A Crossroads for Hybrid Human-Machine decision-making

B Wilson, K Lakshmanan, A Dix, A Rahat, and M Roach

Swansea University, SA1 8EN, UK
b.j.m.wilson@swansea.ac.uk

Abstract. In this paper, we highlight what is missing in the approach to architecting and developing AI models that would mean performance is translated into effective hybrid systems. We conclude people, place and purpose should drive new architectures that support rich interaction through tractable representations that will underpin success.
We call for the data-driven ML community to embrace the consideration of tractable representations in the architecture of algorithms and place a responsibility on HCI researchers to unwrap and expose the significant factors in the design space that are critical for successful hybrid decision-making in the real world.

Keywords: human-machine · AI · interaction · hybrid · mixed-initiative · combination · collaboration · cooperation · symbiosis · decision making

1 How do the human and the computer cross the road?

From computer science immemorial, research has looked at how humans and computers work together, although often assuming that the interaction between the two is unproblematic, out of scope or non-existent. A range of terminologies have described the space over recent decades. *Symbiosis* highlights integration and synergy between humans and machines. *Mixed-initiative* emphasises the distribution of agency between humans and machines. *Cooperation* infers mutual benefit, despite distinct goals. Whereas *collaboration* emphasizes a commonality of goals implying coordinated effort. And *hybrid* suggests the synthesis of processes rather than of results.

The sharpest growth in literature, if we exclude robotics, has been in the last two of these terms as can be seen in Figure 1. We used Semantic Scholar, focusing on papers related to HCI and noted the apparent coincidence of counts for two key terms *collaboration* and *hybrid*. Having re-checked the counts, we felt that the parallel rise in appearance of these terms, was owing to their increasing use as synonyms in that particular space. We infer that these terms describe different aspects of a single approach that sees humans working with computers that are expected to share goals and are capable of a blended route to task completion. Of particular interest is the task of decision-making. In this paper, we use *hybrid decision-making* to mean an approach that aims to blend human

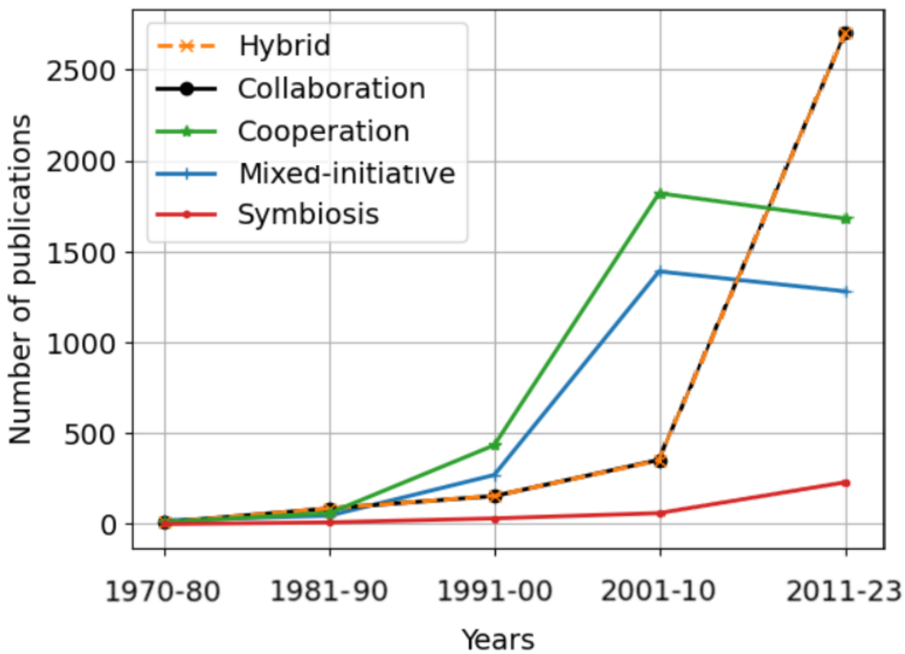

Fig. 1. Trends in terminology

and machine decision processes in pursuit of a shared decision-making goal. And, we deliberately deconstruct two critical strands of human-machine history. First, the deepening insight that machine agents bring to the joint effort. Second, the development of mutually tractable forms of representation that allow each type of agent to make use of the other's insight.

1.1 Deepening algorithmic insight

Algorithmic insight starts with expert systems, developed in the 1970s [10], that codified expert knowledge. A non-expert human user was relegated to feeding the oracle with clues by providing answers to successive questions enabling the system to crank out a synthesised expert opinion. New techniques [22] led to significant growth in Machine Learning (ML) approaches, such as random forests [8], Support-vector machines [2], and neural networks [14,1], which became dominant over codified expert knowledge and opened up new possibilities for human-computer interactions during the 1990s.

