European Journal of Radiology 122 (2020) 108768

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Review

Ethical considerations in artificial intelligence

Nabile M. Safdar[a,c], John D. Banja[b,f], Carolyn C. Meltzer[a,d,e,*]

[a] Department of Radiology and Imaging Sciences, Emory University, Atlanta, Georgia
[b] Department of Rehabilitation Medicine, Emory University, Atlanta, Georgia
[c] Department of Biomedical Informatics, Emory University, Atlanta, Georgia
[d] Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, Georgia
[e] Department of Neurology, Emory University, Atlanta, Georgia
[f] Center for Ethics, Emory University, Atlanta, Georgia

ARTICLE INFO

Keywords:
Artificial intelligence
Machine learning
Ethics
Radiology

ABSTRACT

With artificial intelligence (AI) precipitously perched at the apex of the hype curve, the promise of transforming the disparate fields of healthcare, finance, journalism, and security and law enforcement, among others, is enormous. For healthcare – particularly radiology – AI is anticipated to facilitate improved diagnostics, workflow, and therapeutic planning and monitoring. And, while it is also causing some trepidation among radiologists regarding its uncertain impact on the demand and training of our current and future workforce, most of us welcome the potential to harness AI for transformative improvements in our ability to diagnose disease more accurately and earlier in the populations we serve.

1. Introduction

As in the case of most disruptive technologies, assessment of and consensus on the possible ethical pitfalls lag. New AI applications and start-up companies seem to emerge daily. At the start of 2019, funding in imaging AI companies exceeded $1.2 billion [1]. Yet, questions of algorithm validation, interoperability, translation of bias, security, and patient privacy protections abound.

2. Limitations of AI

2.1. Bias and the black box effect

The confound of selection bias in datasets used to develop AI algorithms is common. Buolamwini and Gebru [2] demonstrated bias in automated facial recognition and the associate datasets, resulting in diminished accuracy in recognizing darker-skinned faces, particularly women. Machine learning (ML) datasets need to be large and often-used clinical trial research databases are largely derived from majority populations. Thus, the resulting algorithms may be more likely to fail when applied to underserved and therefore possibly underrepresented patient groups.

Commercial uses of AI can also result in automation bias, which can diminish the likelihood that the healthcare provider will question an erroneous result due to the tendency to over-rely on automated systems, which after all are typically designed to reduce human error and enhance patient safety. Recent work by Mirsky and colleagues (2019) chillingly showed how malware and AI deep learning tools could be used to mislead expert radiologists on the presence or absence of malignant lung lesions on CT.

2.2. Dealing with rare cases and generalizability

While deep learning has great promise for applications in medicine, some critical issues limit its widespread adoption in clinical settings. Current AI applications in medical imaging are geared towards the most common conditions diagnosed with any given examination. While a state-of-the-art AI application developed with the most rigorous methodology may approach or even exceed human performance in the diagnosis of the most common diagnoses, there is a long tail of uncommon or rare conditions present on a chest radiograph or computed tomography scan that are easily diagnosed by a radiologist but which may not have been present or faithfully labelled in the model's training data. As Andrew Ng recently observed, "Given an image and a medical history, a radiologist can diagnose any of a large number of conditions. Today's ML might be able to reliably diagnose only the top handful." [3] In this circumstance, radiologists must consider mechanisms to avoid patient harm that may occur with over-reliance on tools which are generally accurate on the majority of cases yet may fail altogether in other situations.

[*] Corresponding author at: Emory University Hospital, 1364 Clifton Rd, Suite D-112, Atlanta, GA, 30322, Georgia.
E-mail address: cmeltze@emory.edu (C.C. Meltzer).

https://doi.org/10.1016/j.ejrad.2019.108768
Received 23 July 2019; Accepted 21 November 2019

AI models trained with imaging data acquired from one setting may poorly generalize to other practice settings in other locations with new patients. The use of models based on training data which is not representative of the population, case mix, modalities, and acquisition protocols can compromise performance and confidence in its use, particularly if overfitting has occurred [4]. Re-evaluations of algorithm performance in new settings may require the addition of locally sourced training data. The fullest execution of fiduciary duties as radiologists to our patients requires a practical understanding of these potential pitfalls to ensure the most responsible deployment of AI tools in practice.

