Artificial Intelligence and Humanity's Future:
A Ghost Story

We live in a world populated by digital ghosts of human injustice and cruelty, reanimated in AI software. But is it possible to tell a different, more hopeful story? By Shannon Vallor

Many wonder how the emergence of artificially intelligent machines will impact humanity over the next century – are humans on the verge of being replaced by machine intelligence, or rendered irrelevant, as some have speculated? Are we witnessing in machine agency the birth of a new, alien form of life that we cannot hope to understand? Are we fated to compete with machines for dominance, or to merge with them, as Elon Musk has stated?

These are natural questions to ask for beings that, like most animals, have deep evolutionary reasons to pay attention to anything new in our environment that might be an agent. Agents change things – they take action for purposes that appear to be their own. Typically we look for agency in living things, but not always. We may struggle to see the agency of a mushroom or oak tree, but it can take effort for humans not to see agency in a simmering and unpredictable volcano or a rapacious wildfire. Or a problem-solving machine.

AI includes a broad range of new computing technologies, many components of which have no apparent agency at all. But because AI can be a powerful tool for creating artificial machine agents – from Alexa and Siri to social robots that greet us in hotels, to Internet bots that pose as aggrieved or enthusiastic voters, to generative language models that write stories for us – it's natural to begin to think of AI as a new kind of agency that might radically challenge or overtake our own.

The reality, however, is far more prosaic yet no less challenging for humanity to confront. The challenge is not a new form of life entering our world; AI is not an alien consciousness that asks us to meet, understand and negotiate with it. The challenge is of a wholly different kind, that of living in a world that we have begun to populate with an ever-expanding and replicating army of digital ghosts – ghosts of ourselves.

AI agency = our agency

AI agency is and will remain for the foreseeable future, our agency – only externalised, altered, extended, embedded and embodied in a multitude of new and sometimes surprising forms, virtual and physical. AI systems today – those driven by techniques known as machine learning – work by being fed mountains of human-generated data: records of human movements, observations, measurements, utterances, categories, choices and preferences. The data is fed into highly complex mathematical matrices designed (by humans) to extract patterns and correlations that we can turn into new insights and predictions – or in the case of artificial agents, to generate a range of machine actions that we find useful or interesting.

Notice that even these new forms of agency – the purposeful actions that AI generates – are still ours, as they are constituted from our human ways of seeing, sorting, labelling, and moving in the world. Even when an AI system process generates a surprising new behaviour that serves our aims, it is humans who classify this as an achievement to be enabled rather than an error to be ignored or suppressed, i.e., as an intelligent signal rather than meaningless statistical noise.

Many would rather see in AI today what science fiction has always imagined – machine minds that allow us to view the world through new eyes, the material satisfaction of our desire to encounter alien forms of intelligent life that might show us purposes beyond our own. While this desire can perhaps be fulfilled by more properly recognising other forms of intelligent life on Earth, it would be a profound error to think of AI in the way we rightly think of whales, crows, cephalopods, apes and elephants.

Electronic ghosts

It is far better to think of AI as ghosts. Ghosts, as traditionally imagined, do not point us to alien or inhuman possibilities. Rather, ghosts in art and literature represent our need to more fully reckon with ourselves, our relationships to one another, and the unresolved legacies in our past. From Toni Morrison's novel Beloved to Henry James' The Turn of the Screw, ghosts represent injustices unacknowledged, wounds unhealed, secrets untold, crimes unforgiven, loves unfulfilled, promises unkept.

When AI systems replicate and expose, as they so often do, the powerful patterns of human exclusion, discrimination and cruelty embedded in our own data, we are not seeing a machine spontaneously form racist or sexist or ableist intentions. We are seeing electronic ghosts of our own injustice and cruelty, reanimated in software. When AI facial recognition systems refuse to see black faces, when commercial computer vision AI systems can't be trusted to fairly apply gender labels, and when natural language-processing chatbots spout sexist or genocidal sentiments, we need to understand what is happening. We are not being oppressed by a new, inhuman mechanical evil. We are being haunted by our own ghosts.

Fortunately, ghost stories aren't always stuck in the past, and neither are we. In works such as the film Beetlejuice, ghosts also represent an expansion of possibilities. Ghosts explore the new things we might do with access to virtual bodies and spaces. They reveal the harms that we could confront, rectify and repudiate in a new phase of our existence. They suggest the richer aesthetic, moral and spiritual values we might find in a liminal dimension that allows us to see just a bit further than our own.

The future of AI could, if we choose, be this kind of ghost story. It could be the story of humans that unwittingly create, in the liminal space of software, a relentless army of angry ghosts that keep haunting us with magnified visions of our past cruelties and unkept promises – until we finally reckon more fully with ourselves and our institutions, with what we have been, with what we have failed to be, and with what we can finally be free to become. That's the kind of ghost story I like to read. Maybe it's one that AI can help us to write together.

——————

References:

1. https://www.axios.com/2019/07/18/elon-musk-neuralink-ai-merge
2. https://dl.acm.org/doi/10.1145/3369457.3370913 https://www.theatlantic.com/technology/archive/2020/01/future-politics-bots-drowning-out-humans/604489/
3. https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3
4. https://www.theguardian.com/books/2006/jul/08/fiction.tonimorrison
5. https://www.sparknotes.com/lit/screw/summary/ https://www.wired.com/story/best-algorithms-struggle-recognize-black-faces-equally/
6. https://venturebeat.com/business/researchers-discover-evidence-of-gender-bias-in-major-computer-vision-apis/
7. https://www.technologyreview.com/2020/10/23/1011116/chatbot-gpt3-openai-facebook-google-safety-fix-racist-sexist-language-ai/
8. https://www.imdb.com/title/tt0094721/