A debate ensued about whether to allow automated systems to share the initiative in these interactions [9]. Should humans call up computational power at each step? Should automation take over every task it could, without waiting for an invitation? The case was made for *mixed-initiative* systems. Blending the two capabilities instead of placing the entire burden on one or the other. Alas, the growing capability of machines to generate insight was not accompanied by increased human options for collaboration with them. In fact, in computational intelligence's most celebrated achievement of the 1990s, the victory of a sophisticated and powerful expert system over chess champion Gary Kasparov [25], there was no collaboration. The needle had shifted and the machine was working alone.

From the 2000s, public competitions, large labelled datasets and increasingly available easy-to-use libraries, expand the user-base of experimental ML ever wider. Humans are back in the loop, although they are once again in service of the algorithm - laboriously labelling data. To date, the deepest demonstrations of algorithmic insight come with the advent of deep learning techniques in the 2010s, starting with computer vision [7], and leading to an explosion of work on natural language processing using a range of advanced techniques, for example, vector representations [18] and attention architectures [24]. Once again we see machines pitted against humans, with seminal results achieved by AlphaGo and its successors, demonstrating the capability and benefit of ML [23,13]. The 2020s see language models develop significant capabilities [11].

These advances stimulate a huge public interest in the potential of Artificial Intelligence (AI) to change society in previously unexpected ways [17]. But we note that these advancements have not themselves brought computers any closer to complementarity [12]. What we call for is an ML development approach that consistently and proactively foregrounds the human operational task and makes hybridity a design feature to leverage ever deeper machine insights.

1.2 Mutually tractable representations

In the 1960s Licklider's aspiration of interaction using symbolic, pictorial, speech and handwriting representations [15] was limited by hardware and its interoperability. While we had moved on from teleprinters and paper tape, we had reached as far as the command line interface. By the 1970s, we already had miniaturisation, high-level languages, the mouse, the joystick and a form of graphical user interface. But the process of blending human insight with that of a machine had yet to progress, since the user needed to become an expert. In the 1980's, clones of the IBM PC and the arrival of the Apple Mac helped accelerate the social factors promoting greater usability. Graphical user interfaces became ubiquitous, along with interactive menus and dialogues. In the 2000s interaction design grew in sophistication with touch-screens, advanced graphics and browser technology alongside conversational user interfaces.

All these technologies gave rise to ever richer and more tractable representations. Their potential for enabling humans and computers to express and ingest their distinct insights into a problem is enormous. The most obvious expression of mutually tractable representations is visual analytics, permitting two-way communication of insight between humans and computers. But we have yet to see a more generalised adoption of interaction as a means for each type of agent to contribute what they know in the pursuit of blended decision-making. One action in this space is to evolve interactive transparency of advanced models for the benefit of the user. Another is to look in the other direction and explore how to get the computer to better interpret human intent.

2 They carefully look both ways...

There is an expanding range of new opportunities in hybrid work as a result of these developments. Machines can increasingly originate and rapidly harness more rich, relevant and complex information in the way of decision-making power (deeper insight). At the same time, a computer's capacity to share and dialogue complex results and relationships (tractable representation) is greater than ever before. Despite this, there is little that blends the strengths of humans and computers in a true hybrid decision-making process.

Our collective contribution to tackling this problem can be cultivated if we focus on anticipating real contexts of use. When we think of algorithmic performance, we tend to settle on conventional metrics and develop model designs to optimise performance. When we think of Human-Computer Interaction (HCI) contributions, we think of issues like trust, bias, fairness, privacy, ethics, ubiquity, usability and affect. Each of these perspectives is critically important for realising the benefits of AI - but they are insufficient. **What's missing is an approach to architecting and developing AI models that anticipate and take account of how user-centered they will need to be in deployment for their performance to be translated into real benefit in hybrid systems.** This missing area of work requires us to find ways to translate complex decision-making into complex interactive representations for the combined contribution to be maximised.

