World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Check for updates

Ethical dilemmas in the relationship between professional and patient before the integration of the IA in psychiatric diagnosis

José Fernando Mendez-Licona *

Universidad Libre Seccional Cali, Valle del Cauca-Colombia.

World Journal of Advanced Research and Reviews, 2025, 27(01), 275-279

Publication history: Received on 18 May 2025; revised on 22 June 2025; accepted on 28 June 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.27.1.2472

Abstract

The purpose of this case study is to analyze the ethical and professional challenges that arise in the integration of artificial intelligence for psychiatric diagnosis, with special emphasis on the relationship between the health professional and the patient. The study included the review of scientific literature, analysis of ethical and regulatory standards, as well as the elaboration of a hypothetical clinical case in order to illustrate the issues that are associated with data privacy, as well as the dehumanization of care, aspects of clinical liability and the emergence of biases in algorithms. It was identified that AI is a tool that can improve diagnostic accuracy and personalization of treatments, but this implies risk in the use of sensitive data, the affectation of the therapeutic link and ambiguity in the assignment of responsibilities, as well as the perpetuation of structural inequalities and biases. The multidisciplinary analysis shows the need to balance the potential technological benefits with the protection of human values and ethical aspects in both diagnosis and treatment. It is concluded that AI in psychiatry must be implemented responsibly, ensuring transparency, equity, protection of patients' rights and humanity as the center of the relationship between professional and patient, which makes necessary a clear regulation about safe clinical practice in the digital era.

Keywords: Mental Health Care (51553); Clinical Ethics (36359); Artificial Intelligence (22729); Psychiatry (12007). [DECS/MESH]

1. Introduction

Although artificial intelligence is not a recent development, over the last few decades, and especially in recent years, it has established itself as a revolutionary tool in all fields of knowledge, including the field of medicine, within which it has transformed multiple clinical areas, including psychiatry (1). One of the greatest advantages of this tool is its ability to analyze large amounts of information, which is why it can detect complex patterns, offering new opportunities not only to improve the accuracy of diagnosis, but also to personalize the treatments carried out in the mental health area (2). Despite this, it cannot be ignored that the implementation of this tool presents important ethical, legal and human challenges that require a deep reflection by professionals in all areas of knowledge.

Specifically, when talking about the relationship between mental health professionals and their patients, the fundamental pillar of this professional relationship is composed of trust, as well as empathy and confidentiality, which can be impacted due to the introduction of technologies that automate processes as delicate as psychiatric diagnosis (3). From these approaches arises then the question and the need to analyze how AI tools can be integrated in an ethical and responsible manner, which implies respecting the rights and dignity of patients, while at the same time it requires taking care of the quality of care and professional autonomy.

* Corresponding author: José Fernando Mendez-Licona ORCID: 0009-0000-5475-798X

This dissertation addresses, through a case study, the main dilemmas and considerations that should be taken into account about the ethical use of AI when it is to be used in psychiatric diagnosis, and therefore addresses and focuses on the relationship between professionals and patients; for which key aspects such as data privacy will be explored, as well as the preservation of the therapeutic relationship, also including responsibility in decision-making and ending with some considerations about equity in access and quality of diagnosis (4). Given these conditions, it follows that the aim of this dissertation is to offer a critical, as well as constructive, view, whose purpose is to allow taking advantage of the benefits of AI as a tool, without leaving aside, of course, how the fundamental values under which clinical practice in psychiatry is sustained must be safeguarded.

2. Materials

Fistol, a systematized bibliographic review is carried out, using recent academic and other specialized sources on the subject of artificial intelligence applied to the field of psychiatry, as well as health ethics and digital methodologies used for diagnosis and treatment, including documents such as scientific articles, reviews and technical papers.

Secondly, simulated clinical cases are used, considering examples of virtual patients created using artificial intelligence for teaching and diagnostic evaluation (1) (5), which allow the analysis of interactions and limitations in the use of life in controlled contexts and which are available in the academic literature. Normative documents and ethical frameworks, such as international and local regulations that guide the responsible use of AI tools in the healthcare field, especially those emphasizing privacy, equity, and informed consent, were also used as materials (4, 6, 7).

