

COMMENT

Open Access



# A shift in psychiatry through AI? Ethical challenges

Saskia Wilhelmy<sup>1\*</sup>, Giancarlo Giupponi<sup>2</sup>, Dominik Groß<sup>1</sup>, Klaus Eisendle<sup>3</sup> and Andreas Conca<sup>2</sup>

## Abstract

The digital transformation has made its way into many areas of society, including medicine. While AI-based systems are widespread in medical disciplines, their use in psychiatry is progressing more slowly. However, they promise to revolutionize psychiatric practice in terms of prevention options, diagnostics, or even therapy. Psychiatry is in the midst of this digital transformation, so the question is no longer “whether” to use technology, but “how” we can use it to achieve goals of progress or improvement. The aim of this article is to argue that this revolution brings not only new opportunities but also new ethical challenges for psychiatry, especially with regard to safety, responsibility, autonomy, or transparency. As an example, the relationship between doctor and patient in psychiatry will be addressed, in which digitization is also leading to ethically relevant changes. Ethical reflection on the use of AI systems offers the opportunity to accompany these changes carefully in order to take advantage of the benefits that this change brings. The focus should therefore always be on balancing what is technically possible with what is ethically necessary.

**Keywords** Artificial intelligence, Digital transformation, Psychiatry, Bioethics, Patient–physician relationship

## Background

Digital transformation has taken hold in many areas of society. Increasingly, sophisticated technical innovations enable the continuous use and adaptation of information and communication technology. This technological revolution is omnipresent and generates discourses at the level of society as a whole. Transformation is also evident in medicine, where the era of “medicine 4.0” has already been ushered in, promising greater efficiency in individual patient care, the healthcare system, and

medical research [1]. Buzzwords such as “digitization”, “big data” and “artificial intelligence (AI)” point the way to the future of digital health. While AI-based systems are widely used in disciplines such as radiology [2] or ophthalmology [3], their use in psychiatry, as a “talking” medical discipline, amounts to a Copernican revolution [4]. This revolution brings both new opportunities and ethical challenges.

## Main text

There are reasons why digital transformation is tantamount to a “slow” revolution for psychiatry. Brunn et al. were able to identify challenges that have an influence on the integration of AI applications: skeptical attitudes of psychiatrists toward AI, potential obsolescence of psychiatrists, and potential loss of definitional authority through AI [5]. User acceptance has a pivotal impact on the implementation of AI. Furthermore, technologies reflect and influence social structures—e.g. they shape communication and interpersonal relationships [6]. At the same time, they can give rise to developments that

\*Correspondence:

Saskia Wilhelmy  
swilhelmy@ukaachen.de

<sup>1</sup> Institute for History, Theory and Ethics in Medicine, University Hospital, RWTH Aachen University, Wendlingweg 2, 5074 Aachen, Germany

<sup>2</sup> Academic Teaching Department of Psychiatry, Central Hospital, Sanitary Agency of South Tyrol, Via Lorenz Böhler 5, 39100 Bolzano, Italy

<sup>3</sup> Institute of General Practice and Public Health, Provincial College for Health Professions Claudiana, Lorenz-Böhler-Straße 13, 39100 Bolzano, Italy



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

in turn can become drivers of insecurities and pathologies [7]. This continues to raise questions about the interdependence of technologies and society, thus also for psychiatry and which changes it will be subject to [5]. A survey of psychiatrists on the impact of AI and machine learning (ML) addressed this question. The study found that one in two psychiatrists predicted that their professional field will change significantly in the future. The majority of respondents do not believe that AI/ML could or will ever replace their work as psychiatrists, but that time-consuming work (e.g. documenting) will be transferred to AI/ML systems [8].

In addition to the physical, psychiatry focuses on the psyche and the brain of humans. In diagnostics and therapy, it thus faces the challenge of identifying and taking into account factors that ultimately influence the human psyche and brain. This is a challenge that has not been met primarily by technology. Nonetheless, or precisely because psychiatry is directly intertwined with the social matrix, AI-powered technology has found its way into it. In particular, this has been spurred by phenomena that psychiatry has faced in recent years, leading to calls for supportive or transformative technologies; e.g. regarding the COVID-19 pandemic, natural disasters, or war conflicts [9]. Digitalization has a stake in the crisis-ridden social matrix and at the same time embodies the tool as part of the coping process. This has led to increased engagement in research and clinical implementation of innovative technologies, which come with challenges, e.g. regarding research ethics standards such as transparency or reproducibility of information [10].

