



# 'AI gone mental': engagement and ethics in data-driven technology for mental health

Sarah Carr

**To cite this article:** Sarah Carr (2020) 'AI gone mental': engagement and ethics in data-driven technology for mental health, Journal of Mental Health, 29:2, 125-130, DOI: [10.1080/09638237.2020.1714011](https://doi.org/10.1080/09638237.2020.1714011)

**To link to this article:** <https://doi.org/10.1080/09638237.2020.1714011>



Published online: 30 Jan 2020.



Submit your article to this journal [↗](#)



Article views: 21227



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 31 View citing articles [↗](#)

EDITORIAL



## 'AI gone mental': engagement and ethics in data-driven technology for mental health

### Introduction

In 2017 the tech giant IBM stated that artificial intelligence (AI) will transform the delivery of mental health care over the next five years by helping clinicians better predict, monitor and track conditions, and that “what we say and write will be used as indicators of our mental health and physical wellbeing” (IBM, 2017). Put simply, “AI is the field of computer science that includes machine learning [algorithms], natural language processing, speech processing, robotics and similar automated decision-making” (Hodgson, Berry, Wearne, & Ellis, 2018). As noted in a recent editorial in this journal, psychiatry also seems to believe in the transformative power of such technology (Wykes, 2019), with AI driven mental health interventions featuring in the WPA-Lancet Psychiatry Commission on the Future of Psychiatry (Bhugra et al., 2017). Likewise, the British Secretary of State for Health announced at the NHS Expo 2018 that he is an evangelist for data-driven technology in health, saying “the power of genomics and AI to use the NHS’s data to save lives is literally greater than anywhere else on the planet” (Hancock, 2018). Recently Health Education England issued a report exploring the “digital future of mental health and its workforce”, citing the use of AI methods and applications to use data from the digital monitoring of patients to “provide decision support or prediction” (Foley & Woollard, 2019, p. 6). So, AI and data-driven health technology are rapidly being seen as having “the potential for radical change in terms of service delivery and the development of new treatments” in mental health (Bhugra et al., 2017, p. 775). But as with any new development in mental health, the key stakeholders are surely those on the receiving end – the service users, patients, carers and families. In a Journal of Mental Health editorial ten years ago the use of data in the search for biomarkers was discussed, and “the particular concerns or challenges that biomarker research poses in relation to service user engagement and participation” was explored (Callard & Wykes, 2008, p. 2). This editorial aims to explore some similar questions for research in and use of AI in mental health, including that which uses personal digital monitoring data for AI (see for example Leightley, Williamson, Darby, & Fear, 2019). It will revisit some of the issues for AI previously surfaced for e-mental health such as web-based interventions in this journal in 2012 and 2019 (Schmidt & Wykes, 2012; Wykes, 2019).

### Power, scrutiny and trust

In September 2018, the British Government published an initial voluntary code of conduct for data-driven health and

care technology so that “data-driven technologies must be harnessed in a safe, evidenced and transparent way. We must engage with patients and the public on how to do this in a way that maintains trust” (HM Government, 2018, p. 1). The outline code of conduct presents a set of principles, focusing on “responsibility”, “transparency” and “accountability”, “to result in partnerships that deliver benefits to patients, clinicians, industry and the health and care system as a whole” (Ibid., p. 1). There is an emphasis on partnerships between health and care providers, their patients, service users and staff. But what could this mean for people with mental health problems who are patients and service users? While AI appears to be useful for the detection and treatment of certain physical health conditions, such as sepsis (Komorowski, Celi, Badawi, Gordon, & Faisal, 2018), concerns about power and ethics have been raised about the use of genomic data for population health: “it’s what we do with, and the value we ascribe to, this data that matters” (Price, 2019). The use of power in mental health systems is vital to remember for any data-driven mental health technology, and underlines the need to use AI ethical frameworks and codes of conduct in the field. This is particularly important for “predictive analytics in patient-monitoring devices” and “machine learning... for drug discovery and development” (Rosso, 2018, p. 2). Possibly more than any other group of patients, people with mental health problems can experience particular forms of power and authority in service systems and treatment (UNHRC, 2017). They are the only people with long-term conditions who are subject to compulsory treatment under law (Szmukler, 2015). The implications of these specific power dynamics as well as potential biases in mental health systems must be considered for the ethical development and implementation of any data-driven technology in mental health.

