**COMMENTARY**

# The Future is Knocking: How Artificial Intelligence Will Fundamentally Change Psychiatry

Matthias Brunn [1] · Albert Diefenbacher [2] · Philippe Courtet [1] · William Genieys [3]

Artificial intelligence (AI) is inducing a profound transformation of both the practice and structure of medicine. This implies changes in tasks, where certain processes may be taken over by AI applications, as well as novel ways of collaborating and integrating information. Consider a recent example where AI is used to avoid suicide attempts by using smartphones' native sensors and signal processing techniques [1]. This new suicide prevention technique requires the psychiatrist to acquire new skills (handling and interpreting continuous patient data sent by a dedicated application) and interact with new actors (programmers, data managers, etc.). Further, the abundance of individual patient data may contribute to a shift in conceptualizing care—from the traditional identification of general risk factors towards more tailored prevention strategies in the sense of personalized medicine [1].

Thus, unlike past technologies, AI has the potential to not only enhance medical capacity but also change the way health professionals are organized and embedded into the broader medical context. In particular, the implementation of AI applications is leading to a redistribution and renegotiation of responsibilities—and thus power—both within medicine and in relation with other stakeholders. This article discusses the future impact of AI on psychiatry, highlights the challenges for research, and outlines perspectives for the next generation of psychiatrists.

## Artificial Intelligence: General Background

*AI* is a term in computer science that refers to a system that can reason, learn, and plan, and which exhibits behavior which we associate with biological intelligent systems. *Machine learning* refers to a programming approach in computer science in which the behavior of a program is not fully determined by the code but can adapt its behavior (i.e., learn) based on the input data. *Deep learning* is a particular variant of machine learning which is often modelled on *artificial neural networks*. The latter typically consist of interconnected nodes—representing artificial neurons—with an input layer, hidden layers, and an output layer. In the hidden layers, data from the input layer undergo transformations multiple times [2].

AI, and in particular deep learning techniques, is thus distinct from other, more "linear" technical innovations such as coronary catheterization. The reason is AI's learning capacity that allows for recursive self-improvement. This requires us to shift from viewing these new technologies as simple objects. In fact, in sociologic terms, AI devices may become quasi-social actors with agency and contribute to the re-organization of a wider social system [3].

Using an example from the transport sector, linear technical innovations can make a car faster by improving the engine, which does not change the way cars are perceived and handled. Conversely, self-driving cars and the decisions they make significantly impact the role of human drivers in terms of their responsibilities and interactions with other road users, which potentially challenges the model of who is in charge: the driver or the car. The disruptive quality of AI innovations further lies with its unprecedented speed, surpassing the expectations that we usually base on past experience. Indeed, while still under development in early 2016, effective instant language translation using Internet-based machine learning is now offered free of charge by several providers, with major implications for the labor market [4].

✉ Matthias Brunn
matthias.brunn@umontpellier.fr

[1] University of Montpellier, Montpellier, France

[2] Charité Universitätsmedizin Berlin, Berlin, Germany
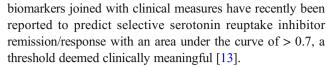
[3] Sciences Po, Paris, France

In the medical field, an emblematic example of the impact of AI on the work of physicians is radiology, which by its very nature relies extensively on machines and where the transformation of the profession already has begun. Current artificial neural networks have accuracy rates that surpass those of human radiologists in tasks such as mammogram reading for breast cancer screening, as reported in a recent large-scale study funded and co-conducted by Google [5]. This raises, on the one hand, concerns about reduced demand for radiologists. On the other hand, careers are built with AI as a vantage point as academic medical centers now host their first generation of radiology professors who are experts on AI in medical imaging.

The dimensions of potential job displacement have recently been estimated for high-income countries and appear significant. Although lower than the average of all occupations, the mean probability of automation for health professionals is currently estimated at 35%, and it rises to 45% for paramedical professions [6]. Across sectors, experts expect these changes to occur within 5–10 years [4].

