BMJ Health & Care Informatics

Biased intelligence: on the subjectivity of digital objectivity

Jeremy T Moreau ,[1] Sylvain Baillet ,[1] Roy WR Dudley[2]

To cite: Moreau JT, Baillet S, Dudley RWR. Biased intelligence: on the subjectivity of digital objectivity. *BMJ Health Care Inform* 2020;**27**:e100146. doi:10.1136/bmjhci-2020-100146

Received 09 March 2020
Accepted 07 July 2020

[1]Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, Montreal, Québec, Canada
[2]Paediatric Surgery, Division of Neurosurgery, Montreal Children's Hospital, Montreal, Québec, Canada

Correspondence to
Jeremy T Moreau;
jeremy.moreau@mail.mcgill.ca

Whether IBM's Watson, Google's DeepMind or Tencent's WeDoctor, the last few years have been characterised by unprecedented levels of research interest and new investments in artificial intelligence (AI) and digital healthcare technology. The number of publications on applications of AI and machine learning to medical diagnosis has dramatically increased since around 2015 (figure 1). Correspondingly, venture capital-backed digital health and AI startups worth over US$1 billion now number in the dozens (figure 1).[1] Yet, this influx of new investment has not been without controversy. Google's recent partnership with national health group Ascension, which gave the company access to the clinical data of around 50 million patients, has been the target of significant mediatic and congressional scrutiny.[2] Likewise, pharmaceutical giant GlaxoSmithKline's (GSK) US$300 million investment in direct-to-consumer genetic testing provider 23andMe has aroused similar concerns.[3] Under the terms of their 4–5 years agreement, GSK gained access to 23andMe's genetic data and became its exclusive collaborator for drug target discovery programmes.[4] While much of the coverage of these partnerships has focused on issues of privacy and consent, we argue that another key consideration lies in the risks associated with exclusive or privileged access to databases of patient information and the development of proprietary diagnostic algorithms.

Why should we care about openness and transparency in AI development? Take the hypothetical case of a tech company developing a new proprietary AI to make prescription recommendations using electronic health record data from a large academic medical centre. Aware of this ongoing programme, a pharmaceutical company decides to make its drugs available at a discounted price to the hospital, resulting in increased prescription of its drugs relative to competitors. Now, without any overt collusion, the tech company's AI may learn that these drugs are more often prescribed by the hospital's physicians and therefore have increased probability of recommending them in the future. Clearly, these recommendations are inappropriate and not based on any medical evidence, yet without the ability to inspect the proprietary AI or the data it was trained on, the possibilities for peer review and scrutiny would be severely limited. Should AIs have their own disclosures? How would such disclosures be regulated and enforced? Would it be desirable to avert healthcare 'data monopolies' with new antitrust legislation? These are questions regulators will need to answer sooner rather than later. While not AI-driven, the recent revelation that popular electronic health record vendor Practice Fusion received kickbacks in exchange for displaying alerts in its software designed to increase prescriptions of opioid analgesics[5] is a chilling reminder of the ability of software vendors to influence treatment decisions. The unmonitored allowance of proprietary healthcare AIs trained on privately held datasets risks providing an avenue for plausible deniability in addition to further hindering the detectability of such complicit partnerships between drug manufacturers and software vendors.

Beyond theoretical scenarios, take also for example a recent study by a group of Google researchers who designed an AI system to read mammograms that outperformed radiologists on a breast cancer identification task.[6] While unintentional and acknowledged by the authors, 95% of the over 90 000 mammograms used in the study were acquired on devices made by a single manufacturer. Would the AI perform as well on images from another manufacturer's systems? What about the 10-year-old mammography system still operating in an under-resourced community? Further studies and clinical trials will be needed to obtain these answers, but this case highlights just how easy it is for systemic biases to be introduced even when no foul play is

