Co-Creation and Co-Design Methodologies to address Social Justice and Ethics in Machine Learning

Alessio Malizia and Silvio Carta
University of Hertfordshire, UK
a.malizia,s.carta{@herts.ac.uk}

Nowadays, we are relying on Machine Learning (ML) algorithms to either make or support operational decisions. In the coming decade, high-tech products are going to rely heavily on ML. However, researchers and practitioners have reported difficulties in anticipating the future behaviour of ML algorithms without knowing what further data will be used for their training. Moreover, developers usually are not entirely aware of how to reflect on social justice while designing ML algorithms.
The question, therefore is: What kind of tools and techniques can help those designing new ML systems to ensure social justice and ethics are adequately taken into account?

Context and Background

Algorithmic bias has been recognised as a relevant issue in ML applications. For example, IEEE and ISO are currently developing standards which cover algorithmic bias, and a new Joint Technical Committee (ISO/IEC-SC42) has been established for the development of standards related to AI. However, mitigating algorithmic bias is far from an easy task. In 2015, a developer highlighted that Google's visual identification algorithm could not accurately distinguish between Black people and gorillas. Three years later, Google had simply switched off the ability to search for gorillas.
The literature on mitigating algorithmic bias has been reported as speculative and rarely based on concrete evidence. Moreover, there is little research on how mitigation strategies work in practice (Morley et al., 2019), for instance, due to the wide adoption of proprietary tools. The current literature is mostly focused on the USA, and few studies focus on the UK or Europe, where governance and circumstances are often quite different.
A first step towards mitigating algorithmic bias consists of tools to help elicit social values and pro-ethically handle value pluralism. Examples of such tools are the Guide to the Ethical Design and Application of Robots and Robotic Systems by the British Standards Institute, or the Responsible Research and Innovation (RRI) methodology employed by the European Commission's Human Brain Project. However, most of these are not easily actionable in practice.
To address such issue, several research projects proposed actionable methodologies. The Center for Democracy and Technology created a Digital Decision Tool, consisting of an interactive flowchart designed to raise concerns regarding bias, fairness, and ethical issues during the development of algorithmic systems. Katell et al. developed an Algorithmic Equity Toolkit based on participatory design methods (2020).

Discursive Strategies, such as workshops and discussion forums, are an interesting class of approaches to mitigate algorithmic bias, which guarantees humans to override automated decisions where necessary, dealing with situations where machines would struggle (Rovatsos, et. al, 2019). Among Discursive Strategies for ML, a novel approach is introduced by Design Fiction, an interdisciplinary method to allow participants creating and reconfiguring concepts into scenarios to expose potential bias and reflect on mitigation strategies (Malizia, 2019). Design Fiction provides opportunities to reveal aspects of how technology will be adopted, becoming a conversation starter to discuss implications, ramifications, and effects of technology in the future.

Co-Creation and Co-Design Methods

Methods of cross-sectoral and transdisciplinary communication —Tools and methods to explore the design and development of ethics-aware ML algorithms will need to be highly inclusive involving engineers, social scientists, policy-makers and citizens. Therefore, exploring methods such as Co-Design (Bødker, 1993), Scenario-based Design (Carroll, 2000) or Design Fiction (Malizia et al., 2019) will allow participants to transform concepts and algorithms into scenarios and prototypes to study the implications of introducing such algorithms in the society.

Rapid dynamic responses to fast-changing technologies — Design thinking and design toolkits based on brainstorming activities involving cards and storyboards have been extensively used to co-design systems but although informative, they are quite limited in rendering the implications on future scenarios. Successful solutions must allow the rapid prototyping of ML algorithms running on near-future scenarios and experiencing the potential implications in terms of fairness, accountability and transparency.

Vision

We propose to introduce co-creation and co-design methodologies to stimulate reflections on accountability, fairness and transparency of ML algorithms at design time before deploying such algorithms in the society and potentially causing exclusion and inequality.

In the 'prime-lining scandal', Amazon made free same-day delivery available to Prime service subscribers in the US but only in some areas, excluding those from, for example, Afro-American residential areas. Amazon subsequently chose to disregard its algorithm and made free same-day delivery available across all areas.

Educating the next generation of ML developers to adopt co-creation and co-design methods will positively affect companies they will be working for, e.g. by being able to launch ML-based products into the market with a lower risk of social issues. Finally, the whole society will indirectly benefit from such approaches by having access to ML-based digital services and applications carrying a lower risk of bias.

References

Bødker, S. (1993). The AT-Project: practical research in cooperative design. DAIMI Report Series, (454).

Carroll, J. M., 2000. Making use: scenario-based design of human-computer interactions. MIT press.

Katell et al. (2020). Toward situated interventions for algorithmic equity: lessons from the field. Proceedings of ACM-FAT* '20 conference.

Malizia, A. (2019). Design Fictions to Mitigate Social Injustice in Possible Futures. Blog@Ubiquity, ACM

Malizia, A, Bond, RR, Turkington, R & Mulvenna, M. (2019), Human and Data-Driven Design Fictions: Entering the Near-Future Zone, in Proceedings of the International Workshop on Rethinking Cognitive Ergonomics: ReCogErg 2019 Rethinking Cognitive Ergonomics. vol. 2539, CEUR-WS, CEUR-WS,

Morley et al. (2019). From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. Science and Engineering Ethics

Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson & Barnes, P. (2020). Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. arXiv preprint arXiv:2001.00973.

Rovatsos et al. (2019). Landscape Summary: Bias In Algorithmic Decision-Making: what is bias in algorithmic decision-making, how can we identify it, and how can we mitigate it?