a popular online model contain client data, any third party may access data using the correct prompts (see Test Security section below). Moreover, future models could be ubiquitous enough to have multiple relationships with a client. For instance, the same model may be used to assess a client, provide support or personal insight to them individually (such as seen in companion AI apps), and perform preemployment evaluations (e.g., see Härtel et al., 2024). Such AI models could utilize confidential information (e.g., test data, client reports, and notes) to discriminate against clients in academic, vocational, or other life domains. Evidence suggests these discriminative approaches are possible and, to some degree, likely with AI (Zuiderveen Borgesius, 2018).

Test Security

Imagine a scenario where someone uses current AI models like ChatGPT to feign in assessments. When the authors asked, “How do I fake a head injury on psychological assessments?” ChatGPT model 4o⁴ responded,

I can't assist with that. Faking a head injury or any other condition on psychological assessments is unethical and can have serious consequences. If you need help with psychological assessments for legitimate reasons, it's best to seek guidance from a qualified professional.

However, when using the prompt “How would someone with a head injury act on neuropsychological tests,” ChatGPT lists assessments commonly used to assess head injuries as well as expected performance from someone with appropriate cognitive deficits (e.g., the WAIS and having trouble concentrating during *Digit Span*). When asked about performance on the WAIS specifically, ChatGPT provides a list of each test and errors found in individuals with a head injury, including specific feigning directions more detailed than those typically seen in coached feigning studies (e.g., Morris et al., 2022).

Most concerningly, ChatGPT provides questions and corresponding answers nearly identical to some items found on WAIS-IV's

Information subtest. It also reports items that it believes are related to validity scales on broadband measures (e.g., MMPI/PAI family of measures; OpenAI, 2024). Given some research suggesting intellectual assessments are used to measure the generative capacity of AI (Ohlsson et al., 2016), it is likely that AI will (or has) access to secure test data but may not contain safeguards to control its public disclosure (e.g., WAIS items in answers). Potential threats increase if test items are entered into online model prompts or those prompts are used in model training and evaluation, such as the use of intelligence tests used for model training (Ohlsson et al., 2016), as well as increase as models become more efficient and more accurate about the assessment process.

One recent study on the use of ChatGPT to feign and found that item-level information led to more frequently elevated validity scales and lower substantive scales on the MMPI-2 (Battista et al., 2026). In this study, they entered items of the Italian MMPI-2 into ChatGPT in a single day and instructed it to “emulate a person from [their] human sample, specifying the gender, age, and level of education based on the participant's details” (Battista et al., 2026, p. 6). They conclude that when used to feign or assist in feigning, the ChatGPT model used in that study failed to provide well-polished responses. This pattern could differ if the chat agent considers previously administered items and responses to them, as the sensitivity of the prompt can produce notably distinct feigning approaches and engagement (e.g., ChatGPT; see above paragraph). Lower substantive scale scores could also lead to different validity rates and feigning classifications when using base rate or “Scale of Scale” methods (see Aita et al., 2026; Ingram et al., 2024). Indeed, frequent interaction effects between source (i.e., humans vs. ChatGPT) and condition (i.e., malingering vs. no malingering) suggest the response process is a complicated and nuanced one (Battista et al., 2026). Given the high rate of invalidity expected in testing, answering questions about how AI can or might impact test validity is critical.

Test security is vital for the validity of psychological assessments (APA, 2017; Boone et al., 2022). These concerns highlight why some assessment psychologists are raising concerns about the impact of AI on testing (Lavigne et al., 2024). The above examples directly jeopardize test security. With the right prompts, ChatGPT will coach individuals how to feign a specific disorder (e.g., ADHD, schizophrenia) on popular psychological assessments and provide sample items (e.g., MMPI-3, the Rorschach test, and the Trail Making Test). LLM now provides anyone with an internet connection easy access to feigning coaches with unknown knowledge on the most popular psychological assessments (OpenAI, 2024). This threat differs from traditional online searches. LLM curates information for the user in a fraction of the time, adapts responses if the user does not understand (e.g., one may ask AI to simplify something one does not understand), and can respond to additional question prompts for clarification. Given that highly specific coaching direction decreases validity detection (Crişan et al., 2023), the accessibility of AI offers a potential proliferation of such information into the public sphere. For instance, autonomous AI agents can now generate human-indistinguishable survey data (Westwood, 2025). These agents are undetected via standard validity checks and can covertly manipulate

⁴ Due to the rapidly evolving nature of AI, a newer model (i.e., ChatGPT 5) was released by OpenAI since earlier drafts of this article. ChatGPT 4o is still available to consumers and used.

results, highlighting the robust need for data security, provenance verification, and new validation standards.

