



UNIVERSITAT DE
BARCELONA

Ethics and Artificial Intelligence in Psychological Assessment

Authors:

Krieger, Virginia^{1,2}, Magallón-Neri, Ernesto^{1,2} and Amador-Campos, Juan Antonio^{1,2}

¹Department of Clinical Psychology and Psychobiology, Faculty of Psychology,
University of Barcelona.

²Institute of Neurosciences at the University of Barcelona, Group for the Study of
Measurement Invariance and Analysis of Change (GEIMAC).

September 26, 2024



Citation: Krieger, V., Magallón-Neri, E., & Amador-Campos, J. A. (2024). *Ética e Inteligencia Artificial en Evaluación Psicológica* [Documento docente]. Facultad de Psicología, Universidad de Barcelona.

Abstract

Artificial intelligence (AI) applications underwent exponential growth in the second decade of the 21st century. Currently, many AI applications can be used in the fields of psychological assessment, intervention and research. This paper addresses the use of AI in psychological assessment and intervention, and also discusses some ideas related to AI bias and the implications for psychological assessment and intervention, issues concerning training for future psychology professionals and modifications and adaptations required in codes of ethics for psychologists.

Keywords: artificial intelligence, psychological assessment and intervention, ethics and professional deontology

Table of contents

1. Artificial intelligence and psychological assessment.....	4
2. Biases in artificial intelligence and human supervision.....	5
3. Ethics and training of future psychology professionals.....	6
4. The psychologist's code of ethics and the use of artificial intelligence.....	8
5. Conclusions.....	10
6. References.....	11

1. Artificial intelligence and psychological assessment

Digital technologies play a role in almost every aspect of our lives and have become indispensable in many personal and professional tasks. In the fields of psychological assessment, intervention and research, these technologies provide access to diverse groups of participants and offer the opportunity to study a wide range of human behaviors in an unobtrusive and real-time manner (Light et al., 2024).

Technological applications in the field of assessment are no long limited to administering tests and instruments by means of electronic devices, and can now detect mental health issues that may require clinical attention (Glenn & Monteith, 2014; Hagstrom & Maranzan, 2019). One of the most recent technologies is artificial intelligence (AI), which can be used to collect, analyze and evaluate large amounts of information (Shatte et al., 2019).

AI is a branch of computer science devoted to creating programs and models that perform operations in a similar way to the human mind, such as learning, perception, creativity and logical reasoning for decision-making (RAE, 2023). AI systems adapt their behavior and analyze the effects of previous actions to work autonomously (European Parliament, 2021).

This type of technology first emerged in the 1940s. However, it was not until the second decade of this century that AI started taking on a significant role in people's everyday lives and society. Some of the most widely used concepts in the field of AI are related to machine learning (ML; Samuel, 1959), which generates computer models capable of learning to solve problems based on input data. Natural language processing (NLP), which focuses on interpreting, interacting with and solving problems using written and spoken language, underpins most of the commercial generative language models with which we interact, including ChatGPT, Gemini and Bard. Another recent AI approach is deep learning (DL; Hinton et al., 2006), which is a learning model based on multilayer neural networks capable of learning more efficiently than linear ML models to perform formal and abstract tasks (Anyoha, 2017; Ribera & Díaz, 2024).

AI has multiple applications in the field of psychology; for example, Cao et al. (2017) developed an ML model to predict mood disorders from metadata collected for keypresses on alphanumeric characters, including duration of a keypress, time since last keypress and

distance from the last key along two axes. The model was able to predict participants' scores on the Hamilton Depression Rating Scale and the Young Mania Rating Scale with over 90% accuracy. Reece and Danforth (2017) analyzed photogenic markers of depression in a study using the Instagram app, while Mandryk et al. (2021) presented a program to predict scores in a self-reported depression questionnaire (Patient Health Questionnaire, PHQ-9) based on the performance of attention, matching and working memory tasks. Performance in these tasks predicted PHQ-9 scores to a significant degree, even when demographic factors that can influence depression such as age and gender were taken into account. ML has also been used to assess personality traits and characteristics from conversations and interactions recorded with mobile apps (Wundrack & Mehl, 2020) and through computer models capable of making judgments about personality traits, apparently with greater accuracy than human judgment (Kosinski et al., 2014; Youyou et al., 2015). Many of these applications have proven useful in mental health (Zhou et al., 2022).