More broadly concerns have been raised about algorithmic decision-making that may be compromised by the collection, quality and analysis of the data used to “train” AI because “if the training data are biased, the AI system risks reproducing that bias” (Borgesius, 2018, p. 11). The use of flawed data for machine learning may carry the risk of AI inventing “new classes which do not correlate with protected characteristics”, as enshrined in equality laws, as well as biased or discriminatory decision making (Ibid., p. 1). In human decision-making unaided by AI, the possibility of discrimination and potential bias has already been evidenced for mental health, particularly for black people who have an increased risk of compulsory detention under the Mental Health Act in Britain (Barnett et al., 2019). Bias is also a

potential issue for algorithms trained by data collected from patient monitoring devices such as online self-report mood trackers and smartphone apps for behavior or activity monitoring (Bauer & Moessner, 2012). Research investigating the quality and acceptability of such monitoring devices for early intervention in psychosis showed that while some mobile health technology can be beneficial for early intervention and clinical decision making, “the monitoring and reporting of adverse events or effects is largely neglected”, including adverse reactions to the digital interventions themselves (such as increased experiences of paranoia) (Bradstreet, Allan, & Gumley, 2019, p. 462). Data mined from such sources to feed algorithms could result in decision-making that lacks precision and may be actively harmful to the person because it does not include adverse effects, reactions or events. This data also excludes the contextual information needed to assess mental health such as the interpersonal, cultural, social, economic and environmental influences that are not currently captured by monitoring devices. Similarly, for population health decisions informed by genomic data, it has been argued that “both as individuals and collectively, we cannot lose sight of the importance of what we already know about the socio-economic, environmental and behavioural determinants of our health” (Price, 2019).

Scrutiny is needed at all times for AI, and it has been strongly argued that humans should not delegate decision-making responsibility to “machines alone” (Fry, 2018), and this position is reflected in recent ethical frameworks for AI development and use, that focus on scrutiny and trust (AI HLEG, 2019; RSA, 2018). Authorities in the field argue that algorithms should be seen as “decision support systems”, with transparent logic trails to show how any conclusion has been reached, so decisions can be scrutinized and challenged (AI HLEG, 2019; House of Lords, 2019; HM Government, 2018). Authors such as political scientist Virginia Eubanks have warned against the unregulated and unscrutinised use of data-driven technology and AI in public services and welfare, giving examples of its negative impacts on workers and the poor in the US. She argues that “digital security guards collect information about us, make inferences about our behaviour and control access to resources” (Eubanks, 2017, p. 5). The British Parliamentary Select Committee on Artificial Intelligence (House of Lords, 2019) highlighted that the general public have negative views of and limited trust in AI and that public engagement for building awareness and trust is required. Research on the use of electronic health records and other health-related data in public services and the insurance industry revealed that “public trust in the use of big data will be dependent on organisations not doing “creepy” things with it” (CII, 2019, p. 6).

There is legislation for algorithmic accountability and the “right to explanation” about algorithmic decision-making for individuals. EU law governing the use of complex algorithms in public life has been strengthened and “the GDPR’s [General Data Protection Regulation] provisions on algorithmic accountability, which include a right to explanation, have the potential to be broader, stronger, and deeper

than the requirements of the preceding [EU] Data Protection Directive” (Kaminski, 2019, p. 189). A recent ground-breaking legal case challenging the British Home Office use of algorithmic visa allocation systems for immigration sought to find out “what the algorithm is and what it does” and whether the decisions were biased and unethical (Artificial Lawyer, 2019a). Accordingly, the Law Society AI Commission has recommended the establishment of a national register of algorithms to enable transparency, accountability and trust, a development which could have considerable implications for the use of AI in mental health (Artificial Lawyer, 2019b). In response to such issues of transparency and trust for health, in their initial code of conduct the British Government stated that “data-driven technologies must be harnessed in a safe, evidenced and transparent way. We must engage with patients and the public on how to do this in a way that maintains trust” (HM Government, 2018, p. 2).