2.1. Methods

A critical documentary analysis was used for the method, which includes an exhaustive review of the scientific literature in order to identify both the advances and the ethical, as well as the technical and social benefits and challenges involved in the use of AI in psychiatry, and those that focus on the professional relationship between the health professional and the patient in making the diagnosis.

Qualitative case studies were also used to develop hypothetical cases based on real and simulated scenarios, taking into account those that raise ethical dilemmas related to privacy in the use of data, dehumanization, clinical responsibility in psychiatry, and biases when AI has been used as a tool for assisted diagnosis.

Likewise, an evaluation of multidisciplinary origin is used in which different clinical, ethical, legal and technological perspectives are integrated to analyze the impact of AI in psychiatric practice for the evaluation of clinical cases and their treatment (8), taking into account the considerations involved for both professionals and patients.

3. Results and discussion

3.1. Data Privacy and Confidentiality in the Use of AI in Psychiatry

Given the sensitive nature of the data and the fact that artificial intelligences require large volumes of personal data, including actual medical records, as well as clinical notes and assessment results, the increased use of this data poses a risk of leaks, unauthorized access and misuse, which can end up affecting both the privacy and security of patients (1).

To this end, it becomes necessary and essential for patients to understand the entire data handling process, which includes explaining to them how their data are collected, how they are stored, and how their data are used in artificial intelligence systems, taking special care to explain to them the risks associated with the use of their data in these tools (4). It should also be considered that the complexity of the algorithms and the lack of transparency of AI tools may make it difficult for patients to give truly informed consent, even if healthcare personnel try to give the broadest and fullest possible explanations of the ways in which this technology operates (9). It should also be borne in mind that the use and safeguarding of data should be done only for the authorized purpose, so that purposes such as marketing or sharing with third parties should be explicitly authorized.

The protocols and regulations contained in the ethical frameworks, whose purpose is to protect the information, must also be considered, which is why it is mandatory to anonymize the data, limit full access only to authorized personnel and guarantee access to traceability about its use (10). Given these conditions, it is essential that the use of artificial intelligence tools in psychiatry complies with global privacy regulations, which implies the adoption of robust technical and organizational measures that seek to prevent security breaches in their hosting and handling (11).

It is also considered that the exchange of data with third parties, including marketing companies, insurers or employers, should always be done with the direct consent of the patient, in order to avoid phenomena such as discrimination and loss of confidence in mental health services (6). This requires the implementation of strict policies and effective control mechanisms, as well as the protocols and regulations contained in the ethical frameworks, whose purpose is to protect the information, which is why it is mandatory to anonymize the data, limit full access only to authorized personnel and guarantee access to the traceability of its use (10). Given these conditions, it is essential that the use of artificial intelligence tools in psychiatry complies with global privacy regulations, which implies the adoption of robust technical and organizational measures that seek to prevent security breaches in their hosting and handling (11).

Regarding ethical practice, privacy and security standards should be as high as possible, which should include the anonymous submission of information, encryption of all data collected, and periodic audits to ensure the transparency of informed consent (1), which should clearly and precisely explain to patients all the conditions related to the use of their data and the associated risks. The active participation of patients in the design and evaluation of artificial intelligence tools should also be encouraged, taking into account that their rights should be ensured and that expectations regarding the use of their information are respected in the development of all phases of AI tools (6).

3.2. Therapeutic Relationship in Psychiatry in the Age of AI

The main challenges in the use of AI in the context of psychiatry as a medical science have to do with the dehumanization of clinical practice and with the danger and bias to which both health professionals and patients are exposed because this tool lacks the ability to interpret and respond adequately to human emotions, mainly because it cannot be genuinely empathic; empathy, active listening and understanding are the basis of the therapeutic bond and the psychotherapeutic process (4). This process of empathy cannot be adequately developed by AI, which is further accentuated by the fact that many patients value feeling listened to and understood by another person, and it is precisely this aspect that is the main source of their feeling of strengthened confidence and willingness to adhere to treatment (12).

