Review Article

Kidney Diseases

Kidney Dis 2019;5:11–17
DOI: 10.1159/000492428

Received: July 23, 2018
Accepted: July 25, 2018
Published online: September 3, 2018

Societal Issues Concerning the Application of Artificial Intelligence in Medicine

Alfredo Vellido

Intelligent Data Science and Artificial Intelligence (IDEAI) Research Center, Universitat Politècnica de Catalunya (UPC BarcelonaTech), Barcelona, Spain

Keywords
Artificial intelligence in medicine · Machine learning · Social impact

Abstract
Background: Medicine is becoming an increasingly data-centred discipline and, beyond classical statistical approaches, artificial intelligence (AI) and, in particular, machine learning (ML) are attracting much interest for the analysis of medical data. It has been argued that AI is experiencing a fast process of commodification. This characterization correctly reflects the current process of *industrialization* of AI and its reach into society. Therefore, societal issues related to the use of AI and ML should not be ignored any longer and certainly not in the medical domain. These societal issues may take many forms, but they all entail the design of models from a human-centred perspective, incorporating human-relevant requirements and constraints. In this brief paper, we discuss a number of specific issues affecting the use of AI and ML in medicine, such as fairness, privacy and anonymity, explainability and interpretability, but also some broader societal issues, such as ethics and legislation. We reckon that all of these are relevant aspects to consider in order to achieve the objective of fostering acceptance of AI- and ML-based technologies, as well as to comply with an evolving legislation concerning the impact of digital technologies on ethically and privacy sensitive matters. Our specific goal here is to reflect on how all these topics affect medical applications of AI and ML. This paper includes some of the contents of the "2nd Meeting of Science and Dialysis: Artificial Intelligence," organized in the Bellvitge University Hospital, Barcelona, Spain. ***Summary and Key Messages:*** AI and ML are attracting much interest from the medical community as key approaches to knowledge extraction from data. These approaches are increasingly colonizing ambits of social impact, such as medicine and healthcare. Issues of social relevance with an impact on medicine and healthcare include (although they are not limited to) fairness, explainability, privacy, ethics and legislation.

Introduction

Medicine, as part of a phenomenon that affects all fields of life sciences, is becoming an increasingly data-centred discipline [1]. Data analysis in medicine has for

Contribution from the 2nd meeting of "Science for Dialysis," organized at the University Hospital of Bellvitge, L'Hospitalet de Llobregat, Barcelona, Spain, on September 28, 2018.

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

Alfredo Vellido, PhD
Intelligent Data Science and Artificial Intelligence (IDEAI) Research Center
Universitat Politècnica de Catalunya (UPC BarcelonaTech)
C/ Jordi Girona, 1–3, ES–08034 Barcelona (Spain)
E-Mail avellido@cs.upc.edu

long been the territory of statisticians, but medical data are reaching beyond the merely quantitative to take more complex forms, such as, for instance, textual information in Electronic Health Records (EHR), images in many modalities, on their own or mixed with other types of signals, or graphs describing biochemical pathways or biomarker interactions [2]. This data complexity is behind the evolution from classical multivariate data analysis towards the nascent field of *data science* [3], which, from the point of view of medicine, embraces a new reality that includes interconnected wearable devices and sensors.

Beyond the more classical statistical approaches, artificial intelligence (AI) and, more in particular, machine learning (ML) are attracting much interest for the analysis of medical data, even if arguably with a relatively low impact yet on clinical practice [4]. It has been acknowledged that AI is experiencing a fast process of commodification (not that this is an entirely new concern, as it was already a matter of academic discussion almost 30 years ago [5]). This characterization is mostly of interest to big IT companies but correctly reflects the current process of *industrialization* of AI, where the academic and industrial limits of research are increasingly blurred, with the main experts in AI and ML on the payroll of private companies. In any case, this means that AI systems and products are reaching the society at large, and, therefore, that societal issues related to the use of AI in general and ML in particular should not be ignored any longer and certainly not in the medicine and healthcare domains.