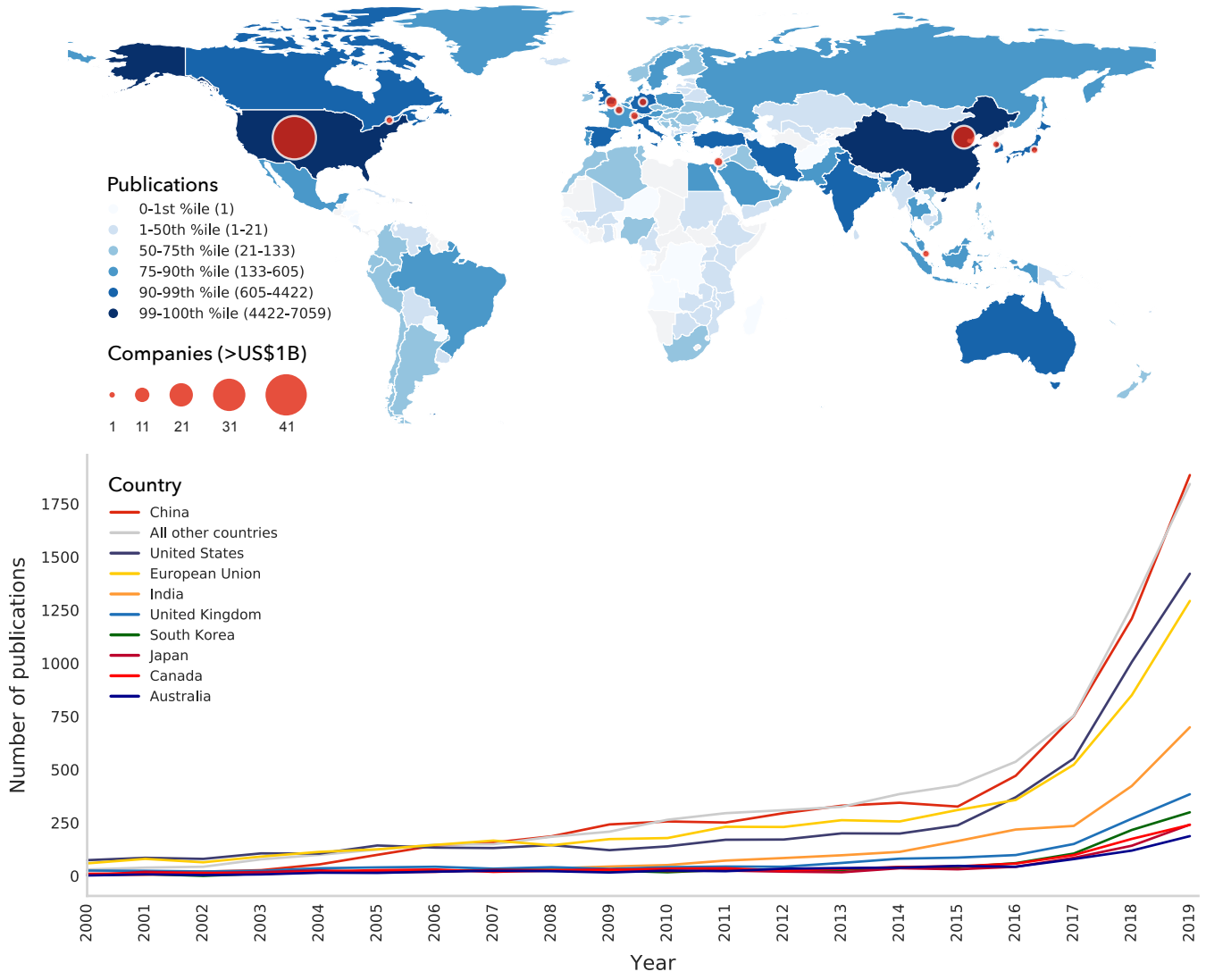

Figure 1 Publications on artificial intelligence (AI)/machine learning applied to medical diagnosis and number of private AI or healthcare startup companies valued at >US$1 billion. Map shows the total number of publications on AI/machine learning applied to medical diagnosis by country from 2000 to 2019. In the legend, numbers in brackets represent number of publications while the colour gradient illustrates percentile categories. The bottom line diagram plots the same data by year and country. Data were extracted from Scopus using the search strategy reported by Liu *et al*.[26] Red dots on the map illustrate the number of venture capital-backed private AI or healthcare startup companies with a valuation of over >US$1 billion.[1]

