

Viewpoint

Trust Me, I'm a Chatbot: How Artificial Intelligence in Health Care Fails the Turing Test

John Powell, PhD, FFPH

Nuffield Department of Primary Care Health Sciences, Medical Sciences Division, University of Oxford, Oxford, United Kingdom

Corresponding Author:

John Powell, PhD, FFPH

Nuffield Department of Primary Care Health Sciences

Medical Sciences Division

University of Oxford

Radcliffe Observatory Quarter

43 Woodstock Road

Oxford, OX2 6GG

United Kingdom

Phone: 44 1865617768 ext 617768

Fax: 44 1865289412

Email: john.powell@phc.ox.ac.uk

Abstract

Over the next decade, one issue which will dominate sociotechnical studies in health informatics is the extent to which the promise of artificial intelligence in health care will be realized, along with the social and ethical issues which accompany it. A useful thought experiment is the application of the Turing test to user-facing artificial intelligence systems in health care (such as chatbots or conversational agents). In this paper I argue that many medical decisions require value judgements and the doctor-patient relationship requires empathy and understanding to arrive at a shared decision, often handling large areas of uncertainty and balancing competing risks. Arguably, medicine requires wisdom more than intelligence, artificial or otherwise. Artificial intelligence therefore needs to supplement rather than replace medical professionals, and identifying the complementary positioning of artificial intelligence in medical consultation is a key challenge for the future. In health care, artificial intelligence needs to pass the implementation game, not the imitation game.

(*J Med Internet Res* 2019;21(10):e16222) doi: [10.2196/16222](https://doi.org/10.2196/16222)

KEYWORDS

artificial intelligence; machine learning; medical informatics; digital health; ehealth; chatbots; conversational agents

Over the last two decades, the concerns of digital health researchers interested in the social impact of the internet have evolved as the technology has matured and new tools have emerged. From a sociotechnical perspective, there were initial preoccupations with the impact of a new, uncontrolled form of mass communication, alongside concerns with the quality of unregulated online information and threats to professions, with medical professionals in particular fearing a loss of authority [1-3]. As Web2.0 developments took hold and the public became producers as well as consumers of health information, researchers began to identify the benefits of online peer-to-peer communication and the sharing of information in virtual communities, social media, and increasingly on health ratings sites [4-7]. With the mass uptake in smartphones, the subsequent rapid developments in mobile health, and the explosion in health apps, we are now exploring the value of low-cost, patient-centered interventions delivered directly to consumers

[8,9]. In addition, we are also gaining a better understanding of the limitations and key issues in their implementation, such as nonadoption and abandonment [10]. As the number one journal in this field, the Journal of Medical Internet Research continues to reflect and illuminate all these debates.

For those of us studying the social science of digital technology in health and health care, one area of research is likely to dominate the next decade: the extent to which the promise of artificial intelligence (AI) in health care will be realized, and the social and ethical issues which accompany it [11-13]. Broadly speaking, we can identify two current strands in the use of AI in health care. Firstly, there are data-facing applications which use techniques such as machine learning and artificial neural networks to derive new knowledge from large datasets, such as improving diagnostic accuracy from scans and other images [14]. Secondly, there are user-facing

applications and intelligent agents which interact with people in real-time, using inferences to provide advice or instruction based on probabilities which the tool can derive and improve over time, such as a chatbot substituting or complementing a health care consultation with a patient [15]. In this article I focus on the latter to consider the approaches of these chatbots, or “robot doctors,” to medical consultation, and specifically the extent to which these technologies will ever pass the celebrated Turing test.

