UDC 17:[004.891.3:159.9](043.2)

H.D. Dziumenko
A.T. Terenchuk

ETHICAL ASPECTS OF AUTOMATED SYSTEMS IN PSYCHOLOGICAL DIAGNOSTICS: AN ANALYTICAL REVIEW

Vinnytsia National Technical University

Анотація

У тезах розглянуто етичні аспекти застосування автоматизованих систем у психологічній діагностиці. Проаналізовано ключові проблеми, пов'язані із захистом персональних даних, прозорістю алгоритмів, алгоритмічною упередженістю та інформованою згодою. Окреслено сферу відповідальності розробників і клінічних фахівців. Запропоновано комплексну систему принципів для етичного впровадження діагностичних систем штучного інтелекту.

Ключові слова: психологічна діагностика, штучний інтелект, машинне навчання, етика, конфіденційність даних, алгоритмічна упередженість, інформована згода.

Abstract

This research paper examines the ethical dimensions of applying automated systems in psychological diagnostics. Key issues related to personal data protection, algorithmic transparency, bias, and informed consent are analysed. The scope of responsibility for developers and clinical practitioners is outlined. A comprehensive set of principles for the ethical deployment of AI-based diagnostic systems is proposed.

Keywords: psychological diagnostics, artificial intelligence, machine learning, ethics, data privacy, algorithmic bias, informed consent.

Introduction

The rapid integration of artificial intelligence (AI) and machine learning (ML) technologies into healthcare and mental health services has opened new horizons for the assessment and diagnosis of psychological disorders. Automated diagnostic systems promise greater consistency, scalability, and accessibility compared to traditional clinical interviews. Systems capable of analysing speech patterns, facial expressions, text, and physiological signals have demonstrated promising results in detecting conditions such as depression, anxiety, schizophrenia, and post-traumatic stress disorder (PTSD) [1-2].

Nevertheless, the deployment of such technologies in sensitive clinical settings raises profound ethical questions that the research community is only beginning to address systematically. Psychological diagnostics occupies a uniquely delicate position at the intersection of mental health, personal identity, and social functioning. A misdiagnosis or privacy breach in this domain can have far-reaching consequences for an individual's employment, legal status, social relationships, and subjective well-being [3].

The growing body of literature on AI ethics in medicine has produced general frameworks, yet targeted analysis of the psychological diagnostics context remains limited. This research paper aims to identify, categorise, and critically evaluate the main ethical challenges that arise when automated systems are applied in psychological diagnostics, and to propose practical set of guiding principles for responsible implementation.

Overview of Automated Systems in Psychological Diagnostics

Contemporary automated diagnostic systems in psychology can be broadly classified into three categories. First, symptom-based questionnaire platforms use digital adaptations of validated instruments (e.g., PHQ-9, GAD-7, MMPI-2) administered via web or mobile interfaces, with automated scoring and preliminary interpretation. Second, multimodal analysis systems employ computer vision, natural language processing (NLP), and audio analysis to detect affective states and psychopathological markers from unstructured data. Third, clinical decision support systems (CDSS) integrate patient history, lab data, and behavioural observations to generate differential diagnoses or risk scores [4].

Each category carries distinct ethical risks that must be understood within the broader context of psychological practice. The degree of automation, the nature of the data collected, and the extent of clinician oversight all significantly modulate the ethical profile of a given system.

Core Ethical Issues

Table 1 presents a comparative analysis of ethical issues in traditional versus automated psychological diagnostics.

Table 1 – Comparative ethical analysis: traditional vs. automated psychological diagnostics

Category	Traditional Diagnostics	Automated Systems
Data Privacy	Limited data storage; personal contact only	Large-scale digital data collection; increased breach risk
Transparency	Clinician explains reasoning directly	Algorithmic 'black box'; limited explainability
Bias & Fairness	Human bias possible but identifiable	Encoded bias in training data; may disadvantage minorities
Informed Consent	Explicit verbal/written consent	Complex consent forms; technical opacity for users
Professional Accountability	Clear practitioner liability	Diffuse responsibility across developers, institutions, users
Accuracy & Validity	Validated clinical instruments with known limitations	Variable validation; risk of over-reliance on output

Data Privacy and Confidentiality

Psychological health data are among the most sensitive categories of personal information recognised by major data protection frameworks, including the General Data Protection Regulation (GDPR) and the Health Insurance Portability and Accountability Act (HIPAA) [5]. Automated diagnostic systems frequently require access to extensive longitudinal datasets to train accurate models, creating tension between utility and privacy.

