Editorial

Public Health Genomics

Public Health Genomics
DOI: 10.1159/000506014

Received: November 27, 2019
Accepted: January 20, 2020
Published online: January 28, 2020

Is Explanation a Marketing Problem? The Quest for Trust in Artificial Intelligence and Two Conflicting Solutions

Stefano Triberti[a, b] Ilaria Durosini[b] Giuseppe Curigliano[c]
Gabriella Pravettoni[a, b]

[a]Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy; [b]Applied Research Division for Cognitive and Psychological Science, IEO, European Institute of Oncology IRCCS, Milan, Italy; [c]Division of Early Drug Development for Innovative Therapy, IEO, European Institute of Oncology IRCCS, Milan, Italy

Over the last decades, artificial intelligence (AI) has become more and more a part of our life, from personal to professional contexts. In the context of care, AI acts as a decision support tool which provides information not achievable by human ability alone [1–3]. Software based on machine learning, deep learning, or decision trees could analyze large amounts of patient data to identify patterns that lead to the identification of diagnosis or treatment. AI testing in a number of medical fields has led to impressive and encouraging results but its clinical value has still to be understood completely [4]. This requires not only to test AI effectiveness as a diagnostic support tool in healthcare context, but also to consider the possible consequences of its implementation in the clinical context.

For instance, while some healthcare professionals declare a notable enthusiasm about the implementation of new technologies, using such tools in everyday clinical practice is another story. Health professionals should continually take important decisions affecting patients' life or death as well as deal with patients' sensitive situations involving complex emotions [5–7] and are also exposed to an increasing risk of controversies and legal actions. Such risk may influence doctors' attitude and their clinical practice [8–10] so that trusting a medical device to take decisions that could affect patients' lives as well as one's own career is no smooth process.

A crucial issue is that a typical AI software operates as a "black box" providing data without a detailed explanation of the machine learning process that has led to results [11, 12]; also, despite the fact that AIs are becoming more and more able to analyze multiple and various kinds of data (e.g., natural language), health professionals could not be sure whether AIs are capable of detecting any kind of meaningful or confounding information (e.g., lying in consultation and clinical tests) [13]. On a practical note, we could not expect healthcare professionals to become experts in AI engineering and informatics. Yet, they deserve to be put in the condition of effortlessly using these devices, without worrying for mistakes [14]. One possible solution is XAI – or eXplainable Artificial Intelligence – a subdiscipline focused on teaching AIs how to explain their own outcomes and processes in an intelligible way to decision makers that are supposed to benefit from AIs' assistance [11]. Relevant to this aim is the realization that XAI is not a technical issue but a problem for the social sciences [15, 16], because building useful explanations within AI capabilities requires understanding how hu-

KARGER

E-Mail karger@karger.com
www.karger.com/phg

Stefano Triberti
Department of Oncology and Hemato-Oncology
University of Milan, Via Festa del Perdono 7
IT–20122 Milan (Italy)
E-Mail stefano.triberti@unimi.it

man professionals think and take decisions in the social context. In other words, the first option towards an evolution of technology is that AI becomes able to "talk human language" and adapt to the real-world contexts of use. Diversely, the alternative option consists in modifying users instead of the tool.

While participating in some conferences on AI and machine learning in healthcare, the authors of the present contribution noticed one specific reaction among some professionals (both engineers and doctors) regarding the trust and adoption issue, which could be summarized in one sentence: "Indeed everyday all of us use technologies we do not know how they work; it is normal. We just have to get used to them!" This is certainly true. Donald Norman, famous psychologist, engineer, and among the founders of the Usability field, demonstrated that users form their own representation of how a given technology works [17] entirely based on the interface – the part of the technology they could actually see and interact with. Thus, users' representation of technology has nothing to do with the designer or engineer's representation, which takes into account hardware or technical functioning and basically corresponds to how the device actually works. Thus, to design technologies that are easy to use, one should not teach common users how technologies actually work; rather, it is important to design interfaces taking into account the way users naturally think and act. In any case, it is true that humans learn to use tools independently of understanding how they actually work, and we could probably track back this attitude to the oldest tools, such as bow and arrow.

If it is natural to use technologies without understanding them, then trust and explanation are a problem for marketing and communication: enterprises designing and selling AI should just exploit persuasion techniques to reassure doctors about the reliability and trustworthiness of their products. Marketing psychology and captology (i.e., the discipline studying persuasion by technologies [18]) could give useful hints to this aim. For example, AI developers could (1) make AI devices and interfaces interesting, involving, and mildly challenging: a fluent and engaging interaction is pleasant per se, reducing critical thinking on the activity to be pursued [19, 20]; (2) give AIs a positive visual appealing with nice and pleasant features, especially on the graphics side: for example, personalization options and the inclusion of human-like virtual characters may improve trust and reduce risk of abandonment [21, 22]; (3) employ reputation and authority heuristic: it is known that people consistently rely on others to make decisions about trustworthiness or credibility of a source [23]; on this basis, having testimonials recognized as experts in the medical field would help AIs to be accepted by other professionals.

