Towards Personalized Guidance and Support for Learning
Proceedings of the 10th European Distance and E-Learning Network Research Workshop, 2018
Barcelona, 24-26 October, 2018
ISBN 978-615-5511-25-7
ISSN: 2707-2819
doi: https://doi.org/10.38069/edenconf-2018-rw-0015

DON'T DO EVIL: IMPLEMENTING ARTIFICIAL INTELLIGENCE IN UNIVERSITIES

Mark Nichols, Wayne Holmes, The Open University, United Kingdom

Summary

Artificial Intelligence (AI) is changing the ways in which we experience everyday tasks, and its reach is extending into education. Promises of AI-driven personalised learning, learner agency, adaptive teaching and changes to teacher roles are increasingly becoming realistic but the ethical considerations surrounding these, and even simpler innovations are far from clear. Various ethical standards are proposed for AI, though these tend to be high-level and generic and do not serve to guide education practice. The multiple agencies concerned with AI analytics are also yet to provide a strong sense of direction. The Open University UK has established an AI working group to explore the contribution AI might make to improving student retention, success and satisfaction. With a specific emphasis on Artificial Intelligence in Education (AIEd), this paper proposes eight principles constituting an open ethical framework for implementing AI in educational settings in ways that empower students and provide transparency.

Introduction

Artificial Intelligence (AI) has rich potential for any organisation, particularly where there are large amounts of data and repetitive, rules-based functions. AI is already a routine aspect of online experience, its uses ranging from standard searches through to shopping. Cortana, Siri and Alexa are popular examples of AI agents designed to provide personal assistance. For the purpose of this paper, we adopt the view that AI describes "computer systems that have been designed to interact with the world through capabilities (for example, visual perception and speech recognition) and intelligent behaviours (for example, assessing the available information and then taking the most sensible action to achieve a stated goal) that we would think of as essentially human" (Luckin, Holmes, Griffiths, & Forcier, 2016; p.14).

Like many universities, The Open University (UK) is taking early steps to explore the contribution AI might make to student retention, success and satisfaction. While initial work is concerned with symbolic AI (the application of data to automate specific tasks) we are conscious that, eventually, AI projects may adopt neural network-based machine learning (more complex computational approaches). Whatever the approach, and whatever the motivation, the application of AI in educational contexts raises critical ethical questions that have yet to be fully considered or addressed.

Artificial Intelligence

The versatility of AI as a technology gives it incredible scope. At the highest level, AI may be thought of as being either "weak" (based on problem-solving, in the sense of *acting intelligently*) or "strong" (based on consciousness, in the sense of *being intelligent*). Although some research and much popular imagination relates to the development and implications of strong AI, in fact the most fruitful discoveries and economic benefit thus far have been demonstrated by weak AI. Critically, however, weak does not mean simple or limited. So-called weak AI gives the potential for autonomous systems which can be given an objective and can decide, based on an algorithm, how best to meet that objective. Weak AI is the form of AI used in drone delivery services, facial recognition, medical diagnoses, legal services, and crime prediction. The potential of this technology is far-reaching.

Wilson and Daugherty (2018) suggest AI can be used for amplifying, interacting and embodying:

- AI can process vast amounts of data to *amplify* human thinking, by providing useful summaries and further data processing to enhance decision-making and creativity.
- AI can *interact* directly with humans. For example, AI interfaces to frequently asked questions might free specialists for higher-level tasks (including picking up matters the AI cannot adequately deal with).
- AI can physically *embody* computational process and extend human capabilities. For example, robots have been developed to interact with humans as carers, companions and teachers.

The various possibilities of AI are poised to reshape multiple global industries, while many countries are jockeying for leadership positions by linking large-scale investment from central agencies and private sources (Sloane, 2018).

AI ethics

The more powerful AI becomes and the more it becomes a standard feature of daily life, the more critical ethical matters become. AI seeks to apply data in objective ways. However, source data is not immune from bias; there is no such thing as "raw data" (Gitelman & Jackson, 2013). Further, the algorithms designed to process data (the choices that have been made by AI engineers) themselves have ethical consequences. Already examples exist where AI systems reflect sexist, racist and other discriminatory behaviour and where the use of AI systems has reinforced and multiplied negative data in a vicious cycle (Crawford, 2016). It is also clear that even small amounts of personal data can be combined through AI algorithms with the potential to undermine democracy, as illustrated by the recent Cambridge Analytica scandal (Parker, 2018). AI draws on data, and thus faces the same ethical issues as does analytics (Prinsloo & Slade, 2016). AI technologies inevitably reflect the motivations of their developers and interact with a world in which automated decisions have ethical consequences.