These societal issues may take many forms, but, more often than not, they entail the design of models from a human-centred perspective, that is, models that incorporate human-relevant requirements and constraints. This is certainly an only partially technical matter.

In this brief paper, we cover, in a non-exhaustive manner, a number of specific societal issues affecting the development of AI and ML methods, such as fairness, privacy and anonymity, and explainability and interpretability, but also some broader societal issues, such as ethics and legislation. Not that these issues should be considered independently; on the contrary, they often overlap in an intricate manner. Let us summarily list them here:

Legislation. The industrialization of AI exposes it to legislation regulating the social domain where it is meant to operate. In some cases, this overlaps issues of privacy and anonymity, such as in AI algorithms used for automated face recognition in public domains. It may also involve more general contexts, such as AI-based autonomous driving or defence weapons. Legislation is also involved in medicine and healthcare practice, and, therefore, we need to ensure that AI and ML technologies comply with current legislation.

Explainability and Interpretability. ML and AI algorithms are often characterized as *black boxes*, that is, methods that generate data models that are difficult (if not impossible) to interpret because the functional form relating the available data (input) to a given outcome (the output) is far too complex. This problem has been exacerbated by the intensity of the current interest in deep learning (DL) methods. Only interpretable models can be explained, and explainability is paramount when decision-making in medicine (diagnosis, prognosis, etc.) must be conveyed to humans.

Privacy and Anonymity. Privacy-preserving ML-based data analysis must deal with the potentially contradictory problem of keeping personal information private while aiming to model it, often to make inferences that will affect a given population. Data anonymity obviously refers to the impossibility of linking personal data with information about the individual that is not meant to be revealed. These are key problems and concerns in the medical and healthcare domains, mainly in the interaction between the public and private sectors.

Ethics and Fairness. Biological intelligence is multifaceted and responds to the environmental pressures of human societies. Ethics are one of those facets for which AI is still fairly unprepared. Interestingly, this topic has become central to AI discussion in recent years. Needless to say, ethics are also a core concern in medicine and healthcare. Such convergence of interests makes it important to create a clear roadmap for the ethical use of AI and ML in medicine. The application of ML and AI in areas of social relevance must also aspire to be *fair*. How do we imbue ML algorithms, which are fairness *agnostic*, with fairness requirements? How do we avoid gender or ethnicity, for instance, unfairly influencing the outcome of a learning algorithm? In the medical domain and in healthcare in particular, where sensible information about the individual may be readily available, how do we ensure that AI- and ML-based decision support tools are not affected by such bias?

We reckon that all of these are relevant aspects to consider in order to achieve the objective of fostering acceptance of AI- and ML-based technologies in the medical and healthcare domains, as well as to comply with an evolving legislation concerning the impact of digital technologies on ethically and privacy sensitive matters. Our specific goal here is to reflect on how all these topics affect medical applications of AI and ML.

This paper reflects some topics addressed in the "2nd Meeting of Science and Dialysis: Artificial Intelligence," organized at the Bellvitge University Hospital, Barcelona, in the Catalonia region of Spain.

Societal Issues of AI and ML Application

Legislation

Human societies are regulated by bodies of legislation. While remaining within the academic realm, AI and ML developments have stayed fairly oblivious to legal concerns, but the moment these technologies start occupying the social space at large, their impact on people is likely to hit a few legal walls. One widely discussed case is the use of AI as the basis for autonomously driving vehicles. When a human is in charge of any decision-making at the wheel of a vehicle, legal responsibilities are quite clearly drawn. The quick industrial development of semi-autonomous vehicles, leading towards the objective of fully autonomous driving, has stretched the seams of current legislation, though.