In this sense, speaking about the role of the health professional, it should be considered in the first instance that professionals should integrate AI as a complementary tool and should not delegate to them completely the responsibility within the therapeutic and diagnostic processes (11). This highlights the fundamental quality of critical thinking and approach and of the close human relationship, which enhances the closeness of the patient to the process.

There is also a risk, in the simulation of conversations, of the ethical responsibility and uniqueness of the patient, reinforced with the use of conversational simulation that can be performed in contact with artificial intelligences (8). It is therefore necessary for health professionals to be transparent about the limitations that AI may have, emphasizing that, although it simulates conversations and emotions in the use of its language, its use is not equivalent to truly human care.

3.3. Responsibility and Decision Making in the Use of AI

Although the increased speed of diagnosis and treatment that can be achieved through artificial intelligence tools, as well as the accuracy of diagnosis, has already been mentioned, this does not leave aside some questions about the legal and ethical responsibility that must be considered with respect to clinical decision making. For this reason, it should be taken into account that in the event of an erroneous diagnosis or an inadequate treatment based on artificial intelligence indications, the first thing to consider is that it is complex to determine whether the responsibility should fall on the health professional (13).

In addition to the above, it should be noted that there is still no regulatory framework in some countries, which makes it very difficult to impute responsibilities and the legal protection that can be applied to both patients and professionals. This indicates that for the use of AI it should be seen as a tool that supports the diagnosis but does not issue it completely, for which it is argued that it does not replace the clinical judgment and much less the final decision making by a health professional; being his responsibility to critically evaluate all the recommendations within the clinical context, the patient's history and those subjective factors that AI cannot capture completely, because it cannot completely read the emotional and mental context of the patient, an action that a health professional should perform (7).

It should also be considered that the use of AI could make it difficult for the health professional to explain diagnostic decisions, affecting not only the relationship with his patient, but also directly implicating his professionalism (7). This is why continuous monitoring and ethical evaluation is necessary to ensure a safe and responsible use of AI.

3.4. Biases and Diagnostic Fairness in Psychiatry using AI

Because these digital tools learn with information from historical data, which are contaminated with social, demographic and cultural biases, they come to directly reflect some inequalities that have existed in the health system for some years and in some regions (14). This includes the omission of some representative features of ethnic minorities, people with low resources or some vulnerable groups, which may be incorrectly diagnosed due to ignorance of the particularities of the culture or community in which they live and their social and individual background.

This is compounded by the lack of diversity that can be perceived in the data sets, which limits the ability of distal models to generalize and provide diagnoses that are accurate for all individuals, since they read all data in a general way, encompassed within a social or cultural phenomenon. This inability to diagnose accurately would come to affect populations with less access to health services or who possess some clinical characteristics that are atypical, which could even be ignored or even misinterpreted by digital AI tools.

To this must also be added the fact that many of the ideas do not explain the processes they develop to make decisions or to give their perspective on the issues that are put to them (15), which makes it difficult for health professionals to understand how they made diagnostic decisions.

4. Conclusion

Although artificial intelligence represents an innovative tool with great potential to improve both the accuracy and personalization of psychiatric diagnosis, when integrating it into psychiatric diagnosis and treatment it must be carefully managed in order to preserve the quality and humanity of care, key factors for the good performance of healthcare professionals.

Issues related to the protection of privacy and confidentiality of data should also be considered, given the sensitive nature of the information collected and handled by psychiatric professionals, which implies the need to create precise and strict protocols that guarantee transparent informed consent.

It should also be considered that the therapeutic relationship between professional and patient is one of the key factors in the development of mental health care, which is why digital tools cannot replace human attention for now, being empathy and trust in the health professional a key factor for the development and diagnosis within a good psychiatric treatment.