### Patient and public involvement in mental health AI research

One way of building trust is to engage service users and patients in the research and development of AI for mental health through public and patient involvement (PPI) but at present there is a question about the extent to which this research is subject to PPI as well as research ethical standards. Similar concerns have already been voiced about the adequate research testing of digital mental health interventions such as smartphone apps and web-based therapies (Wykes, 2019), which are potential sources of data for algorithms. Researchers examining wearable biometric monitoring devices (BMDs) have asserted that “coupled with the progress of artificial intelligence (AI), the thousands of data points collected from BMDs may help in informing diagnosis, predicting patient outcomes, and helping care professionals select the best treatment for their patients” (Tran, Riveros, & Ravaud, 2019, p. 1). In their review, The James Lind Alliance Priority Setting Partnership (JLAPSP) concluded that “the evidence base for digital mental health interventions, including the demonstration of clinical effectiveness and cost effectiveness in real-world settings, remains inadequate” (Hollis et al., 2018, p. 1). Health Education England have also noted that such an “effectiveness evidence vacuum” for AI as well as digital interventions for monitoring and treatment in mental health, with concerns about the risk of “spurious claims and overhyped technologies that fail to deliver for patients” (Foley & Woollard, 2019, p. 31). The JLAPSP highlighted the importance of PPI to improve research in this area:

If research is to be of value to decision makers, including people with lived experience of mental health problems, health and social care providers, and health care commissioners and policy makers, the identification and framing of research questions must involve people affected by these decisions (Hollis et al., 2018, p. 7)

This conclusion on PPI for digital mental health intervention research is equally applicable, and related, to AI

research and development in mental health, but the extent to which PPI will or is taking place is as yet largely unknown. Research and development of AI in for health and mental health is being conducted outside universities with a considerable amount of AI research activity taking place in the parallel sphere of private companies and venture capital funded new tech start-ups (Rosso, 2018). Many of these companies do not have the same systematic practice of research ethical approval or tradition of PPI as universities or the NHS. Reportedly, ethical frameworks for AI research remain significantly under developed in market-driven contexts (New Mind, no date).

In their paper on predictive modeling in e-mental health, Becker et al. (2018) reveal an example of some flawed thinking around PPI. Predictive modeling is a type of machine learning in AI that uses big and personal data to find patterns to predict future events. The authors argue for a “common language framework” and shared research goals so that “clinical researchers and members of the data mining community increasingly join forces to build predictive models for health monitoring, treatment selection and treatment personalization” (Becker et al., 2018, p. 57). However, they privilege the therapist’s needs as those that the data science community should understand. At no point in the paper is the patient recognised as the individual whose personal data is being used and who needs to consent to its use; as a key collaborator in research or as the ultimate beneficiary to whom both therapists and data scientists are accountable (HM Government, 2018). Such exclusion of the patient runs counter to established practice of PPI in general mental health research as well as ethical frameworks for the development of AI systems and applications in health and beyond. The European Commission High-Level Expert Group on AI emphasized that “stakeholder participation and social dialogue” is an ethical imperative for AI development (AI HLEG, 2019).

The Royal Society for the encouragement of Arts, Manufactures and Commerce (RSA) has asserted that citizens need to be engaged in the ethical use of AI for automated decision making. Their proposals are broader than PPI, but nonetheless form an important argument for involvement, based in the need to build trust and accountability in machine learning systems, particularly because AI is “being increasingly used to make predictions about the likelihood of future events occurring” (RSA, 2018, p. 6). This includes prediction of mental health problem onset or relapse by clinicians using personal data (Becker et al., 2018), which has significant ethical implications for data ownership and fully informed consent for data use (including information on risk), algorithmic accountability and the right to explanation. The RSA propose a working definition of ethical AI that addresses these and other issues. It is one in which “AI...is designed and implemented based on the public’s values, as articulated through a deliberative and inclusive dialogue between experts and citizens” (RSA, 2018, p. 9). For mental health, such a proposal needs careful consideration because the optimistic implication is that the public’s values will always be benign and inclusive, but

discrimination and fear still characterize attitudes towards people living with mental health problems, and public values can be influenced by socio-economic and political environments (Henderson et al., 2014; Thornicroft, 2009).