## Artificial Intelligence and Psychiatry

Potential applications of AI in psychiatry can be broadly grouped into two categories [7]. One is natural language processing, which enables computers to understand, interpret, and manipulate human language. Research in this field has advanced significantly thanks to the vast amount of text available on the Internet along with exponentially increasing computing power. For example, researchers used natural language processing to analyze the speech patterns of 34 individuals at high risk for psychosis in a proof-of-concept exploration. Their predictive analytics algorithm outperformed clinical ratings in the prediction of psychosis, a finding cross-validated in a larger sample [8]. Another application are chatbots—digital conversational agents that use AI methods via text and/or voice to mimic human behavior through evolving dialogue. They are seen as a means to provide mental health care in regions with low access to medical care or to persons who have difficulties disclosing their feelings to a human being [9]. Chatbots providing cognitive behavioral therapy in a non-clinical college population were shown to be effective in reducing symptoms of depression and anxiety [10].

The second category of applications involves AI for the integration of diverse biomarkers (clinical, imaging, genetics, etc.) in classifying certain disorders [11]. In the case of dementia, for example, a recent review on deep learning techniques for the early detection and automated classification of Alzheimer's disease has found accuracies of up to 98.8% using a combination of MRI, PET, and CSF as markers [12]. In the case of depression, machine learning–based algorithms that included functionally validated pharmacogenomic

biomarkers joined with clinical measures have recently been reported to predict selective serotonin reuptake inhibitor remission/response with an area under the curve of > 0.7, a threshold deemed clinically meaningful [13].

While many other examples exist, the current state of research on AI applications in psychiatry and its transfer into clinical practice is not nearly as advanced as in other fields such as medical imaging. A recent review on this question concludes that there is a high potential for AI in mental health care but most studies are still in a stage of early proof-of-concept [14]. The research trend is strong however, with a 250% increase in articles on "artificial intelligence and psychiatry" between 2015 and 2019 in PubMed. It seems likely that this trend will continue and foster the introduction of AI applications into routine psychiatric care.

## Challenges for the Discipline

The challenges that such applications hold for psychiatrists can be related to at least four dimensions. The first concerns the attitudes of psychiatrists towards AI. An international survey by Doraiswamy et al. with 791 respondents found that most psychiatrists were skeptical that AI could perform complex psychiatric tasks as well as or better than human doctors, and only 4% respondents thought it was likely that future technology would make their jobs obsolete. This led the authors to suggest that psychiatrists may underestimate the speed of progress and therefore lack preparedness [15]. These findings are relevant since, on the one hand, it is well understood that attitudes towards a new technology will determine its degree of adoption [16]. On the other hand, as a consequence, this could mean that AI applications for mental health will be implemented anyways—but without psychiatrists. This is the case with the abovementioned chatbot, which has been developed by a private corporation outside the medical field. While this may not automatically result in competition (when the chatbot is used e.g. with refugees who have very limited access to regular mental health care), it raises important professional questions.

This leads to the second dimension, which is the potential obsolescence of psychiatrists. Given their specific skillset—including, notably, complex social skills—it seems likely that psychiatrists may actually be relatively well sheltered from job displacement. Indeed, psychiatry requires greater integration of cultural and psychosocial factors than other, more pattern-based disciplines [17]. Hence, in a perspective where competencies that are complementary to machine prediction will become more valuable in the future while competencies that are substitutes for machine prediction will become less valuable [18], psychiatrists could capitalize on the potential benefits of AI in psychiatric practice. These include, as proposed by Kim et al. and illustrated by our examples, (1) potentially

better diagnoses (particularly in initial patient evaluation including early recognition, and assessment of psychopharmacology) and (2) potentially better outcomes due to improved capacity to select viable treatments (either psychotherapeutic or psychopharmacological treatment, as well as professional social support systems) based on diagnostic criteria, thus reducing human error or vulnerability to bias [19]. Without having sought to, psychiatry may become a model for other disciplines—where the primary focus is largely on communication and interaction with the patient, and where intuition, empathy, and abstraction are valued and sheltered assets. By using these virtues, AI may actually help attract more students to the discipline.