The use of technological tools such as AI in psychological assessment and intervention should be supervised by professionals with adequate training and knowledge of the instruments and programs used. Such tools must be used for their intended purpose and be underpinned by proven validity and reliability (Landers & Behrend, 2023), which is rare in many of the services offered, for which this information is not available. Information about the tests used and the people who will evaluate and report the data obtained from the tests is also frequently unavailable. With respect to intervention programs, these pose similar challenges, with the added problem that the developers of these programs are often not psychology professionals. Even when the programs are developed by psychologists (Díaz-Asper et al., 2024), there is a lack of data on the effectiveness of the intervention and it is impossible to ensure that these instruments and programs are used properly.

2. Biases in artificial intelligence and human supervision

Researchers, ethics bodies and committees, and professional associations are becoming increasingly concerned about biases and algorithmic injustice (or algorithmic unfairness) in AI-based decision-making processes applied to individual cases in the field of assessment and intervention and the potential unintended consequences (Hagstrom & Maranzan, 2019).

Applications that use AI in psychological assessment and intervention can present several biases:

1. **Algorithmic bias.** Algorithms used in AI applications may have inherent biases due to the way in which they were designed or programmed. This adds to the “black box” nature of some algorithms, which lack access to the source code and thus the inner workings. For example, an algorithm could be used to assess a video of a job interview or clinical session based on variables such as speech content, tone of voice and response latency. If the algorithm relies solely on these data and does not consider the patient’s mental health record, it could offer erroneous recommendations or diagnoses.
2. **Training bias.** If the datasets used to train algorithms are not representative of the population, lack validity or reliability, or are biased, the accuracy of the predictions, conclusions and recommendations generated will be affected.
3. **Cultural bias.** AI applications may reflect cultural and social biases, depending on the social conventions of the people who created them (adaptation), and this can lead to biased outcomes in terms of psychological assessment and treatment.

Acknowledging these biases is key to actively working toward mitigating them through careful selection of algorithm training data, review of the algorithms used, and critical evaluation of the results generated by AI applications in the field of psychology. It is also worth noting that AI applications should serve as tools to complement the evaluations and clinical judgments of trained professionals with a view to avoiding potential biases and errors in decision-making.

It is important to stress that AI systems used in psychological assessment may also have unintended negative effects. Mayer et al. (2020) observed that the introduction of AI into the management systems of a company’s executive department led to a loss of employees’ critical thinking skills and autonomy. Therefore, it is crucial to understand the impact of introducing new AI systems in specific situations, especially in fields that involve people, to prevent these problems. Recently, a feasibility assessment process for implementing AI-based decision-making systems called PAAI was proposed (Buschmeyer et al., 2023). This

evaluation system considers the entire social environment in which the system is applied and the equipment and human capital required for the task, measures the psychological impact, and considers both internal resources (e.g., intelligence) and external resources (e.g., social support) to mediate the psychological load generated, whether positive (activation and experience flow) or negative (mental overload and stress), in accordance with standards of satisfaction and performance (ISO, 2019). This evaluation approach would represent a significant improvement in the integration procedures of AI-based decision-making systems involving interactions with people.

3. Ethics and training for future psychology professionals

In 1986, Hartman commented on the increasing use of computer programs for assessment and intervention among psychologists, pointed out growing concerns about ethical and professional issues, and highlighted the importance that “psychologists understand the consequences of automating their services” (p. 463). Such concerns about the ethics of applying AI are even more relevant today, as the use of AI in psychological assessment and intervention presents many challenges and ethical considerations.

AI applications are used both to conduct and to evaluate psychological and educational assessments (Yan et al., 2020), intervention programs, and research to shed light on the effectiveness of educational and psychological interventions (D’Mello et al., 2022). AI applied to the field of psychological assessment and intervention is developing so rapidly that practicing professionals and training programs are struggling to stay up to date. This raises the question of what aspects should be considered in the education of future psychologists and continuous training for graduates to ensure that these applications are used properly.