As discussed earlier, mental health systems still operate using legal compulsion (Szmukler, 2015) so there is a question about how this might effect data collection from patients and service users for AI predictive modeling where “a wealth of fine and coarse-grained data is collected, from heart-rate sensors, physical activity sensors, and other mobile applications, to assess the dynamics of symptoms, affect, behaviour and cognition over time” (Becker et al., 2018, p. 57). Complications have been highlighted for NHS mental health care because “some of the technologies...have profound implications in terms of the level of surveillance that they place on the patient” (Foley & Woollard, 2019, p. 31). A French study on patients’ views of wearable monitoring devices and AI for healthcare found that “35% of patients would refuse to integrate at least one existing or soon-to-be available intervention using biometric monitoring devices and AI-based tools in their care. Accounting for patients’ perspectives will help make the most of technology without impairing the human aspects of care, generating a burden or intruding on patients’ lives” (Tran et al., 2019, p. 1). Although this research did not include those living with mental health problems, it indicates a possible future difficulty concerning the acceptability and potential uptake of monitoring devices in mental health care. This then raises questions about the use of monitoring data for AI in the context of compulsory treatment under the Mental Health Act in Britain. For example, if a person is subject to a community treatment order (Szmukler & Appelbaum, 2008) and does not consent to share their data would they be compelled to do so, or be designated non-complaint if they exercised their legal right not to consent? Would issues of consent and compulsion vary between data collected through active self-reporting and behavioural or activity data harvested through monitoring devices? Questions like these should be addressed through specific mental health service user, patient and carer involvement in public discussions about the ethics of AI and through PPI in research.

European AI ethics guidelines indicate that for the principles of “respect for human autonomy, prevention of harm, fairness and explicability” to be realized, particular attention must be paid to “situations involving more vulnerable groups such as children, people with disabilities and others that have historically been disadvantaged or are at risk of exclusion, and to situations which are characterised by asymmetries of power” (AI HLEG, 2019, p. 2). This suggests that those with psychosocial disabilities should in fact be prioritized in PPI as “vulnerable persons [who] should receive greater attention and be included in the development, deployment and use of AI systems” (Ibid., p. 12). A crucial question for AI in mental health must be “what does “duty of care” mean when applied to those who are developing algorithms for use in healthcare and medical research?” (Wellcome Trust, 2018, p. 8).



## The role of the “domain expert” in AI research

In the “inclusive dialogue” between experts and citizens recommended by the RSA, where do service users and patients fit, as members of both groups (Rose, Fleischmann, & Tonkiss, 2003)? Although their citizenship has been compromised and contested (Sayce, 2015), people with mental health problems are both citizens and members of the public. They also possess particular experiential knowledge and expertise (Beresford, 2003) in relation to inclusive dialogue and for AI development in mental health. This is where the service user, patient or carer should be positioned as a “domain expert” in AI research. The involvement of the domain expert in the development of AI programmes has been actively recommended by the coalition of AI organisations that issued a report entitled “The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation” (Brundage, Avin, & Clark, 2018). The authors state that AI developers must “actively [seek] to expand the range of stakeholders and domain experts involved in discussions of the challenges” (Brundage et al., 2018 quoted in RSA, 2018, p. 9). However, a systematic review on health care analytics and the application of data mining for AI found that 32% of research reviewed did not “utilise expert opinion in any form” (Islam, Hasan, Wang, Germack, & Noor-E-Alam, 2018, p. 1). The review authors highlighted the “interdisciplinary nature of study and domain expert knowledge” (Ibid., p. 34) and concluded that “lack of prescriptive analysis in practice and integration of domain expert knowledge in the decision-making process emphasises the necessity of future research” (Ibid., p. 1). As with the seeming lack of public engagement in debates on AI, there should be concerns about the participation of patients, service users and carers as domain experts in AI mental health research projects, as well as in PPI more widely.

Despite some of the apparent flaws in AI health and mental health research to date, the equivalent of service user and patient participation is happening as the positive impact of involvement of domain experts in research studies has already been evidenced for other areas such as youth work and criminal justice. Arguably transferrable models are being developed. For example, the authors of one US study on the application of algorithms to predict risk in gang violence found that “when analysing ... data from marginalised communities, algorithms lack the ability to accurately interpret off-line context, which may lead to dangerous assumptions about the implications for marginalised communities” (Frey, Patton, Gaskell, & McGregor, 2018, p. 1). To reduce this risk, the researchers involved young people with experience of gang violence in the study as “domain experts”, who they concluded

... must be involved in the interpretation of unstructured data, solution creation and many other aspects of the research process. This goes beyond harvesting and capturing domain expertise. The involvement of domain experts in various areas of social and data science research, including mechanisms for accountability and ethically sound research practices, is a critical piece of truly creating algorithms trained to support and protect marginalised youth and communities (Frey et al., 2018, p. 12–13).