Alan Turing, the British mathematician and theoretical computer scientist, is widely regarded as the founding father of AI. He proposed that for a machine to be considered intelligent it should provide responses to a blinded interrogation that are indistinguishable from those given by a human comparator [16]. In other words, the interrogator should not be able to tell whether the machine or the human was responding. If we extrapolate this thought experiment to current health care, we can pose the question of whether AI-based medical consultations (conversational agents and medical chatbots) will ever be considered intelligent by Turing’s standard. Of course, context is important, and if a patient is asking a simple factual question that requires a binary response, for example, then even current AI systems can mimic a human interlocutor with high accuracy. However, we know that medical consultations are complex [17], that many medical decisions require value judgements, and that the doctor-patient relationship requires empathy and understanding to arrive at a shared decision [18]. The practice of medicine is as much an art as a science, and patients may choose a path which is not necessarily the one that logic would determine. Even the pioneers of evidence-based medicine defined their normative approach as:

the conscientious and judicious use of current best evidence from clinical care research in the management of individual patients [19].

Conscience and the ability to weigh competing personal values are not strengths of AI. A key skill for medical professionals is the ability to deal with uncertainty alongside considering patients’ preferences. What doctors often need is wisdom rather than intelligence, and we are a long way away from a science of artificial wisdom.

It is doubtful whether AI will ever pass the Turing test for complex medical consultations, but this is to misunderstand the place of AI in future medical care. AI should complement rather

than replace medical professionals. As various studies into the future of work have shown, automation in the workplace will not eliminate all human tasks [20]. Chatbot approaches have many potential benefits, including the potential to allow clinicians to have more time for delivering empathic and personalized care [15]. Perhaps, as a senior clinical informatics leader in the UK has suggested, “AI will allow doctors to be more human” [13]. However, as has been well established for many innovations in health care, especially digital ones, the key challenges for health systems seeking to harness the benefits of the technology are not just related to its effectiveness but also to the wider issues of its integration and implementation [10,12,21]. We need to understand how to integrate the tools and practices of AI within the work and culture of professionals and organizations, to investigate factors related to adoption, nonadoption, and abandonment [10,12], and investigate the work required to sustain innovation [22]. Factors which will influence the implementation of AI tools include those related to people, such as professional and public attitudes, trust, existing work practices, training needs, and the risks of deskilling and disempowerment; those related to the health system, such as leadership and management, the positioning of clinical responsibility and accountability, and the possibility of harm, alongside issues of regulation and service provision (including scalability and the possibility of providing two-tier services with or without AI); those related to the data, such as issues of data security, privacy, consent and ownership; and those related to the tool itself, such as transparency of the algorithm, issues of reliability and validity, and algorithmic bias [12,21,23]. To take an example, in an early study of an algorithm-based triage tool in primary care, we showed that physicians lacked trust in the ability of the machine to take clinical risks and worried about issues of governance and accountability, such that the sensitivity of the tool, in terms of the urgency of triage, was consistently set at a threshold which would increase urgent clinical workload rather than reduce it [24].

Identifying the complementary positioning of AI tools in health care in general, and in particular for their use in the medical consultation, is a key challenge for the future. We need to understand how to integrate the precision and power of AI tools and practices with the wisdom and empathy of the doctor-patient relationship. In health care, it is more important that artificial intelligence passes the implementation game rather than the imitation game.

Acknowledgments

JP first discussed applying the Turing test to AI in health care in 2016 and had subsequent discussions with colleagues in Oxford and elsewhere. JP is funded by the National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care Oxford at Oxford Health National Health Service Foundation Trust.

Conflicts of Interest

None declared.

References

1. Hardey M. Doctor in the house: the Internet as a source of lay health knowledge and the challenge to expertise. *Sociology of Health & Illness* 2001 Dec 25;21(6):820-835. [doi: [10.1111/1467-9566.00185](https://doi.org/10.1111/1467-9566.00185)]