Again, any application of AI and ML in actual medical practice is bound to generate discussion about its legal boundaries and implications. A pertinent example is the recent (May 2018) implementation of the European Union directive for General Data Protection Regulation (GDPR). This directive mandates a *right to explanation* of all decisions made by "automated or artificially intelligent algorithmic systems" [6]. According to Article 13 of the directive, the right to explanation implies that the "data controller" is legally bound to provide requesting citizens with "meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing [automated decision making, as described in its Article 22] for the data subject" [6]. AI and ML may be the tools used to provide such automated decision making, and, therefore, it places these technologies in a legal spotlight. Some guidelines for GDPR-compliant ML development have recently been provided by Veale et al. [7].

The implications of GDPR for the use of AI and ML in medicine and healthcare are not too difficult to appreciate. Any AI- or ML-based medical decision support system (MDSS) whose purpose it is to assist the medical experts in their decision-making will be explicitly providing a (semi)automated decision on an individual (for instance, diagnosis, prognosis or recommendations on treatment concerning individual patients, perhaps even in life-threatening conditions). The data controller in this case will be the medical expert (from nurses to specialists [8]) and the institution this expert belongs to.

Note that this piece of legislation (of compulsory application in all countries belonging to the European Union) requires something very specific from the AI and ML technologies (or, more accurately, from the people designing, implementing and using them): interpretable and explainable models, as discussed in the next section. A medical expert or any healthcare system employee using these technologies must be able to interpret how they reached specific decisions (say, why an ML model diagnosed a brain tumour as a metastasis and not a high-grade glioma) and must be able to explain those decisions to any human affected by them. In the implementation of the artificial kidney as one of the most promising technologies in nephrology, we should be concerned, for instance, about the possibility of an opaque AI- or ML-based alarm system not being able to explain the basis for a false alarm that might endanger the life of the dialysis patient.

At a higher level, and on the basis of legal safeguards such as the GDPR, a healthcare system might decide not to implement an opaque MDSS in clinical practice, despite its perceived effectiveness, only to avoid the prospect of unsustainable litigation costs caused by the false-positive and -negative cases or the incorrect estimations and predictions churned by these automated systems.

In the light of this discussion, we recommend that medical experts and healthcare practitioners should keep in mind the need to balance the effectiveness of AI- and ML-based technologies and their adherence to current legislation. Beyond GDPR and its relation to interpretability, this issue overlaps with some of the others we will discuss in the following sections, such as ethics, fairness, and privacy and anonymity.

Interpretability and Explainability

Biological brains have not necessarily evolved the means to explain themselves. Arguably, this has only happened in species with social behaviour (although it could also be argued that social behaviour can only happen in species whose brains are capable of *explaining themselves* through some form of communication). In the human species, natural language performs that communicative or explanatory function.

AI was originally conceived as an attempt to reproduce aspects of biological intelligence, but self-explanatory capabilities were never a key aspect to consider. If the biological brain was meant to be understood as a form of

information-processing system, so was AI, and the idea of *social* AI is relatively new, for instance in the form of intelligent agents and multi-agent systems [9]. Only recently, the interpretability and explainability of AI and ML systems has come to the forefront of research in the field [10]. One key reason for this is the breakthrough created by DL technologies. DL is an augmented version of traditional artificial neural networks. The latter were long ago maligned as *black box* opaque models. DL models risk being considered augmented black boxes. Interpretability in this context can be seen as a human-computer interaction problem. We humans must be able to understand and interpret the outcome of an AI or ML model. That is, we need to ensure that even a very complex model can be explained (usually to other humans). A human brain, colossally more complex, has developed natural language to convey some level of explanation of its inner workings. Similar attempts with AI and ML are still very limited. Despite recent and thorough attempts to address the issue of how to characterize interpretability in ML [11], such attempts only highlight the tremendous difficulty involved in the scientific pursue of truly interpretable ML models.

