Review of: "Collaborative Intelligence: A scoping review of current applications"

Richard Harper[1]

1 Lancaster University

Potential competing interests: No potential competing interests to declare.

The 'Ironies of AI' - a review of 'Collaborative Intelligence: a scoping review of current applications' Schleiger, E., et al.

The paper sets out to investigate where AI and humans can complement each other. The paper explains that when collaborating AI and humans can do more. The paper argues that the activities constitutive of AI human collaboration are those that have these three features:

'shared objectives';

an agreed 'task' reacted to dynamically (i.e., as the situation alters or changes);

and where the joint efforts of the machine and human lead to 'improvement in performance/novelty/productivity and /or quality' of the output.

The paper goes on to argue that these features also need to go alongside a situation where the AI and the human can be thought of as 'agents', though what is meant by this is somewhat unclear. This term 'agent' is nevertheless important as it points towards the notion that intelligence is at play; agency *implies* intelligence. It is this that is at the heart of the paper - an indirectly conceived notion of what is behind or expressed in the acts that AI and humans do; their *agential intelligence*.

The paper summarises the kinds of collaborative systems with this 'agentiality' reported in the literature. Most of these are prototypes, with some being robots, others combining virtual systems with the human. Given the transformation in experience and functionality that large language models (LLMs such as found in Chat-GPT) have recently created, it is disappointing that these are not mentioned in the paper. Of course, delays on reporting the use of LLMs may be expected.

However, I am not sure what I learn from the paper - I feel as if it doesn't provide an insight that justifies the endeavour reported. I learn about numbers of papers reporting different types of AI-human activities but not much more. It makes for a somewhat frustrating read.

Part of the problem here, I think, is that the paper is not careful enough with terms and in the case of AI-human collaborations, being precise about terms isn't just a matter of clarity. Sorting out meanings can lead to insight.

For example, it seems to me one needs care when distinguishing types of AI-human collaborations. The deployment of what one might suggest are tacit or assumed meanings that corral types one way or another can be a little risky. This applies here with the use of the term intelligence and its connection with agency.

In the general literature and public discourse, the term intelligence is sometimes used in ways that doesn't seem apposite, or at least not when it is used alongside agency. A person might have intelligence but doesn't always act intelligently, after all; their agency does not always result in intelligent acts. Similarly, a machine that happens to incorporate some techniques that are labelled 'AI' (or machine learning, an equally common term these days) doesn't always produce outcomes that are in themselves always intelligent. In this paper, the things that are reported and which the authors want to claim constitute AI human collaboration are activities that by their own definition (or rather the definition they take from the literature), do not, in my view, adequately define or constraint the activities to situations where the term intelligence applies. To be sure, the activities require 'shared objectives', an agreed 'task' reacted to dynamically, and improvements in performance/novelty/productivity and /or quality of the output. But one doesn't need intelligence in either a machine or a person to pass these three criteria in every case. And so too in the papers reported. Hence, I am sure what I learn. Is the categorization good? Misleading? Highlighting important distinctions?

Nevertheless, What these criteria do bring to mind is human factors research and for this I am grateful. In the view from this discipline (human factors), there is no assumption that machines are intelligent, only that they can process more effectively that humans. But even so research does say interesting things and this leads me to reflect that *Schleiger et al* might have missed an opportunity to say something interesting themselves.

Human factors research shows there is a tipping point beyond which the distribution of tasks between person and machine reaches an imbalance. A machine can take on too much, and this can result in a situation where the person cannot take on tasks as and when needed, being effectively *de-skilled* by the role of the machine. Given that there is always a likely hood that a person will need to take on some of the machine's role as machines can be guaranteed to break down, this is something to take seriously.

It seems to me that this 'rule' will apply to AI machines too. It is almost certain that at some point a person will have to act as improvised solvers for the unreliability or failure of AI machines. What I would like to know is how critical these problems might be in the different contexts in which AI systems are now being found? This would matter greatly in safety critical situations, less so in simple administrative tasks. Even so, does the fact that some AI systems function like black boxes make this a greater problem? How do organisations plan for AI failure? And so on.

I am not saying that this is what *Shleiger et al* should have looked at so much as I am wanting to suggest that reviews of how technology are used can gain value when they offer insight about problems and dilemmas that are showing themselves, or are likely to, such as this problem of over-automating that human factors research identified. Could there be something similar as regards AI?

As it happens, the topic of how to manage the balance of processing powers between people and machines was beautifully explored by the human factors researcher, Lisanne Bainbridge, in her paper, *The Ironies of automation*. It is

the most cited ergonomics paper of all time. Perhaps *Shleiger et al* should aim equally high? Why not? The field of AI/human collaboration is calling out for it - a paper on the *Ironies of AI.*

Bainbridge, Lisanne (1983-11-01). "Ironies of automation". Automatica. 19 (6): 775–779. doi:10.1016/0005-1098(83)90046-8.