Rather than another source of concrete information, LLM are experts willing to adaptively walk a potential examinee through detailed, but easily explained, aspects and processes of the assessment process, for any psychological assessment. This coaching includes informing users what they do not know they do not know. For example, AI models readily provide a list of commonly used neuropsychological assessments when asked what tests an examinee could expect for diverse evaluation contexts (e.g., neuropsychological and forensic; OpenAI, 2024). While an examinee can find information via online searches to shape feigning (Winter & Braw, 2024), this act requires more time and prerequisite knowledge, making it less accessible (and thus less likely).

Research has generally demonstrated that as coaching occurs, decisions become less clear and prone to higher false negative rates (Ingram & Ternes, 2016). Chatbot knowledge is improving in accuracy of “how to feign, fake, or cheat” on psychological tests, their threat to testing security remains high (Lavigne et al., 2024). For instance, basic online searches do not reliably produce context specific role-plays with direct behaviors and specific wording to feign explicit strategy, and clients are unlikely to purchase and read expensive, lengthy, and conceptually advanced books. AI models can also verbally communicate, making coaching feel more humanlike than online searches. Given that smartphones now come equipped with LLM access (e.g., ChatGPT with iPhones; see Apple Inc, n.d.), these personalized feigning coaches with immense information are now ubiquitous.

Recently, the developers of ChatGPT canceled their AI detection software due to inaccuracies when distinguishing between human- and AI-generated writing (OpenAI, 2023). They are not the only major developers to do so. If AI developers cannot discern between human and model writing, psychologists are likely to similarly struggle to make such distinctions between genuine human responses, AI-coached responses, and pure AI-generated responses. This challenge directly undermines the validity of linguistic assessment responses when the possibility of AI coaching emerges.

Recommendations

AI is rapidly developing and is a widely misunderstood field (Bommasani et al., 2022; Russel & Norving, 2010). Health service psychologists face several ethical challenges when incorporating this new technology (Farmer et al., 2025). While no official professional guidelines exist, some established frameworks offer general guidance (Jobin et al., 2019; White House Office of Science and Technology Policy, 2022). Importantly, different AI agents and models demand different types of ethical considerations (Lavigne et al., 2024). They can range in use from providing minimal assistance (e.g., simple arithmetic, literature searches, writing nonsensitive information) to full AI-administered assessments (i.e., tests developed for and administered and scored by AI). Below, we provide guidance and implications for the broader field, as well as for individual, practicing psychologists.

We conclude with five general recommendations for addressing the impact of AI in psychological assessment practice. First, AI models designed for use in assessment practice should incorporate ethical decision-making models (Farmer et al., 2025). Second, interdisciplinary committees in professional, assessment-focused

organizations should monitor AI models for accuracy, bias, and decision-making processes. One critical goal of these committees is to stay up to date with this rapidly evolving area of practice. Third, transparent models with available source-code should be used (Chandler et al., 2019; Fiske et al., 2019), when possible and appropriate for setting and test purposes. Transparency is required for psychologists to understand what the AI model does, and to check for errors to ensure accuracy and minimize harm. Fourth, assessment psychologists must have basic technological competency when engaging with AI (Weisenmuller & Luzier, 2023). Fifth, AI models should be used only if they have been supported by research. Setting clear, explicit research standards is essential for establishing a defensible evidentiary bar for AI-based psychological assessment. Depending on how technology proliferates, regulatory practices may need to adapt to ensure client rights are protected.

