PRZEGLĄD PSYCHOLOGICZNY – THE REVIEW OF PSYCHOLOGY
2026, TOM 69, NR 1, 171–177
DOI: 10.31648/przegldpsychologiczny.13030

Principles for the Use of Artificial Intelligence in Psychology.

A Review of the Global Psychology Alliance Document: Top 10 Principles for Understanding Artificial Intelligence in Psychology

Beata Krzywosz-Rynkiewicz[1]
University of Warmia and Mazury in Olsztyn, Department of Clinical, Developmental and Educational Psychology
https://orcid.org/0000-0003-4306-4875

Krzysztof Mariusz Ciepliński
Vizja University in Warsaw, Faculty of Human Sciences
https://orcid.org/0000-0003-1100-6419

In recent years numerous scholarly publications have pointed to the important role of artificial intelligence in psychology and in the field of mental health. AI-based technologies are used, among other things, to promote changes in health-related behavior (Aggarwal et al., 2023), and in the diagnosis, monitoring, and treatment of mental health conditions (Cruz-Gonzalez et al., 2025; Kolding et al., 2024). Dakanalis et al. (2024) describe artificial intelligence as a "game-changer" in psychiatric care. AI is also being applied in psychological and psychiatric education (Prégent et al., 2025), and may help reduce clinicians' workload by automating the preparation of medical documentation (McCrudden et al., 2026; Olson et al., 2025). Another recurring theme in the literature is the relationship between AI-generated diagnostic tools and traditional, standardized psychological tests (Kirthivasan & Bhardwaj, 2026).

The use of AI-based tools in the everyday work of practicing psychologists and researchers is therefore not a matter for some distant future, but an issue of immediate relevance. The development of artificial intelligence nevertheless

[1] Correspondence: beata.rynkiewicz@uwm.edu.pl.

raises important questions for psychology concerning professional responsibility, ethics, the validity of interventions, client safety, and the role of the psychologist in a world increasingly saturated with digital technologies.

A response to these challenges is the document *Top 10 Principles for Understanding Artificial Intelligence in Psychology*, developed within the framework of the Global Psychology Alliance (GPA) (Global Psychology Alliance, 2026). The GPA is an international initiative bringing together psychological organizations and experts from various regions of the world, established by the American Psychological Association to respond to problems that exceed the capacity of national communities acting alone. The document on artificial intelligence was prepared by a team of authors representing diverse professional traditions, institutions, and cultural contexts. The contributors include, among others, representatives of the American Psychological Association, the Pontificia Universidad Católica del Perú, the Jamaican Psychological Society, University College Dublin, the Finnish Psychological Association, and the Counsellor Council of India. The work was coordinated by Maria Elena Humphrey of the Central American Union of Psychology Associations, and the document was reviewed by experts from Canada, Colombia, Pakistan, Hungary, Taiwan, and the Philippines.

From the perspective of the Polish psychological community, it is particularly significant that the Polish Psychological Association, as a member of the Global Psychology Alliance, participates in this international collaboration. The value of the GPA document lies not only in its content but also in the manner of its creation: it is the product of international reflection on how to protect the dignity, safety, and rights of people who use psychological knowledge amid digital transformation.

The document is neither a legal code nor a detailed procedural manual for the use of AI. It is normative and conceptual in character: it maps out the main areas of risk, sets out basic principles for the responsible use of AI, and establishes a framework for further professional, educational, and institutional regulation.

Content of the Document

The document presents ten principles for understanding and applying artificial intelligence in psychology. These are organized into three main areas: (1) ethics, transparency, and accountability; (2) professional competence and human oversight; and (3) access, fairness, and inclusion. This structure captures well the key questions that a psychologist using AI tools should ask, namely:

1) Is the use of the tool ethically justified?
2) Does the psychologist understand not only its benefits but also its limitations?
3) Is professional judgment being preserved?
4) Does the technology avoid deepening inequality and increasing the risk of harm?