There are signs that researchers have started to address this gap. Cabitza and Natali [4] anticipate and take account of user-centeredness in deployment. In another contribution, Cabitza et al take pains to consider specifics of how human-machine working is achieved [3]. Cai et al [5] clearly demonstrate interactivity that provides a window onto machine insight, enabling users to see strengths and weaknesses in their algorithmic tools. Andrew Ng has made an emphatic point of the need to squeeze out real benefit, *"our job isn't only to achieve high average test accuracy - our job is to solve the problem at hand"* [19]

This year, Inkpen et al [12] call attention to the need to tune algorithmic outputs to the situated use they are intended to support in a process of advancing 'complimentarity'. Zajac et al [26] make an evaluation of 'AI in the wild' and conclude that more work should be put into anticipating both the technical and social challenges that operational use inevitably throws up. This approach is exemplified by Sendak et al [21], who start their algorithmic development process by considering what the end-users can realistically make use of and focus on ensuring situated processes allow that machine insight to be leveraged. These contributions show that we are not the first to make the case for co-design, interaction, insight convergence or iterative, situated evaluation. Our provocation aims to ground and amplify the call - to place the exploration of hybridity in decision-making at the forefront of research efforts, and have both algorithmic development and interaction design serve this end.

The latest technology to create widespread excitement illustrates the need for our call to action. Chat GPT is an example of complex algorithm design and success in terms of providing a tractable representation for humans. However, due to its infancy, it has limitations in representing back to the computer what the user's intentions are. As humans, we are left repeatedly trying distinct inputs in the hope of prompting an improved output. Our expectation is that the experience of using technologies like chat-GPT will be transformed so that users are not required to craft the perfect input by trial and error to get what they want out. While that evolutionary need is obvious, and the design work is likely to take place, we have to think analogously about all interaction modalities, how they relate to the situated task we aim to enhance and how they can be as effective in drawing humans and machines into convergences of insight.

Our contention is that we have to focus on getting the balance right between insight and tractable representation. There are implications for agency and empowerment in designing hybrid systems. Tractability informs agency and getting the balance of agency right will drive the level of influence the human can have on the combined decision. We know that our history is of systems designed to put one or the other agent in a dominant position. But we postulate that there is a much larger space between these extremes, where the level of influence or authority on the combined decision is more calibrated according to the value added.

We argue that translating the current interest in hybrid into real benefit means we should anticipate moving beyond individual mechanisms of deference [16], aggregating [6], balancing [20], and transparency [21] into designing for a potentially much richer peer space where humans and machines begin to create collaborative partnerships. For this to happen, there must be opportunities for greater dialogue between them. This means much further development of machines to allow their access to such representations.

3 ...hold hands and step forward

Our call to action is for the data-driven Machine Learning community to embrace consideration of tractable representations, to identify and design mutually tractable representations that enable an interactive balance between relevant insight and appropriate influence in hybrid systems. This means going beyond developing capability in decision-making and machine insight and developing appropriate, tractable representations of both contributions. And it places a responsibility on HCI researchers to unwrap and expose the significant factors in the design space that are critical for success in hybrid decision-making in the real world.

Acknowledgements This work was funded by Swansea University EPSRC grant EP/S021892/1. For the purpose of Open Access, the author has applied a CC BY license to any Author Accepted Manuscript (AAM) version arising from this submission.

References

1. Bishop, C.M.: Neural networks and their applications. Review of Scientific Instruments **65**(6), 1803–1832 (Jun 1994). https://doi.org/10.1063/1.1144830
2. Burges, C.J.: A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery **2**(2), 121–167 (Jun 1998). https://doi.org/10.1023/A:1009715923555
3. Cabitza, F., Campagner, A., Ronzio, L., Cameli, M., Mandoli, G.E., Pastore, M.C., Sconfienza, L.M., Folgado, D., Barandas, M., Gamboa, H.: Rams, hounds and white boxes: Investigating human–AI collaboration protocols in medical diagnosis. Artificial Intelligence in Medicine **138**, 102506 (Apr 2023). https://doi.org/10.1016/j.artmed.2023.102506, https://www.sciencedirect.com/science/article/pii/S0933365723000209
4. Cabitza, F., Natali, C.: Open, Multiple, Adjunct. Decision Support at the Time of Relational AI. In: Schlobach, S., Pérez-Ortiz, M., Tielman, M. (eds.) Frontiers in Artificial Intelligence and Applications. IOS Press (Sep 2022). https://doi.org/10.3233/FAIA220204
5. Cai, C.J., Reif, E., Hegde, N., Hipp, J., Kim, B., Smilkov, D., Wattenberg, M., Viegas, F., Corrado, G.S., Stumpe, M.C., Terry, M.: Human-Centered Tools for Coping with Imperfect Algorithms During Medical Decision-Making. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 1–14. CHI '19, Association for Computing Machinery, New York, NY, USA (May 2019). https://doi.org/10.1145/3290605.3300234
6. De Fauw, J., Ledsam, J.R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., et al.: Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature medicine **24**(9), 1342–1350 (2018). https://doi.org/https://doi.org/10.1038/s41591-018-0107-6
7. Deng, L., Yu, D.: Deep Learning: Methods and Applications. Foundations and Trends in Signal Processing **7**(3–4), 197–387 (Jun 2014). https://doi.org/10.1561/2000000039
8. Ho, T.K.: Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition. vol. 1, pp. 278–282 vol.1 (Aug 1995). https://doi.org/10.1109/ICDAR.1995.598994
9. Horvitz, E.: Principles of mixed-initiative user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems the CHI Is the Limit - CHI '99. pp. 159–166. ACM Press, Pittsburgh, Pennsylvania, United States (1999). https://doi.org/10.1145/302979.303030
10. Horvitz, E.J., Breese, J.S., Henrion, M.: Decision theory in expert systems and artificial intelligence. International Journal of Approximate Reasoning **2**(3), 247–302 (Jul 1988). https://doi.org/10.1016/0888-613X(88)90120-X
11. Hughes, A.: ChatGPT: Everything you need to know about OpenAI's GPT-4 tool (Jun 2023), https://www.sciencefocus.com/future-technology/gpt-3/

