Proposal for SIGCHItaly Workshop on "Teaching HCI for AI: Co-design of a Syllabus"
Rome, 7 July 2020

Vision Statement for Teaching HCI4AI Syllabus

B.L. William Wong
Interaction Design Centre, Middlesex University London, UK
w.wong@mdx.ac.uk

AI and ML systems are being increasingly used in a range of problem environments. Some are low consequence environments where mistakes or wrong assumptions made by the technology is of little consequence, e.g. recommendations of books to purchase when browsing an on-line store. However, AI and ML technology is increasingly being used in complex and high consequence environments, where mistakes or inappropriate assumptions can lead to loss of life, infringement of our privacy, or longer prison sentences because of embedded prejudices and therefore impact our rights as individuals in democratic societies. We identified some of these ethical and practical problems as accidental discrimination, the mosaic effect, and algorithmic opacity (Duquenoy, Gotterbarn, Kimppa, Patrignani, & Wong, 2018; Paudyal & Wong, 2018). For the purposes of the HCI4AI statement, we will limit our discussion to issues regarding the problem of algorithmic opacity and the need for computational transparency as one approach to mitigating such problems.

In recent times, we have investigated methods for computational transparency to make the black box algorithms visible to inspection (Hepenstal, Kodagoda, Zhang, Paudyal, & Wong, 2019); translating the ethical and legal regulations and principles into user requirements and system specifications (Duquenoy et al., 2018); and methods to model the cognitive work of users of the algorithms (Paudyal & Wong, 2018) in order to enable recoverability from and resilience to automation surprises.

We have observed that many AI and ML systems are black boxes. The inner workings of the algorithms and how they produce the results are often unknown to the user. Users are unable to ascertain or challenge the results. Nadine Sarter and her colleagues have studied the problem of black box automation and how such black boxes can lead to a phenomenon called 'automation surprise' (Sarter, Woods, & Billings, 1997). This occurs when the actions performed by the automation lead to unexpected outcomes – what is the system doing? why is the system giving contradicting information? or why is the AI system recommending that black men be given longer prison sentences that are longer than white men for the same kinds of crimes? Because of algorithmic opacity users are unable to inspect, challenge, ascertain or verify if and how the outcomes produced by the algorithm are sensible or correct. If users cannot inspect or verify the outcomes, they will not be able to trust it, and probably not maximise its benefits.

Therefore, what should a curriculum for HCI4AI look like? What should be its focus and priorities? We propose that a HCI4AI curriculum should emphasise the development of knowledge and skills that would enable designers and developers to create human-machine teams that use AI / ML technologies (McDermott et al., 2018). Such teams should operate on the basis of human-machine symbiosis (Licklider, 1960), leveraging off the capabilities of

the other in the context of their environment, creating a joint cognitive system (Hollnagel & Woods, 2005). Whether the purpose is to develop AI-based control systems or to enable dynamic interaction with information, such a design approach is to ensure that the algorithm-based system will have the variety of capabilities such that together with the human, will be able to manage the variety of demands presented by the work in the environment (Ashby, 1958). Such systems are also less likely to suffer brittle-ness and fail when unanticipated conditions are presented.

In addition to the standard HCI and AI curriculum, a HCI4AI curriculum could include:

1. **Understanding the phenomenon** of human-machine symbiosis as basis for joint cognitive systems to create human-machine teams.
    i. AI / ML: how should algorithm- and probability-based computation behave to augment human cognition?
    ii. Human cognition: what are the needs of the macro-cognitive functions like situation awareness, sense making, problem solving, and decision making in the context of using AI technology to augment human intelligence?
    iii. Cognitive and machine (computational) biases
2. **Designing for algorithmic transparency**: An understanding of the need for system visibility and what this means in terms of the designing features for visibility, e.g. as guided by the Algorithmic Transparency Framework (Hepenstal et al., 2019). This is a model that describes what a user needs from black box algorithms: visibility of the functional relationships mapped against the goals and constraints of the system; explanations of how results from algorithms are arrived at; explanations that are interpretable by the user in a manner that enables users to question and challenge the results; and context in which to interpret the explanations.
3. **Privacy by Design**: The need to understand human rights, their basis, and why the abuse of human rights can lead to societies where smart technologies can be abused, e.g. to control the behaviour and freedoms of a population. The key to this is understanding the need to respect human rights as a design goal with constraints, and how the constraints can force the design of novel designs, e.g. location-based vs distance-based COVID-19 contact tracing apps.
4. **Resilience**: The need for resilience of the system / organisation / human to failure; e.g. how can we tell if the algorithms are making errors? If or when errors are made, how should the system be designed to enable safe recovery?
5. **Cognitive Engineering**
    i. **Methods for modelling the algorithms** i.e. what it does and how it works in a user/task-relevant manner, e.g. *cognitive work analysis (Vicente, 1999)*
    ii. **Methods that describe how a user thinks and reasons** and what is required to understand how to control or manage the algorithms, and diagnose when things go wrong, e.g. *cognitive task analysis (Klein, Calderwood, & MacGregor, 1989)*
    iii. **Methods for representing the cognitive work on a user interface**, e.g. *Representation Design* (Bennett & Flach, 2011), Ecological Interface Design (Burns & Hajdukiewicz, 2004)

References

Ashby, W. R. (1958). Requisite variety and its implications for the control of complex systems. *Cybernetica, 1*(2), 83-99.

Bennett, K. B., & Flach, J. M. (2011). *Display and Interface Design: Subtle Science, Exact Art*. Boca Raton: CRC Press, Taylor and Francis Group.

Burns, C. M., & Hajdukiewicz, J. R. (2004). *Ecological Interface Design*. Boca Raton, FL: CRC Press.

Duquenoy, P., Gotterbarn, D., Kimppa, K. K., Patrignani, N., & Wong, B. L. W. (2018). Addressing Ethical Challenges of Creating New Technology for Criminal Investigation: The VALCRI Project In G. Leventakis & M. R. Haberfeld (Eds.), *Societal Implications of Community-Oriented Policing Technology* (pp. 31-38): SpringerBriefs in Policing.

Hepenstal, S., Kodagoda, N., Zhang, L., Paudyal, P., & Wong, B. L. W. (2019). Algorithmic Transparency of Conversational Agents. In *Proceedings of ACM IUI ATEC '19, Algorithmic Transparency in Emerging Technologies, 20 March 2019, Los Angeles, USA.* (pp. Accepted for publication). New York, NY: ACM Press.

Hollnagel, E., & Woods, D. D. (2005). *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*. Boca Raton, FL 33487-2742: CRC Press, Taylor and Francis Group, LLC.

Klein, G. A., Calderwood, R., & MacGregor, D. (1989). Critical decision method for eliciting knowledge *IEEE Transactions on Systems, Man and Cybernetics, 19*(3), 462-472.

Licklider, J. C. R. (1960). Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics HFE*, 4–11.

McDermott, P., Dominguez, C., Kasdaglis, N., Ryan, M., Trahan, I., & Nelson, A. (2018). *Human-Machine Teaming Systems Engineering Guide*. Bedford, MA: The MITRE Corporation.

Paudyal, P., & Wong, B. L. W. (2018). Algorithmic opacity: making algorithmic processes transparent through abstraction hierarchy. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting HFES 2018, Philadelphia, 1-5 October 2018* (pp. 192-196). Santa Monica, CA: Human Factors and Erognomics Society.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation Surprises. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (2nd ed., pp. 1926-1943). New York: John Wiley and Sons.

Vicente, K. J. (1999). *Cognitive Work Analysis: Toward safe, productive, and healthy computer-based work*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.