The First Group of Principles: Ethics, Transparency, Informed Consent, and Accountability

The document emphasizes that the use of AI in psychology should be grounded in well-established ethical principles, such as respect for the dignity of the person, accountability, justice, concern for well-being, and the avoidance of harm. Psychologists should understand not only the technical capabilities of AI but also its impact on human decision-making, trust, dependence on technology, social relationships, and the sense of agency. Artificial intelligence cannot be treated as a neutral tool because its functioning always depends on the assumptions adopted, the data used, the way it is designed, and the context of its use. An example would be an application supporting mood self-assessment that, on the basis of a few brief answers from the user, suggests a "high risk of depression" and encourages the user to book a paid consultation. If the user does not know that the recommendation was generated automatically, based on limited data and without a clinical diagnosis, they may treat it as a credible psychological assessment. In such a case, the problem is not the technology itself, but the lack of transparency, informed consent, and responsible professional oversight.

Transparency is therefore of great importance in the use of AI. Psychologists should select only AI tools that provide reasonably clear information about their data sources, limitations, and data-processing methods. Clients using psychological services, research participants, students, or other recipients of a psychologist's work should know when an AI system is being used in the process. This is especially relevant in situations where a user might mistakenly believe they are communicating with a human being, or that the generated content reflects the independent judgment of a professional. The requirement of transparency is thus intended to protect trust in the psychologist and in psychology as a profession.

Closely linked to transparency is the principle of informed consent. The document extends the classical understanding of consent to include new aspects arising from the use of AI. A client, patient, research participant, or trainee should be informed whether their data will be collected, analyzed, stored, or processed by an AI tool. Consent should encompass knowledge of the potential benefits and risks, the system's limitations, and any possible use of personal data. The authors also stress that consent must not be obtained through manipulative design solutions that make it harder for the user to make an informed decision.

The next principle concerns accountability and governance. The use of AI does not shift professional responsibility from the psychologist onto the technology. Even when a system generates an analysis, suggestion, interpretation, supporting diagnosis, or recommendation, the psychologist remains responsible for how that information is used. This is particularly important in diagnosis, crisis intervention, psychotherapy, expert opinion-giving, scientific research, and education. An example would be the use of a program that, after analyzing a client's statements, flags a "high suicide risk". The psychologist cannot simply accept or ignore such information automatically. It must be verified through

conversation with the client, assessed in light of the clinical context, and followed by the appropriate safety procedures. If the system is wrong, responsibility for the decision does not rest with the algorithm but with the professional and the institution that approved its use. The document therefore reminds us that AI operates within a complex network of stakeholders: system developers, vendors, institutions, employers, regulators, and users. This multiplicity of actors, however, must not be allowed to dilute accountability. Psychologists should know who controls the tool, what data are being used, what oversight procedures exist, and how errors or harms can be addressed.

The Second Group of Principles: Professional Competence, Scientific Grounding, Human Oversight, and Harm Prevention

The document indicates that psychologists should possess a sufficient understanding of the AI tools they use. This does not mean they must be programmers or machine-learning specialists. It does mean having the ability to critically evaluate: what the tool was designed for, what data it operates on, what its limitations are, whether it has been validated, and whether it is fit for a specific psychological purpose. An example would be a school psychologist using an application that, on the basis of a short questionnaire, identifies students "at risk of anxiety disorders". The psychologist does not need to know the underlying code, but should know whether the tool has been studied in children and adolescents, in which country it was developed, what norms it uses, how often it produces false positives, and whether it can be applied in the Polish school context. Without this knowledge, the result should be treated only as a signal warranting further assessment. Digital competence and the ability to critically evaluate AI are thus becoming part of the psychologist's contemporary professional competence.

The principle of human oversight and professional judgment is of great importance. AI may support the psychologist's work, but it should not replace their decisions. The results generated by a system require interpretation in the context of the person, the situation, culture, language, life history, the purpose of the assessment, or the nature of the intervention. An example would be a system that, after analyzing a teenager's statements, classifies them as "low risk" of a mental health crisis because words directly referring to self-harm do not appear. A psychologist who is aware of the family context, previous instances of self-harm, and the way the student communicates may judge the result to be misleading. Professional judgment then requires rejecting the AI's recommendation and pursuing further conversation and protective action. The document draws attention to the risk of so-called automation bias — excessive trust in automatically generated recommendations. In psychology, this is especially dangerous because a seemingly objective result may obscure the complexity of human experience. Psychologists must retain both the right and the duty to question, correct, or reject a result produced by AI.