involved. Nonetheless, AI presents a tremendous opportunity to reduce barriers to care in low-resource settings around the world.[7 8] Unfortunately, current trends in AI research and private funding (figure 1) suggest the existence of strong geographical bias. A select group of countries, including notably China and the USA, are responsible for most of the research and investment in AI-assisted medical diagnostics. Unless representative samples of patients are included, the likelihood of these tools providing equal benefits outside of their countries of origin is limited. Collaboration and exchange of data and experience between healthcare systems on a global scale is needed if we are to benefit from truly generalisable and equitable AI systems. Exploratory research and development of AI systems on small single-centre sample datasets is necessary for identifying promising applications, but we suggest that a similar framework of 'levels of evidence' as proposed by Woo *et al*[9] for biomarkers in translational neuroimaging could be applied more broadly for all AI system development in medicine. This framework suggests that early exploratory AI model development should be followed by progressively more comprehensive assessments of generalisability across larger and more diverse research contexts and population samples. Models for which initial results can be satisfactorily reproduced across larger multicentric studies and in diverse groups of patients then become strong candidates for translation into real-world clinical practice.

AI systems often—even to the ignorance of their creators—replicate the societal biases extant within the data they are trained on. In our own study,[10] we found that the models we had trained on data from over 60 000 patients from a national cancer registry to predict meningioma malignancy and survival predicted worse survival for black and uninsured patients. Another study which developed an algorithm to predict no-show appointments in paediatric orthopaedic clinics likewise identified that insurance type was a significant factor in predicting the rate of no-shows.[11] While these predictions are factually representative of the data, the predicted outcomes are much more reflective of social and economic realities than they are of any biology. Other previously reported examples of bias include a melanoma diagnosis algorithm that did not factor skin colour or the use of genomic databases in which minorities are under-represented.[12] Patient age is another factor from which disparities may arise. It has for example been reported that Babylon Health's *GP at Hand* system, which offers online consultations and an AI-driven symptom checker, has attracted on average younger and healthier patients as compared with in-person general practice clinics.[13] Barriers to care driven by difficulties with adapting to new technologies are one issue, but the relative lack of training data in certain age groups could also lead to AI systems becoming more proficient at identifying the health issues more frequently experienced by the groups of patients for which they hold more data. These cases underscore the importance for healthcare practitioners to critically assess the predictions of putatively 'objective' machine learning systems. They are also a reminder that while technological solutions will undoubtedly form part of our efforts for better care delivery, other systemic issues remain just as, if not more, critical to address.

While there is a strong argument to be made in favour of federating de-identified health data in national or even international databases to allow for the development of healthcare AI systems,[14 15] we argue that these data should be considered a public good. As discussed above, there is a real risk that allowing exclusive or privileged access to databases of patient information may allow for intentional or unintentional bias to be introduced in health AI systems. Treating patient data as a commodity could also create perverse incentives for companies to invest more in acquiring datasets or developing products in countries or among groups of individuals with greater purchasing power. This could be of particular concern for direct-to-consumer health products. The Apple Heart Study[16] investigated the ability of an optical pulse sensor and smartwatch application to identify atrial fibrillation. The study recruited an impressive sample of over 400 000 individuals; however, all participants were from the USA and owners of Apple smartphone and smartwatch devices. Moreover, the paper and data sharing statement for this study notably state that the data are 'not available to be shared' and that 'Apple sponsored the study and owns the data'.[16] While the challenges involved in balancing commercial and research interests cannot uniquely be attributed to this study, the possibility that socioeconomic (and consequentially demographic) groups were not equally represented does raise concerns given the cost of these devices. There is a need for greater advocacy for the inclusion of data from historically underserved communities in datasets that will be used to train the next generation of health AI systems. While the scale of potential consequences is hard to estimate given the paucity of systems currently in real-world use, we must take the initiative to ensure that underserved communities are adequately represented in health AI developments. Requiring the systematic evaluation and reporting of health AI systems' performance in diverse population subsamples as a condition in the approval process for commercialisation is one step regulators could take in this direction.

