Psychological Medicine

cambridge.org/psm

Correspondence

[‡]The online version of this article has been updated since original publication. A notice detailing the changes has also been published at https://doi.org/10.1017/S0033291721000684.

Cite this article: Gauld C, Micoulaud-Franchi J-A, Dumas G (2020). Comment on Starke et al.: 'Computing schizophrenia: ethical challenges for machine learning in psychiatry': from machine learning to student learning: pedagogical challenges for psychiatry. *Psychological Medicine* 1–3. https://doi.org/10.1017/S0033291720003906

Received: 1 September 2020
Accepted: 29 September 2020

Key words:
Artificial intelligence; computational psychiatry; education; epistemology; ethics; health care

Author for correspondence:
Gauld Christophe,
E-mail: gauldchristophe@gmail.com

CAMBRIDGE UNIVERSITY PRESS

Comment on Starke et al.: 'Computing schizophrenia: ethical challenges for machine learning in psychiatry': from machine learning to student learning: pedagogical challenges for psychiatry[‡]

Christophe Gauld[1,2], Jean-Arthur Micoulaud-Franchi[3,4] and Guillaume Dumas[5,6]

[1]Department of Psychiatry, University of Grenoble, Avenue du Maquis du Grésivaudan, 38 000 Grenoble, France; [2]UMR CNRS 8590 IHPST, Sorbonne University, Paris 1, France; [3]University Sleep Clinic, Services of functional exploration of the nervous system, University Hospital of Bordeaux, Place Amélie Raba-Leon, 33 076 Bordeaux, France; [4]USR CNRS 3413 SANPSY, University Hospital Pellegrin, University of Bordeaux, Bordeaux, France; [5]Precision Psychiatry and Social Physiology Laboratory, CHU Sainte-Justine Research Center, Department of Psychiatry, University of Montreal, Quebec, Canada and [6]Human Brain and Behavior Laboratory, Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, Florida, USA

Introduction

We agree with Starke et al. (2020): despite the current lack of direct clinical applications, artificial intelligence (AI) will undeniably transform the future of psychiatry. AI has led to algorithms that can perform more and more complex tasks by interpreting and learning from data (Dobrev, 2012). AI applications in psychiatry are receiving more attention, with a 3-fold increase in the number of PubMed/MEDLINE articles on IA in psychiatry over the past 3 years ($N = 567$ results). The impact of AI on the entire psychiatric profession is likely to be significant (Brown, Story, Mourão-Miranda, & Baker, 2019; Grisanzio et al., 2018; Huys, Maia, & Frank, 2016; Torous, Stern, Padmanabhan, Keshavan, & Perez, 2015). These effects will be felt not only through the advent of advanced applications in brain imaging (Starke et al., 2020) but also through the stratification and refinement of our clinical categories, a more profound challenge which 'lies in its long-embattled nosology' (Kendler, 2016).

These technical challenges are subsumed by ethical ones. In particular, the risk of non-transparency and reductionism in psychiatric practice is a burning issue. Clinical medicine has already developed the overarching ethical principles of respect for autonomy, non-maleficence, beneficence, and justice (Beauchamp & Childress, 2001). The need for the principle of Explainability should be added to this list, specifically regarding the issues involved by AI (Floridi et al., 2018). Explainability concerns the understanding of how a given algorithm works (Intelligibility) and who is responsible for the way it works (Accountability). We totally agree with Starke et al. (2020) that Explainability is essential and constitutes a real challenge for future developments in AI. In addition, however, we think that this ethical issue requires dedicated pedagogical training that must be underpinned by a solid epistemological framework.

The practice of young physicians depends primarily on core educational principles (Chen, Joshi, & Ghassemi, 2020; Pinto Dos Santos et al., 2019). Raising awareness about ethics requires sound training covering the multiple aspects of AI, from its history and underlying principles to the challenges of current applications and even its promotion for the future of psychiatry. Furthermore, young scientists and physicians must be trained in the interdisciplinary challenges that lie ahead of them by becoming fully versed in the philosophy of ethics, computer science, cognitive neuroscience, computational psychiatry, and clinical practice. They should learn how to identify which technology can help in a given clinical context, to interpret and understand the results (with the potential errors, biases, or clinical inapplicability), and ultimately to explain those results both to patients and other health professionals (McCoy et al., 2020). Training physicians in this way to be at ease in both AI and medicine has already become the norm in some prestigious institutions, as attested by the partnerships between the University of Toronto and the Vector Institute or between Harvard Medical School and MIT. Concerted efforts to develop such dual competencies will enable a new academic ecosystem to emerge that will bring together AI and clinical practice in psychiatry.