The emergence of technological innovations has thus triggered a dynamic in medicine in which the assessment of the use of such systems constantly oscillates between opposites: opportunities and hope on the one hand, risks and skepticism on the other [11]. A conclusive evaluation, e.g. with regard to possible risks or benefits, is often not possible, but rather a constant evaluation of the technology used is required. This is essential due to the rapid technological progress, which has also led to an acceleration and complexity of knowledge in medicine: today, medicine has a half-life of about 1–2 years, in the future it will certainly be even shorter [12].

#### **Ethical considerations on AI**

Technological upheavals, such as the introduction of AI in society, simultaneously generate ethical challenges, which the European Union addressed in 2019 by introducing general ethical guidelines for the development, deployment and use of AI [13]: “Its central concern is to identify how AI can advance or raise concerns to the good life of individuals, whether in terms of quality of life, or human autonomy and freedom necessary for a

democratic society” [13]. These guidelines concern the society as a whole, which is why they are formulated in an open manner, with the indication that they can be adapted and evaluated depending on the scope of application of AI. In addition to fundamental rights (such as respect for human dignity), the guidelines specify four non-hierarchical ethical principles that should be considered: respect for human autonomy, prevention of harm, fairness, and explicability [13]. These principles, which serve to protect humans interacting with AI, reflect ethical values that are also relevant in medicine when dealing with patients and can be found in the “principlism” established by Beauchamp and Childress. Their principles include (1) respect for human autonomy, (2) non-maleficence, (3) beneficence, and (4) justice [14]. Unlike Beauchamp and Childress, the European Commission uses the principles mentioned to be taken into account as fixed values and not to be weighed against each other. When considering and evaluating AI applications in psychiatry, it makes sense to consult not only general but also medical-specific ethical guidelines—especially when AI is used with vulnerable groups. To this end, various ethical criteria in the use of technology in medicine provide guidance for conducting an ethical evaluation, e.g. with regard to self-determination, safety, privacy, or fairness [15, 16].

#### **Ethical challenges in psychiatry: doctor–patient interaction and AI**

But what ethical challenges arise from the use of AI in psychiatry? This question aims at the ethical acceptability of using AI systems in this context. Various stakeholders (such as patients, relatives, or medical, nursing, and technical staff) involved in psychiatry and in the digitization process play a role. To illustrate the changes in the interpersonal interaction of these stakeholders, the doctor–patient relationship is considered as an example.

#### **Psychiatrist’s perspective**

Physicians have always had sovereign power in medical diagnosis and treatment. This expertise will undoubtedly be strengthened by the use of AI-based systems for the time being in terms of optimization. Thus, it is already possible to provide more objectified and more complex diagnostics as well as personalized prognosis [17]—for example, referring to biomarkers (e.g. clinical, imaging, genetics), psycho-markers (e.g. personality traits, cognitive functioning), and social markers (type of social media use) in classifying certain mental disorders [17–19]. In the near future, psychiatrists will consciously and transparently shape their mediating role between the AI-generated expertise and the ethical decision-making process in the sense of patient autonomy.

**Table 1** AI involvement in psychiatry: a selection of ethical challenges

	Ethical challenges	Type of AI involvement
Prevention	Harm prevention/safety [21], efficacy [22], data security [23]	Suicide prevention (e.g. based on the electronic health records) [24]
Diagnosis prognosis	Transparence (“black box”), data privacy [4], patient autonomy [18, 19]	Use of complex/multidimensional biomarkers [22], use of diagnostic apps
Therapy	Lack of human contact, safety [19]	Chatbots/robots as “therapists” [24]
Education	Responsibility [23], lack of guidance and training [21]	Digital competence in the field of AI necessary
Research	Participants’ data security, informed consent, data privacy, safety [25]	Study participation, data collection (e.g. monitoring)