In the medical domain, AI and ML models are often part of MDSS. Their potential and the possible barriers to their adoption have been investigated in the last decade [12]. The paradox is that these methods, despite their advantages, are far from universal acceptance in medical practice. Arguably, one of the reasons is precisely (lack of) interpretability, expressed as "the need to open the machine learning black box" [13]. As already mentioned, DL-based technologies can worsen the problem, despite having already found their way into biomedicine and healthcare [14, 15]. In medicine, this has clear implications: if an ML-based MDSS makes decisions that cannot be comprehensibly explained, the medical expert can be put in the uncomfortable position of having to vouch for the system's trustworthiness, transferring the trust on a decision that she or he cannot explain to either the patient or to other medical experts. This does not mean to say that efforts have not been made to imbue MDSS with knowledge representations that are comprehensible to humans. Examples include rule-based representations, usually compatible with medical reasoning [16]; and nomograms, commonly used by clinicians for visualizing the relative weights of symptoms on a diagnosis or a prognosis [17].

AI- and ML-based systems may have quantifiable goals and may still be useless unless they conform to clinical guidelines. Note that computer-based systems, such as MDSS, are often seen by clinicians as an extra burden in their day-to-day practice [18]. The problem may appear when the MDSS conflicts with guidelines of medical practice [19], something bound to happen unless those guidelines are somehow fed as *prior knowledge* to the intelligent systems. In this scenario, interpretability might be seen as an opportunity to make model performance and compliance with guidelines compatible goals.

The role of ML in healthcare has been described as acting "as a tool to aid and refine specific tasks performed by human professionals" [20]. Note that this means that interpretability should not be considered here a fully technical issue dissociated from the cognitive abilities of the human interpreter. As acknowledged by Dreiseitl and Binder [12] when discussing the weak levels of adoption of MDSS at the point of care, researchers often sidestep practical questions, such as whether adequate "explanations [are] given for the system's diagnosis"; "the form of explanation [is] satisfactory for the physicians using the system"; or "how intuitive is its use."

An effort should be made to integrate medical expert knowledge into the AI and ML models or use prior expert knowledge in formal frameworks for machine-human interaction in the pursuit of interpretability and explainability. The data analyst must play a proactive role in seeking medical expert verification. In return, the medical expert should ensure that the analysis outcomes are interpretable and usable in medical practice.

Privacy and Anonymity

Technological advances and the widespread adoption of networked computing and telecommunication systems are flooding our societies (and mostly governments and technology providers) with data. The physical society bonds are being swiftly amplified by our use of virtual social networks. In this scenario, data privacy and anonymity have become main social concerns and have triggered legal initiatives, such as the European GDPR discussed in previous sections.

Needless to say, privacy and anonymity have been a core concern for healthcare systems for far longer than for society at large. The current adoption of EHRs in medical practice enhances this issue, as sensitive patient data are uploaded in digital form to networked systems with varying levels of security systems in place. An interesting review on security and privacy in EHRs can be found in the study by Fernández-Alemán et al. [21]. The strong links between privacy and anonymity, on one side, and

legislation, on the other, are clearly described in this study, although it is also acknowledged that "there has been very little activity in policy development involving the numerous significant privacy issues raised by a shift from a largely disconnected, paper-based health record system to one that is integrated and electronic" [21].

This is not an issue ignored by the AI and ML communities. As early as 2002, data confidentiality and anonymity in data mining medical applications were already discussed in journals of these fields [22], highlighting the responsibilities of *data miners* to human subjects. Privacy-preserving models and algorithms have been discussed in some detail [23]. A commonplace situation for data analysts in clinical environments is the need to analyse data that are distributed among multiple clinical parties. These parties (e.g., hospitals) may have privacy protocols in place that prevent merging data from different origins into centralized locations (in other words, prevent data "leaving" a given hospital). The AI and ML communities have already worked on producing decentralized analytical solutions to bypass this bottleneck [24].