3. Data ownership, security, and patient privacy

3.1. Data ownership

ML is data-hungry. Deep learning is data-ravenous. Scientists developing deep learning applications will gladly take hundreds of thousands of cases to develop and test new tools, especially given their familiarity with databases like Imagenet, which now has over 14 million images. The desire to create and market new AI applications in medicine has created a demand and marketplace for patient-derived data. However, ownership and the rights to use these data are complex and vary by jurisdiction, sometimes depending on the degree to which the data has been de-identified or anonymized [5,6]. In Europe, patients have both strong ownership and usage rights of their own data, while healthcare providers in North America may own the “physical” evidence associated with patient data [7,8]. This variability leads to several unsettled questions, especially in an increasingly globalized AI marketplace where a model may be trained on data from one country but marketed in another continent altogether.

Healthcare entities have been able to use and share de-identified records for research and development, either with patient consent or waivers. With newer, more protective regulations like General Data Protection Regulation [9], will patients continue to tolerate such use? What is to prevent companies from seeking less well-regulated environments with poorly-defined or enforced data ownership regulations? Which mechanisms will EU citizens use to enforce their rights when treated in other jurisdictions like the US, where waivers of consent for the use of their data may still be granted? Alternatively, will the GDPR disincentivize US companies from doing business with EU companies if the former must comply with the GDPR’s expectations? While it is unlikely that patients will successfully be able to claim ownership of intellectual property derived from the use of their data, these issues will likely be settled in both courts of law and public opinion.

3.2. Security and patient privacy

Privacy protections typically concern regulating access to a person or what is known about him/her [10]. Such access may involve the individual’s right to bodily privacy, personal information, property or place of habitation, or control of his/her name, image, or likeness [11]. In radiologic practice, privacy typically involves controlling access to protected health information. As such, informational privacy is protected by data security measures that ensure reasonable steps are taken to prevent protected health information from being accessed by individuals who have no right to it or need to know it. Especially in AI applications, clinical and business entities must protect such data from hackers as well as be careful not to place protected data on insecure or vulnerable servers [12]. But the use of large databases – a staple of machine learning models – illustrates how ML and privacy protection can be at cross purposes [13]. As noted above, ML “feasts” on big data in order to be better “educated” by valid and generalizable training materials [14]. Thus as large amounts of training materials for ML applications are gathered from multiple and diverse sources (e.g., medical and insurance records, pharmaceutical data, genetic data, and social media), it becomes easier to trace that data to patient referents and (intentionally or unintentionally) defeat the goals of privacy [13].

On May 25, 2018, the European Union (EU) began enforcing its GDPR, which one scholar called “the world’s most sweeping privacy law” [15]. Strictures of the GDPR require that EU companies, and companies having business relationships with them, must be explicitly truthful to individuals about how their private data will be used. Companies must also limit data collection to only what is strictly necessary, with imposed temporal limits on how long they will keep the protected data. Notably, entities in possession of private information must be able to tell people what data they have on them, how the data will be used, and if the individual so requests, be able to alter or delete that data. Failure to comply with these regulations may result in steep fines [9].

While superficially reassuring, the promise of data de-identification to protect privacy may be entirely unrealistic. Traditional manual approaches to de-identification of free text may achieve an inspection rate of approximately 18,000 words per hour but still miss items, not to mention are very expensive [16]. And while automated approaches to de-identifying free text have been developed, even the most accurate appear unable to remove all the types of protected health information (PHI). For example, successful de-identification of medical imaging data stored in the DICOM (Digital Imaging and Communications in Medicine) standard requires deletion or overwriting metadata found in sometimes obscure, private tags which may contain PHI. Doing so will require knowhow on the ways that the metadata are registered on individual machines, as well as an inspection of images for possible “pixel-burned-in” PHI. Inspectors or data controllers must also avoid images that could be reconstructed to create a facial visualization. Even a pixilated photograph is no guarantee that the subject cannot be reidentified [17,18]. Similar challenges exist for non-imaging data stored in electronic medical records [19].

Consequently, the GDPR may come to require that companies inform patients that they cannot promise true de-identification [20]. As such, that admission may place an even greater emphasis on truthfully disclosing what data a company has on a patient and how it will be used in ML applications. As long as the threat of re-identification remains real, patients will only feel a heightened interest and perhaps anxiety in controlling what is known about them.

4. Next steps

We applaud the recent collaboration of professional scientific groups and the National Institutes of Health in identifying key research priorities that will help to address some of the concerns addressed herein [21]. These include developing image reconstruction and automated image annotation methods that limit the effect of human bias, ways to penetrate the “black box” of machine learning algorithms (i.e., “explainable AI), and safe, validated ways to de-identify patient data for large-scale image dataset sharing. An in-depth multisociety European and North American statement on the ethics of AI in radiology is currently in draft stage and calls on the community to develop and adhere to a uniform code of ethics so as to provide a reliable ethical framework as the technology rapidly advances [8].