There seems unanimous recognition of the importance of ethical principles to direct AI practices, such that the consideration of AI ethics is both timely and urgent (Boddington,

ISBN 978-615-5511-25-7

2017). The UK government is seeking to establish a "Centre for Data Ethics" (Sloane, 2018), while multiple agencies are also investigating the issues. The Leverhulme Centre for the Future of Intelligence, OpenAI, the Ada Lovelace Institute, DeepMind Ethics and Society, the Oxford Future of Life Institute, and the Partnership on AI are among the major agencies already funded to explore the boundaries within which AI should be practised (CFI, 2018; Gardam, 2017; Hern, 2017; OpenAI, 2015; Parker, 2018; Partnership on AI, n.d.). While each agency has clearly laudable goals, thus far the breadth of agency interest has developed more questions than answers. Further, the agencies tend to over-represent AI developers, venture capitalists and the corporate perspective, raising questions over whether they will be thinking broadly and critically enough (Sloane, 2018). Also unclear is the way in which any ethical principles might and will inform legislation, without which the principles themselves may not have sufficient bite to direct practice.

AIED: Artificial Intelligence in Education

Educational potential of AI

Returning to the three areas of potential for AI suggested by Wilson and Daugherty (2018), it is possible to imagine some of the potential for weak AI in education by drawing on their amplifying and interacting categories. For example, AI can process and provide learning analytics data in timely ways to better support students (*amplifying*). Educational data mining has rich potential for improving education support (Luckin et al., 2016). AI bots can directly answer frequently asked questions using voice or chat interfaces (*interacting*). Even in general discussion forums, AI can monitor and respond to FAQ-style queries. AI-driven robots have been used to support children's learning, especially to address the social needs of children on the autism spectrum (*embodying*) (Scassellati, 2007).

AI bots such as Ada (Hussain & Baggaley, 2017) and Jill Watson (McFarland, 2016) are already illustrative of the interactive applications for AI in education. Eventually AI will become more widely applied to such functions as automating grading (for example, with essay grading and feedback tools such as WriteToLearn and Open Essayist), supporting teachers (for example, through chatbot Teaching Assistants), supporting students (via a lifelong AI learning companion), assisting students with special needs through automatic materials adaptation, and otherwise freeing the teacher to focus on high-value tasks such as motivating students and providing additional support and tuition to those who need it (Lynch, 2018).

AIED has particular potential for assisting students requiring additional support (achievement gaps), either by providing additional AI-assisted tuition or by freeing up teacher time (Luckin et al., 2016). However, it is unlikely that AI will replace teachers, even when strong AI becomes available (ibid.). Despite some initial success, Intelligent Tutoring Systems are best suited to limited domains and are expensive to produce (Ferster, 2017).

Ethics in AIED

The lack of definite reference points for approved AI ethics frameworks extends into AIED. Holmes argues that "around the world, virtually no research has been undertaken, no

ISBN 978-615-5511-25-7

guidelines have been provided, no policies have been developed, and no regulations have been enacted to address the specific ethical issues raised by AIED" (Holmes, 2018; para.3). The need for ethical frameworks for AIED is further amplified by the potential for universities, as research-oriented organisations, to experiment on students. To give just one example, Jill Watson (a Virtual Teaching Assistant developed by an academic at Georgia Tech in a large-scale computer science module) was deliberately disguised when implemented. A time-delay was also added to responses, and the bot was given a human-like pseudonym. The comments posted in response to the article that brought Jill Watson to public attention (McFarland, 2016) demonstrate the ethical tension:

> *"Contemptible. Lying to the students about whom they are working with is absolutely inexcusable. They are not paying tuition to be taught by a machine--or, rather, if they are, they have a right to know it."*
>
> *"Why be prejudice[d] against the robot if it is performing as well as a human?"*

It appears that Jill Watson was seen by its developers as just a *technical challenge*, rather than as an intervention with ethical consequences. Even though most of the students appeared to be happy with the experiment, serious questions must be asked about the appropriateness of using deception in the more general application of AI in education.

