# Human-Centred AI: The New Zeitgeist

Yvonne Rogers, UCLIC, University College London

Hancock's article on avoiding autonomous agent actions reads like a literary piece of work replete with esoteric phrasing and the odd bit of Latin thrown into the mix. It is cleverly crafted prose, with poignant metaphors warning us of a foreboding future of malfunctioning, dysfunctioning and failing autonomous machines that could wreak havoc in society and even destroy us if we don't do something about them sharp. One of the most dramatic analogies is to cast autonomous systems like the evolution of a ring of volcanic islands rising from the ocean; abrupt and explosive rather than a slow gradual evolution. In contrast, we humans are viewed as taking the "littoral role of beaches and riparian shorelines", receding as the volcanos spew forth.  The oceanic imagery conjured up certainly does paint an apocalyptic future. I imagined hearing the soundtrack of War of the Worlds as I read it.

So, what can and should we do about the predicted sudden eruption of autonomous machines before it is too late? Hancock suggests spending billions of dollars on training up a new cadre (sic) of machine forensic psychologists who would be able to see into the brains of our machines more clearly than we can now and understand better why they choose particular actions. All well and good and a necessary call to arms.

But this is already beginning to happen in the broader field of AI – albeit at not quite a grandiose scale. There is a growing body of Human-Computer Interaction (HCI) researchers, computer scientists, philosophers and psychologists addressing and confronting AI's (sic) autonomy – working out how to replace it with alternative transparent algorithms that will allow humans (and other machines) to inspect, understand and put to right the machine learning and machine decision-making the algorithms are programmed to execute. Making them fairer, accountable, explainable and unbiased has become universally accepted goals. There are numerous frameworks, white papers and policies that have been published on how to achieve this. For example, earlier this year the EU published its Regulation that amongst its detailed pages of guidance, rules and restrictions, recommends **banning** AI systems that cause or are likely to cause "physical or psychological" harm to people. These includes autonomous systems that recognize people's faces and facial expressions without their knowledge; automatically deciding whether to permit or not them loans, credit, jobs and so on. Katharine Miller (2021) from Stanford's pioneering HAI centre has just published an article on the future of work, where she argues for replacing ideas of AI-driven automation with alternative values that encourage a more human-centered workplace.

What surprised me most about Hancock's latest piece against autonomous tech – given it was submitted to the Journal of Human-Computer Interaction – was the lack of coverage about what kinds of control, interactions and interfaces should we be considering and designing for technology, if we assume we don't want it to be fully autonomous. There was one sentence midway through the piece that claimed we are *"witnessing a crucial watershed in human-machine interaction and teaming."* But nothing further about what these are.

Instead, for most of the article, Hancock conducts a lengthy SWOT analysis about what would happen if we do nothing, do something or try our best to avert an impending catastrophe should the military and the arms industry be left to their own devices. But if he had looked more at what we are doing in our field he would see that there is currently a burgeoning debate about human-centred AI (HAI). By this we mean designing AI systems that enhance human capacities and improve their experiences rather than replacing them through automation. It is in the DNA of HCI to suggest, warn and advise about how to design and build any kind of system (sic) that humans interact with or are threatened by.

Change is already afoot. Governments are taking heed. As are tech companies, local groups and research activists. Furthermore, Ben Shneiderman has set up this year a website that contains a diversity of resources for Human-Centered AI, including research groups, organizations, events, courses and tools. It has over 1500 researchers signed up to it and the numbers are increasing. The goal is to support the growing community of those who promote and work on HAI. I am also part of a large European research network on humane-AI, comprising over 50 partners. Our aim is again to develop *“trustworthy, ethical AI that enhances human capabilities and empowers citizens and society to effectively deal with the challenges of an interconnected globalized world.”*

Rather than fixate on the threat autonomous systems impose on society we explore how to *develop* new AI systems that humans can work, create or solve problems with. We are also beginning to build new HAI interfaces, and evaluating novel kinds of multimodal interactions to enable humans to have a better understanding and be empowered by such systems. For example, there has been much interest in how to design chatbots, agents and robots that we can collaborate with, that can make proactive suggestions to us, and so on. The aim is to create systems (some autonomous, some semi-autonomous, and some that are completely under our control) that can augment human cognition.

Hence, rather than revisiting longstanding dystopian (or utopian) visions of AI and autonomy, we are reimagining human-machine interaction in all its guises. We are an eclectic and relatively nascent discipline full of inclusive voices. We listen, we act, we do. Our work is exciting, enabling and empowering. That is not to say that we are also only too aware of the dark side of tech (Rogers et al, 2021), the creeping creepiness of AI apps, the fear of self-driving cars losing control and so on. We have emerging frameworks and research programmes in place to address these worrying concerns. Our involvement in the creation and critique of the design of AI technologies demonstrates how society can benefit from having many kinds of human-machine interaction at our fingertips rather than focussing on the consequences of a seismic shift in machine autonomy. Let’s look on the bright side.

**References**

Human-Centered AI website: https://hcai.site/

Humane-AI net: https://www.humane-ai.eu/

Katharine Miller (2021) Future of Work: Beyond Bossware and Job-Killing Robots https://hai.stanford.edu/news/future-work-beyond-bossware-and-job-killing-robots

Yvonne Rogers, Margot Brereton, Paul Dourish, Jodi Forlizzi, and Patrick Olivier (2021) The Dark Side of Interaction Design. Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, Article 152, 1–2. DOI: https://doi.org/10.1145/3411763.3450397

Regulation of The European Parliament and of The Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206