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Declaration of Competing Interest

None.

References

[1] Harris S, Parekh S. Funding Analysis of Companies Developing Machine Learning

Solutions for Medical Imaging. Signify Research. [Internet]. Funding Analysis of Companies Developing Machine Learning Solutions for Medical Imaging. Signify Research. [cited 2019 Jun 2]. Available from: https://s3-eu-west-2.amazonaws.com/signifyresearch/app/uploads/2019/01/31125920/Funding-Analysis-of-Companies-Developing-Machine-Learning-Solutions-for-Medical-Imaging_Jan-2019.

[2] J. Buolamwini, T. Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, (2018), pp. 77–91.

[3] A. Ng, AI Is the New Electricity: the Disruptive Power of AI Applications in Medical Imaging, RSNA Spotlight Course, San Francisco, 2019 May 31.

[4] S.H. Park, K. Han, Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction, Radiology 286 (3) (2018) 800–809.

[5] M.A. Hall, K.A. Schulman, Ownership of medical information, JAMA 301 (12) (2009) 1282–1284.

[6] L. Cartwright-Smith, E. Gray, J.H. Thorpe, Health information ownership: legal theories and policy implications, Vand. J. Ent. & Tech. L. 19 (2016) 207.

[7] A. Tang, R. Tam, A. Cadrin-Chênevert, W. Guest, J. Chong, J. Barfett, et al., Canadian Association of Radiologists white paper on artificial intelligence in radiology, Can. Assoc. Radiol. J. 69 (2) (2018) 120–135.

[8] Geis et al. R. Ethics of AI in Radiology: the European and North American Multisociety Statement [Internet]. American College of Radiology. [cited 2019 Jun 23]. Available from: https://www.acrdsi.org/-/media/DSI/Files/PDFs/Ethics-of-AI-in-Radiology-Statement_RFC.

[9] GDPR Recitals and Articles [Internet], European Union, 2019 [cited 2019 Jun 17]. Available from: https://ico.org.uk/global/page-not-found/.

[10] W.J. Winslade, J.W. Ross, Privacy, confidentiality, and autonomy in psychotherapy, Neb L Rev. 64 (1985) 578.

[11] T. Beauchamp, J. Childress, Principles of Biomedical Ethics, Seventh, 2013.

[12] M. Brundage, S. Avin, J. Clark, H. Toner, P. Eckersley, B. Garfinkel, et al., The malicious use of artificial intelligence: Forecasting, prevention, and mitigation, arXiv preprint (2018) arXiv:180207228.

[13] R.A. Ford, W. Price, I. Nicholson, Privacy and accountability in black-box medicine, Mich Telecomm & Tech. L Rev. 23 (2016) 1.

[14] Meyer D. AI Has a Big Privacy Problem And Europe's GDPR Is About to Expose It | Fortune [Internet]. Fortune. [cited 2019 Jun 17]. Available from: https://fortune.com/2018/05/25/ai-machine-learning-privacy-gdpr/.

[15] Roberts J. The GDPR Is in Effect: Should U.S. Companies Be Afraid? [Internet]. Yahoo FInance. [cited 2019 Jun 30]. Available from: https://finance.yahoo.com/news/gdpr-effect-u-companies-afraid-033029332.html.

[16] I. Neamatullah, M.M. Douglass, L.H. Lehman, A. Reisner, M. Villarroel, W.J. Long, et al., Automated de-identification of free-text medical records, BMC Med. Inform. Decis. Mak. 8 (July (1)) (2008) 32.

[17] M. Milchenko, D. Marcus, Obscuring surface anatomy in volumetric imaging data, Neuroinform 11 (January (1)) (2013) 65–75.

[18] Newman L. AI Can Recognize Your Face Even If You're Pixelated | WIRED [Internet]. Wired. [cited 2019 Jun 30]. Available from: https://www.wired.com/2016/09/machine-learning-can-identify-pixelated-faces-researchers-show/.

[19] M.A. Rothstein, Is deidentification sufficient to protect health privacy in research? Am. J. Bioeth. 10 (9) (2010) 3–11.

[20] L. Na, C. Yang, C.-C. Lo, F. Zhao, Y. Fukuoka, A. Aswani, Feasibility of reidentifying individuals in large national physical activity data sets from which protected health information has been removed with use of machine learning, JAMA Network Open 1 (8) (2018) e186040–e186040.

[21] C.P. Langlotz, B. Allen, B.J. Erickson, J. Kalpathy-Cramer, K. Bigelow, T.S. Cook, et al., A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/the Academy Workshop, Radiology 291 (3) (2019) 781–791.