There is a new and disruptive element of the privacy and anonymity discussion in AI and ML applications in medicine that must be considered: the *en masse* landing of big IT corporations in the medical field, many of them proposing or integrating AI elements (some examples would be Microsoft's Hanover project, IBM's Watson Oncology, or Google's DeepMind), together with a myriad of AI-based medically oriented start-ups [25]. The involvement of IT companies in health provision raises the bar for privacy and anonymity issues that were already on the table due to the pressure of insurance companies, especially in the most liberalized national health systems. An illustrative example of the complexities and potential drawbacks of this involvement can be found in *Nature* journal's report of the UK Information Commissioner's Office declaration that the operator of three London-based hospitals "had broken civil law when it gave health data to Google's London-based subsidiary DeepMind" [26]. These data were meant to be the basis for models to test results for signs of acute kidney injuries, but privacy and protocols of identification were breached in a large-scale transference of patients' data from the hospitals to the private company. According to the Royal Statistical Society's executive director, three lessons are to be extracted from this particular case of application to the medical domain: (1) due to society's increasing data trust deficit, data transference transparency and openness should be guaranteed; (2) data transference should be proportional to the medical task at hand (in this case, the development of models for the detection of signs of acute kidney injury); and (3) governance (not just legislation) mechanisms of control of data handling, management and use should be strengthened or created when necessary. He also makes a key statement when saying that "innovations such as artificial intelligence, machine learning […] offer great opportunities, but will falter without a public consensus around the role of data" [26].

Ethics and Fairness

The time-honoured ultimate aspiration of AI is to replicate biological intelligence in silico. Biological intelligence, though, is the product of evolution and, as such, is multi-faceted and at least to some extent the product of environmental pressures of human societies. Ethics, as a compass for human decision-making, are one of those facets and could be argued to provide the foundations for the legislative regulation of societies, whose importance for medical applications of AI and ML has already been discussed in this paper.

The truth though is that the AI and ML fields are still fairly unprepared to address this pressing matter [27]. Interestingly, this topic has become central to AI discussion only in recent years, once it has also become a central topic in global research agendas [28]. In what sense might ethics be part of the AI and ML equation and in what sense do we want these technologies be imbued with ethical considerations, beyond the overlap with bodies of regulation and legislation? Let us provide an illustrative example: the ongoing debate on the use of AI as part of autonomous weapons systems in defence and warfare. Unmanned autonomous vehicles, at least partially driven by AI, are being used for targeted bombing in areas of conflict. The ethical issues involved in human decisions concerning the choice of human targets in war periods are quite clearly delineated by international conventions, but who bears ethical responsibility in the case of targets at least partially chosen by AI-driven machines? This type of problem currently drives not-for-profit organization campaigns, such as those undertaken by Article 36 [29], "to stop killer robots" [30].

Needless to say, ethics are also a core concern in medicine and healthcare that has attracted much academic discussion [31]. Can AI- and ML-supported tools address the basic biomedical ethical principles of respect for autonomy, non-maleficence, beneficence and justice? Should they, or should this be left to the medical practitioners? Medical practitioners, though, do not usually de-

velop the AI and ML tools for medical application. Should they at least ensure that AI and ML developers do not transgress these principles in the design of such tools? According to Magoulas and Prentza [32], it is humans and not systems who can identify ethical issues, and, therefore, it is important to consider "the motivations and ethical dilemmas of researchers, developers and medical users of ML methods in medical applications."

Such convergence of interests makes it important, in any case, to create a clear roadmap for the ethical use of AI and ML in medicine that involves players both from the fields of medicine and AI.

The concept of *fairness* may be considered as subjective as the concept of ethics and, perhaps, more vaguely defined. If distinguishing what is fair and what is not in a human society is difficult and often controversial, trying to embed the concept of fairness in AI-based decision-making might be seen as a hopeless endeavour. Nevertheless, the use of ML and AI in socially relevant areas should at least aspire to be *fair*. As stated by Veale and Binns [33], "real-world fairness challenges in ML are not abstract, […] but are institutionally and contextually grounded."