However, if training in such dual competency is to be efficient, it needs strong epistemological foundations. Indeed, without minimal instruction in core concepts, models and theories, such a pedagogical program could end up in the massive misuse of algorithms. Epistemology is specifically the science of the nature and grounds of knowledge that

Table 1. Pedagogical challenges for modern psychiatry

Levels of development	Theoretical goals	Practical goals
Clinical	Knowing good practices	Use, evaluate, critically interpret and explain AI
Academic	Bring engineers, scientists, and physicians together in interdisciplinarity networks	Curricula both in data science and biomedical engineering and in ethics – epistemology
Collaborative	Promoting a horizontal organization without partitioning between medical and AI ecosystems	Encouraging scientific, ethical, and epistemological programs

While medical education is currently facing the progress of AI, ethics and epistemology offer two structuring frameworks to constrain the associated issues and to allow the development of relevant educational programs.

coordinates sets of concepts, models and theories, thus allowing knowledge to be structured. While clinical observations form the building blocks of medicine, clinical reasoning corresponds to how physicians establish diagnoses and prognoses and solve therapeutic problems. AI certainly offers new opportunities for assembling these bricks, but it must be accompanied by a new epistemological framework to structure all the emerging ethical and clinical issues. Therefore, while we fully support the pedagogical integration of AI in medicine, we also argue strongly for the teaching of medical epistemology. Learning to formalize medical theories based on AI – and not just models based on AI – seems as important as attempting to apply continually evolving and changing techniques (Muthukrishna & Henrich, 2019).

To translate the ethical principles proposed by Starke et al. (2020) into clinical practice, such an epistemological and methodological framework could be built upon the principles recently proposed by McCoy et al. (2020) and by Torous et al. (2015) with regard to pedagogy in AI and neuroscience, respectively. McCoy et al. (2020) propose instilling durable fundamental concepts about AI while avoiding technical specifics, stating that it is 'important for students to have a robust conceptual understanding of AI and the structure of clinical data science'. They go on to suggest that there is a need to 'introduce frameworks for approaching ethical considerations, both clinically and at a systems level. The content of such consideration should allow students to appreciate fairness, accountability, and transparency as core AI.' Torous et al. (2015) propose including neuroscientific methods in psychiatry residency training to face the challenge of Explainability. In particular, they encourage real-time 'circuit-specific discussions of brain-symptom relationships across the care of psychiatric patients,' to enable medical students to understand and select 'psychological and biologically informed treatments,' keeping in mind the necessity to render new models and theories in psychiatry intelligible. This approach may help in understanding the stratification of psychiatric categories, the definition of micro-phenotypes and the dynamics of clusters in the context of staging models, i.e. ebb and flow core symptoms of a psychiatric manifestation (McGorry & Nelson, 2019), as opposed to the macro-phenotypes defined by the traditional rigid categorical diagnoses provided by the DSM. It may even help in circumventing the Bernardian dichotomy between psychiatric manifestations (phenotypes) and causal mechanisms (endotypes), and in structuring the definition of diagnostic, prognostic and predictive biomarkers (McGorry et al., 2014).

An epistemological framework for teaching AI in medicine could also convey some educational principles for clinical practice (Table 1):

(1) At the clinical level, teach the medical students how to use and evaluate, to interpret critically and to explain AI results, especially by knowing which practices are acceptable (Poldrack, Huckins, & Varoquaux, 2019) and the difference between prediction and inference (Bzdok, Engemann, Grisel, Varoquaux, & Thirion, 2018).
(2) At the academic level, develop curricula both in data science and biomedical engineering and in ethics and epistemology to develop cross-disciplinary collaboration between engineers, scientists and physicians, with the help of interdisciplinary networks.
(3) At the collaborative level, promote a horizontal organization without partitioning between medical and AI ecosystems by encouraging scientific, ethical and epistemological programs.

On one hand, psychiatry is beginning to appropriate the 5P values of medicine (Personalized, Preventive, Participative, Predictive and Pluri-expert/Populational) and the 5 V vision of data (Volume, Velocity, Variety, Veracity, and Value). On the other, AI is now embracing both neuroscience and the 4EA approach (Embodied, Embedded, Enacted, Extended, Affective). Following the lead of the Society for Neuroscience and the *Neuroscience Core Concepts* (McNerney, Chang, & Spitzer, 2009), we assume that the future of medical epistemology will be based on a theoretical and methodological framework that requires to-ing and fro-ing between how scientists appropriate the concepts of philosophy (AI in epistemology) and how philosophers appropriate those of science (epistemology in AI) in a dynamic, grounded and evolutive manner (Pradeu & Carosella, 2006). Tomorrow's physicians should be aware of these epistemological backgrounds so that well-planned, practical AI systems may be developed that foster trust and confidence between them and their patients.