In the context of primary care, the commoditisation of personal medical data runs counter to public expectations of the confidential nature of the physician-patient relationship.[17] In this regard, we believe that there is an urgent need for greater transparency and public discourse on how and for what purpose health data are exchanged. Even if data are de-identified, patients should ultimately have the ability to know and decide who has access to their data and for what purpose these data are being used. The question of 'ownership' of medical data remains ill-defined from a legal perspective in many jurisdictions[18 19] in spite of the frequent mismatch between patient expectations and actual data usage. In the UK, the now scrapped NHS *care.data* programme raised significant concerns with respect to the provision of health data to the insurance industry, for instance.[17] Beyond the technological challenges lies the issue of maintaining public confidence.[14] While a cancer patient may support sharing data for research into developing a diagnostic AI system that could allow for earlier disease detection, this same patient may not agree with any data being used to train an AI system designed to calculate life insurance premiums. In a recent survey of patient attitudes towards sharing data from electronic health records for research purposes, only 4% of patients recruited from two US academic medical centres declined to share data with researchers from the home institution, while 28% declined to share with other non-profit institutions and 47% were not willing to share data with for-profit institutions.[20] Allowing for the responsible use of aggregated health data to develop AI-driven diagnostic tools has considerable potential to benefit patients, but we must ensure that mechanisms allowing for ethical oversight and independent validation remain available. Beyond preventing exclusive private ownership of patient data, this also means requiring a minimum level of transparency in disclosing what data were used to train health AI systems and actively informing patients about the use of their data. Open data and transparent reporting of data sources used in AI development will allow for the necessary accountability to ensure that algorithm developers build generalisable health AI systems

that minimise bias and respect public expectations of medical data usage.

In spite of the challenges, there is growing recognition of the necessity for intentional design of equitable AI systems.[21 22] Human-centred AI, a perspective that argues that AI systems must be designed for social responsibility with an understanding of sociocultural context,[23 24] has been gaining traction among AI researchers. There have, moreover, been encouraging steps towards policy discussion and legislation to protect personal information while requiring transparency, fairness and accountability for processors of personal data.[25] These are promising developments, but we cannot stop here. In the end, sensitivity, specificity and other metrics tell only part of the story. While we can and should attempt to build performant AI systems that emulate ethical decision making, we must remember that human-designed AI remains biased by the same social, cultural and political biases that shaped the data these systems were trained on. The physician's role as an advocate for patients' interests is as important today as it has ever been. We will increasingly come to rely on AI-assisted diagnosis and prognosis in the years to come, but treatment recommendations must remain conscious of societal context and continue to represent a shared decision-making process between physician and patient.

Contributors JTM: drafting the manuscript, literature review, data analysis, figures. SB: supervision, funding and administration, critically revising the manuscript. RWRD: supervision, funding and administration, critically revising the manuscript.

Funding No funding was received specifically for this work. JTM has received training awards from the Canada First Research Excellence Fund awarded to McGill University for the Healthy Brains, Healthy Lives initiative, the Fonds de Recherche du Québec-Santé and the Foundation of Stars. SB was supported by a Discovery Grant from the Natural Science and Engineering Research Council of Canada (436355-13), the NIH (1R01EB026299-01) and a Tier-1 Canada Research Chair in Neural Dynamics of Brain Systems.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Jeremy T Moreau http://orcid.org/0000-0002-6545-7525
Sylvain Baillet http://orcid.org/0002-6762-5713

