Books, Bias and other ways to say AI

Vision Statement: Teaching HCI for AI: Co-design of a Syllabus

Alan Dix
Computational Foundry, Swansea University
http://alandix.com/

Summary

I am a HCI educator and researcher who has also worked at the intersection of AI and HCI including some of the earliest work on algorithmic bias. Currently I am writing a new edition of an introduction to Artificial Intelligence and wish to bring in a more human-centred aspect to the treatment of the subject.

Personal Background

Although principally known as a HCI researcher and educator, I have also had long-standing interests at the intersections between AI and HCI. This has included commercial work on applying genetic algorithms to early submarine design in the late 1990s and intelligent internet interfaces using data recognisers [DB00] during the dot-com years. However, the majority has been in the academic sphere including an early introductory AI textbook co-authored with Janet Finlay [FD96] and one of the earliest (or possibly the earliest) works to highlight the potential dangers of gender and ethnic bias in black-box machine learning algorithms published in 1992 [Di92]. I have also been engaged with the establishment of the human-like computing (HLC) initiative in the UK and used HLC paradigms in my own work, in particular cognitively inspired models of regret, which boost the rate of active learning by a factor of 5-10.

In my current role as Director of the Computational Foundry at Swansea University I am part of a team establishing a joint AI Centre with Grenoble (on hold due to Covid!) and also part of a doctoral training centre in 'people first' AI/Big Data (http://people-first.best/) which is nurturing 55 PhD students over the coming years. These researchers will be engaged in industry and third-sector co-sponsored projects that push the fundamental boundaries of AI, ML and data science in human-centred domains from health to manufacturing 4.0.

Over the last two years I have also resumed my interest in understanding bias with various talks, keynotes and tutorials. This is partly with a theoretical and policy agenda, particularly trying to communicate the message that it is not simply a matter of de-biasing training data, but that potentially unethical bias is fundamental to the nature of data-driven generalisation. It also has a technical side working towards a 'kitbag' of techniques and heuristics for explainable AI.

Touchstone phrases

Over the years in discussions particularly about the interaction between HCI and AI, I have found myself returning to two re-phrasings of AI:

Alien Intelligence –The nature of 'intelligence' in AI is often far from the way we think as a human. Being explicit about this can often be helpful when discussing AI with non-computer people. It can also open up critical questions regarding circumstances when this does or does not matter, such as explainability.

Appropriate Intelligence – Many or maybe most AI-based systems are designed to be used alongside people. In such a socio-technical (or symbiotic) system, it is the way that AI works as part of an interaction that is critical, for example, it is often more important to understand and deal with failure situations than to be right as often as possible.

In recent years, when considering explainability these two come together in concepts of *'sufficient reason*' realising that much of our own decision-making is tacit, unconscious and effectively alien – what is necessary is that we can give sufficient explanation to ourselves or others for the task at hand.

Current foci related the workshop

HCI in AI Education

Twenty-five years ago Janet Finlay and I wrote a short textbook, An Introduction to Artificial Intelligence [FD96], focused particularly on those studying on conversion masters courses for whom the Bible-like tomes available at the time were not suitable. For many years there were few major changes in AI, but that has changed in recent years due partly to increasing volumes of data and partly to faster and more scalable computation.

I have agreed to produce a second edition of this (see [Di20]). For this book I am aiming not just to add these recent developments, but also introduce more human elements that place these developments in a broader systems and societal context. My expectation is that this will not dominate the text (not backdoor HCI!), but rather inform many aspects, sometimes coming to the forefront (ethics, case studies of intelligent interfaces, bias, accountability and explainability), and sometimes comprising a background thread.

Although I already have some ideas of how this will play out, I will be really interested in the main themes and topics that emerge from this workshop. Crucially this is about a text book for AI courses, so a really chance to ensure that critical human elements are not missed.

It will be another year before this is finished and published, but in the short term the first edition has been released free as a PDF and over the coming months I will be writing portions, and creating online materials (slides, videos, case studies). I will be really pleased to hear from those teaching this kind of material; if you would like me to prioritise particular topics. My aim is to support colleagues in the new academic term which will be challenging for many as the Covid crisis continues to dominate and transform the ways we teach.

AI for HCI

In parallel with the human-centred AI textbook, I've also agreed to do a shorter "AI for HCI" book aimed at UX/IxD practitioners and HCI researchers. Although this is addressing the opposite issue: communicating the potential

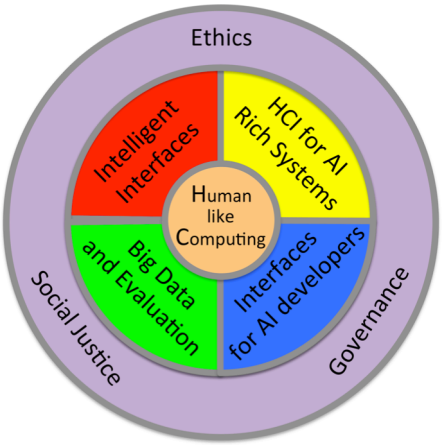

Figure 1. AI challenges in HCI

and problems of AI to HCI and UX community. I also expect synergies.

Figure 1 summarises the planned content. As is evident, there are clear areas of overlap between HCI for AI and AI for HCI including ethics and human-like computing. On the left of figure 1 are areas where AI can be used within HCI deployment (e.g. intelligent interfaces) and development (e.g. big-data analysis of large-scale trace data; these are potential applications areas for case studies . On the right are areas where HCI can be used within AI deployment (e.g. autonomous cars) and development (e.g. AI explainability tools)

References

[DB00] A. Dix, R. Beale and A. Wood (2000). Architectures to make Simple Visualisations using Simple Systems. Proceedings of Advanced Visual Interfaces - AVI2000, ACM Press, pp. 51-60. https://www.alandix.com/academic/papers/avi2000/

[Di92] A. Dix (1992). Human issues in the use of pattern recognition techniques. In Neural Networks and Pattern Recognition in Human Computer Interaction Eds. R. Beale and J. Finlay. Ellis Horwood. 429-451. https://alandix.com/academic/papers/neuro92/

[Di20] A. Dix (2020). Artificial Intelligence – humans at the heart of algorithms (accessed 25/6/2020). https://alandix.com/aibook/second-edition/

[FD96] J. Finlay and A. Dix (1996). An Introduction to Artificial Intelligence. UCL Press / Taylor and Francis, ISBN 1-85728-399-6