Let us illustrate this with an example: gender bias can be added to an ML model by just biasing the choice with which the data used to train the model are selected. Caliskan et al. [34] have recently shown that semantics derived automatically using ML from language corpora will incorporate human-like stereotyped biases. As noted by Veale and Binns [33], lack of fairness may sometimes be the inadvertent result of organisations not holding data on sensitive attributes, such as gender, ethnicity, sexuality or disability, due to legal, institutional or commercial reasons. Without such data, indirect discrimination-by-proxy risks are being increased.

In the medical domain and in healthcare in particular, where sensible information about the individual may be readily available, how do we ensure that AI- and ML-based decision support tools are not affected by such bias? Fairness constraints can be integrated in learning algorithms, as shown in a study by Celis et al. [35]. Given that fairness criteria are reasonably clean-cut in the medical context, such constraints should be easier to integrate than in other domains. Following Veale and Binns [33], fairness may be helped by trusting third parties with the selective storage of those data that might be necessary for incorporating fairness constraints into model-building in a privacy-preserving manner. A recent proposal of a "continuous framework for fairness" [36] seeks to subject decision makers to fairness constraints that can be operationalized in an algorithmic (and therefore in AI and ML) setting, with such constraints facilitating a trade-off between individual and group fairness, a type of trade-off that could have clear implications in medical domains from access to drugs and health services to personalized medicine.

Conclusions

AI and ML have, for decades, been mostly investigated and developed within the academic environment, with some inroads into broader social domains. Over the last years, though, these fields are experiencing an intense process of industrialization that comes with societal strings attached. Many of these should concern medical and healthcare practice and have been brought to attention and discussed in this paper. We have considered legislation, ethics and fairness, interpretability and explainability and privacy and anonymity, but further issues, such as robustness and safety, economics and accessibility, or complex data management, could have also been considered. Our closing remark is a call for the collaboration between the AI-ML and medicine-healthcare communities in the pursuit of methods, protocols, guidelines and data analysis pipelines that explicitly take into consideration all these societal issues.

Statement of Ethics

No ethical approval was required for the writing of this article.

Disclosure Statement

There is no conflict of interest affecting the contents of this article.

Funding Sources

This research was funded by Spanish MINECO TIN2016-79576-R research project.