It can be strongly argued that the concept of the “domain expert” and this type of participatory and inclusive approach in research and development for AI should transfer to mental health, as it could lead the to “more robust understandings of context ... language, culture and events” (Frey et al., 2018, p. 1) vital for ethical decision making.

## Conclusion

AI is increasingly being seen by psychiatrists, psychologists, politicians and tech companies as having a significant role in future mental health treatment and care, with developments in the field being driven by their particular agendas. However, it appears that key stakeholders are currently excluded from the discussions about AI in mental health – patients, service users, carers, and families. If rights-based guidelines for ethical AI (AI HLEG, 2019) are to be implemented in mental health, then the implications of the United Nations Convention on the Rights of Persons with Disabilities (UNCRPD) needs to be considered. The UNCRPD have said that it is essential to involve disabled people (including those with psychosocial disabilities [Szmukler, Daw, & Callard, 2014]) and their representative organisations in developments and decision-making that will affect their lives:

Often, persons with disabilities are not consulted in the decision-making about matters relating to or affecting their lives, with decisions continuing to be made on their behalf. Consultation with persons with disabilities has been acknowledged as important in the last few decades, thanks to the emergence of movements of persons with disabilities demanding recognition of their human rights and their role in determining those rights. The motto “nothing about us without us” resonates with the philosophy and history of the disability rights movement, which relies on the principle of meaningful participation. (UNCRPD, 2018, p. 2)

General ethical frameworks for AI emphasize the importance of the participation of the public in discussions about the use of AI in everyday life. If AI is to be increasingly used in the treatment and care of people with mental health problems then patients, service users and carers should participate as experts in its design, research and development. Their data will be used to train and drive many of the AI applications designed for predictive modeling, to inform clinical decisions and to determine the timing and types of intervention. There are risks of replicating existing and even creating new inequalities in health and mental health as well as risks that new forms of coercion or compulsory treatment could emerge. Scrutiny, transparency and algorithmic accountability are essential. AI research and development for mental health care cannot continue without PPI and full participation in research. The area is still emerging and as Health Education England predict, the impact timescale for AI applications in mental health is 3–10+ years (Foley & Woollard, 2019, p. 6). It is not too late to involve patients, service users and carers as domain experts in AI research and discussions about the ethical use of AI. It is therefore time to assess the situation, to question those who are driving this transformative agenda forward and to listen to

excluded experts – those whose lives these technologies will ultimately affect.


## Disclosure statement

No potential conflict of interest was reported by the author.