12. Inkpen, K., Chappidi, S., Mallari, K., Nushi, B., Ramesh, D., Michelucci, P., Mandava, V., VepŘek, L.H., Quinn, G.: Advancing Human-AI Complementarity: The Impact of User Expertise and Algorithmic Tuning on Joint Decision Making. ACM Transactions on Computer-Human Interaction (Mar 2023). https://doi.org/10.1145/3534561
13. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D.: Highly accurate protein structure prediction with AlphaFold. Nature **596**(7873), 583–589 (Aug 2021). https://doi.org/10.1038/s41586-021-03819-2
14. Lawrence, S., Giles, C., Tsoi, A.C., Back, A.: Face recognition: A convolutional neural-network approach. IEEE Transactions on Neural Networks **8**(1), 98–113 (Jan 1997). https://doi.org/10.1109/72.554195
15. Licklider, J.C.R.: Man-Computer Symbiosis. IRE Transactions on Human Factors in Electronics **HFE-1**(1), 4–11 (Mar 1960). https://doi.org/10.1109/THFE2.1960.4503259
16. Madras, D., Pitassi, T., Zemel, R.: Predict Responsibly: Increasing Fairness by Learning to Defer. ICLR 2018 Workshop Submission (Feb 2018), https://openreview.net/forum?id=HyjVKLJwz
17. Malik, K.: Fantasy fears about AI are obscuring how we already abuse machine intelligence. The Observer (Jun 2023), https://www.theguardian.com/commentisfree/2023/jun/11/big-tech-warns-of-threat-from-ai-but-the-real-danger-is-the-people
18. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient Estimation of Word Representations in Vector Space (Sep 2013). https://doi.org/https://doi.org/10.48550/arXiv.1301.3781
19. Ng, A.: Batch aI news (2020), https://www.deeplearning.ai/
20. Russell, S.: The history and future of AI. Oxford Review of Economic Policy **37**(3), 509–520 (Sep 2021). https://doi.org/10.1093/oxrep/grab013
21. Sendak, M., Elish, M.C., Gao, M., Futoma, J., Ratliff, W., Nichols, M., Bedoya, A., Balu, S., O'Brien, C.: "The Human Body Is a Black Box" supporting clinical decision-making with deep learning. In: FAT* '20 : Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. pp. 99–109 (2020). https://doi.org/https://doi.org/10.1145/3351095.3372827
22. Shavlik, J.W., Dietterich, T.G.: Readings in Machine Learning. Morgan Kaufmann (1990)
23. Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., Hassabis, D.: Mastering the game of Go without human knowledge. Nature **550**(7676), 354–359 (Oct 2017). https://doi.org/10.1038/nature24270
24. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is All you Need. In: Advances in Neural Information Processing Systems. vol. 30. Curran Associates, Inc. (2017). https://doi.org/https://doi.org/10.48550/arXiv.1706.03762
25. Weber, B.: Computer Defeats Kasparov, Stunning the Chess Experts. The New York Times (May 1997), https://www.nytimes.com/1997/05/05/nyregion/computer-defeats-kasparov-stunning-the-chess-experts.html

26. Zając, H.D., Li, D., Dai, X., Carlsen, J.F., Kensing, F., Andersen, T.O.: Clinician-Facing AI in the Wild: Taking Stock of the Sociotechnical Challenges and Opportunities for HCI. ACM Transactions on Computer-Human Interaction **30**(2), 1–39 (Apr 2023). https://doi.org/10.1145/3582430