On the other hand, clear regulatory mechanisms and guidelines should be established to establish responsibilities in clinical decision making, explicitly stating everything related to both the ethics in the use of AI for treatment and diagnosis in mental health, as well as the roles and responsibilities framed in its use and the consequences of it.

It is also important to take into account that another ethical aspect has to do with the technical training of mental health professionals, which is a key aspect for the responsible integration of AI and the use of its benefits and limitations. All these aspects raise the need for a multidisciplinary approach that should include technical, ethical, legal and social factors that guarantee technological innovation contributing to a more accessible care, at the same time that is fairer and focused on the patient's welfare.

References

[1] Ocando-Parra LC. Artificial intelligence in psychiatry: innovations, challenges and future diagnosis and treatment. Bibliographic review. Rev. Latinoamericana de Hipertensión. 2025; 20(1): p. 85-94.

[2] López-Zúñiga LA, Rodríguez-Zúñiga MA. THE ETHICS OF USING ARTIFICIAL INTELLIGENCE IN THE PSYCHOLOGICAL EVALUATION AND DIAGNOSIS OF PATIENTS IN DURANGO, MEXICO. Rev. Ciencia Latina. 2024; 8(4): p. 423-46.

[3] Clínica Psiquiátrica Privada José Ingenieros. La Importancia del Profesional Médico. [Online]; 2024. Available at: https://clinicajoseingenieros.com.ar/iaymedicina/.

[4] Consejo General de la Psicología de España. INFOCOP. [Online]; 2025. Available at: https://www.infocop.es/beneficios-y-riesgos-de-la-ia-en-la-atencion-a-la-salud-mental-segun-mhe/.

[5] Baile-Ayensa JI. Paciente con depresión creado por inteligencia artificial de libre acceso para la enseñanza de Psicología. Estudio preliminar de su validez. Rev. Tec. Ciencias y Educ. 2024; 27: p. 7-42.

[6] UNESCO. INteligencia Artificial. [Online]; 2023. Available at: https://www.unesco.org/es/artificial-intelligence/recommendation-ethics/cases.

[7] Biscaia-Fernández JM, Mohedano-del Pozo RB. Artificial intelligence in the prevention of suicidal behaviour: technical aspects and ethical-legal considerations. Rev. Bioética y Derecho. 2024; 59: p. 181-203.

[8] López-Santín JM, Alvarado-Serón P. Digital mental health: a critical approach from ethics. Rev. Asoc. Esp. Neuropsiq. 2018; 38(134): p. 359-79.

[9] Suárez-Villa M, Borré-Ortiz YM, Expósito MJ, Pérez-Almagro MC. Educational practices in health: fundamental axis in the development of early childhood in integral development centers. Rev. Salud Uninorte. 2021; 36(1): p. 325-41.

[10] Serrano-Gutiérrez L. Facultad de Ciencias Económicas y Empresariales. [Online]; 2021. Available at: https://repositorio.comillas.edu/rest/bitstreams/436674/retrieve.

[11] García-López A, Girón-Luque F, Roselli D. The Integration of Artificial Intelligence in Healthcare: Ethical and. Universitas Méd. 2022; 64(3): p. 11-9.

[12] Wei M. Psychology Today. [Online]; 2024. Available at: https://www.psychologytoday.com/co/blog/las-personas-prefieren-medicos-humanos-y-psiquiatras-o-ia.

[13] Aroca-Ríos MC, Calderón-Velásquez MF. Universidad Autónoma de Bucaramanga. [Online]; 2024. Available at: https://repository.unab.edu.co/bitstream/handle/20.500.12749/27598/Tesis.pdf?sequence=1&isAllowed=y.

[14] Lanzagorda-Ortega D, Carrillo-Pérez D, Carrillo-Esper R. Artificial intelligence in medicine: present and future. Gac. Méd. Méx. 2023; 158(supl.1): p. S(1)17-21.

[15] Rojas O. Confidencialidad en psiquiatría. Rev. Méd. Clínica Los Condes. 2012; 23(5): p. 634-9.