## References

- AI HLEG. (2019). *Ethics Guidelines for Trustworthy AI*. Brussels: European Commission.
- Artificial Lawyer. (2019a, October 31). UK Government faces court over 'biased' visa algorithm. *Artificial Lawyer*. Retrieved from <https://www.artificiallawyer.com/2019/10/31/uk-government-faces-court-over-biased-visa-algorithm/> (Accessed 11 November 2019).
- Artificial Lawyer. (2019b, June 4). Create a National Register of Algorithms – Law Society AI Commission. *Artificial Lawyer*. Retrieved from <https://www.artificiallawyer.com/2019/06/04/create-a-national-register-of-algorithms-law-society-ai-commission/> (Accessed 11 November 2019).
- Barnett, P., Mackay, E., Matthews, H., Gate, R., Greenwood, H., Ariyo, K., ... Smith, S. (2019). Ethnic variations in compulsory detention under the Mental Health Act: A systematic review and meta-analysis of international data. *The Lancet Psychiatry*, 6(4), 305–317. doi:10.1016/S2215-0366(19)30027-6
- Bauer, S., & Moessner, M. (2012). Technology-enhanced monitoring in psychotherapy and e-mental health. *Journal of Mental Health*, 21(4), 355–363. doi:10.3109/09638237.2012.667886
- Becker, D., van Breda, W., Funk, B., Hoogendoorn, M., Ruwaard, J., & Riper, H. (2018). Predictive modeling in e-mental health: A common language framework. *Internet Interventions*, 12, 57–67. doi:10.1016/j.invent.2018.03.002
- Beresford, P. (2003). *It's Our Lives: A short theory of knowledge, distance and experience*. London: OSP/Citizen Press.
- Bhugra, D., Tasman, A., Pathare, S., Priebe, S., Smith, S., Torous, J., ... Ventriglio, A. (2017). The WPA-Lancet Psychiatry Commission on the Future of Psychiatry. *The Lancet. Psychiatry*, 4(10), 775–818. doi:10.1016/S2215-0366(17)30333-4
- Borgesius, F.Z. (2018). *Discrimination, artificial intelligence and algorithmic decision-making*. Strasbourg: Council of Europe.
- Bradstreet, S., Allan, S., & Gumley, A. (2019). Adverse event monitoring in mHealth for psychosis interventions provides an important opportunity for learning. *Journal of Mental Health*, 28(5), 461–466. doi:10.1080/09638237.2019.1630727
- Brundage, M., Avin, S., & Clark, J. (2018). *The malicious use of artificial intelligence: Forecasting, prevention and mitigation*. Oxford: University of Oxford.
- Callard, F., & Wykes, T. (2008). Mental health and perceptions of biomarker research – possible effects on participation. *Journal of Mental Health*, 17(1), 1–7. doi:10.1080/09638230801931944
- CII. (2019). *Shaping the Future of Medical Records and Protection Insurance*. London: Chartered Insurance Institute.
- Eubanks, V. (2017). *Automating Inequality*. New York: St Martin's Press.
- Foley, T., & Woollard, J. (2019). *The digital future of mental healthcare and its workforce*. London: Health Education England.
- Frey, W., Patton, D., Gaskell, M., & McGregor, K. (2018). Artificial intelligence and inclusion: Formerly gang-involved youth as domain experts for analyzing unstructured Twitter data. *Social Science Computer Review*, 38(1), 1–15. doi:10.1177/0894439318788314
- Fry, H. (2018). *How to be Human in the Age of the Machine*. London: Black Swan.
- Hancock, M. (2018, September 6). My vision for a more tech-driven NHS. *Gov.uk*. Retrieved from <https://www.gov.uk/government/speeches/my-vision-for-a-more-tech-driven-nhs> (Accessed 2 August 2019).
- Henderson, R. C., Corker, E., Hamilton, S., Williams, P., Pinfold, V., Rose, D., ... Thornicroft, G. (2014). Viewpoint survey of mental health service users' experiences of discrimination in England 2008–12. *Social Psychiatry and Psychiatric Epidemiology*, 49(10), 1599–1608. doi:10.1007/s00127-014-0875-3
- HM Government. (2018). *Code of conduct for data-driven health and care technology*. London: DHSC.
- Hodgson, L., Berry, N., Wearne, N., & Ellis, M. (2018, July 6). AI and insurance: Planning for an intelligent future. *Insurance Law Tomorrow*. Retrieved from <https://www.insurancelawtomorrow.com/2018/07/ai-and-the-insurance-planning-for-an-intelligent-future/> (Accessed 2 August 2019).
- Hollis, C., Sampson, S., Simons, L., Davies, E. B., Churchill, R., Betton, V., ... Tomlin, A. (2018). Identifying research priorities for digital technology in mental health care: Results of the James Lind alliance priority setting partnership. *The Lancet Psychiatry*, 5 (10), 845–854. doi:10.1016/S2215-0366(18)30296-7