AIED gives rise to an indeterminate range of questions (Holmes, 2018; Luckin et al., 2016), such as:

- What are the criteria for ethically acceptable AIED, and how might these differ across private organisations (developers of AIED products) and public authorities (schools and universities involved in AIED research)?
- What controls of data should be in place, and what opt-out options are appropriate?
- What is the ethical balance between providing data provision options, and withholding the benefits that might come from using that data?
- How transparent should decision-making algorithms and source data be?
- To what extent Is the data used to train the AI representative, and open to bias?
- Who is responsible and accountable for the AI's performance and the outcomes it leads to?

Such questions must be considered drawing on the expertise of technologists, social scientists, philosophers and pedagogues, in addition to senior administrators and students. Open and transparent implementation frameworks for AIED are urgently required.

AI at The Open University UK

The Open University inaugurated an AI Working Group in early 2018. The purpose of the group is to "provide strategic direction, leadership, design and collaborative working to ensure that the OU appropriately embraces AI technologies to benefit our students and also determine how the OU will become a leading purveyor of AI in Higher Education within the

ISBN 978-615-5511-25-7

UK" (Open University, 2018; para.1). At present membership consists of representatives across "IT, Business Improvement", the faculties, student experience units, the "Learning, Teaching and Innovation" unit, and the sponsors of various Proof of Concept (PoC) projects. Awareness of the ethical issues likely to be raised by implementation prompted the development of a set of principles as an immediate activity.

Proof of Concept projects

Several PoC projects are underway, including chatbot and unstructured (qualitative) data analysis. The two chatbot PoCs will assess two different solutions, one student-facing addressing FAQs and one staff-facing to address People Services (HR) queries. The chatbots will provide first-tier support for standard queries, initially aiding self-navigation and self-service but extending its range of support as it is trained on real queries. The qualitative data analysis PoC will be used to summarise over two million free text data fields of student queries from web forms and forums each year using semantic mining, and provide an interface summarising the data to assist with evidence-based decision-making.

These two PoC project-types give rise to different ethical issues. The first relates to the ethical use of data and the responsibility of an organisation to transparently inform the user that they are engaging with an AI agent. The second highlights the ethical issues surrounding the analysis of sensitive user data.

Principles for implementation: AI at the Open University

In its early days, the OU AI Working Group determined that it needed to properly understand the ethical issues surrounding what is, for the OU, a new way of applying technology to its operations. A distinction is made across development and implementation, so that technical experimentation can take place freely.

The principles draw on four pre-existing sets, applying them for the context of the Open University. The Future of Life Institute Asilomar AI Principles, IEEE Code of Ethics, House of Lords overarching principles for AI and Google AI objectives all inform these implementation principles (Asilomar, 2017; House of Lords Select Committee on Artificial Intelligence, 2018; IEEE, 2018; Pichai, 2018), as do the summaries of AI ethics proposed by IBM CEO Ginni Rometty (Purpose, Transparency, Skills) (DeNisco Rayome, 2017) and IBM Watson CTO Rob High (Trust, Respect, Privacy) (Hiner, 2018). Each of these frameworks emphasise the importance of openness and informed implementation, from the perspective of the organisation (deliberate objectives and staff training) and end users. Additionally, specific mention is required for the use and protection of data both legislatively in the General Data Protection Regulation (GDPR) and OU-specific policies including those related to analytics.

This extensive review gave rise to the following eight principles, all of which will underpin the research and implementation of AI at The Open University UK.

1. We will fully comply with the GDPR and all OU data policies.

ISBN 978-615-5511-25-7

2. We will have full and transparent understanding of the incumbent (as is) and planned (to be) process.
    a. We will engage with all relevant internal stakeholders.
    b. We will process map and validate both as-is and to-be states, consistent with our reference architecture.
3. We will improve the user (internal or external) experience.
    a. We will incorporate appropriate service KPIs for AI solutions.
    b. We will deliberately monitor and act on customer feedback using analytics (ongoing) and a post-implementation survey.
    c. We will maintain a list of lessons learned from each AI implementation.
4. We will provide transparent disclosure at the point of interface.
    a. We will ensure that the user knows they are in an AI environment.
    b. We will ensure the user can learn about the algorithm and the data it makes use of, in a general sense.
    c. We will provide a contact point for anyone seeking additional information.
5. We will ensure a smooth transition to a human when needed.
6. We will ensure all people working in the context of an AI solution are fully trained.
7. We will provide AI solutions with appropriate machine learning algorithms and evaluation points.
8. We will ensure each AI implementation has a named owner and point of contact.