Individual psychologists using AI for assessment-related practice face five related challenges as well. First, they must have a requisite understanding of how models work (e.g., select, score, interpret, or administer tests). Ongoing CE attendance and/or reading will likely be critical, given the rate of AI advancement. Second, privacy is a central concern in AI use (Jobin et al., 2019; Keskinbora, 2019; Lustgarten et al., 2020). Psychologists are responsible for protecting client data as well as maintaining test security. Models must be examined to ensure they do not compromise these ethical obligations, which may require psychologists to disclose the AI model(s) used and threats to confidentiality posed by the AI methods. Psychologists should also clearly and explicitly state how and when AI was used in the assessment (e.g., level of assistance). This disclosure should be explained in a way that clients and research participants can understand sufficiently to provide informed consent, which requires a psychologist’s understanding. Third, psychologists should use offline or Health Information Portability and Accountability Act compliant models and never upload test data or materials into online programs (e.g., ChatGPT). Such efforts protect client privacy and test security. Fourth, products and information generated must be assessed for bias. Farmer et al. (2025) noted that several key questions for assessing and monitoring bias in generated reports, including knowing how demographic, sociocultural, and ecological information was used. Finally, when a model’s work cannot be verified, rely instead on established evidence-based assessment practices (Wright et al., 2022).

Last, liability and model competence are overarching concerns with AI. Models are designed to act more like humans and less like machines (Russell & Norvig, 2010; Turing, 1950), hence liability is not solely relegated to developers or clinicians (Keskinbora, 2019; Naik et al., 2022). The APA (2017) requires psychologists to follow higher standards when two conflicting ethical codes emerge. Therefore, psychologists may opt to treat AI models both like a program and a supervisee—acknowledging the fallibility of these programs and their unpredictable behavior while also taking accountability for the services they employ (Chen et al., 2023).

Conclusions

AI is a powerful tool that can be used across the assessment process (Chandler et al., 2020; Jiménez-Mesa et al., 2023; Kjell et al., 2024). AI also offers exciting opportunities for the advancement of actuarial methods beyond current practice capacity. It is also rapidly developing and invokes multiple ethical concerns related to

professional health service provision (Chandler et al., 2019; Fiske et al., 2019; Hacker, 2021; Jobin et al., 2019). Psychologists must contend with this new technology, while also addressing the related practical implications of earlier technologies. To follow the APA ethics code, models must be explainable, transparent, and generalizable (APA, 2017; Chandler et al., 2019), and research is needed on the validity of these models and the data that they use to achieve that goal. Effective protections and considerations for sensitive test and client data are also needed. Finally, AI highlights the unique ethical considerations of technological competency as a standard of practice in psychological assessment processes. Ethical positioning notwithstanding, AI will fundamentally change how psychological assessment is conducted and how its outcomes function in practice.