The principle of psychological safety and harm prevention indicates that AI can generate new risks: erroneous recommendations, false information, excessive

personalization, the reinforcement of emotional dependence on the system, stigmatization, inappropriate crisis-related messaging, or the mislabeling of suicide risk or relapse. In the field of mental health, the consequences of such errors can be serious. Psychologists should therefore monitor the effects of using AI, anticipate possible harms, and discontinue the use of a tool when the risks outweigh the benefits.

The document also strongly emphasizes the need for validation and an evidence base for AI tools. AI tools are not self-validating. Their usefulness depends on the quality of the data, the accuracy of the model, the reliability of the results, the context of use, and the population for which they have been tested. Evidence drawn from one culture, language group, or care system cannot automatically be assumed to transfer to other conditions. If a given tool lacks independent validation, a psychologist should treat it, at most, as an exploratory or organizational aid — not as a basis for decisions affecting people's lives, health, rights, or opportunities.

The Third Group of Principles: Algorithmic Bias, Fairness, Access, and Inclusion

The authors point out that AI systems may replicate or reinforce inequalities present in training data and social practices. Algorithmic bias may relate to gender, age, language, ethnic origin, socioeconomic status, disability, place of residence, or cultural context. In psychological practice, this can lead to misdiagnosis, inappropriate recommendations, unequal access to services, or the perpetuation of stereotypes. An example would be an AI tool that supports the diagnosis of emotional difficulties, trained primarily on data from English-speaking adult users in large cities. When applied to adolescents from small towns or to speakers of minority languages, it may misinterpret their ways of expressing emotions as "low insight" or "resistance". As a result, the system may recommend less appropriate forms of help. Psychologists should therefore check whether a given tool has been validated for the specific linguistic, cultural, and age group in question, and more generally should ask whether it performs equally well across different groups and whether it does not systematically disadvantage people who are already at risk of marginalization.

The final principle concerns access, equity, and inclusion. AI is sometimes presented as a tool that democratizes access to psychological help, since it can increase the availability of services and lower their cost. The document, however, treats this claim with caution. Access to AI tools depends on digital infrastructure, language, technological competence, cost, internet connectivity quality, accessibility for people with disabilities, and cultural fit. Without reflection on these factors, AI may not reduce inequality but deepen it. An example would be a free "first-line psychological support" chatbot available only through a mobile application, in English only, and requiring a constant internet connection. For students in large cities, it may be a convenient supplement to existing help, but for older adults, residents of areas with poorer digital connectivity, people with disabilities, or those who do not speak English, it becomes yet another

barrier. In such a case, AI formally expands the range of services on offer, but in practice only broadens access for selected groups. Psychologists should therefore assess whether a given solution genuinely expands access to help or primarily benefits people and groups who already possess technological resources.

The document closes with a glossary of basic concepts, including artificial intelligence, generative AI, large language model, training data, black box, algorithmic bias, automation bias, and AI hallucinations. This is an important practical element, as it standardizes the language of the debate and can make the document easier to use for psychologists who do not work with technology daily.

Value, Limitations, and the Need for Further Regulation

The value of the GPA document lies, above all, in its identification and discussion of the most important areas of a psychologist's responsibility regarding artificial intelligence. It shows that AI is not merely a technical tool, but an element embedded in professional relationships, decision-making processes, service organizations, scientific research, and education. The document reminds us that the core values of psychology — respect for the dignity of the person, their well-being, accountability, fairness, the validity of interventions, and protection from harm — remain relevant in the digital environment.