Third, while our examples so far have largely remained within the hospital walls, we must consider an important trend in the outside realm—the increasing use and role of social media and its particular implications for AI in psychiatry. The reason lies with the very nature of social media—to harness and share emotions—which means that tremendous amounts of data (which can be text, but also shared content or "likes") are available in real time and can be linked to emotional states [20]. As a consequence, social media have become a space where prediction and diagnosis of mental disorders can take place outside of the traditional medical domain. Facebook has employed suicide risk screening of users since 2017 which likely relies upon deep learning algorithms, although no details on the program have been provided by the company. In terms of medical diagnosis, a recent study using machine learning has demonstrated that content shared on Facebook can predict the documented onset of depression with fair accuracy (area under the curve = 0.69) [21]. The overall evidence base for the efficacy of this type of algorithms appears thus solid, and the authors of a recent application study underscore that that the question is not anymore *if* technologies like this will be implemented, but *how* [22], to which we may add *by whom*.

Finally, the impact of AI on psychiatry could concern the very foundations of the discipline: the definition of mental illness. This point reflects recurrent critiques of diagnostic classifications in psychiatry, drawing on issues such as reliability [23] and diffuse symptom expression in early stages of illness [24]. In this context, the potential role of AI could at first be to "refine" nosological schemes in the sense of personalized medicine. In this scenario, using traditional diagnostic labels, AI can assist in identifying novel biomarkers which, in turn, may allow to better identify variation in phenotypes and targeted treatment options [25, 26]. More radically then, some argue that AI could be a means to abandon diagnostic labels altogether and rely on alternative concepts such as functional domains as proposed in the Research Domain Criteria [27]. Using this correlation matrix is seen as key to exposing hidden, previously unknown data features and to the creation of more accurate clinical outcome prediction [28].

Alternatively, AI could allow identifying entirely new demarcations that result from the full integration of the vast data domains (clinical, social, neurobiology, imaging, genetics, etc.) available [25]. This could rely on autoencoders—neural networks that can automatically learn how to extract features from diverse data sources. For example, a study linking GPS data to depression has demonstrated that autoencoders, using raw input data, are more precise in predicting depressive states than models with predefined mobility features (such as total distance covered or number of places visited) [29].

## Challenges and Questions for Research

These examples highlight the challenges and implications of deployment of AI for psychiatrists. They underscore the need of a more detailed and structured understanding of the dynamic that AI unfolds in terms of social and organizational factors. We group these challenges in three broad categories and suggest avenues to explore them further.

The first concerns the internal organization of physicians in terms of disciplines, tasks, and training. Indeed, we have noted different attitudes towards the implementation of AI which implies that there is a continuum of how the adoption of AI is perceived among different medical disciplines. We can further hypothesize that there is differential adoption within psychiatry, for example along the lines of practice traditions (biological psychiatry, social psychiatry, etc.). What are the factors shaping such a continuum? How do these factors affect the balance among groups? Will there be an accentuation of silo-thinking, or can AI foster trans-disciplinary, transprofessional collaboration? These questions evoke a strand of literature which views the medical profession as constituted by different segments (largely equivalent to disciplines) with distinctive identities and goals. They organize activities which will secure an institutional position, and the organization of the profession shifts with the competition and conflict of these segments in movement [30, 31].