- House of Lords. (2019). *Select Committee on Artificial Intelligence Report of Session 2017-19: AI in the UK: Ready, willing and able?* London: House of Lords.
- IBM. (2017). With AI, our words will be a window into our mental health. *IBM Research*. Retrieved from <https://www.research.ibm.com/5-in-5/mental-health/> (Accessed 2 August 2019).
- Islam, M., Hasan, M., Wang, X., Germack, H., & Noor-E-Alam, M. (2018). A systematic review of healthcare analytics: Application and theoretical perspective of data mining. *Healthcare*, 6(2), 54. doi:10.3390/healthcare6020054
- Kaminski, M.E. (2019). The right to explanation, explained. *Berkeley Technology Law Journal*, 34, 189–128.
- Komorowski, M., Celi, L. A., Badawi, O., Gordon, A. C., & Faisal, A. A. (2018). The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. *Nature Medicine*, 24(11), 1716–1720. doi:10.1038/s41591-018-0213-5
- Leightley, D., Williamson, V., Darby, J., & Fear, N. (2019). Identifying probable post-traumatic stress disorder: Applying supervised machine learning to data from a UK military cohort. *Journal of Mental Health*, 28 (1), 34–41. doi:10.1080/09638237.2018.1521946
- NewMind. (no date). *NewMind Research Roadmap*. Retrieved from <http://www.newmindnetwork.org.uk/media/1399/newmind-research-roadmap-final.pdf> (Accessed 2 August 2019).
- Price, C. (2019, November 7). Genomic medicine: A tool for population health? *The Kings Fund*. Retrieved from [https://www.kingsfund.org.uk/blog/2019/11/genomic-medicine-population\\_health?utm\\_source=twitter&utm\\_medium=social&utm\\_term=thekingsfund](https://www.kingsfund.org.uk/blog/2019/11/genomic-medicine-population_health?utm_source=twitter&utm_medium=social&utm_term=thekingsfund) (Accessed 11 November 2019).
- Rose, D., Fleischmann, P., & Tonkiss, F. (2003). *User and carer involvement in change management in a mental health context: Review of the literature*. London: NCCSDO.
- Rosso, C. (2018, June 13). The future of AI in health care. *Psychology Today UK*. Retrieved from <https://www.psychologytoday.com/gb/blog/the-future-brain/201806/the-future-ai-in-health-care> (Accessed 2 August 2019).
- RSA. (2018). *Artificial intelligence: Real public engagement*. London: RSA.
- Sayce, L. (2015). *From psychiatric patient to citizen revisited*. Basingstoke: Palgrave Macmillan.
- Schmidt, U., & Wykes, T. (2012). E-mental health – a land of unlimited possibilities. *Journal of Mental Health*, 21(4), 327–331. doi:10.3109/09638237.2012.705930
- Szmukler, G. (2015). Compulsion and “coercion” in mental health care. *World Psychiatry*, 14(3), 259–261. doi:10.1002/wps.20264
- Szmukler, G., & Appelbaum, P.S. (2008). Treatment pressures, leverage, coercion, and compulsion in mental health care. *Journal of Mental Health*, 17(3), 233–244. doi:10.1080/09638230802052203
- Szmukler, G., Daw, R., & Callard, F. (2014). Mental health law and the UN Convention on the rights of persons with disabilities international. *International Journal of Law and Psychiatry*, 37(3), 245–252. doi:10.1016/j.ijlp.2013.11.024
- Thornicroft, G. (2009). *Shunned: Discrimination against people with mental illness*. Oxford: Oxford University Press.
- Tran, V., Riveros, C., & Ravard, P. (2019). Patients' views of wearable devices and AI in healthcare: Findings from the ComPaRe e-cohort. *Npj Digital Medicine*, 2, Article 53. doi:10.1038/s41746-019-0132-y
- UNCRPD. (2018). Committee on the Rights of Persons with Disabilities General comment No. 7 (2018) on the participation of persons with disabilities, including children with disabilities, through their

- representative organizations, in the implementation and monitoring of the Convention. Geneva: UNCRPD.
- UNHRC. (2017). *Report of the Special Rapporteur on the right of everyone to the enjoyment of the highest attainable standard of physical and mental health*. New York: UNHRC
- Wellcome Trust. (2018). *Ethical, social and political challenges of artificial intelligence in health*. London: Wellcome Trust.
- Wykes, T. (2019). Racing towards a digital paradise or a digital hell? *Journal of Mental Health*, 28(1), 1–3. doi:[10.1080/09638237.2019.1581360](https://doi.org/10.1080/09638237.2019.1581360)

Sarah Carr  
Senior Fellow in Mental Health Policy, University of  
Birmingham, Edgbaston, Birmingham  
 [S.E.Carr@bham.ac.uk](mailto:S.E.Carr@bham.ac.uk)

Received 8 August 2019; revised 12 November 2019; accepted  
15 November 2019

© 2020 Informa UK Limited, trading as Taylor & Francis Group