REFERENCES

1 CB Insights. The complete list of unicorn companies, 2020. Available: https://www.cbinsights.com/research-unicorn-companies [Accessed 22 Jan 2020].
2 Blumenthal D. Why Google's Move into Patient Information Is a Big Deal, 2019. Available: https://hbr.org/2019/11/why-googles-move-into-patient-information-is-a-big-deal [Accessed 26 Jan 2020].
3 Hendricks-Sturrup RM, Prince AER, Lu CY. Direct-To-Consumer genetic testing and potential Loopholes in protecting consumer privacy and Nondiscrimination. *JAMA* 2019;321:1869–70.
4 GlaxoSmithKline. GSK and 23andMe sign agreement to leverage genetic insights for the development of novel medicines, 2020. Available: https://www.gsk.com/en-gb/media/press-releases/gsk-and-23andme-sign-agreement-to-leverage-genetic-insights-for-the-development-of-novel-medicines/ [Accessed 26 Jan 2020].
5 Spector M, Hals T. Exclusive: OxyContin maker Purdue is 'Pharma Co X' in U.S. opioid kickback probe - sources. *Reuters*.
6 McKinney SM, Sieniek M, Godbole V, *et al*. International evaluation of an AI system for breast cancer screening. *Nature* 2020;577:89–94.
7 Wahl B, Cossy-Gantner A, Germann S, *et al*. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *BMJ Glob Health* 2018;3:e000798.
8 Reddy CL, Mitra S, Meara JG, *et al*. Artificial intelligence and its role in surgical care in low-income and middle-income countries. *Lancet Digit Health* 2019;1:e384–6.
9 Woo C-W, Chang LJ, Lindquist MA, *et al*. Building better biomarkers: brain models in translational neuroimaging. *Nat Neurosci* 2017;20:365–77.
10 Moreau JT, Hankinson TC, Baillet S, *et al*. Individual-Patient prediction of meningioma malignancy and survival using the surveillance, epidemiology, and end results database. *NPJ Digit Med* 2020;3:12.
11 Robaina JA, Bastrom TP, Richardson AC, *et al*. Predicting no-shows in paediatric orthopaedic clinics. *BMJ Health Care Inform* 2020;27:e100047.
12 Topol EJ. High-Performance medicine: the convergence of human and artificial intelligence. *Nat Med* 2019;25:44–56.
13 Sujan M, Furniss D, Grundy K, *et al*. Human factors challenges for the safe use of artificial intelligence in patient care. *BMJ Health Care Inform* 2019;26:e100081.
14 Godlee F. What can we salvage from care.data? *BMJ* 2016;354:i3907.
15 Cresswell K, McKinstry B, Wolters M, *et al*. Five key strategic priorities of integrating patient generated health data into United Kingdom electronic health records. *J Innov Health Inform* 2019;25:254–9.
16 Perez MV, Mahaffey KW, Hedlin H, *et al*. Large-Scale assessment of a Smartwatch to identify atrial fibrillation. *N Engl J Med* 2019;381:1909–17.
17 Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data Ran into trouble. *J Med Ethics* 2015;41:404–9.
18 Cios KJ, Moore GW. Uniqueness of medical data mining. *Artif Intell Med* 2002;26:1–24.
19 Moberly T. Should we be worried about the NHS selling patient data? *BMJ* 2020;368:m113.
20 Kim J, Kim H, Bell E, *et al*. Patient perspectives about decisions to share medical data and biospecimens for research. *JAMA Netw Open* 2019;2:e199550.
21 Zou J, Schiebinger L. AI can be sexist and racist - it's time to make it fair. *Nature* 2018;559:324–6.
22 Matheny ME, Whicher D, Thadaney Israni S. Artificial intelligence in health care: a report from the National Academy of medicine. *JAMA* 2019. doi:10.1001/jama.2019.21579. [Epub ahead of print: 17 Dec 2019].
23 Riedl MO. Human-centered artificial intelligence and machine learning. *Hum Behav Emerg Technol* 2019;1:33–6.
24 Our values. Stanford HAI. Available: https://hai.stanford.edu/about/values [Accessed 21 Jun 2020].
25 Forcier MB, Gallois H, Mullan S, *et al*. Integrating artificial intelligence into health care through data access: can the GDPR act as a beacon for policymakers? *J Law Biosci* 2019;6:317–35.
26 Liu X, Faes L, Kale AU, *et al*. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. *Lancet Digit Health* 2019;1:e271–97.