Perspectives on Digital literacy and HexAI

DR Ann Borda
Health and Biomedical Informatics Centre
Melbourne Medical School
The University of Melbourne VIC 3010
AUSTRALIA
www.findanexpert.unimelb.edu.au/display/person197899

Background

This position paper arises from a workshop event at the Alan Turing Institute (ATI) in January 2019 entitled *Automating the Crowd* (https://www.turing.ac.uk/events/automating-crowd) in which I was a co-organiser. The event focused on the relationship between automation and human work as a longstanding area of concern and reflection, and further raised questions around public policy and across disciplinary boundaries. In particular, the use of 'crowdsourced' efforts and the increasing diversity of contributors in human-machine interactions presented certain challenges. The workshop was followed by a public panel *The Real People Behind Artificial Intelligence (AI)* (https://www.turing.ac.uk/events/real-people-behind-artificial-intelligence) opening up and addressing questions about whose work underpins automation, whether this is work at all, and if so what kind?

A particular issue of debate was the growing move in different sector areas, such as finance, education, and biomedicine toward the need for 'explainable' AI systems whose decision-making procedures are accessible to humans. AI systems are most often comprised of "black box" algorithms, meaning that they are not open to scrutiny from users. So, whilst users see the output of an algorithm (e.g. search results), they are not granted access to the input (e.g., the data that has been collected about a user) nor the decision-making process that filtered the output.

The U.S. Defense Advanced Research Projects Agency (DARPA) Explainable AI (XAI) Program is acknowledged as an influential initiative in this move to develop AI systems that can engage (end) users in a process in which the mechanisms and "decisions" of the AI are explained (Gunning 2017). The concept of explainability itself, however, sits at the intersection of several areas of research and concern, namely; transparency, causality, bias, fairness and safety, and these have further implications on approaches to the debate (Hagras 2018).

Digital Literacy

Among the emergent themes in the ATI workshop context was the topic of digital literacy and fair representation in the use and development of explainable AI enabled platforms. Currently, algorithms can reinforce existing biases, for example, through a so-called "filter bubble" (Carrington 2018). The context of digital literacy, and associated algorithmic literacy, is both a broad and a deeply layered topic with reaches at personal and government levels, as well as national policy.

Focusing on digital literacy, for example, The *Organisation for Economic Co-operation and Development (*OECD) assesses a range of skills, such as literacy and numeracy, in its Survey of Adult Skills (PIAAC) (http://www.oecd.org/skills/piaac/). These skills are of integral importance to OECD countries for work and education advancement (Van Deursen et al 2014; OECD 2016). With the progress in reproducing human skills, the development of AI capabilities will have particularly far-reaching implications for this advancement, and consequently, the structure of the economy and workforce skills will need to be radically transformed in the near future.

In this regard, a recent OECD study uses PIAAC to assess what skills might be needed in an AI supported workforce (Elliott 2017). The study is only exploratory, but the findings suggest that

current AI techniques are close to allowing computers to perform at Level 3 on PIAAC in literacy and numeracy—this is at or above the proficiency of 89% of adults in OECD countries. The reported OECD average of 11% of adults are above the level that AI is close to reproducing. However, in the highest performing countries, adults with tertiary education reached 37% in literacy and 36% in numeracy, for Japan and Sweden, respectively, and these results exceed the average (OECD 2016; Elliott 2017).

The future of work in AI is the current focus of a T20 Taskforce that responds to shared economic and market issues across G20 countries (Lyons et al 2019). In the T20 Report, strides towards developing a framework of measures for the *multidimensionality* of digital literacy in the 'Age of AI' are outlined. Chetty et al. (2018) highlight the importance of creating a fluid definition for digital literacy that is responsive to the changing needs of employers in the AI context. In this sense, multidimensionality goes beyond the core domains of digital skills, such as information usage and communication literacy, content creation, digital citizenship, and industry-related competencies, for example (e.g., UNESCO 2018; World Economic Forum 2018).