2. Eysenbach G, Powell J, Kuss O, Sa E. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *JAMA* 2002 May 22;287(20):2691-2700. [doi: [10.1001/jama.287.20.2691](https://doi.org/10.1001/jama.287.20.2691)] [Medline: [12020305](https://pubmed.ncbi.nlm.nih.gov/12020305/)]
3. Ziebland S. The importance of being expert: the quest for cancer information on the Internet. *Soc Sci Med* 2004 Nov;59(9):1783-1793. [doi: [10.1016/j.socscimed.2004.02.019](https://doi.org/10.1016/j.socscimed.2004.02.019)] [Medline: [15312914](https://pubmed.ncbi.nlm.nih.gov/15312914/)]
4. Hardey M. 'E-health': the internet and the transformation of patients into consumers and producers of health knowledge. *Info., Comm. & Soc* 2001 Oct;4(3):388-405. [doi: [10.1080/713768551](https://doi.org/10.1080/713768551)]
5. Powell J, McCarthy N, Eysenbach G. Cross-sectional survey of users of Internet depression communities. *BMC Psychiatry* 2003 Dec 10;3(1):1-7. [doi: [10.1186/1471-244x-3-19](https://doi.org/10.1186/1471-244x-3-19)]
6. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *BMJ* 2004 May 15;328(7449):1166 [FREE Full text] [doi: [10.1136/bmj.328.7449.1166](https://doi.org/10.1136/bmj.328.7449.1166)] [Medline: [15142921](https://pubmed.ncbi.nlm.nih.gov/15142921/)]
7. van Velthoven MH, Atherton H, Powell J. A cross sectional survey of the UK public to understand use of online ratings and reviews of health services. *Patient Educ Couns* 2018 Sep;101(9):1690-1696 [FREE Full text] [doi: [10.1016/j.pec.2018.04.001](https://doi.org/10.1016/j.pec.2018.04.001)] [Medline: [29666022](https://pubmed.ncbi.nlm.nih.gov/29666022/)]
8. Powell J, Hamborg T, Stallard N, Burls A, McSorley J, Bennett K, et al. Effectiveness of a web-based cognitive-behavioral tool to improve mental well-being in the general population: randomized controlled trial. *J Med Internet Res* 2012 Dec 31;15(1):e2 [FREE Full text] [doi: [10.2196/jmir.2240](https://doi.org/10.2196/jmir.2240)] [Medline: [23302475](https://pubmed.ncbi.nlm.nih.gov/23302475/)]
9. Rathbone AL, Clarry L, Prescott J. Assessing the Efficacy of Mobile Health Apps Using the Basic Principles of Cognitive Behavioral Therapy: Systematic Review. *J Med Internet Res* 2017 Nov 28;19(11):e399 [FREE Full text] [doi: [10.2196/jmir.8598](https://doi.org/10.2196/jmir.8598)] [Medline: [29187342](https://pubmed.ncbi.nlm.nih.gov/29187342/)]
10. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond Adoption: A New Framework for Theorizing and Evaluating Nonadoption, Abandonment, and Challenges to the Scale-Up, Spread, and Sustainability of Health and Care Technologies. *J Med Internet Res* 2017 Nov 01;19(11):e367 [FREE Full text] [doi: [10.2196/jmir.8775](https://doi.org/10.2196/jmir.8775)] [Medline: [29092808](https://pubmed.ncbi.nlm.nih.gov/29092808/)]
11. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. *PLoS Med* 2018 Nov 6;15(11):e1002689 [FREE Full text] [doi: [10.1371/journal.pmed.1002689](https://doi.org/10.1371/journal.pmed.1002689)] [Medline: [30399149](https://pubmed.ncbi.nlm.nih.gov/30399149/)]
12. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial Intelligence and the Implementation Challenge. *J Med Internet Res* 2019 Jul 10;21(7):e13659 [FREE Full text] [doi: [10.2196/13659](https://doi.org/10.2196/13659)] [Medline: [31293245](https://pubmed.ncbi.nlm.nih.gov/31293245/)]
13. Academy of Medical Royal Colleges. London; 2019 Jan 28. Artificial Intelligence in Healthcare URL: <https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/> [accessed 2019-10-16]
