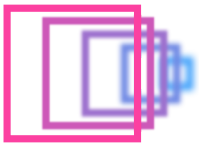


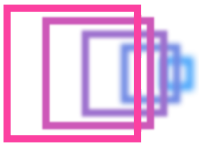
Experiential AI

Editorial

Toward Fairness, Morality and Transparency in Artificial Intelligence through Experiential AI

Drew Hemment, University of Edinburgh
Vaishak Belle, University of Edinburgh
Ruth Aylett, Heriot-Watt University
Dave Murray-Rust, University of Edinburgh
Larissa Pschetz, University of Edinburgh
Frank Broz, Heriot-Watt University

To appear in Leonardo, 2019. 10.1162/LEON_a_01795



Experiential AI

Toward Fairness, Morality and Transparency in Artificial Intelligence through Experiential AI

The new research theme Experiential AI, at Edinburgh Futures Institute, responds to concerns around the societal impacts of artificial intelligence (AI) and proposes a cross-disciplinary AI research and practice between art and science [1].

The last year has been a watershed for the conjunction of art and AI, reflected in Ars Electronica introducing a new category in their Prix [2]. Elsewhere, capabilities of AI were often overclaimed in reporting on the first artwork “created by an algorithm” to be sold at Christie’s [3]. Increasingly, artists are experimenting with machine learning algorithms as subject and tool. Something of an emerging machine learning aesthetic reveals and manifests distortions in the ways algorithms interpret the world. The misshapen imagery that can result, named by one proponent the “Francis Bacon effect” [4], arguably enables the character of machine reasoning and vision to be made explicit and its artifacts tangible [5].

Other artists have addressed the social and ethical consequences of algorithms. CV Dazzle by Adam Harvey presents hair styling and makeup as camouflage from face-detection technology [6], and Mushon Zer-Aviv questions the construction of prejudice and normalcy in The Normalizing Machine [7].

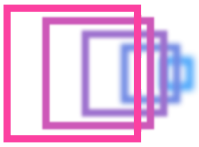
The fairness and morality of AI systems, a topic so far little explored by the current generation of artists working with machine learning algorithms, is already under investigation by AI scientists. Because most prediction systems look for frequent patterns, algorithms trained on data embodying specific historical and cultural biases produce, unsurprisingly, systems exhibiting these same biases. Thus, AI researchers are devising definitions that enable predictions to respect fairness constraints with respect to gender, race and other “protected attributes” despite biases in data.

Fairness, is however, one part of a larger picture. AI algorithms are frequently used for recruitment and trading stocks, but also in self-driving cars and domestic robots that act physically on the environment and with people. What biases might these robotic applications infer from long-term interactions with us? Conversely, what values would we like them to embody? Should we strive for a shared computational framework enabling machines to reason about their actions and ethical implications? How can we make such systems sufficiently transparent that we can understand and critique their reasoning?

The design of moral machines needs to account for the contingency of human value systems across contexts, cultures and demographic groups. The logic of computation shapes political and public discourse in its image, in ways that can be inimical to societal values. There are, therefore, long-standing concerns about translating morals from the domain of human ethics and politics into a framework using numeric representations. We must both consider machines that can engage in ethical reasoning and simultaneously recommend social domains that are not suited to automated decision-making.

Fairness, morality and transparency will be the theme of the first programme in Experiential AI. Participating artists will experiment with new ideas, data and technologies and engage both AI practitioners and publics in envisioning futures for ethical and responsible AI. Their works can allow audiences to explore future scenarios and experience various aspects of the moral dimension, making algorithmic mechanisms vividly apparent.

As a methodology and approach, experiential AI can question data harvesting, algorithms and the outcomes of their application, and how a system is understood. Art can create experiences around social impacts and consequences of technology, giving audiences direct experience of philosophical and computational principles. It can challenge what the algorithms are and how they are used. This can lead to significant new works and create insights to feed into the design of these technologies.



Experiential AI

References and Notes

1. D. Hemment et al., “Experiential AI,” AI Matters 5, No. 1 (2019).
2. www.ars.electronica.art/prix/en/categories/artificial-intelligence-life-art.
3. www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx.
4. www.twitter.com/wxswxs/status/1040518889849925633.
5. See Ref. [1].
6. www.cvdazzle.com.
7. www.mushon.com/tnm.

Authors

Drew Hemment, University of Edinburgh
Vaishak Belle, University of Edinburgh
Ruth Aylett, Heriot-Watt University
Dave Murray-Rust, University of Edinburgh
Larissa Pschetz, University of Edinburgh
Frank Broz, Heriot-Watt University