There is a recognition, however, that the technical aspects of digital literacy are most often emphasized in existing frameworks. Miller, Howe and Sonenberg (2017) argue that explainable AI solutions need to meet the needs of the users, an area that has been well studied in philosophy, psychology, and cognitive science. A critical need, therefore, for the G20 is to expand efforts to create a more dynamic and evolving definition for digital literacy to include softer human skills and higher-order cognitive digital skills related to current and emerging AI technologies (Lyons et al, 2019).

National policy and Explainable AI

Individual governments such as the US, Canada, UK, New Zealand, the European Union, and Australia, among others, are responding to the implications of AI across policy, labour sectors, and for society as a whole. In March 2017, the Government of Canada announced the launch of the *Pan-Canadian AI Strategy* which is reported as the first fully-funded strategy of its kind, followed by announcements of a variety of forms of AI strategies by 18 countries, such as France, Mexico, the UAE, and China. A study by the Canadian charitable research organization CIFAR provides perspectives and summaries on several national AI strategies (Dutton 2018). Funding for research or pilot programs to create explainable and transparent AI are among the identified policy areas.

Canada, New Zealand, China, France, Singapore and the European Union have programs underway addressing explainable AI, largely as part of ethical standards and pilot developments. New Zealand and Finland have also directly aligned AI policy with literacy considerations, although other countries more generally address skills and the future of work. The UK has a strong foundation in this alignment. In the House of Lords Select Committee on Artificial Intelligence Report of Session 2017–19, it is stated that there is a need to improve digital understanding and data literacy across society, "as these are the foundations upon which knowledge about AI is built. This effort must be undertaken collaboratively by public sector organisations, civil society organisations (such as the Royal Society) and the private sector." (UK House of Lords Committee on Artificial Intelligence 2018).

In the connections across literacy, work and explainable AI, national policy strategies have highlighted domain specific challenges. New Zealand's AI strategy recommends that medical treatment paths, for instance, will need to be more transparent than one that manages online shopping recommendations (AI Forum New Zealand 2018). Hence, some AI decision-making techniques are more amenable to explanation than others in this regard. Singapore launched *AI Singapore* (www.aisingapore.org/) - a five-year, S$150 million national program to enhance Singapore's capabilities in AI. "Explainable AI as a Service for Community Healthcare" is one of the

funded strands in which advanced AI prototype devices are built for deployment and testing in a community setting – to allow AI results to be used in precision medicine, preventive advice and automatic lifestyle coaching such as food logging, for example.

In research towards building explainable-AI systems for application in health and medicine requires maintaining a high level of learning performance for a range of machine learning and human-computer interaction techniques (Flaxman and Vos 2018; Langlotz et al 2018). There is further an inherent tension between machine learning performance (predictive accuracy) and explainability (Holzinger et al 2017) which will need to be resolved – particularly as a trust concern. For example, by explaining the reasoning behind a patient's likelihood of readmission to hospital, physicians could have a stronger basis for accepting or rejecting predictions and recommendations.

Future research on a domain basis, such as what might be related to the healthcare sector, could be critical in expanding the contextual and human-centred relationships needed to more fully understand the literacy capabilities required to support new workforce skills – as well as applied uses and development of Explainable AI. *Interpretable* machine learning, for instance, is one advancement being used to support explanations of machine learning models to humans with domain knowledge. There are further new frameworks evolving to provide meaningful explanations for predictions, such as TED (Teaching Explanations for Decisions) which augments training data to include explanations elicited from domain users (Hind et al 2019). Relevant policy areas can also be better informed through such case study approaches and the multidimensional (human-centred) considerations that are arising in the process.

REFERENCES:

AI Forum New Zealand (2018). Artificial Intelligence: Shaping a Future New Zealand. May 2018. Retrieved from: https://aiforum.org.nz/wp-content/uploads/2018/07/AI-Report-2018_web-version.pdf

Carrington, V. (2018). The Changing Landscape of Literacies: Big Data and Algorithms. Digital Culture & Education 10: 67–76.