14. Shen J, Zhang CJP, Jiang B, Chen J, Song J, Liu Z, et al. Artificial Intelligence Versus Clinicians in Disease Diagnosis: Systematic Review. *JMIR Med Inform* 2019 Aug 16;7(3):e10010 [FREE Full text] [doi: [10.2196/10010](https://doi.org/10.2196/10010)] [Medline: [31420959](https://pubmed.ncbi.nlm.nih.gov/31420959/)]
15. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. *J Med Internet Res* 2019 Apr 05;21(4):e12887 [FREE Full text] [doi: [10.2196/12887](https://doi.org/10.2196/12887)] [Medline: [30950796](https://pubmed.ncbi.nlm.nih.gov/30950796/)]
16. Turing AM. Computing Machinery and Intelligence. *Mind, New Series* 1950 Oct;59(236):433-460 Published by Oxford University Press on behalf of the Mind Association [FREE Full text]
17. Innes AD, Champion PD, Griffiths FE. Complex consultations and the 'edge of chaos'. *Br J Gen Pract* 2005 Jan;55(510):47-52 [FREE Full text] [Medline: [15667766](https://pubmed.ncbi.nlm.nih.gov/15667766/)]
18. Barry MJ, Edgman-Levitan S. Shared decision making - The pinnacle of patient-centered care. *N Engl J Med* 2012 Mar 01;366(9):780-781. [doi: [10.1056/NEJMp1109283](https://doi.org/10.1056/NEJMp1109283)] [Medline: [22375967](https://pubmed.ncbi.nlm.nih.gov/22375967/)]
19. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ* 1996 Jan 13;312(7023):71-72 [FREE Full text] [doi: [10.1136/bmj.312.7023.71](https://doi.org/10.1136/bmj.312.7023.71)] [Medline: [8555924](https://pubmed.ncbi.nlm.nih.gov/8555924/)]
20. Autor DH. Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic Perspectives* 2015 Aug;29(3):3-30. [doi: [10.1257/jep.29.3.3](https://doi.org/10.1257/jep.29.3.3)]
21. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful implementation and adoption of large-scale health information technology. *J Am Med Inform Assoc* 2013 Jun;20(e1):e9-e13 [FREE Full text] [doi: [10.1136/amiajnl-2013-001684](https://doi.org/10.1136/amiajnl-2013-001684)] [Medline: [23599226](https://pubmed.ncbi.nlm.nih.gov/23599226/)]
22. Pope C, Halford S, Turnbull J, Prichard J, Calestani M, May C. Using computer decision support systems in NHS emergency and urgent care: ethnographic study using normalisation process theory. *BMC Health Serv Res* 2013 Mar 23;13:111 [FREE Full text] [doi: [10.1186/1472-6963-13-111](https://doi.org/10.1186/1472-6963-13-111)] [Medline: [23522021](https://pubmed.ncbi.nlm.nih.gov/23522021/)]
23. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q* 2004;82(4):581-629 [FREE Full text] [doi: [10.1111/j.0887-378X.2004.00325.x](https://doi.org/10.1111/j.0887-378X.2004.00325.x)] [Medline: [15595944](https://pubmed.ncbi.nlm.nih.gov/15595944/)]
24. Poote AE, French DP, Dale J, Powell J. A study of automated self-assessment in a primary care student health centre setting. *J Telemed Telecare* 2014 Apr;20(3):123-127. [doi: [10.1177/1357633X14529246](https://doi.org/10.1177/1357633X14529246)] [Medline: [24643948](https://pubmed.ncbi.nlm.nih.gov/24643948/)]

Abbreviations

AI: artificial intelligence

Edited by G Eysenbach; submitted 11.09.19; peer-reviewed by B Xie; accepted 12.10.19; published 28.10.19

Please cite as:

Powell J

Trust Me, I'm a Chatbot: How Artificial Intelligence in Health Care Fails the Turing Test

J Med Internet Res 2019;21(10):e16222

URL: <http://www.jmir.org/2019/10/e16222/>

doi: [10.2196/16222](https://doi.org/10.2196/16222)

PMID: [31661083](https://pubmed.ncbi.nlm.nih.gov/31661083/)

©John Powell. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 28.10.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.