The following aspects could be useful for both the ethical training of future psychologists and the professional development of graduates:

- 1. Understanding and knowledge among both users and professionals.** Both psychologists and users need a clear understanding of the benefits and limitations of

AI applications. They should be aware of how AI can be used in assessment, diagnosis, intervention and treatment within a rigorous professional and ethical framework.

2. **Equality of access.** It is essential and a top priority to ensure that AI applications are accessible to diverse populations to avoid excluding minority or disadvantaged groups.
3. **Consent, privacy and confidentiality in data protection.** Individuals must be informed about how their data will be used, and their explicit consent must be sought before AI is used to assess or treat them. It is essential to guarantee that people's personal and sensitive information is protected and handled confidentially. AI applications must adhere to the data protection regulations of the country in which they are used (European Union regulation, EU, 2024).
4. **Transparency in procedures and explanations of results.** Users, patients and professionals must be able to understand how and why an AI application reaches specific conclusions or recommendations. The AI application should offer clear and understandable explanations about the decision-making process.
5. **Detection and mitigation of bias.** AI applications must be designed to prevent and correct any type of bias that could impact the assessment and treatment of users or patients, thus ensuring fairness and justice in outcomes.
6. **Validity and effectiveness.** AI applications must be valid, effective and based on scientific evidence. Their efficacy and effectiveness should be tested in controlled studies before being implemented in clinical practice. It is crucial to critically evaluate these qualities in AI tools before they are used in professional practice.
7. **Professional responsibility and skills acquisition.** Ultimate responsibility for assessment and intervention decisions should lie with psychologists when using AI applications. Such applications should serve as complementary tools and must not replace professional judgment. Professionals should also receive training and practical instruction on specific technological tools, such as relevant AI software and platforms for assessment and intervention.
8. **Supervision and evaluation.** Continuous monitoring and evaluation of AI applications is crucial to ensure that they are effective and accurate, and adjustments

should be made where necessary. Moreover, it is important to establish guidelines and criteria for the proper use of AI in various contexts, including assessment, intervention and research.

9. **Impact on the therapeutic relationship.** The use of AI applications must not disrupt the therapeutic relationship between the psychologist and the patient, which is key to the success of assessments and treatments. Psychological interventions require a high level of empathy and human understanding, which AI applications are currently unable to replicate.
10. **Ethical regulation.** Professional associations and organizations should establish specific guidelines within codes of ethics regarding the use of AI in psychological assessment, intervention and research.
11. **Interdisciplinary collaboration.** Psychologists and AI experts must collaborate on the development and evaluation of the benefits, effectiveness and validity of AI tools. This partnership should take a multidisciplinary approach that integrates insights from other fields such as ethics, computer science and sociology.
12. **Continuous training and learning.** Professionals should be encouraged to actively pursue continuous training and stay up-to-date on new technologies and AI developments that can benefit psychological assessment, intervention and research.

4. The psychologist's code of ethics and the use of artificial intelligence

The use of clinical AI applications should adhere to the principles established in codes of professional ethics and deontology (Light et al., 2024; Skorburg et al., 2024; Villas Olmeda & Camacho Ibáñez, 2022). The *Deontological Code of the Psychologist*, published by the General Council of Psychology of Spain, establishes the ethical standards and principles that must be followed by psychology professionals. However, this code was written before the widespread use of AI in the field of psychology, and therefore does not specifically address AI technologies. Table 1 outlines several principles and articles from the current code, along with suggestions on how it could be applied or expanded to include the use of AI.

Table 1.

Several principles from the current deontological code, along with suggestions on how it could be applied to include the use of AI.