Point 2 picks up on the importance of understanding the role AI will play in the running of the organisation, ensuring informed implementation and overall alignment with how things are done. Point 3 acknowledges that AI solutions must sometimes be given opportunity to prove themselves, and that proof will often need to be evidenced after implementation. In support of point 4, the OU will apply standard text alongside each interface similar to:

> *You are now engaging with an AI interface. Click [here] for more information. To instead engage with a person, click [here].*

Clicking for more information will reveal text similar to the following:

> *This AI function seeks to [objective]. To do so, it makes use of [data] and [algorithm]. Contact [point of contact] for more information.*
>
> *Example: This AI function seeks to provide you with immediate access to relevant policies. To do so, it makes use of your question and checks it against a database of policy information. It will store your query anonymously as part of its ongoing improvement. Contact HR-admin for more information.*

ISBN 978-615-5511-25-7

Such information might also be linked to from the profile of a forum chat bot, appropriately named to ensure transparency. Points 5 and 6 are natural extension of point 2, ensuring that the end-user has a means of escalating their query should the AI not provide an unsatisfactory or questionable result and that the staff associated with such escalations are aware of the AIs working and limitations. Finally, points 7 and 8 guarantee the longevity of, and accountability for, the performance of the AI.

The eight principles for the implementation of Artificial Intelligence technologies at The Open University are designed to reflect robust ethical considerations. Once endorsed by the AI working group, only those instances of AI that comply with the eight principles will be eligible for production. A successful PoC can advance into change management activities that structure ethical concerns into the outcome.

Conclusion

Crawford and Calo remark that analysis of AI "needs to draw on philosophy, law, sociology, anthropology and science-and-technology studies, among other disciplines. It must also turn to studies of how social, political and cultural values affect and are affected by technological change and scientific research" (2016; p.313). If AI is also to be used responsibly in education, a pedagogical analysis will also be necessary. Until formal ethical frameworks emerge for the implementation of AI in education, it is important that universities design AI solutions mindful of their limitations and potential to do harm. The Open University implementation principles aim to ensure that any AI implemented in the university is transparent, and that owners are accountable for their ongoing compliance with ethical issues from the user's perspective. Even once an ethical framework is developed and adopted, it must be continuously updated to anticipate and reflect the breadth of capability for AI.

References

1. Asilomar (2017). *AI Principles – Future of Life Institute.* Retrieved August 17, 2018, from https://futureoflife.org/ai-principles/
2. Boddington, P. (2017). *Towards a code of ethics for artificial intelligence.* Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-60648-4
3. CFI (2018). *Trust and transparency.* Retrieved August 13, 2018, from http://lcfi.ac.uk/projects/ai-trust-and-society/trust-and-transparency/
4. Crawford, K. (2016, June 25). Artificial Intelligence's white guy problem. The New York Times [Blog post]. Retrieved August 6, 2018, from https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html?_r=1
5. Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature*, 311–313. Retrieved from http://www.nature.com/news/there-is-a-blind-spot-in-ai-research-1.20805
6. DeNisco Rayome, A. (2017, January 17). 3 guiding principles for ethical AI, from IBM CEO Ginni Rometty. TechRepublic [Blog post]. Retrieved from

ISBN 978-615-5511-25-7

https://www.techrepublic.com/article/3-guiding-principles-for-ethical-ai-from-ibm-ceo-ginni-rometty/