Marketing experts could certainly come up with additional useful strategies to promote AI usage, but is this really a solution? And if it is, is this the solution we should hope for?

Especially in the context of care, the usage of technologies without a detailed understanding of the machine learning process that leads to clinical outcomes can determine some obstacles to professional practice: (1) Employing the marketing strategy would promote a medical/technological culture according to which the best tools are those you can use without asking questions; in other words, reducing doctors' critical thinking on the resources they use on a daily basis. Secondarily, the promotion of a medical culture in which technologies can be used as "black boxes" can lead to overconfidence in doctors (i.e., relying on technology for tasks other than those it was originally designed to perform) and to deskilling (i.e., reducing user's abilities because activities can now be performed by machines; [24, 25]). In such a scenario, doctors would not be able to intervene effectively when AI does not work or is under maintenance.

(2) Despite the fact that AI software uses a large amount of patient data to identify patterns that lead to the identification of diagnosis or treatment, errors may still be made (e.g., miscalculations, inaccuracies, misinterpretations, overestimations, underestimations…). Doctors who use AI despite the fact that they are "black boxes" could find themselves in severe difficulties in understanding and justifying those errors and so, to defend themselves from possible legal actions or disciplinary measures.

(3) A limited understanding of how the AI arrived at conclusions (especially unexpected or counterintuitive ones) may result in healthcare professionals having a partial understanding of the diagnosis as a whole, with enhanced difficulty to explain it to patients and caregivers in the context of a collaborative approach and shared decision-making. In the long run, this could impact treatment efficacy and also the credibility of the professional and of the entire healthcare organization involved.

Generally, we can say that using medical devices (AI or not) without having a clear understanding of their contribution to practice means deceiving the patient, who trusts the medical expert and his or her professional choices. However, reducing trust to a matter of persuasion and communication-marketing would mean deceiving both the doctor and the patient, because the former is not reminded of the importance of being in complete control of the diagnostic and therapeutic process.

Future research, development, and implementation of AI for medical practice should take into account the risks inherent in the choice outlined here. Despite the fact that developing AI able to explain its own processing (so, able to effectively collaborate with human doctors) is the hard way, it is also the option that points towards authentic empowerment of healthcare, because health professionals are then able to use their tools with awareness. Furthermore, AI for healthcare development should not forget patients and caregivers, namely the ultimate benefiters, so that AI too can be used not as a mere device but (through proper explanation in the collaborative decision-making) a tool for patient engagement in their own healthcare journey [26–28].

On the one hand, XAI researchers should probably give more and more importance to the real-life contexts of AI utilization, to design interface devices based on awareness of professionals' knowledge, needs, and abilities; on the other hand, the formation of future doctors should continue to cultivate health professionals' critical thinking to face the risks of overconfidence and deskilling. In this scenario, as envisioned by recent treatises on the topic [29, 30], it would be possible that AI becomes not only a resource to improve diagnosis and treatment identification, but a technology that helps doctors to recover the "time lost" in technical tasks. Such time could be given back to medical consultation, so as to improve active listening, shared decision-making, and a truly patient-centered approach to medical practice.

Acknowledgements

This work was partially supported by the Italian Ministry of Health with Ricerca Corrente and 5x1000 funds for IEO European Institute of Oncology IRCCS. S.T. and G.P. were also supported by MIUR – Italian Ministry of University and Research (Departments of Excellence Italian Law n.232, 11th December 2016) for University of Milan. I.D. was supported by Fondazione Umberto Veronesi.

Statement of Ethics

This contribution did not involve human or animal participants.

Disclosure Statement

The authors declare that they have no conflicts of interest.

Funding Sources

This contribution received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions

All persons who meet authorship criteria are listed as authors. All authors attest that they have participated sufficiently in the work reported to take public responsibility for the material reported. S.T. and I.D. prepared an initial draft of the manuscript. G.C. and G.P. supervised the process and contributed important intellectual content to the ideas presented. All authors read, edited, and approved the final manuscript.

References

1 Fazal MI, Patel ME, Tye J, Gupta Y. The past, present and future role of artificial intelligence in imaging. Eur J Radiol. 2018 Aug;105: 246–50.

2 Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017 Jun;2(4):230–43.

3 Horgan D, Romao M, Morré SA, Kalra D. Articifial interlligence: power for civilisation – and for better healthcare. Public Health Genomics. 2019. https://doi.org/10.1159/000504785.