**Patient’s perspective**

In recent years, patients have matured into medical “lay experts” who use digital tools and the Internet in particular to acquire knowledge and derive actions or treatments. For example, AI-powered apps that are easily accessible to smartphone users expand patient empowerment in this regard and shape trust by making physicians’ actions verifiable [18]. How the free will to decide can be guaranteed, however, remains a central topic of the situational as well as the developing doctor–patient relationship. Not only physicians and patients grow and learn, but also ML or even Deep Learning (DL) are trainable technologies and, like humans, must be continuously subjected to the learning process [20]. As a consequence, this can also improve the interaction and trust relationship with AI.

For psychiatry, various ethical challenges (Table 1) to which the doctor–patient relationship is subject arise or intensify not only in the areas of prevention, diagnosis/prognosis, and therapy, but also in the areas of education and research.

**Conclusions**

AI systems are currently one of the most important emerging technologies. Digital technologies should be considered not only as tools, but also as an acquired part of their users’ identity (e.g. viewing the smartphone as a “mobile identity”, i.e. a close, identity-forming connection between a person and technology) [26]. Accordingly, the goal should be not only to get lost in transhumanistic optimization, but also to follow the path of transformation. As we are in the midst of this change, the question is no longer “whether” technology should be used, but “how” we can use it to meet goals of progress or improvement. The focus should therefore always be on weighing technological possibility against ethical necessity. The European Commission’s ethical guidelines provide initial, but not exhaustive guidance, and the four principles of

biomedical ethics provide a more concrete patient-centered and medical-practice view for AI systems in psychiatry. However, when considering the benefits and risks of using AI systems, it should always be checked whether the technology also stands up to ethical evaluation. In this context, Jotterand/Bosco (2020) basically claims that technological solutions should only be applied in medicine if they incorporate the ethical imperative of humanity and thus fulfill three requirements: technology serves human purposes, respects personal identity, and promotes human interaction [27].

**Abbreviations**

- AI Artificial intelligence
- COVID-19 Coronavirus disease 2019
- DL Deep learning
- ML Machine learning

**Author contributions**

Conceptualization and design: SW and AC; initial draft preparation: SW and AC; editing and review: all authors. All authors read and approved the final manuscript.

**Funding**

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

**Availability of data and materials**

Not applicable.

**Declarations**

**Ethics approval and consent to participate**

Because of the type of article (Commentary), there was no need for obtaining ethical approval.

**Consent for publication**

Not applicable.

**Competing interests**

The authors have no competing interests to report.

Received: 19 July 2023 Accepted: 24 October 2023  
Published online: 02 November 2023