An important strength of the document is its international character. Incorporating perspectives from different countries helps avoid narrowing the discussion to a single legal system, culture, or service delivery model. The document highlights the significance of global inequalities in access to technology, cultural differences, and infrastructural limitations. This is particularly important given that many AI tools are developed under conditions dominated by English, large technology markets, and data drawn from select populations.

A limitation of the document, however, is its general character. It does not contain detailed procedures relating to the Polish legal, professional, and institutional context. It does not determine which AI tools may be admissible in psychological diagnosis and assistance, psychotherapy, crisis intervention, expert opinion-giving, education, or scientific research. Nor does it specify minimum validation criteria, standards for documenting AI use, rules for handling sensitive data, or procedures for responding to system errors. For this reason, the document should be treated as a framework for reflection rather than as a ready-made practice standard.

For Polish psychologists, the GPA document may serve as an important starting point for discussions aimed at developing local regulations. Detailed recommendations are needed regarding, among other things: data protection; informed consent; documentation of AI use; the scope and permissibility of tools in diagnosis and therapy; principles for using AI in work with children and adolescents; and professional accountability in high-risk situations. Future regulations should combine ethical, legal, scientific, and practical perspectives, so that artificial intelligence can support psychology without compromising its fundamental values.

References

Aggarwal, A., Tam, C. C., Wu, D., Li, X., & Qiao, S. (2023). Artificial intelligence–based chatbots for promoting health behavioral changes: Systematic review. *Journal of Medical Internet Research*, *25*, Article e40789. https://doi.org/10.2196/40789

Cruz-Gonzalez, P., He, A. W.-J., Lam, E. P., Ng, I. M. C., Li, M. W., Hou, R., Chan, J. N.-M., Sahni, Y., Vinas Guasch, N., Miller, T., Lau, B. W.-M., & Sánchez Vidaña, D. I. (2025). Artificial intelligence in mental health care: A systematic review of diagnosis, monitoring, and intervention applications. *Psychological Medicine*, *55*, Article e18. https://doi.org/10.1017/S0033291724003295

Dakanalis, A., Wiederhold, B. K., & Riva, G. (2024). Artificial intelligence: A game-changer for mental health care. *Cyberpsychology*, *Behavior and Social Networking*, *27*(2), 100–104. https://doi.org/10.1089/cyber.2023.0723

Global Psychology Alliance. (2026). *Top 10 principles for understanding artificial intelligence in psychology*. https://www.apa.org/international/networks/global-psychology-alliance/principles-artificial-intelligence.pdf

Kirthivasan, M., & Bhardwaj, A. B. (2026). ChatGPT (AI) vs. standardized psychological testing. In R. Sharma, F. Rabby, R. Bansal, T. Sachdeva, J. Mrabet (Eds.), *Confronting mental health stigma with AI and machine learning* (pp. 203–237). Wiley. https://doi.org/10.1002/9781394347308.ch9

Kolding, S., Lundin, R. M., Hansen, L., & Østergaard, S. D. (2024). Use of generative artificial intelligence (AI) in psychiatry and mental health care: A systematic review. *Acta Neuropsychiatrica*, *37*, Article e37. https://doi.org/10.1017/neu.2024.50

McCrudden, K. E., Swirbul, M. S., Peake, E. E., Rodio, M. J., & Padmanabhan, A. (2026). AI-powered documentation for mental health providers: Retrospective observational mixed methods study. *JMIR Formative Research*, *10*, Article e84628. https://doi.org/10.2196/84628

Olson, K. D., Meeker, D., Troup, M., Barker, T. D., Nguyen, V. H., Manders, J. B., Stults, C. D., Jones, V. G., Shah, S. D., Shah, T., & Schwamm, L. H. (2025). Use of ambient AI scribes to reduce administrative burden and professional burnout. *JAMA Network Open*, *8*(10), Article e2534976. https://doi.org/10.1001/jamanetworkopen.2025.34976

Prégent, J., Chung, V. H. A., El Adib, I., Désilets, M., & Hudon, A. (2025). Applications of artificial intelligence in psychiatry and psychology education: Scoping review. *JMIR Medical Education*, *11*, Article e75238. https://doi.org/10.2196/75238