The risks are multifaceted. Data breaches can expose highly intimate information. Re-identification attacks on seemingly anonymised datasets remain a genuine concern [6]. Furthermore, secondary uses of diagnostic data by commercial entities without explicit consent represent a systemic ethical failure that regulators have only partially addressed. Federated learning approaches and differential privacy techniques offer partial technical solutions, but their adoption in clinical AI remains limited.

Algorithmic Transparency and Explainability

Many high-performing diagnostic AI systems are based on deep neural networks whose internal decision logic is opaque even to their developers – a phenomenon commonly described as the 'black box' problem. In psychological diagnostics, this opacity is ethically problematic for at least two reasons. First, a clinician who cannot understand the reasoning behind an AI recommendation may be unable to critically evaluate it, effectively ceding clinical judgement to an algorithm. Second, when an automated system produces an incorrect or harmful output, the absence of an interpretable reasoning chain impedes accountability and appeals [7].

Explainable AI (XAI) methods such as LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) offer post-hoc interpretations of model outputs. However, there is ongoing debate about whether such post-hoc explanations accurately reflect the model's true decision process. The development of inherently interpretable models for clinical applications represents an important research priority.

Algorithmic Bias and Fairness

AI systems learn statistical patterns from historical data. When that data reflects existing social inequalities – such as underdiagnosis of depression in men or differential symptom presentation across ethnic groups – the resulting model systematically disadvantages certain populations. Studies have demonstrated that commercial affect recognition systems perform significantly worse for individuals with darker skin tones, raising concerns about racial bias in multimodal diagnostic tools [8].

Bias can enter the diagnostic pipeline at multiple points: data collection, labelling, feature selection, model training, and threshold setting. Addressing bias requires proactive equity audits at each stage, diverse and representative training datasets, and ongoing post-deployment monitoring stratified by demographic variables.

Informed Consent and Patient Autonomy

Ethical clinical practice requires that patients provide meaningful informed consent prior to any diagnostic procedure. In the context of automated systems, achieving genuine informed consent is complicated by the technical complexity of the tools involved, the difficulty of communicating uncertainty and limitations in lay terms, and the power asymmetry between institutions deploying AI and individual service users [9].

Particular attention must be paid to vulnerable populations, including children, individuals with severe cognitive or emotional impairment, and those from low-literacy backgrounds, who may be unable to evaluate consent documents adequately. Dynamic consent frameworks, which allow individuals to update their preferences over time, are emerging as a promising approach but require substantial institutional infrastructure to implement.

Professional Responsibility and Accountability

When an automated system contributes to a harmful diagnostic outcome, responsibility may be distributed among multiple parties: the software developer, the healthcare institution, the clinician who acted on the recommendation, and potentially the regulatory body that certified the product. This diffusion of responsibility can create accountability gaps in which no single actor is held fully liable for harm [10].

Professional regulatory bodies in psychology and psychiatry are beginning to develop guidance on the use of AI-assisted diagnostic tools. Core principles emerging from this work include the requirement that a qualified clinician must retain ultimate diagnostic responsibility and that AI outputs must be treated as decision support rather than authoritative verdicts.

Ethical Framework for Responsible Implementation

Drawing on the analysis above, Table 2 presents an integrated ethical framework mapping bioethical principles to specific requirements and implementation mechanisms for automated diagnostic systems.

Table 2 – Ethical framework for automated psychological diagnostic systems

Ethical Principle	Requirement	Implementation Mechanism
Autonomy	Respect user agency and self-determination	Informed consent; opt-out mechanisms
Non-maleficence	Avoid harm from misdiagnosis	Mandatory human review of AI outputs
Beneficence	Maximize diagnostic benefit	Continuous accuracy monitoring; bias audits
Justice	Equal quality of care across populations	Diverse training data; equity impact assessments
Explainability	Clinicians understand system reasoning	Explainable AI (XAI) methods; audit trails
Data Protection	Secure personal health information	GDPR/HIPAA compliance; data minimization

The framework adapts the four classical principles of bioethics – autonomy, non-maleficence, beneficence, and justice – developed by Beauchamp and Childress [11], and supplements them with principles of explainability and data protection that are particularly salient for AI-based systems. This synthesis acknowledges that no single principle operates in isolation; trade-offs between, for instance, beneficence (maximising diagnostic accuracy through data sharing) and privacy (restricting data use) must be negotiated contextually.

Regulatory alignment represents another critical dimension. The European Union's AI Act classifies AI systems used in psychological assessment as high-risk, imposing mandatory requirements for conformity assessment, technical documentation, transparency, and human oversight before market deployment [12]. Developers and healthcare institutions must treat compliance with such frameworks not as a bureaucratic burden but as a baseline ethical floor rather than a ceiling.

Conclusions

The integration of automated systems into psychological diagnostics offers significant potential benefits, including improved accessibility, greater consistency, and the capacity to detect subtle patterns undetectable by the unaided clinician. However, this potential can only be realised ethically if the fundamental challenges of data privacy, algorithmic transparency, bias, informed consent, and accountability are addressed rigorously.

This analysis demonstrates that these challenges are not merely technical problems amenable to engineering solutions; they are deeply normative issues that require interdisciplinary collaboration among psychologists, AI researchers, ethicists, legal scholars, and patient advocates. The ethical framework presented in Table 2 offers a structured starting point for such collaboration, mapping established bioethical principles to concrete implementation requirements.

Future research should focus on: (1) empirical evaluation of bias and fairness in existing deployed systems; (2) development and validation of domain-specific XAI methods for psychological diagnostics; (3) longitudinal studies of patient outcomes following AI-assisted versus traditional diagnosis; and (4) participatory design methodologies that incorporate the perspectives of mental health service users in the development of diagnostic tools.

Ultimately, the question is not whether automated systems should be used in psychological diagnostics, but under what conditions, with what safeguards, and subject to whose oversight. Answering this question responsibly demands the same intellectual rigour and ethical commitment that the profession of psychology has always brought to the assessment of the human mind.

REFERENCES

1. Torous J., Keshavan M., Gutheil T. Promise and perils of digital psychiatry. Asian Journal of Psychiatry. 2014. Vol. 10. P. 1–3.
2. Cummins N., Scherer S., Krajewski J., Schnieder S., Epps J., Quatieri T.F. A review of depression and suicide risk assessment using speech analysis. Speech Communication. 2015. Vol. 71. P. 10–49.
3. Char D.S., Shah N.H., Magnus D. Implementing machine learning in health care – addressing ethical challenges. New England Journal of Medicine. 2018. Vol. 378. No. 11. P. 981–983.
4. Bzdok D., Meyer-Lindenberg A. Machine learning for precision psychiatry: Opportunities and challenges. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2018. Vol. 3. No. 3. P. 223–230.
5. Price W.N., Cohen I.G. Privacy in the age of medical big data. Nature Medicine. 2019. Vol. 25. No. 1. P. 37–43.
6. Rocher L., Hendrickx J.M., De Montjoye Y.A. Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communications. 2019. Vol. 10. Research paper 3069.
7. Doshi-Velez F., Kim B. Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. 2017.
8. Buolamwini J., Gebru T. Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research. 2018. Vol. 81. P. 77–91.
9. Mittelstadt B.D., Allo P., Taddeo M., Wachter S., Floridi L. The ethics of algorithms: Mapping the debate. Big Data & Society. 2016. Vol. 3. No. 2. P. 1–21.
10. Topol E.J. High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine. 2019. Vol. 25. No. 1. P. 44–56.
11. Beauchamp T.L., Childress J.F. Principles of Biomedical Ethics. 8th ed. New York: Oxford University Press, 2019. 512 p.
12. European Parliament and of the Council. Regulation (EU) 2024/1689 on Artificial Intelligence (AI Act). Official Journal of the European Union. 2024.

Дзюменко Георгій Дмитрович – студент групи 1ПІ-22б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, e–mail: heorhiidziumenko@gmail.com.

Теренчук Анатолій Тимофійович – кандидат технічних наук, доцент, старший викладач кафедри програмного забезпечення, Вінницький національний технічний університет, м. Вінниця, e-mail: anateren59@gmail.com.

Dziumenko Heorhii D. – student of group 1PI-22b, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: heorhiidziumenko@gmail.com.

Terenchuk Anatolii T. – Candidate of Technical Sciences, Associate Professor, Senior Lecturer of the Software Engineering Department, Vinnytsia National Technical University, Vinnytsia, e-mail: anateren59@gmail.com.