References

1 Leonelli S: Data-Centric Biology: A Philosophical Study. University of Chicago Press, 2016.
2 Bacciu D, Lisboa PJ, Martín JD, Stoean R, Vellido A: Bioinformatics and medicine in the era of deep learning; in Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN). Bruges, Belgium, i6doc.com, 2018, pp 345–354.
3 Provost F, Fawcett T: Data science and its relationship to big data and data-driven decision making. Big Data 2013;1:51–59.
4 Deo RC: Machine learning in medicine. Circulation 2015;132:1920–1930.
5 Cornwall-Jones K: The Commercialization of Artificial Intelligence in the UK. Doctoral dissertation, University of Sussex, UK, 1990.
6 Goodman B, Flaxman S: European Union regulations on algorithmic decision making and a "right to explanation." AI Mag 2017;38.
7 Veale M, Binns R, Van Kleek M: Some HCI priorities for GDPR-compliant machine learning. arXiv preprint arXiv:1803.06174, 2018.
8 O'Connor S: Big data and data science in health care: what nurses and midwives need to know. J Clin Nurs 2017, DOI: 10.1111/jocn.14164.
9 Castelfranchi C: The theory of social functions: challenges for computational social science and multi-agent learning. Cogn Syst Res 2001;2:5–38.
10 Vellido A, Martín-Guerrero JD, Lisboa PJG: Making machine learning models interpretable; in: Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2012). Bruges, Belgium, i6doc.com, 2012, pp 163–172.
11 Doshi-Velez F, Kim B: Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608, 2017.
12 Dreiseitl S, Binder M: Do physicians value decision support? A look at the effect of decision support systems on physician opinion. Artif Intell Med 2005;33:25–30.
13 Cabitza F, Rasoini R, Gensini GF: Unintended consequences of machine learning in medicine. JAMA 2017;318:517–518.
14 Mamoshina P, Vieira A, Putin E, Zhavoronkov A: Applications of deep learning in biomedicine. Mol Pharm 2016;13:1445–1454.
15 Ravì D, Wong C, Deligianni F, Berthelot M, Andreu-Pérez J, Lo B, Yang GZ: Deep learning for health informatics. IEEE J Biomed Health 2017;21:4–21.
16 Rögnvaldsson T, Etchells TA, You L, Garwicz D, Jarman I, Lisboa PJ: How to find simple and accurate rules for viral protease cleavage specificities. BMC Bioinformatics 2009;10:149.
17 Van Belle V, Van Calster B, Van Huffel S, Suykens JAK, Lisboa PJ: Explaining Support Vector Machines: a color based nomogram. PLoS One 2016;11:e0164568.
18 Ash JS, Berg M, Coiera E: Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. JAMA 2004;11:104–112.
19 Hoff T: Deskilling and adaptation among primary care physicians using two work innovations. Health Care Manage Rev 2011;36:338–348.
20 Reid MJ: Black-box machine learning: implications for healthcare. Polygeia 2017;April 6.
21 Fernández-Alemán JL, Señor IC, Lozoya PÁO, Toval A: Security and privacy in electronic health records: a systematic literature review. J Biomed Inform 2013;46:541–562.
22 Berman JJ: Confidentiality issues for medical data miners. Artif Intell Med 2002;26:25–36.
23 Aggarwal CC, Philip SY: A general survey of privacy-preserving data mining models and algorithms; in Aggarwal CC, Philip SY (eds): Privacy-Preserving Data Mining. Boston, MA, Springer, 2008, pp 11–52.
24 Scardapane S, Altilio R, Ciccarelli V, Uncini A, Panella M: Privacy-preserving data mining for distributed medical scenarios; in Esposito A, Faudez-Zanuy M, Morabito FC, Pasero E (eds): Multidisciplinary Approaches to Neural Computing. Springer, 2018, pp 119–128.
25 The Medical Futurist: Top Artificial Intelligence Companies in Healthcare to Keep an Eye On. January 31, 2017. http://medicalfuturist.com/top-artificial-intelligence-companies-in-healthcare (accessed June 2018).
26 Shah H: The DeepMind debacle demands dialogue on data. Nature 2017;547:259.
27 Moor JH: The nature, importance, and difficulty of machine ethics. IEEE Intell Syst 2006;21:18–21.
28 Ladikas M, Stemerding D, Chaturvedi S, Zhao Y: Science and Technology Governance and Ethics: A Global Perspective from Europe, India and China. Springer, 2015.
29 Article 36. http://www.article36.org.
30 Campaign to stop killer robots. https://www.stopkillerrobots.org.
31 Beauchamp T, Childress J: Principles of Biomedical Ethics, ed 7. New York, Oxford University Press, 2013.
32 Magoulas GD, Prentza A: Machine learning in medical applications; in Paliouras G, Karkaletsis V, Spyropoulos CD (eds): Machine Learning and Its Applications. Advanced Course on Artificial Intelligence. Berlin, Heidelberg, Springer, 1999, pp 300–307.
33 Veale M, Binns R: Fairer machine learning in the real world: mitigating discrimination without collecting sensitive data. Big Data Society 2017;4:2053951717743530.
34 Caliskan A, Bryson JJ, Narayanan A: Semantics derived automatically from language corpora contain human-like biases. Science 2017;356:183–186.
35 Celis LE, Straszak D, Vishnoi NK: Ranking with fairness constraints. arXiv preprint arXiv:1704.06840, 2017.
36 Hacker P, Wiedemann E: A continuous framework for fairness. arXiv preprint arXiv:1712.07924, 2017.