Current regulations	Application adapted to AI
1. Professional competence and continuing education. The code mandates that psychologists keep their knowledge and skills current by means of ongoing training (articles 6, 16, 17 and 18).	This should include staying informed about the use and potential implications of AI in their professional practice.
2. Confidentiality, privacy and data protection. The code emphasizes the importance of confidentiality and the protection of client data (articles 34, 39, 40, 44 and 46).	The code could be expanded to incorporate specific guidelines on data management in the context of AI. AI applications should adhere to these guidelines to ensure that all processed data are treated with confidentiality.
3. Informed consent, transparency and explanations. The code requires that psychologists obtain informed consent from their clients and explain the nature of the assessment or intervention (articles 25, 29, 34, 35, 37 and 44).	The code could be expanded to informing clients about the use of AI in assessment, intervention and research and the implications. The processes of generating, training and using AI models, as well as the results, should also be explained to clients.
4. Ethical integrity and responsible use. The code recommends integrity and ethics in psychological practice (articles 35, 37, 40, 44 and 48).	It is essential that psychologists explicitly ensure that AI is used ethically and does not harm clients, while also being aware of the potential risks.
5. Fairness, equity and non-discrimination. The code emphasizes the importance of non-discrimination and fair treatment (Article 10).	Standards should be established to ensure that the use of AI is free from bias and actively promotes equity. Continuous monitoring of AI systems is necessary to prevent any biases that could lead to discrimination of any kind.
6. Supervision, evaluation and professional responsibility. The code stipulates that psychologists are	It is essential that the code include principles to ensure professional supervision of the outcomes generated by AI applications. This means that

accountable for their professional decisions (articles 12, 17, 18, 24, 29 and 48).	professional responsibility cannot be entirely delegated and that the information provided by AI must be carefully monitored. Additionally, it is important to incorporate legal and professional responsibilities regarding the use of AI.
--	---

5. Conclusions

Artificial intelligence is a set of technologies with enormous potential for society and the field of psychology. In the specific discipline of psychological assessment, AI must be used within a framework of rigorous professionalism and ethical commitment. Implementing AI poses a wide range of challenges and issues that must be addressed through an ethical approach in line with evolving legislative regulations and standards that are currently under discussion by regulatory bodies at European and national level.

References

- Anyoha, R. (2017). *The History of Artificial Intelligence*. SITN, Harvard Kenneth C Griffith Graduate school of arts and sciences. <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>
- Buschmeyer, K., Hatfield, S., & Zenner, J. (2023). Psychological assessment of AI-based decision support systems: tool development and expected benefits. *Frontiers in Artificial Intelligence*, 6:1249322. <https://doi.org/10.3389/frai.2023.1249322>
- Cao, B., Zheng, L., Zhang, C., Yu, P. S., Piscitello, A., Zulueta, J., Ajilore, O., Ryan, K., & Leow, A. D. (2017, August). *Deepmood: Modeling mobile phone typing dynamics for mood detection* [Conference session]. Proceedings of the 23rd ACM International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada.
- Colegio Oficial de Psicólogos (España). *Código deontológico del psicólogo*. <https://www.cop.es/index.php?page=CodigoDeontologico>
- D'Mello, S. K., Tay, L., & Southwell, R. (2022). Psychological measurement in the information age: Machine-learned computational models. *Current Directions in Psychological Science*, 31(1), 76–87. <https://doi.org/10.1177/09637214211056906>
- Diaz-Asper, C., Hauglid, M. K., Chandler, C., Cohen, A. S., Foltz, P. W., & Elvevåg, B. (2024). A framework for language technologies in behavioral research and clinical applications: Ethical challenges, implications, and solutions. *The American Psychologist*, 79(1), 79–91. <https://doi.org/10.1037/amp0001195>
- Glenn, T., & Monteith, S. (2014). New measures of mental state and behavior based on data collected from sensors, smartphones, and the Internet. *Current Psychiatry Reports*, 16, 523. <http://doi.org/10.1007/s11920-014-0523-3>
- Hagstrom, S. L., & Maranzan, K. A. (2019). Bridging the gap between technological advance and professional psychology training: A way forward. *Canadian Psychology/Psychologie canadienne*, 60(4), 281-289. 289. <https://doi.org/10.1037/cap0000186>