4 Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf. 2019 Mar;28(3):231–7.

5 Durosini I, Tarocchi A, Aschieri F. Therapeutic Assessment with a Client with Persistent Complex Bereavement Disorder: A Single-Case Time-Series Design. Clin Case Stud. 2017;16(4):295–312.

6 Arnaboldi P, Riva S, Crico C, Pravettoni G. A systematic literature review exploring the prevalence of post-traumatic stress disorder and the role played by stress and traumatic stress in breast cancer diagnosis and trajectory. Breast Cancer (Dove Med Press). 2017 Jul;9:473–85.

7 Fioretti C, Mazzocco K, Pravettoni G. Psychological Support in Breast Cancer Patients: A Personalized Approach. In: Veronesi U, Goldhirsch A, Veronesi P, Gentilini O, Leonardi M, editors. Breast Cancer. New York: Springer International Publishing; 2017. pp. 841–7.

8 Nahed BV, Babu MA, Smith TR, Heary RF. Malpractice liability and defensive medicine: a national survey of neurosurgeons. PLoS One. 2012;7(6):e39237.

9 Rothberg MB, Class J, Bishop TF, Friderici J, Kleppel R, Lindenauer PK. The cost of defensive medicine on 3 hospital medicine services. JAMA Intern Med. 2014 Nov;174(11):1867–8.

10 Sekhar MS, Vyas N. Defensive medicine: a bane to healthcare. Ann Med Health Sci Res. 2013 Apr;3(2):295–6.

11 Adadi A, Berrada M: Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). IEEE Access 2018;6: 52138–52160.

12 Castelvecchi D. Can we open the black box of AI? Nature. 2016 Oct;538(7623):20–3.

13 Fantini F, Banis A, Dell'Acqua E, Durosini I, Aschieri F. Exploring Children's Induced Defensiveness to the Tell Me a Story Test (TEMAS). J Pers Assess. 2017 May-Jun;99(3):275–85.

14 Iannello P, Perucca V, Riva S, Antonietti A, Pravettoni G. What do physicians believe about the way decisions are made? A pilot study on metacognitive knowledge in the medical context. Eur J Psychol. 2015 Nov;11(4):691–706.

15 Miller T. Explanation in artificial intelligence: insights from the social sciences. Artif Intell. 2019;267:1–38.

16 Miller T, Hower P, Sonenberg L, Explainable AI. Beware of Inmates Running the Asylum. IJCAI (U S). 2017.

17 Norman DA. Some observations on mental models. In Mental Models 1983;15-22.

18 Fogg BJ. Captology: The study of computers as persuasive technologies. In Conference on Human Factors in Computing Systems - Proceedings 1997.

19 Appel M, Richter T. Transportation and need for affect in narrative persuasion: A mediated moderation model. Media Psychol. 2010;13(2):101–35.

20 Vanwesenbeeck I, Ponnet K, Walrave M. Go with the flow: how children's persuasion knowledge is associated with their state of flow and emotions during advergame play. J Consum Behav. 2016;15(1):38–47.

21 Desmet P, Hekkert P. Framework of product experience. International journal of design, 1(1)2007. National Science Council, Taipei, 30-Apr-2007.

22 Triberti S, Chirico A, La Rocca G, Riva G. Developing emotional design: emotions as cognitive processes and their role in the design of interactive technologies. Front Psychol. 2017 Oct;8:1773.

23 Metzger MJ, Flanagin AJ, Medders RB. Social and heuristic approaches to credibility evaluation online. J Commun. 2010;60(3):413–39.

24 Lu J. Will Medical Technology Deskill Doctors? Int Educ Stud. 2016;9(7):130–4.

25 Triki A, Weisner MM. Lessons from the literature on the theory of technology dominance: possibilities for an extended research framework. J Emerg Technol Account. 2014;11(1):41–69.

26 Graffigna G, Barello S, Triberti S. Patient Engagement: A consumer-centered model to innovate healthcare. Berlin: De Gruyter Open; 2015. https://doi.org/10.1515/9783110452440.

27 Aschieri F, De Saeger H, Durosini I. Therapeutic assessment and collaborative: empirical evidence. Prat Psychol. 2015;21:307–17.

28 Renzi C, Riva S, Masiero M, Pravettoni G. The choice dilemma in chronic hematological conditions: why choosing is not only a medical issue? A psycho-cognitive perspective. Crit Rev Oncol Hematol. 2016 Mar;99:134–40.

29 Topol E. Deep Medicine. Basic Books; 2019.

30 Pravettoni G, Triberti S: Il Medico 4.0. EDRA, 2019.