7. Ferster, B. (2017, January 21). Intelligent Tutoring Systems: What happened? eLearning Industry [Blog post]. Retrieved August 29, 2018, from https://elearningindustry.com/intelligent-tutoring-systems-what-happened
8. Gardam, T. (2017, December 11). Social well-being and data ethics. Ada Lovelace Institute [Blog post]. Retrieved August 10, 2018, from https://www.adalovelaceinstitute.org/social-well-being-and-data-ethics-tim-gardams-speech-to-techuk-digital-ethics-summit/
9. Gitelman, L., & Jackson, V. (2013). Introduction. In L. Gitelman (Ed.), *"Raw data" is an oxymoron* (pp. 1–14). Cambridge, Massausetts; London, England: The MIT Press. https://doi.org/10.1080/1369118X.2014.920042
10. Hern, A. (2017, October 4). DeepMind announces ethics group to focus on problems of AI. Technology | The Guardian [Blog post]. Retrieved August 10, 2018, from https://www.theguardian.com/technology/2017/oct/04/google-deepmind-ai-artificial-intelligence-ethics-group-problems
11. Hiner, J. (2018, March 2). IBM Watson CTO: The 3 ethical principles AI needs to embrace. TechRepublic [Blog post]. Retrieved from https://www.techrepublic.com/article/ibm-watson-cto-the-3-ethical-principles-ai-needs-to-embrace/
12. Holmes, W. (2018). The ethics of Artificial Intelligence in education. University Business. Retrieved August 10, 2018, from https://universitybusiness.co.uk/Article/the-ethics-of-artificial-intelligence-in-education-who-care
13. House of Lords Select Committee on Artificial Intelligence (2018). *AI in the UK: Ready, willing and able?* Report session 2017-19. London. Retrieved from https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf
14. Hussain, A., & Baggaley, D. (2017). *Bolton College used IBM Watson to build a virtual assistant that enhances teaching, learning and information access – Watson.* Retrieved August 17, 2018, from https://www.ibm.com/blogs/watson/2017/08/bolton-college-uses-ibm-watson-ai-to-build-virtual-assistant-that-enhances-teaching-learning-and-assessment/
15. IEEE (2018). *IEEE – IEEE Code of Ethics.* Retrieved August 13, 2018, from https://www.ieee.org/about/corporate/governance/p7-8.html
16. Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education.* London. Retrieved from https://static.googleusercontent.com/media/edu.google.com/en//pdfs/Intelligence-Unleashed-Publication.pdf

ISBN 978-615-5511-25-7

17. Lynch, M. (2018, May 5). 7 roles for Artificial Intelligence in education. The Tech Advocate [blog post]. Retrieved August 13, 2018, from https://www.thetechedvocate.org/7-roles-for-artificial-intelligence-in-education/

18. McFarland, M. (2016, May 11). What happened when a professor built a chatbot to be his teaching assistant. The Washington Post [Blog post]. Retrieved August 13, 2018, from https://www.washingtonpost.com/news/innovations/wp/2016/05/11/this-professor-stunned-his-students-when-he-revealed-the-secret-identity-of-his-teaching-assistant/?noredirect=on&utm_term=.70facd03940d

19. Open University, The (2018). *AI Working Group terms of reference.* Internal document.

20. OpenAI (2015). *About OpenAI.* Retrieved August 13, 2018, from https://openai.com/about/

21. Parker, I. (2018, April 26). UK wants to lead the world in tech ethics…but what does that mean? Ada Lovelace Institute [Blog post]. Retrieved August 10, 2018, from https://www.adalovelaceinstitute.org/uk-wants-to-lead-the-world-in-tech-ethicsbut-what-does-that-mean/

22. Partnership on AI (n.d.). *FAQ – The Partnership on AI.* Retrieved August 10, 2018, from https://www.partnershiponai.org/faq/

23. Pichai, S. (2018, June 7). AI at Google: our principles. The Keyword, Google [Blog post]. Retrieved from https://blog.google/technology/ai/ai-principles/

24. Prinsloo, P., & Slade, S. (2016). Student vulnerability, agency, and learning analytics: An exploration. *Journal of Learning Analytics, 3*(1), 159–182. Retrieved from http://oro.open.ac.uk/46172/

25. Scassellati, B. (2007). How social robots will help us to diagnose, treat, and understand autism. In S. Thrun, R. Brooks, & H. Durrant-Whyte (Eds.), *Robotics Research* (pp. 552–563). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-48113-3_47

26. Sloane, M. (2018, July 6). Making artificial intelligence socially just: Why the current focus on ethics is not enough. LSE, The London School of Economics and Political Science [Blog post]. Retrieved from http://blogs.lse.ac.uk/politicsandpolicy/artificial-intelligence-and-society-ethics/

27. Wilson, J. R., & Daugherty, P. R. (2018). How humans and AI are working together in 1,500 companies. *Harvard Business Review, July-August.* Retrieved August 13, 2018, from https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces

ISBN 978-615-5511-25-7