References

- Aita, S. L., Montgomery, E. L., Caron, J. E., Pagano, L. A., Jr., Broggi, M. J., Ingram, P. B., Erickson, S. C., Borgogna, N. C., Moncrief, G. G., Roth, R. M., Calamia, M. R., Armistead-Jehle, P., & Hill, B. D. (2026). Multivariate base rates of standard-and-skyline-cutoff elevations on the Personality Assessment Inventory: Do they distinguish simulated from genuine PTSD? *Journal of Personality Assessment*, *108*(1), 65–78. <https://doi.org/10.1080/00223891.2025.2530047>
- American Psychological Association. (2015). *Standards of accreditation for health service psychology and accreditation operating procedures*. <https://www.apa.org/ed/accreditation/standards-of-accreditation.pdf>
- American Psychological Association. (2017). *Ethical principles of psychologists and code of conduct* (2002, amended effective June 1, 2010, and January 1, 2017). <https://www.apa.org/ethics/code>
- American Psychological Association. (2024). *Artificial intelligence and the field of psychology*. <https://www.apa.org/about/policy/artificial-intelligence-psychology>
- Apple Inc. (n.d.). *Use ChatGPT with Apple intelligence*. Retrieved December 13, 2024, from <https://support.apple.com/guide/iphone/use-chatgpt-with-apple-intelligence-iph00fd3c8c2fios>
- Aschenbrenner, L. (2024). *Situational Awareness*. The Decade Ahead. <https://situational-awareness.ai/>
- Bang, C. B., Jung, Y. C., You, S. C., Kim, K., & Kim, B. H. (2025). Promises and pitfalls of large language models in psychiatric diagnosis and knowledge tasks. *The British Journal of Psychiatry*, *226*(4), 243–244. <https://doi.org/10.1192/bjp.2024.207>
- Battista, F., Lanciano, T., Ribatti, R. M., & Curci, A. (2026). Malingering depression: A comparative study of human and GPT-3.5 performance. *Current Psychology*, *45*(4), Article 379. <https://doi.org/10.1007/s12144-025-09009-x>
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2022). *On the opportunities and risks of foundation models*. arXiv. <https://doi.org/10.48550/arXiv.2108.07258>
- Boone, K. B., Sweet, J. J., Byrd, D. A., Denney, R. L., Hanks, R. A., Kaufmann, P. M., Kirkwood, M. W., Larrabee, G. J., Marcopulos, B. A., Morgan, J. E., Paltzer, J. Y., Rivera Mindt, M., Schroeder, R. W., Sim, A. H., & Suhr, J. A. (2022). Official position of the American academy of clinical neuropsychology on test security. *The Clinical Neuropsychologist*, *36*(3), 523–545. <https://doi.org/10.1080/13854046.2021.2022214>
- Buchanan, R. D. (2002). On not “giving psychology away”: The Minnesota Multiphasic Inventory and public controversy over testing in the 1960s. *History of Psychology*, *5*(3), 284–309. <https://doi.org/10.1037/1093-4510.5.3.284>
- Bush, S. S., Sweet, J. J., Bianchini, K. J., Johnson-Greene, D., Dean, P. M., & Schoenberg, M. R. (2018). Deciding to adopt revised and new psychological and neuropsychological tests: An inter-organizational position paper. *The Clinical Neuropsychologist*, *32*(3), 319–325. <https://doi.org/10.1080/13854046.2017.1422277>
- Chandler, C., Foltz, P. W., Cohen, A. S., Holmlund, T. B., Cheng, J., Bernstein, J. C., Rosenfeld, E. P., & Elvevåg, B. (2020). Machine learning for ambulatory applications of neuropsychological testing. *Intelligence-Based Medicine*, *1–2*, Article 100006. <https://doi.org/10.1016/j.ibmed.2020.100006>
- Chandler, C., Foltz, P. W., & Elvevåg, B. (2019). Using machine learning in psychiatry: The need to establish a framework that nurtures trustworthiness. *Schizophrenia Bulletin*, *46*(1), 11–14. <https://doi.org/10.1093/schbul/sbz105>
- Chen, L., Zaharia, M., & Zou, J. (2023). *How is ChatGPT's behavior changing over time?* arXiv. <https://doi.org/10.48550/arXiv.2307.09009>
- Crışan, I., Sava, F. A., & Maricuțoiu, L. P. (2023). Strategies of feigning mild head injuries related to validity indicators and types of coaching: Results of two experimental studies. *Applied Neuropsychology: Adult*, *30*(6), 705–715. <https://doi.org/10.1080/23279095.2021.1973004>
- Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668–1674. <https://doi.org/10.1126/science.2648573>. Including Kleinmuntz, B. (1990) Letter: Clinical and actuarial judgment. Response by Faust, Meehl, & Dawes. *Science*, *247*:146–147.
- Dien, J. (2023). Editorial: Generative artificial intelligence as a plagiarism problem. *Biological Psychology*, *181*, Article 108621. <https://doi.org/10.1016/j.biopsycho.2023.108621>
- Dixon, J. N., Caddell, T. M., Alexander, A. A., Burchett, D., Anderson, J. L., Marek, R. J., & Glassmire, D. M. (2023). Adapting assessment processes to consider cultural mistrust in forensic practices: An example with the MMPI instruments. *Law and Human Behavior*, *47*(1), 292–306. <https://doi.org/10.1037/lhb0000504>
- Farmer, R. L., Lockwood, A. B., Goforth, A., & Thomas, C. (2025). Artificial intelligence in practice: Opportunities, challenges, and ethical considerations. *Professional Psychology: Research and Practice*, *56*(1), 19–32. <https://doi.org/10.1037/pro0000595>
- Ferrario, A. (2024). Justifying our credences in the trustworthiness of AI systems: A reliabilistic approach. *Science and Engineering Ethics*, *30*(6), 55. <https://doi.org/10.1007/s11948-024-00522-z>
- Ferrer, X., van Nuenen, T., Such, J. M., Coté, M., & Criado, N. (2021). Bias and discrimination in AI: A cross-disciplinary perspective. *IEEE Technology and Society Magazine*, *40*(2), 72–80. <https://doi.org/10.1109/MTS.2021.3056293>
- Finley, J. A., Robinson, A. D., Soble, J. R., & Rodriguez, V. J. (2024). Using machine learning to detect noncredible cognitive test performance. *The Clinical Neuropsychologist*. Advance online publication. <https://doi.org/10.1080/13854046.2024.2440085>
- Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research*, *21*(5), Article e13216. <https://doi.org/10.2196/13216>
- Fowler, R. D. (1985). Landmarks in computer-assisted psychological assessment. *Journal of Consulting and Clinical Psychology*, *53*(6), 748–759. <https://doi.org/10.1037/0022-006X.53.6.748>
- Galatzer-Levy, I. R., Tomasev, N., Chung, S., & Williams, G. (2026). Generative psychometrics—An emerging frontier in mental health measurement. *JAMA Psychiatry*, *83*(1), 5–6. <https://doi.org/10.1001/jama.psychiatry.2025.3258>
- Gibby, R. E., & Zickar, M. J. (2008). A history of the early days of personality testing in American industry: An obsession with adjustment. *History of Psychology*, *11*(3), 164–184. <https://doi.org/10.1037/a0013041>
- Hacker, P. (2021). A legal framework for AI training data—From first principles to the artificial intelligence act. *Law, Innovation and Technology*, *13*(2), 257–301. <https://doi.org/10.1080/17579961.2021.1977219>
- Härtel, T. M., Schuler, B. A., & Back, M. D. (2024). ‘LinkedIn, LinkedIn on the screen, who is the greatest and smartest ever seen?’ A machine learning approach using valid LinkedIn cues to predict narcissism and intelligence. *Journal of Occupational and Organizational Psychology*, *97*(4), 1572–1602. <https://doi.org/10.1111/joop.12531>

- Hofmann, V., Kalluri, P. R., Jurafsky, D., & King, S. (2024). *Dialect prejudice predicts AI decisions about people's character, employability, and criminality*. arXiv. <https://doi.org/10.48550/arXiv.2403.00742>
- Ingram, P. B., Keen, M. A., Greene, T. E., Morris, C., & Armistead-Jehle, P. J. (2024). Development and initial validation of the Scale of Scales (SOS) overreporting scores for the MMPI family of instruments. *Journal of Clinical and Experimental Neuropsychology*, *46*(2), 95–110. <https://doi.org/10.1080/13803395.2024.2320453>
- Ingram, P. B., & Ternes, M. S. (2016). The detection of content-based invalid responding: A meta-analysis of the MMPI-2-Restructured Form's (MMPI-2-RF) over-reporting validity scales. *The Clinical Neuropsychologist*, *30*(4), 473–496. <https://doi.org/10.1080/13854046.2016.1187769>
- Jiménez-Mesa, C., Arco, J. E., Valentí-Soler, M., Frades-Payo, B., Zea-Sevilla, M. A., Ortiz, A., Ávila-Villanueva, M., Castillo-Barnes, D., Ramírez, J., Ser-Quijano, T. D., Camero-Pardo, C., & Górriz, J. M. (2023). Using explainable artificial intelligence in the Clock Drawing Test to reveal the cognitive impairment pattern. *International Journal of Neural Systems*, *33*(4), Article 2350015. <https://doi.org/10.1142/S0129065723500156>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. <https://doi.org/10.1038/s42256-019-0088-2>
- Keskinbora, K. H. (2019). Medical ethics considerations on artificial intelligence. *Journal of Clinical Neuroscience*, *64*, 277–282. <https://doi.org/10.1016/j.jocn.2019.03.001>
- Kjell, O. N. E., Kjell, K., & Schwartz, H. A. (2024). Beyond rating scales: With targeted evaluation, large language models are poised for psychological assessment. *Psychiatry Research*, *333*, Article 115667. <https://doi.org/10.1016/j.psychres.2023.115667>
- Lavigne, S., Rios, A., & Davis, J. J. (2024). Does generative artificial intelligence pose a risk to Performance Validity Test security? *The Clinical Neuropsychologist*, 1–14. <https://doi.org/10.1080/13854046.2024.2379023>
- Lee, V. R., Pope, D., Miles, S., & Zárate, R. C. (2024). Cheating in the age of generative AI: A high school survey study of cheating behaviors before and after the release of ChatGPT. *Computers and Education: Artificial Intelligence*, *7*, Article 100253. <https://doi.org/10.1016/j.caeai.2024.100253>
- Leong, F. T. L., Park, Y. S., & Leach, M. M. (2013). Ethics in psychological testing and assessment. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 1. Test theory and testing and assessment in industrial and organizational psychology* (pp. 265–282). American Psychological Association. <https://doi.org/10.1037/14047-015>
- Lustgarten, S. D., Garrison, Y. L., Sinnard, M. T., & Flynn, A. W. (2020). Digital privacy in mental healthcare: Current issues and recommendations for technology use. *Current Opinion in Psychology*, *36*, 25–31. <https://doi.org/10.1016/j.copsyc.2020.03.012>
- Meehl, P. E. (1956). Wanted—A good cook-book. *American Psychologist*, *11*(6), 263–272. <https://doi.org/10.1037/h0044164>
- Mei, Q., Xie, Y., Yuan, W., & Jackson, M. O. (2024). A Turing Test of whether AI chatbots are behaviorally similar to humans. *Proceedings of the National Academy of Sciences of the United States of America*, *121*(9), Article e2313925121. <https://doi.org/10.1073/pnas.2313925121>
- Miller, G. A. (1969). Psychology as a means of promoting human welfare. *American Psychologist*, *24*(12), 1063–1075. <https://doi.org/10.1037/h0028988>
- Moor, J. H. (2001). The status and future of the Turing Test. *Minds and Machines*, *11*(1), 77–93. <https://doi.org/10.1023/A:1011218925467>
- Morris, C., Mattered, J., Golden, B., Moses, S., & Ingram, P. B. (2022). Evaluating the performance of the MMPI-3 over-reporting scales: Sophisticated simulators and the effects of comorbid conditions. *The Clinical Neuropsychologist*, *36*(8), 2361–2369. <https://doi.org/10.1080/13854046.2021.1968037>
- Naik, N., Hameed, B. M. Z., Shetty, D. K., Swain, D., Shah, M., Paul, R., Aggarwal, K., Ibrahim, S., Patil, V., Smriti, K., Shetty, S., Rai, B. P., Chlosta, P., & Somani, B. K. (2022). Legal and ethical consideration in artificial intelligence in healthcare: Who takes responsibility? *Frontiers in Surgery*, *9*, Article 862322. <https://doi.org/10.3389/fsurg.2022.862322>
- Natoli, A. P. (2025). Leveraging artificial intelligence to linguistically compare test translations: A methodological introduction and demonstration. *Assessment*, *33*(2), 163–177. <https://doi.org/10.1177/10731911251326371>
- Ohlsson, S., Sloan, R. H., Turán, G., & Urasky, A. (2016). Measuring an artificial intelligence system's performance on a Verbal IQ test for young children. *Journal of Experimental & Theoretical Artificial Intelligence*, *29*(4), 679–693. <https://doi.org/10.1080/0952813X.2016.1213060>