The second category of challenges includes the link of innovation and leadership. Will AI facilitate the surge of new leaders who will challenge the established order? These changes may be as disruptive as the technological innovations. What would be the profile of this new generation of leaders and managers and, more broadly, what are the competencies that characterize them? One way to conceptualize the role of innovation in social change is the notion of entrepreneurs [32], theorizing such individuals as driven by an intrinsic motivation for innovation, but also by the desire for power. In pursuing a certain vision of desired behaviors, by use of innovations, entrepreneurs engage in "creative destruction" of established rules, orders, and values. Understanding these aspects will inform the debate on change management in health organizations such as hospitals, for which the literature

emphasizes the facilitating role of strong leadership in creating a context receptive for change [16].

The third category of challenges concerns the interaction with professions and institutions outside the medical domain. Indeed, the innovations discussed in our illustrations also rely significantly on the expertise of people such as software engineers. In the case of chatbots, for example, the majority of research is presented at engineering conferences, outside of traditional medical publication outlets [9]. At the same time, the expertise on AI applications is currently heavily concentrated around private corporations. Will this mean a redistribution of competencies between the medical sphere and others, drawing on new sources of legitimacy such as programming and data management skills? This echoes the idea that the power of physicians is rooted in their ability to define a coherent set of knowledge around the art and science of medicine, strongly linked to the capacity of the profession to organize itself in an effective manner [33]. As such, medicine in general and psychiatry in particular are then in constant competition with other occupations and professions claiming the legitimacy to address social phenomena such as disease and illness [34].

## Perspectives for Future Psychiatrists and Their Teachers

While we have so far raised many critical questions, we believe that there are numerous ways to prepare psychiatry for the advent of AI. This includes a broad and timely discussion of challenges in order to develop an evidence base for informed decisions. Addressing the challenges requires research on the qualitative aspects of AI's role in psychiatry as well as empirical and conceptual work on the link between innovation and social change—from the level of frontline implementation up to the realm of national policy making. In this process, academic medicine will play a double role. On the one hand, academic medical centers are the place where most innovations are tested, implemented, or even generated. On the other hand, medical schools and their affiliated facilities are key to the socialization of the profession and thus strongly structure its internal organization. Academic medicine is setting norms—which will be altered by, and adapted to, the transformative nature of AI.

One crucial challenge in this process is that of privacy and ethical issues, which is particularly relevant in the vulnerable type of population that psychiatric patients represent. The tradeoff between potential prevention to be gained from AI applications and the related data privacy problems should become a key theme in training. This is an opportunity to emphasize a trademark of medical deontology—the intimacy of the patient-practitioner dyad and the medical secret—in opposition to the current practice of digital corporations which

consists in "running codes and apologies" (i.e., first establishing new techniques without seeking ex ante consensus on privacy and other issues, and then dealing with the effects later). Including user experience in such teaching units, in particular in domains such as suicide prevention, seems a particularly worthwhile avenue to pursue and could also enhance the "data literacy" [2] of the population in question.

In order to be effective, such specific training contents should be as concrete as possible. One way to do so is by associating trainees hands-on in the design, operation, and evaluation of AI applications to psychiatry. One example is a study mentioned at the beginning of this article, initiated and operated by two psychiatric teaching hospitals in France and Spain. It combines mobile-health and AI methods to avoid suicide attempts by using smartphones' native sensors, advanced machine learning, and signal processing techniques in order to identify suicide risk [1]. Mobilizing scientists and clinicians from various domains, it is an illustration of linking research and training; residency programs should benefit from such projects by proposing short-term rotations or longer research assignments. Again, this training content could and should be enhanced by including user experience. In discussing directly with patients and/or non-symptomatic "targets" of AI applications, psychiatric trainees are likely to gain the most sustainable understanding of the issues at play. This would allow them to become literate in AI and turn them into empowered stakeholders—in a setting, as we have argued, where many segments inside and outside the medical profession will make their claim. While we focus here on the question of residency, the foundations of this literacy must, however, be laid much earlier—during statistics courses at medical school or, even better, at college level. To this end, AI techniques can be introduced as one possibility among others to address issues in population health or evidence-based medicine [35]. Across all stages of education, hackathons (coding competitions of small teams around a given theme) have recently emerged as a means of engaging a variety of profiles (students, entrepreneurs, scientists, etc.) in an alternative format that makes medical innovation education more accessible and easily adoptable for academic medical centers [36]. This format appears particularly well suited for training in the domain of AI and, in addition, provides the opportunity to assess participants' performance in a team-based and goal-oriented environment.