## References

- Ioppolo G, Vazquez F, Hennerici MG, Andrés E. Medicine 4.0: new technologies as tools for a society 5.0. *J Clin Med*. 2020. <https://doi.org/10.3390/jcm9072198>.
- Fazekas S, Budai BK, Stollmayer R, Kaposi PN, Bérczi V. Artificial intelligence and neural networks in radiology—basics that all radiology residents should know. *Imaging*. 2022;14:73–81.
- Anton N, Doroftei D, Curteanu S, Catălin L, Ilie O-D, Tărcoveanu F, Bogdănici CM. Comprehensive review on the use of artificial intelligence in ophthalmology and future research directions. *Diagnostics*. 2022;13:100. <https://doi.org/10.3390/diagnostics13010100>.
- Hariman K, Ventriglio A, Bhugra D. The future of digital psychiatry. *Curr Psychiatry Rep*. 2019;21:88. <https://doi.org/10.1007/s11920-019-1074-4>.
- Brunn M, Diefenbacher A, Courtet P, Genieys W. The future is knocking: how artificial intelligence will fundamentally change psychiatry. *Acad Psychiatry*. 2020;44:461–6.
- Baumann Z. *Liquid modernity*. Oxford: Polity Press; 2000.
- Ehrenberg A. *La Société du Malaise. II Mental et le Social [The malaise society. The mental and the social]*. Paris: Odile Jacob; 2010.
- Doraiswamy PM, Blease C, Bodner K. Artificial Intelligence and the future of psychiatry: insights from a global physician survey. *Artif Intell Med*. 2020;102: 101753. <https://doi.org/10.1016/j.artmed.2019.101753>.
- Ćosić K, Popović S, Šarlija M, Kesedžić I. Impact of human disasters and COVID-19 pandemic on mental health: potential of digital psychiatry. *Psychiatr Danub*. 2020;32:25–31.
- Tornero-Costa R, Martínez-Millana A, Azzopardi-Muscat N, Lazeri L, Traver V, Novillo-Ortiz D. Methodological and quality flaws in the use of artificial intelligence in mental health research: systematic review. *JMIR Ment Health*. 2023. <https://doi.org/10.2196/42045>.
- Lee EE, Torous J, De Choudhury M, Depp CA, Graham SA, Kim H-C, Paulus KJH, Jeste DV. Artificial intelligence for mental healthcare: clinical applications, barriers, facilitators, and artificial wisdom. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2021;6:856–64.
- Colacino C. *Medicine in a Changing World*. 2017. <https://hms.harvard.edu/news/medicine-changing-world>. Accessed 10 July 2023.
- High Level Expert Group on Artificial Intelligence. *Ethics guidelines for trustworthy AI*. European Commission. 2019. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. Accessed 10 July 2023.
- Beauchamp TL, Childress JF. *Principles of biomedical ethics*. 8th ed. Oxford: Oxford University Press; 2019.
- Karimian G, Petelos E, Evers SMAA. The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review. *AI Ethics*. 2022;2:539–51.
- Marckmann G. *Ethische Aspekte von eHealth [Ethical aspects of eHealth]*. In: Fischer F, Krämer A, editors. *eHealth in Deutschland Anforderungen und Potenziale innovativer Versorgungsstrukturen*. Springer. Berlin: Requirements and potentials of innovative care structures; 2016.
- Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim H-C, Jeste DV. Artificial intelligence for mental health and mental illnesses: an overview. *Curr Psychiatry Rep*. 2019;21:116. <https://doi.org/10.1007/s11920-019-1094-0>.
- Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, Carvalho AF, Keshavan M, Linardon J, Firth J. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry*. 2021;20:318–35.
- Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol*. 2018;14:91–118.
- Lovejoy CA, Arora A, Buch V, Dayan I. Key considerations for the use of Artificial Intelligence in healthcare and clinical research. *Future Healthc J*. 2022;9:75–8.
- Fiske A, Henningsen P, Buys A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *J Med Int Res*. 2019;21: e13216. <https://doi.org/10.2196/13216>.
- Kretzschmar K, Tyroll H, Pavarini G, Manzini A, Singh I. Can your phone be your therapist? Young People's ethical perspectives on the use of fully automated conversational agents in mental health support. *Biomed Inform Insights*. 2019. <https://doi.org/10.1177/1178222619829083>.
- Lejeune A, Le Glaz A, Perron P-A, Sebt J, Baca-Garcia E, Walter M, Lemey C, Berrouguet S. Artificial intelligence and suicide prevention: a systematic review. *Eur Psychiatry*. 2022;65:1–22.
- Fakhoury M. Artificial intelligence in psychiatry. *Adv Exp Med Biol*. 2019;1192:119–25.
- Jacobson NC, Bentley KH, Walton A, Wang SB, Fortgang RG, Millner AJ, Coombs G, Rodman AM, Coppersmith DDL. Ethical dilemmas posed by mobile health and machine learning in psychiatry research. *Bull World Health Organ*. 2020;98:270–6.
- Lou J, Han N, Wang D, Pei Y. Effects of mobile identity on smartphone symbolic use an attachment theory perspective. *Int J Environ Res Public Health*. 2022. <https://doi.org/10.3390/ijerph192114036>.
- Jotterand F, Bosco C. Keeping the “human in the loop” in the age of artificial intelligence: accompanying commentary for “correcting the brain?” by Rainey and Erden. *Sci Eng Ethics*. 2020;26:2455–60.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