Chetty, K., et al. (2018). Bridging the digital divide: Measuring digital literacy. Economics: The Open-Access, Open-Assessment E-Journal. Kiel Institute for the World Economy (IfW) 12(2018-23): 1-20. Retrieved from: http://dx.doi.org/10.5018/economics-ejournal.ja.2018-23

Dutton, T. (2018). Building an AI World: CIFAR Report on National and Regional AI Strategies. CIFAR. Retrieved from: https://www.cifar.ca/docs/default-source/ai-society/buildinganaiworld_eng.pdf

Elliott, S.W. (2017). *Computers and the Future of Skill Demand*. Paris: OECD Publishing. Retrieved from: http://dx.doi.org/10.1787/9789264284395-en.

Flaxman, A.D. and Vos, T. (2018). Machine learning in population health: Opportunities and threats. PLOS Medicine 15(11): e1002702. Retrieved from: https://doi.org/10.1371/journal.pmed.1002702

Gunning, D. (2017). Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA). Retrieved from: https://www.darpa.mil/program/explainable-artificial-intelligence

Hagras, H. (2018). Toward Human-Understandable, Explainable AI. Computer 51 (9): 28-36, 2018. doi:10.1109/MC.2018.3620965

Hind, M., Wei, D., Campbell, M., Codella, N. C. F., Dhurandhar, A., Mojsilovic, A., Ramamurthy, K. N., and Varshney, K. R. (2019). TED: Teaching AI to explain its decisions. In AAAI/ACM conference on Artificial Intelligence, Ethics and Society 2019. Retrieved from: https://arxiv.org/abs/1811.04896 [15 Jun 2019 (this version, v2]

Holzinger, A. et al. (2017). What do we need to build explainable AI systems for the medical domain? Retrieved from: arXiv:1712.09923v1 [cs.AI] 28 Dec 2017.

Langlotz, C.P. et al. (2018). A Roadmap for Foundational Research on Artificial Intelligence in Medical Imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop. Radiology 291:781–791. https://doi.org/10.1148/radiol.2019190613

Lyons, A. et al. (2019). Leaving No One Behind: Measuring the Multidimensionality of Digital Literacy in the Age of AI and other Transformative Technologies. T20 TaskForce, Japan 2019. 31 March 2019. Retrieved from: https://t20japan.org/policy-brief-multidimensionality-digital-literacy/

Miller, T., Howe, P. and Sonenberg, L. (2017). Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences. In Proc. IJCAI Workshop Explainable Artif. Intell. Melbourne, Australia. Retrieved from: https://arxiv.org/abs/1712.00547 [5 Dec 2017 (this version, v2]

OECD (2016). *Skills Matter: Further Results from the Survey of Adult Skills*. OECD Skill Studies, OECD Publishing, Paris. Retrieved from: http://dx.doi.org/10.1787/9789264258051-en.

OECD. (2018). Social and emotional skills: Well-being, connectedness and success. Paris, France: Directorate for Education Skills, OECD. Retrieved from: http://www.oecd.org/education/school/UPDATED%20Social%20and%20Emotional%20Skills%20-%20Well-being,%20connectedness%20and%20success.pdf%20(website).pdf

UNESCO. (2018). A global framework of reference on digital literacy skills for indicators 4.4.2. Information Paper No. 51. Montreal, Canada: UNESCO Institute for Statistics. Retrieved from: http://uis.unesco.org/sites/default/files/documents/ip51-global-framework-reference-digital-literacy-skills-2018-en.pdf

United Kingdom. House of Lords Select Committee on Artificial Intelligence. (2018) Report of Session 2017-19 - AI in the UK: ready, willing and able? . HL Paper 100. Retrieved from: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10002.htm

van Deursen, A. J. A. M., Helsper, E. J., & Eynon, R. (2014). Measuring digital skills: From digital skills to tangible outcomes project report. London, England: The London School of Economics and Political Science. Retrieved from: http://www.lse.ac.uk/media-and-communications/assets/documents/research/projects/disto/Measuring-Digital-Skills.pdf

World Economic Forum (2018). The future of jobs report 2018. Geneva, Switzerland. Retrieved from: http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf