Machine learning algorithms in healthcare

Submission from: Lev Tankelevitch, Alice Ahn, Rachel Paterson, Matthew Reid, Jason Yun, Parth Patel, Alexander Hitchcock, Dror Nir

Executive summary

- Machine learning algorithms are being developed for many aspects of the healthcare sector (including prevention, diagnosis, prognosis, treatment, demand management and resource allocation), often in the form of decision support tools to aid and refine the work of human professionals.

- It is necessary to develop standards for the design, documentation, and evaluation of algorithms and their interpretability. These standards should be tailored to the specific area of application, constrained by pragmatic, ethical, and legal concerns, and developed via multidisciplinary and cross-sector collaborations.

- Regulatory appraisal policies for machine learning technology need to be updated and guided by standards for algorithms and their interpretability, and should involve experts in machine learning, healthcare, and policymaking, with the aim of striking a balance between safety, efficacy, and innovation.

- Privacy and consent regimes should include specification about data collection and storage, as well as the use and end-goals of data, and should be coupled with comprehensive auditing systems to track data processing.

- To develop sound policies around legal liability with machine learning technology, there is a need to clarify the standard of care for healthcare professionals, with consideration for standards for algorithms and their interpretability as well as the level of expertise required of healthcare professionals to work with the technology.

- The issue of discrimination and bias based on demographics is as important in healthcare as it is in areas like criminal justice, but may be more challenging due to the complicated role of demographic information in health outcomes. Detecting and preventing bias is tied to work on standards for algorithms and their interpretability, and requires nuanced consideration of demographics.

- Implementation of algorithms within the healthcare system requires the rapid and effective digitisation of the NHS, as well as workforce development to establish the expertise necessary for working with machine learning technology. The development of the Health Informatics professional field in the UK, the NHS Digital Academy, and the Alan Turing Institute are ideal opportunities to ensure that a dedicated area for machine learning expertise exists within the UK healthcare sector, and should be guided by foresight analysis.

- Dialogue with the public is necessary to obtain confidence and trust, and to develop sound policies around all aspects of machine learning technology.

1. About the submitting organisation

1.1. Polygeia is an independent, non-party, and non-profit think-tank focusing on health and its intersection with technology, politics, and economics. Our aim is to produce high-quality research on global health issues and policies. With branches in Oxford, Cambridge, London, and New York, our work has led to policy reports, peer-reviewed publications, and
2. The use of algorithms in healthcare

2.1. The rapid development of algorithms for decision-making has been widely touted as revolutionary\(^1,2\). Most, if not all, algorithms used today belong to the field of *machine learning*, a form of artificial intelligence. Broadly defined, machine learning algorithms learn statistical relationships within data, and can be applied to novel data to make “decisions” such as categorisation or prediction\(^3\). For example, in medical diagnosis, an algorithm can learn the relationship between the features of ultrasound scans and whether they belong to patients with cancer, and then, given new ultrasound scans, can be used to predict the likelihood of cancer in new patients\(^1\).

2.2. In clinical practice, tools like the one described above are being developed to refine diagnostic procedures by rapidly analysing medical images already collected as standard clinical practice\(^4-6\). Machine learning is also at the forefront of precision medicine, helping to identify patient subgroups and target treatments appropriately\(^7-9\). In epidemiology, it is being applied to public health data to detect and track infectious disease outbreaks. It is also being used to enhance medical monitoring, and to optimise demand management and resource allocation in healthcare systems\(^10-13\). Algorithms in these applications are not autonomous agents, but rather act as decision support tools to aid and refine the work of human professionals.

3. Standards for algorithms and their interpretability

3.1. There is a need to define standards for the design, documentation, and evaluation of algorithms both in the private and public sector. These standards should outline the quantity and quality of data needed for designing and validating algorithms, the types of metrics to be used for evaluating performance, the specific role of human oversight during real-world use of algorithms, and consideration for any performance instabilities (e.g., due to algorithms’ potential capacity to refine themselves during regular use, or due to changes in the data provided to the algorithm across time).

3.2. *Algorithm interpretability* — the transparency of an algorithm’s “reasoning” about its decisions — is another important area needing standardisation. Those who design algorithms understand their mathematical basis. However, once an algorithm is presented with data, even experts often do not understand how it takes the variety of information into account in each specific application\(^14,15\). The machine learning field is aware of this issue, with a diverse set of approaches being taken to enable interpretability\(^14,15\). Some of these are algorithm-specific, though there is also development of generalised tools, which can be applied to any algorithm, and may pave the way for interpretability standards\(^14,15\). However, researchers have also raised concerns that progress in the field is hampered by the lack of consensus around how to define, develop, and evaluate “interpretability”\(^14,15\). There is a need to pinpoint how much and what kind of information is sufficient for users to understand an algorithm’s decision.

3.3. These standards are broadly relevant, but should be tailored to each sector (e.g., healthcare), to divisions within each sector (e.g., primary care), and to specific policy areas (e.g., legal liability). They should be informed by, and likewise inform, pragmatic, ethical, and legal concerns in policy areas including regulatory appraisal, privacy and consent, legal liability, discrimination, workforce development, and public engagement. Prioritising the applied, “real-world” context of algorithms is essential not only for effective policymaking but also for guiding academic and industry development of the technology.
3.4. To ensure rapid progress for the healthcare sector, the development of standards requires tight multidisciplinary and cross-sector collaborations between academia, industry, the healthcare system, patients and policymakers. This would ensure that standards are both useful for policymaking and industry development, and safe and effective for the healthcare sector. To this end, the Future of Healthcare project at the University of Oxford is a leading example. A collaboration between the Oxford Internet Institute, the Department of Engineering Science, and the Oxford Martin School, the project seeks to systematically characterise tasks in primary healthcare in the context of current and potential automation to inform machine learning research and policymaking. The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems is another example of forward-thinking collaboration, with a focus on building an ethical framework for the technology. Such multidisciplinary and cross-sector projects are essential for the design, evaluation, and interpretability of algorithms for real-world use, with the final aim being the development of a set of standards for industry and within each relevant policy area.

4. **Regulatory appraisal of algorithms**

4.1. Regulatory appraisal of algorithms is a pertinent issue given the tighter regulatory requirements for medical software introduced in the EU, and should be considered by the Medicines and Healthcare Products Regulatory Agency (MHRA), the National Institute for Health and Care Excellence (NICE), and the British Standards Institute (BSI). Considering the unique aspects of this technology, these agencies should closely collaborate with any bodies focusing on machine learning, such as the proposed Commission on Artificial Intelligence, and the European Agency for Robotics and AI.

4.2. The development of regulatory appraisal policies should be guided by concrete standards for the design, documentation, and evaluation of algorithms and their interpretability (see sections 3). Both the development and execution of regulatory appraisal should involve those with sufficient combined expertise in machine learning, healthcare, and policymaking, with the aim of striking a balance between safety and efficacy on the one hand, and innovation and progress on the other.

5. **Privacy, and consent**

5.1. As machine learning becomes widespread in the healthcare sector, more health data will be collected from an increasing variety of sources. A person’s consent regarding their data should be a specific and informed indication that they agree both to the storage and, if relevant, to the processing of their data. Patients must be able to understand what their data is used for, and are also more likely to provide consent if the benefits of their data are clearly explained. For policymaking, it is also necessary to clarify the extent of algorithm interpretability required for explanations of data usage, especially if the consenter will be directly impacted by any analysis.

5.2. Rather than providing blank consent for all data storage and usage, a better process may be to consent to specific uses of data, at the relevant times. However, it is not always possible to predict all possible use cases for personal data. Instead, a broad consent can be given which allows patients to delegate future decision-making on how their data is used to a body such as a Data Access Committee (DAC) or Ethics Committee.

5.3. From the research and development end, there is a need for systems that can precisely and rapidly track data usage, particularly in the increasingly common case of data transfer between healthcare facilities and industry partners. Such systems are currently being developed by companies (e.g., Google DeepMind), but they should be monitored by independent bodies, like the proposed Council of Data Ethics, to ensure that standards are met. We hope that these questions will be addressed by cross-sector initiatives such as
6. Legal liability

6.1. The likely status of many algorithms as decision support tools for healthcare professionals raises the question of legal liability, an issue implied in the EU’s General Data Protection Regulation, and recently raised by MEPs. Who is liable for a misdiagnosis involving an algorithm’s suggestion — the hospital, the doctor, or the manufacturer? In healthcare, a likely source of liability will be the duty of care imposed by the law of negligence. Healthcare professionals have the duty to use reasonable care and skill in diagnosing and treating patients, and could be held liable for negligent use of machine learning technology, depending on the prevailing standard of care. It is essential to determine what this standard should be.

6.2. The standard of care is tied to the amount and type of human oversight required to operate machine learning technology, and the extent to which professionals are expected to understand algorithmic decision-making (i.e., algorithm interpretability — see section 3). For professionals to fulfil the duty of care, they must verify an algorithm’s output against their own expertise, a task difficult to accomplish without insight into the algorithm itself. Additionally, healthcare facilities must have the resources in place to enable professionals to carry out their duty of care, which is also dependent on the expected performance standards and regulatory appraisal of algorithms (see sections 3-4).

6.3. Clarifying the standard of care should therefore involve a two-pronged approach: determining the level of performance and interpretability required of machine learning tools, and the level of expertise required of healthcare professionals to work with such tools — issues which are tied to the development of standards for algorithms, and which should be resolved via multidisciplinary collaborations between academia, industry, the healthcare system, and policymakers (see section 3).

7. Discrimination and bias

7.1. The issue of bias in algorithms based on demographic factors has been raised in sectors such as criminal justice, and consumer finance. In healthcare, it remains a key issue, and presents its own challenges.

7.2. The performance of algorithms greatly depends on the data used to develop them. Biases that are present in the data will be reflected by algorithms. Such biases, intentional or not, may be inherent to the data, may appear during data collection, or may be introduced during human processing. Thus, the detection of biases is also strongly tied to the development of standards for the design and evaluation of algorithms, including their interpretability (see section 3), and there is ongoing work on developing such standards within the context of discrimination.

7.3. Eliminating biases, however, is more challenging because multiple aspects of the data are often inextricably linked: removing one variable (e.g., age) may remove other useful information (e.g., income), ultimately weakening the algorithm. Thus, there is a balance between ensuring fairness and maintaining performance. In healthcare, this is further complicated by the diverse impact of demographics: information like age, gender, and race may play one role in diagnosis and treatment, but another in demand and resource management. Indeed, research suggests that careful use of demographic information in algorithms may ensure more fairness.

7.4. Consideration of the role of demographics is required for each specific application of
machine learning within healthcare, and should be factored into the development of standards for algorithms. Policymaking should include multidisciplinary and cross-sector input from industry, governmental, academic, and independent bodies.\textsuperscript{16,30,41–47}

8. Infrastructure and workforce development

8.1. We reinforce the many calls for rapid and effective digitisation of the NHS to ensure that the opportunities that machine learning affords can be realised.\textsuperscript{2,24,48–50} Variability in digitisation across communities may exacerbate differences in healthcare access, as trusts lagging in digitisation will be unable to benefit from technology that can optimise resources, and improve access.\textsuperscript{49} We also emphasise calls for interoperability of digital systems across regions, as the flow of information can allow trusts without complete digitisation to still benefit from aspects of the technology.\textsuperscript{49,51} Any long-term national strategy for a digital NHS should dedicate a focus area to the use of algorithms.\textsuperscript{49}

8.2. We also echo calls for workforce development.\textsuperscript{48–50,52,53} Given the demand in healthcare for specialised skills and social intelligence, and the current limits of technology without human oversight, healthcare is expected to be largely resistant to job automation in the near future. Instead, the key employment issue in healthcare is meeting workforce demands in terms of numbers and the digital skills necessary for working safely and effectively with machine learning technology.\textsuperscript{50,53,54}

8.3. Chief clinical information officers should have sufficient expertise in machine learning, and ensure tight coordination with the proposed Commission on Artificial Intelligence, and Council for Data Ethics.\textsuperscript{2} The Federation for Informatics Professionals, and the NHS Digital Academy (both currently in development) should include a well-defined focus area for machine learning to help ensure that healthcare professionals are well-equipped to manage the technology (an issue relevant to policies around liability – see section 6).\textsuperscript{49,50,52,55}

8.4. The move to upskill the healthcare workforce should be guided by extensive foresight analysis, including consideration of standards for algorithms and their interpretability.\textsuperscript{41,56} A comprehensive examination of the healthcare sector is needed to assess which, when, and how jobs will be impacted (e.g., see Oxford University’s Future of Healthcare project, focusing on primary care).\textsuperscript{17}

9. Public engagement

9.1. Engaging the public in dialogue about machine learning, particularly in the high-consequence healthcare sector, is key to managing public perception, ensuring patient acceptance, and developing effective policies around the technology.\textsuperscript{2,21} Surveys suggest that the public may fear the risks associated with the use of AI in sensitive areas like healthcare (e.g., 53% of people would not trust robot-assisted surgery), or that algorithms will replace the desired direct contact with clinicians.\textsuperscript{57} More research is needed to understand whether the public’s perception stems from a lack of understanding of how algorithms will be employed.

9.2. Concerns may also be justified — there are many open questions about liability, regulations, and transparency. Though we are confident that they will be resolved, the process of policymaking and innovation should engage the public in dialogue along the way.\textsuperscript{21} Ongoing public debates (e.g., those by the Royal Society or the British Science Association) are important for communicating the nuances of the technology.\textsuperscript{58,59} We also reiterate previous calls for the proposed Commission on Artificial Intelligence to include a significant public discussion component.\textsuperscript{2,21}

9.3. Another barrier to patient trust is the confidence that the public places in companies holding
Written evidence submitted by Polygeia (ALG0043)

their data. A survey found that although a majority of the public are supportive of data sharing in healthcare, 20% are concerned that external organisations cannot be trusted to properly store the data40. The media reports around DeepMind’s partnership with the NHS suggest that proactive communication is necessary to ensure that the public has a chance to understand and voice their concerns about any public-private agreements47. The Wellcome Trust’s new Understanding Patient Data website, which explains health data usage, is also useful to this end26.

April 2017

References


29. Dodson, H. Revealed: Google AI has access to huge haul of NHS patient data. New Scientist (2016).
Written evidence submitted by Polygeia (ALG0043)

52. Ecorys UK. *Digital skills for the UK economy*. (UK Department for Business, Innovation & Skills; Department for Culture, Media & Sport, 2016).
53. Frey, C. B. *et al.* Technology at Work v2. 0: The future is not what it used to be. *CityGroup Univ. Oxf.* (2016).
59. Odell, S. Machine Learning and the Public.
60. Ipsos MORI. *Commercial access to health data*. (Wellcome Trust, 2016).