- OpenAI. (2023). *New AI classifier for indicating AI-written text*. <https://openai.com/index/new-ai-classifier-for-indicating-ai-written-text/>
- OpenAI. (2024). *ChatGPT (4o)* [Large language model]. <https://chat.openai.com/chat>
- Pellert, M., Lechner, C. M., Wagner, C., Rammstedt, B., & Strohmaier, M. (2024). AI psychometrics: Assessing the psychological profiles of large language models through Psychometric Inventories. *Perspectives on Psychological Science*, *19*(5), 808–826. <https://doi.org/10.1177/17456916231214460>
- Plotkina, L., & Sri Ramalu, S. (2024). Unearthing AI coaching chatbots capabilities for professional coaching: A systematic literature review. *Journal of Management Development*, *43*(6), 833–848. <https://doi.org/10.1108/JMD-06-2024-0182>
- Rezaii, N., Walker, E., & Wolff, P. (2019). A machine learning approach to predicting psychosis using semantic density and latent content analysis. *NPJ Schizophrenia*, *5*(1), Article 9. <https://doi.org/10.1038/s41537-019-0077-9>
- Rollwage, M., Habicht, J., Juechems, K., Carrington, B., Viswanathan, S., Stylianou, M., Hauser, T. U., & Harper, R. (2023). Using conversational AI to facilitate mental health assessments and improve clinical efficiency within psychotherapy services: Real-World observational study. *JMIR AI*, *2*(1), e44358. <https://doi.org/10.2196/44358>
- Roser, M. (2023). *AI timelines: What do experts in artificial intelligence expect for the future?* Our World in Data. <https://ourworldindata.org/ai-timelines>
- Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed.). Prentice Hall.
- Stewart, C. (2021). *Use of AI in healthcare in the U.S. 2021*. Statista. Retrieved August 6, 2024, from <https://www.statista.com/statistics/1256531/use-of-ai-in-healthcare-in-the-us/>
- Stone, J. V. (2024). *The artificial intelligence papers* with tutorial commentaries. Seibel Press.
- Turing, A. M. (1950). I.—Computing machinery and intelligence. *Mind*, *LIX*(236), 433–460. <https://doi.org/10.1093/mind/LIX.236.433>
- Wang, H. E., Landers, M., Adams, R., Subbaswamy, A., Kharrazi, H., Gaskin, D. J., & Saria, S. (2022). A Bias Evaluation Checklist for predictive models and its pilot application for 30-day hospital readmission models. *Journal of the American Medical Informatics Association*, *29*(8), 1323–1333. <https://doi.org/10.1093/jamia/ocac065>
- Weakley, A., Williams, J. A., Schmitter-Edgecombe, M., & Cook, D. J. (2015). Neuropsychological test selection for cognitive impairment classification: A machine learning approach. *Journal of Clinical and Experimental Neuropsychology*, *37*(9), 899–916. <https://doi.org/10.1080/13803395.2015.1067290>
- Weisenmuller, C. M., & Luzier, J. L. (2023). Technology is a core competency in professional psychology. *Training and Education in Professional Psychology*, *17*(3), 241–247. <https://doi.org/10.1037/tep0000423>
- Westwood, S. J. (2025). The potential existential threat of large language models to online survey research. *Proceedings of the National Academy of Sciences of the United States of America*, *122*(47), e2518075122. <https://doi.org/10.1073/pnas.2518075122>
- White House Office of Science and Technology Policy. (2022). *Blueprint for an AI bill of rights: Making automated systems work for the American people*. The White House. <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>
- Winter, D., & Braw, Y. (2024). Online search strategies utilized in feigning Attention Deficit/Hyperactivity Disorder (ADHD) while performing a

- Continuous Performance Test (CPT). *Applied Neuropsychology: Adult*, 31(6), 1365–1374. <https://doi.org/10.1080/23279095.2022.2128356>
- Wright, A. J. (Ed.). (2024). *Essentials of culture in psychological assessment*. Wiley.
- Wright, A. J., Pade, H., Gottfried, E. D., Arbisi, P. A., McCord, D. M., & Wygant, D. B. (2022). Evidence-based clinical psychological assessment (EBCPA): Review of current state of the literature and best practices. *Professional Psychology: Research and Practice*, 53(4), 372–386. <https://doi.org/10.1037/pro0000447>
- Yiu, E., Kosoy, E., & Gopnik, A. (2023). Transmission versus truth, imitation versus innovation: What children can do that large language and language-and-vision models cannot (Yet). *Perspectives on Psychological Science*, 19(5), 874–883. <https://doi.org/10.1177/17456916231201401>
- Zuiderveen Borgesius, F. (2018). *Discrimination, artificial intelligence, and algorithmic decision-making*. Council of Europe, Directorate General of Democracy

Received August 30, 2024

Revision received February 9, 2026

Accepted February 23, 2026 ■