- Hartman, D. E. (1986). On the use of clinical psychology software: Practical, legal, and ethical concerns. *Professional Psychology: Research and Practice*, 17(5), 462–465. <https://doi.org/10.1037/0735-7028.17.5.462>
- Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast-learning algorithm for deep belief nets. *Neural Computation* 18(7), 1527-54. <https://doi.org/10.1162/neco.2006.18.7.1527>.
- ISO International Organization for Standardization (2019). ISO 9241-210:2010. *Ergonomics of Human-System Interaction: Human-Centered Design for Interactive Systems*. Available online at: <https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-1:v1:en>
- Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., & Graepel, T. (2014). Manifestations of user personality in website choice and behaviour on online social networks. *Machine Learning*, 95, 357–380. <http://doi.org/10.1007/s10994-013-5415-y>
- Landers, R. N., & Behrend, T. S. (2023). Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models. *The American Psychologist*, 78(1), 36–49. <https://doi.org/10.1037/amp0000972>
- Light, L. L., Panicker, S., Abrams, L., & Huh-Yoo, J. (2024). Ethical challenges in the use of digital technologies in psychological science: Introduction to the special issue. *American Psychologist*, 79(1), 1–8. <https://doi.org/10.1037/amp0001286>
- Mandryk R. L., Birk, M. V., Vedress, S., Wiley, K., Reid, E., Berger, P., & Frommel, J. (2021) Remote Assessment of Depression Using Digital Biomarkers from Cognitive Tasks. *Frontiers in Psychology*, 12:767507. <https://doi.org/10.3389/fpsyg.2021.767507>
- Mayer, A.S., Strich, F., & Fiedler, M. (2020). Unintended consequences of introducing AI systems for decision making. *MIS Quarterly Executive*, 19(4), 6. <https://aisel.aisnet.org/misqe/vol19/iss4/6>
- Parlamento Europeo (2021). ¿Qué es la inteligencia artificial y cómo se usa? Artículo 20200827STO85804. Consultado en

<https://www.europarl.europa.eu/topics/es/article/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa>

Real Academia de la Lengua Española REA (2023). Diccionario de la lengua española. <https://dle.rae.es/inteligencia>

Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, 6(1), Article 15. <https://doi.org/10.1140/epjds/s13688-017-0110-z>

Ribera, M., & Díaz O. (2024). *ChatGPT y educación universitaria: posibilidades y límites de ChatGPT como herramienta docente*. Octaedro IDP/ICE. Universidad de Barcelona.

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers, *IBM Journal of Research and Development*, 3(3), 210-229, <https://doi.org/10.1147/rd.33.0210>.

Shatte, A. B. R., Hutchinson, D. M., & Teague, S. J. (2019). Machine learning in mental health: a scoping review of methods and applications. *Psychological Medicine*, 49(9), 1426–1448. <https://doi.org/10.1017/S0033291719000151>

Skorburg, J. A., O'Doherty, K., & Friesen, P. (2024). Persons or data points? Ethics, artificial intelligence, and the participatory turn in mental health research. *The American Psychologist*, 79(1), 137–149. <https://doi.org/10.1037/amp0001168>

Unión Europea, UE (2024). REGLAMENTO (UE) 2024/1689 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 13 de junio de 2024 por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.o 300/2008, (UE) n.o 167/2013, (UE) n.o 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:L_202401689

Villas Olmeda, M., & Camacho Ibáñez, J. (2022) *Manual de ética aplicada en inteligencia artificial*. Anaya.

- Wundrack, R., & Mehl, M. R. (2020). El sonido de la personalidad. *Mente y Cerebro*, 104, 10-15.
- Yan, D., Rupp, A. C., & Foltz, P. W. (2020). *Handbook of Automated Assessment: Theory into practice*. Taylor & Francis; CRC Press.
- Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the national Academy of Sciences of the United States of America*, 112, 1036–1040. <http://doi.org/10.1073/pnas.1418680112>
- Zhou, S., Zhao, J., & Zhang, L. (2022). Application of Artificial Intelligence on Psychological Interventions and Diagnosis: An Overview. *Frontiers in Psychiatry*, 13, 811665. <https://doi.org/10.3389/fpsyt.2022.811665>