References

Beauchamp, T. L., & Childress, J. F. (2001). *Principles of biomedical ethics* (5th Edn). New York, NY: Oxford University Press.

Brown, C., Story, G. W., Mourão-Miranda, J., & Baker, J. T. (2019). Will artificial intelligence eventually replace psychiatrists? *The British Journal of Psychiatry*, 1–4. https://doi.org/10.1192/bjp.2019.245.

Bzdok, D., Engemann, D., Grisel, O., Varoquaux, G., & Thirion, B. (2018). Prediction and inference diverge in biomedicine: Simulations and real-world data. Preprint. Bioinformatics. https://doi.org/10.1101/327437.

Chen, I. Y., Joshi, S., & Ghassemi, M. (2020). Treating health disparities with artificial intelligence. *Nature Medicine*, *26*(1), 16–17. https://doi.org/10.1038/s41591-019-0649-2.

Dobrev, D. (2012). A definition of artificial intelligence. *ArXiv*:1210.1568 [*Cs*], October. http://arxiv.org/abs/1210.1568.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., … Vayena, E. (2018). AI4People – an ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5.

Grisanzio, K. A., Goldstein-Piekarski, A. N., Wang, M. Y., Ahmed, A. P. R., Samara, Z., & Williams, L. M. (2018). Transdiagnostic symptom clusters and associations with brain, behavior, and daily function in mood, anxiety, and trauma disorders. *JAMA Psychiatry*, *75*(2), 201–209. https://doi.org/10.1001/jamapsychiatry.2017.3951.

Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, *19*(3), 404–413. https://doi.org/10.1038/nn.4238.

Kendler, K. S. (2016). The nature of psychiatric disorders. *World Psychiatry*, *15*(1), 5–12. https://doi.org/10.1002/wps.20292.

McCoy, L. G., Nagaraj, S., Morgado, F., Harish, V., Das, S., & Celi, L. A. (2020). What do medical students actually need to know about artificial intelligence? *Npj Digital Medicine*, *3*(1), 1–3. https://doi.org/10.1038/s41746-020-0294-7.

McGorry, M. K., Goldstone, S., Amminger, P., Allott, K., Berk, M., Lavoie, S. (2014). Biomarkers and clinical staging in psychiatry. *World Psychiatry*, *13*(3), 211–223. https://doi.org/10.1002/wps.20144.

McGorry, P. D., & Nelson, B. (2019). Transdiagnostic psychiatry: Premature closure on a crucial pathway to clinical utility for psychiatric diagnosis. *World Psychiatry*, *18*(3), 359–360. https://doi.org/10.1002/wps.20679.

McNerney, C. D., Chang, E.-J., & Spitzer, N. C. (2009). Brain awareness week and beyond: Encouraging the next generation. *Journal of Undergraduate Neuroscience Education*, *8*(1), A61–A65.

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*(3), 221–229. https://doi.org/10.1038/s41562-018-0522-1.

Pinto Dos Santos, D., Giese, D., Brodehl, S., Chon, S. H., Staab, W., Kleinert, R., … Baeßler, B. (2019). Medical students' attitude towards artificial intelligence: A multicentre survey. *European Radiology*, *29*(4), 1640–46. https://doi.org/10.1007/s00330-018-5601-1.

Poldrack, R. A., Huckins, G., & Varoquaux, G. (2019). Establishment of best practices for evidence for prediction: A review. *JAMA Psychiatry*, *77*(5), 534–540. https://doi.org/10.1001/jamapsychiatry.2019.3671.

Pradeu, T., & Carosella, E. D. (2006). On the definition of a criterion of immunogenicity. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(47), 17858–17861. https://doi.org/10.1073/pnas.0608683103.

Starke, G., De Clercq, E., Borgwardt, S., & Elger, B. S. (2020). Computing schizophrenia: Ethical challenges for machine learning in psychiatry. *Psychological Medicine*, 1–7. https://doi.org/10.1017/S0033291720001683.

Torous, J., Stern, A. P., Padmanabhan, J. L., Keshavan, M. S., & Perez, D. L. (2015). A proposed solution to integrating cognitive-affective neuroscience and neuropsychiatry in psychiatry residency training: The time is now. *Asian Journal of Psychiatry*, *17*, 116–121. https://doi.org/10.1016/j.ajp.2015.05.007.