

Building trust in artificial intelligence and new technologies in mental health

Bessie O'Dell,^{1,2} Katherine Stevens,^{1,2} Anneka Tomlinson,^{1,2} Ilina Singh,¹ Andrea Cipriani ^{1,2,3}

In 2019, the Topol review was published on behalf of the secretary of state for health and social care in the UK, preparing the healthcare workforce to deliver the digital future.¹ A multidisciplinary team of experts, including clinicians, researchers, ethicists, computer scientists, engineers and economists, reviewed the available data and projected into the future (ie, next 20 years) two key questions: what impact technological developments (including genomics, artificial intelligence (AI), digital medicine and robotics) will have on the roles and functions of National Health System clinical staff? How could this innovation (ie, biosensors, electronic patient record, smartphone apps, digital infrastructure and virtual reality) ensure safer, more productive, more effective and more personalised care for patients? It is now widely recognised that data science and information technologies enable understanding of the uniqueness of each individual and the ability to deliver healthcare on a far more timely, efficient and tailored basis.

Mental health is a top priority in the UK national research agenda² and presents a unique opportunity because it is in the next wave of adoption of digital health and innovation technologies.³ The Topol review noted that innovation can 'bring a new emphasis on the nurturing of the precious interhuman bond, based on trust, clinical presence, empathy and communication'. Patients must occupy a central role when assessing and implementing any new technologies.⁴ This is especially true in our field, as people can be more vulnerable due to brain or cognitive disorders, for example, and the patient-clinician interaction has long been at the core of the therapeutic relationship. In this relationship, 'trust' has a specific role, as recently highlighted by a European Commission White Paper.⁵

What is 'trust'? In its most basic form, 'trust' is a willingness to rely on something or someone. However, how exactly we conceptualise 'trust' will depend on the theoretical lens that we are using—and as such the concept of 'trust' has been applied differently across a variety of disciplines, ranging from sociology and psychology to philosophy and economics. The importance of trust in human interactions has been repeatedly recognised in the last 50 years: 'the entire fabric of our day-to-day living, of our social world, rests on trust, as almost all of our decisions involve trusting someone else'.⁶ Trust not only helps to facilitate collaboration among people, but also underpins the formation and maintenance of social relationships.⁷ More recently, accounts of trust have developed to delineate the more generic idea of 'trust' from trust specifically developed in digital contexts and/or involving artificial agents—including a concept identified as e-trust.⁸ A related, yet distinct notion, is that of 'trustworthiness'—the extent to which something or someone is deserving of trust or confidence.⁹ The terms 'trust' and 'trustworthiness' are often conflated in bodies of the literature discussing these concepts, even though implicitly much of it is primarily about trustworthiness, not trust.¹⁰

Why then, are interhuman bonds so important when considering the adoption of technology? Interhuman bonds (eg, between a patient and clinician) and the trusting relationships underpinning them are key to the successful adoption and implementation of digital health and innovation technologies. For example, research has shown that in inherently uncertain territories (such as the use of virtual doctors, as well as other algorithmic decision-makers), people will favour human judgement. Therefore, in medical decision-making, people may be unwilling to use even the best possible algorithm.¹¹ Patients may therefore be more likely to defer to, or rely on (ie, trust) clinicians to mediate their interactions with digital health and innovation technologies. Such deference could be particularly pronounced in mental health, given both the vulnerability of patients, and the importance of patient-clinician

interaction in the therapeutic relationship. This highlights the need for both clinicians and technology to be deemed 'trustworthy' (deserving of trust) in mental health settings; and, indeed, for clinicians to also consider the technology they bring into the clinical relationship to be trustworthy.

Not paying sufficient attention to the importance of interhuman bonds based on trust is detrimental to the development and adoption of technologies. Taking AI technology as an example, recent years have seen an exponential growth in the number of AI algorithms and projects published in the medical literature. AI systems have consistently demonstrated as being more beneficial than clinical care without a tool,^{12 13} and are key to the delivery of personalised, evidence-based care. However, this academic interest in AI technologies does not appear to translate well to clinical settings, where the 'clinical impact in terms of patient outcomes remains to be demonstrated'.¹⁴ Keane and Topol have considered how the lack of uptake in AI technologies (despite their potential) could be due to a so-called 'AI chasm', in which there is an overemphasis on the technical aspects of the proposed algorithms, with insufficient attention given to the factors that affect the interaction with their human users.¹⁵ Others have also called for a better focus on what has been deemed the 'softer' or more 'qualitative' impacts of AI technologies in clinical care.¹⁶ In order to better aid in the translation of novel technologies from research into clinical settings and address this 'AI chasm', relevant stakeholders (including researchers and developers) should therefore take into account the more 'qualitative' impacts of these technologies. This includes potential impacts on interhuman bonds, trust and trustworthiness.