The implementation of such dedicated teaching will not necessarily require substantial resources. Instead, psychiatric departments will need to activate trans-disciplinary expertise from data science, engineering, and ethics, or consider to generate such expertise in-house. The latter represents a key bottleneck in the preparation of future psychiatrists for AI, and this extends to the fact that role models among senior staff for the "good" use of AI applications are still very rare. The "training of the trainers," as well as continuous education at

large, therefore represents another key priority for academic psychiatry centers. There are currently only few institutions offering short or online courses in this domain, and psychiatrists should further or even initiate their set-up. In this context, as discussed above, the cooperation with private organizations should require a very careful examination of their interests and ethical standards, and medical boards may provide a venue to assist with such arbitrations.

Taken together, this article has outlined several promises and challenges of AI for psychiatry. This includes changes to the tasks, professional identity, and remit of psychiatrists, inherently linked to questions of socialization and training. Considering psychiatry as part of a wider social system, rather than navigating inside a disciplinary "bubble," may prove very helpful in addressing these challenges. In the meantime, while not shaping most of routine practice yet, the tracks for a major transformation are being laid via the ongoing development and testing of AI applications. And it is high time to develop psychiatrists into empowered stakeholders of this transformation.

## Compliance with ethical standards

## References

1. Berrouiguet S, Barrigón ML, Castroman JL, Courtet P, Artés-Rodríguez A, Baca-García E. Combining mobile-health (mHealth) and artificial intelligence (AI) methods to avoid suicide attempts: the Smartcrises study protocol. BMC Psychiatry. 2019;19:277.

2. Kellmeyer P. Big brain data: on the responsible use of brain data from clinical and consumer-directed neurotechnological devices. Neuroethics [Internet]. 2018 [cited 2019 Dec 13]; Available from: https://doi.org/10.1007/s12152-018-9371-x

3. Latour B. Reassembling the social: an introduction to actor-network-theory. Oxford: Oxford University Press; 2007.

4. Baldwin R. The globotics upheaval: globalization, robotics, and the future of work. Oxford: Oxford University Press; 2019.

5. McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafian H, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020;577:89–94.

6. Nedelkoska L, Quintini G. Automation, skills use and training. OECD Social, Employment and Migration Working Papers [Internet]. 2018 [cited 2019 Oct 11]; Available from: https://www.oecd-ilibrary.org/employment/automation-skills-use-and-training_2e2f4eea-en

7. Meyer-Lindenberg A. Künstliche Intelligenz in der Psychiatrie – ein Überblick. Nervenarzt. 2018;89:861–8.

8. Corcoran CM, Carrillo F, Fernández-Slezak D, Bedi G, Klim C, Javitt DC, et al. Prediction of psychosis across protocols and risk cohorts using automated language analysis. World Psychiatry. 2018;17:67–75.

9. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatr. 2019;64:456–64.

10. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health [Internet]. 2018 [cited 2020 Feb 28];5. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6315222/

11. Vieira S, Pinaya WHL, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: methods and applications. Neurosci Biobehav Rev. 2017;74:58–75.

12. Jo T, Nho K, Saykin AJ. Deep learning in Alzheimer's disease: diagnostic classification and prognostic prediction using neuroimaging data. Front Aging Neurosci [Internet]. 2019 [cited 2019 Nov 26];11. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6710444/

13. Athreya AP, Neavin D, Carrillo-Roa T, Skime M, Biernacka J, Frye MA, et al. Pharmacogenomics-driven prediction of antidepressant treatment outcomes: a machine-learning approach with multi-trial replication. Clin Pharmacol Ther. 2019;106:855–65.

14. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim H-C, et al. Artificial intelligence for mental health and mental illnesses: an overview. Curr Psychiatry Rep. 2019;21:116.

15. 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.

16. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q. 2004;82:581–629.

17. Doraiswamy PM, Blease C, Bodner K. Artificial intelligence and the future of psychiatry: insights from a global physician survey. Cornell University arXiv:190712386 [Internet]. 2019 [cited 2019 Sep 4]; Available from: http://arxiv.org/abs/1907.12386

18. Li D, Kulasegaram K, Hodges BD. Why we needn't fear the machines: opportunities for medicine in a machine learning world. Acad Med. 2019;94:623–5.

19. Kim JW, Jones KL, D'Angelo E. How to prepare prospective psychiatrists in the era of artificial intelligence. Acad Psychiatry. 2019;43:337–9.

20. Boullier D, Lohard A. Opinion mining et Sentiment analysis: Méthodes et outils. Marseille: OpenEdition Press; 2012.

21. Eichstaedt JC, Smith RJ, Merchant RM, Ungar LH, Crutchley P, Preoțiuc-Pietro D, et al. Facebook language predicts depression in medical records. PNAS. 2018;115:11203–8.

22. Coppersmith G, Leary R, Crutchley P, Fine A. Natural language processing of social media as screening for suicide risk. Biomed Inform Insights. 2018;10:1178222618792860.

23. Grove WM, Andreasen NC, McDonald-Scott P, Keller MB, Shapiro RW. Reliability studies of psychiatric diagnosis. Theory and practice. Arch Gen Psychiatry. 1981;38:408–13.

24. Wigman JTW, van Os J, Thiery E, Derom C, Collip D, Jacobs N, et al. Psychiatric diagnosis revisited: towards a system of staging and profiling combining nomothetic and idiographic parameters of momentary mental states. PLoS One. 2013;8:e59559.

25. Durstewitz D, Koppe G, Meyer-Lindenberg A. Deep neural networks in psychiatry. Mol Psychiatry. 2019;24:1583–98.

26. Menke A. Precision pharmacotherapy: psychiatry's future direction in preventing, diagnosing, and treating mental disorders. Pharmgenomics Pers Med. 2018;11:211–22.

27. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research Domain Criteria (RDoC): toward a new classification framework for research on mental disorders. AJP. 2010;167:748–51.

28. Tai AMY, Albuquerque A, Carmona NE, Subramanieapillai M, Cha DS, Sheko M, et al. Machine learning and big data: implications for disease modeling and therapeutic discovery in psychiatry. Artif Intell Med. 2019;99:101704.

29. Mehrotra A, Musolesi M. Using autoencoders to automatically extract mobility features for predicting depressive states. Proc ACM Interact Mob Wearable Ubiquitous Technol. 2018;2:127:1–127:20.

30. Bucher R, Strauss A. Professions in process. Am J Sociol. 1961;66: 325–34.

31. Gritzer G, Arluke A. The making of rehabilitation: a political economy of medical specialization, 1890-1980. Berkeley: University of California Press; 1985.

32. Schumpeter JA. Capitalism, socialism and democracy. London: Routledge; 2010.

33. Freidson E. Profession of medicine: a study of the sociology of applied knowledge. Chicago: University of Chicago Press; 1988.

34. Abbott A. The system of professions: an essay on the division of expert labor. Chicago: University of Chicago Press; 1988.

35. Kolachalama VB, Garg PS. Machine learning and medical education. NPJ Digit Med [Internet]. 2018 [cited 2019 Oct 18];1. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6550167/

36. Wang JK, Pamnani RD, Capasso R, Chang RT. An extended hackathon model for collaborative education in medical innovation. J Med Syst. 2018;42:239.