In conclusion, trust underpins interhuman bonds. These bonds are central to clinical care in mental health, and this is equally the case when digital health and innovation technologies are used.¹⁷ Ultimately, it is the role of the clinician to bridge the gap between the technology and their patient(s), and only in doing so can digital health and innovation technologies be better used in mental health.¹⁸

Twitter Andrea Cipriani @And_Cipriani

Contributors AC and BO drafted the editorial. All other authors critically revised the text. All authors approved the final version of the article.

Funding AC is supported by the National Institute for Health Research (NIHR) Oxford Cognitive Health Clinical Research Facility, by an NIHR Research Professorship (grant RP-2017-08-ST2-006), by the NIHR Oxford and Thames Valley Applied Research

¹Department of Psychiatry, University of Oxford, Oxford, UK

²Oxford Precision Psychiatry Lab, Oxford Health Biomedical Research Centre, Oxford, UK

³Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK

Correspondence to Professor Andrea Cipriani, Department of Psychiatry, University of Oxford, Oxford OX1 2JD, UK; andrea.cipriani@psych.ox.ac.uk

Collaboration and by the NIHR Oxford Health Biomedical Research Centre (grant BRC-1215-20005).

Disclaimer The views expressed are those of the authors and not necessarily those of the UK National Health Service, the National Institute for Health Research or the UK Department of Health.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; internally peer reviewed.

© Author(s) (or their employer(s)) 2022. No commercial re-use. See rights and permissions. Published by BMJ.



To cite O'Dell B, Stevens K, Tomlinson A, et al. *Evid Based Ment Health* 2022;**25**:45–46.

Received 6 April 2022
Accepted 8 April 2022
Published Online First 20 April 2022

Evid Based Ment Health 2022;**25**:45–46.
doi:10.1136/ebmental-2022-300489

ORCID iD
Andrea Cipriani <http://orcid.org/0000-0001-5179-8321>

REFERENCES

- 1 The Topol review an independent report on behalf of the Secretary of state for health and social care, 2019. Available: <https://topol.hee.nhs.uk/>
- 2 Wellcome Trust. Mental health. Available: <https://wellcome.org/what-we-do/mental-health> [Accessed 4 Apr 2022].
- 3 HM Government. Life sciences vision: build back better: our plan for growth. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1013597/life-sciences-vision-2021.pdf [Accessed 4 Apr 2022].
- 4 NHS. Technology to improve service. Available: <https://www.england.nhs.uk/improvement-hub/wp-content/uploads/sites/44/2017/11/ILG-1.6-Use-of-Technology-to-Improve-Services.pdf> [Accessed 4 Apr 2022].
- 5 European Commission. 'White Paper on AI: A European approach to excellence and trust' COM(2020) 65 final, 2020. Available: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf [Accessed 4 Apr 2022].
- 6 Rotter JB. Generalized expectancies for interpersonal trust. *Am Psychol* 1971;**26**:443–52.
- 7 Sutcliffe A, Wang D. Computational modelling of trust and social relationships. *JASSS* 2012;**15**:3.
- 8 Taddeo M, Floridi L. The case for e-trust. *Ethics Inf Technol* 2011;**13**:1–3.
- 9 Hardin R. Trustworthiness. *Ethics* 1996;**107**:26–42.
- 10 Hardin R. Conceptions and explanations of trust. In: Cook KS, ed. *Russell SAGE Foundation series on trust, vol. 2. trust in society*. Russell Sage Foundation, 2001: 3–39.
- 11 Dietvorst BJ, Bharti S. People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error. *Psychol Sci* 2020;**31**:1302–14.
- 12 Choi D-J, Park JJ, Ali T, et al. Artificial intelligence for the diagnosis of heart failure. *NPJ Digit Med* 2020;**3**:54.
- 13 Somashekhar SP, Sepúlveda M-J, Puglielli S, et al. Watson for oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board. *Ann Oncol* 2018;**29**:418–23.
- 14 DECIDE-AI Steering Group. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. *Nat Med* 2021;**27**:186–7.
- 15 Keane PA, Topol EJ. With an eye to AI and autonomous diagnosis. *NPJ Digit Med* 2018;**1**:40.
- 16 Kudina O, de Boer B. Co-designing diagnosis: towards a responsible integration of machine learning decision-support systems in medical diagnostics. *J Eval Clin Pract* 2021;**27**:529–36.
- 17 Wisniewski H, Torous J. Digital navigators to implement smartphone and digital tools in care. *Acta Psychiatr Scand* 2020;**141**:350–5.
- 18 Hong JS, Sheriff R, Smith K, et al. Impact of COVID-19 on telepsychiatry at the service and individual patient level across two UK NHS mental health trusts. *Evid Based Ment Health* 2021;**24**:161–6.