What if we chose new metaphors for artificial intelligence?

When we talk about artificial intelligence (AI), we often use metaphors. Even the term 'AI' relies upon a metaphor for the human quality of intelligence, and its development is regularly described as a 'race'. While metaphors are useful in highlighting some features of their subject, they can be so powerful that it becomes difficult to imagine or discuss their subject in other terms. Here, we examine some challenges presented by the central metaphor of 'intelligence', and whether metaphors for AI and its development emphasise competition at the cost of cooperation. Perhaps new metaphors could help us to articulate ambitious visions for AI, and new criteria for success.

Metaphors play a remarkable role in human history. They provide useful shortcuts to help us understand complex concepts, as well as powerful images of the world and how it could or should be in future. Whether unintentionally framing subjects or deliberately mobilising arguments, metaphors open some ways of thinking while closing others down. So, while they are an integral part of language and communication, specific choices of metaphors are worth reflection and care in how they are used.

AI is an umbrella term which refers to a wide range of technologies. It includes 'expert systems' – whereby humans encode their own knowledge and experience into rules – as well as 'machine learning' systems that identify patterns in data to generate rules by themselves. Discussions of AI are replete with metaphors for both the technology and its development.

© bizvector / AdobeStock

Potential impacts and developments

The choice of the term 'intelligence' is a legacy of early scholarship and ambitions in the discipline. However, it poses some enduring difficulties for the definition of the technology. First, since human intelligence is itself a subjective and contested concept, the concept of AI is also destined for constant debate and reinterpretation. Second, by defining AI with reference to how we evaluate its apparent performance (intelligence), rather than by what it does (applications) or how it does it (techniques), AI can refer to almost any technology – from thermostats to 'terminators' – whether they exist or not. And third, since the various AI applications have such a diverse range of impacts, using the same word to refer to all of them can amplify the appearance of disagreements and make debates less productive. Technologists have long recognised these limitations, and tend to prefer more precise alternatives such as 'machine learning' or, simply, 'statistics'. Nonetheless, AI retains its usage in public and policy settings.

Metaphors have linked minds and machines for centuries. From hydraulics to telegraphs and computers, we have conceptualised the brain with reference to the key technologies of our times. Today, we also reverse these metaphors to explain technologies in terms of human functions. For example, 'artificial neural networks' (ANNs) invoke the neural networks in our brains. The metaphors of machine 'vision', 'learning', 'recognition', and 'understanding' suggest that machines fulfil the same functions as humans, in the same kind of way. While this is misleading, the comparison is so well established that, since the Turing test, AI advancement has been continually measured and evaluated against human performance of the same tasks. Contemporary assessments and ambitions for AI tend to focus on trustworthiness and trust which, as metaphors for the qualities of the AI and our relationship with it, can anthropomorphise the technology and divert accountability from those responsible for its use when something goes wrong.

EN

Neuroscientists are concerned that metaphors reduce our brains to the status of computers, and make it difficult to imagine other – perhaps better – conceptualisations of what a technology does, and how. Likewise, metaphoric thinking elevates our software to the status of our minds. This poses several risks for AI development. First, it might tempt us to over-estimate the capabilities of AI tools and entrust them with tasks that they are not competent to perform. Indeed, this is the danger at the heart of many of the highest-risk AI applications. Second, when something goes wrong, we might be tempted, as alluded to above, to assign fault to the machine itself, rather than the people and organisations that inappropriately delegate tasks to them. Third, and perhaps most importantly, by reinforcing the idea of equivalence between what humans and computers do and how they do them, we position them in competition to perform the same kinds of tasks, rather than in cooperation to perform complementary tasks. The engrained language of AI as doing things 'like humans' imposes a potent conceptualisation for our future relationship with machines. It shapes how we articulate our ambitions, prioritise our development paths, and evaluate our progress.

We also find several powerful metaphors in debates about the international dimensions of AI development. Perhaps the most prominent is that of the 'global AI race', often positioning the EU as struggling for a bronze medal behind the USA and China. This provides an intuitive framing for AI development at global scale. However, a race implies a single 'finish line', which fails to capture that AI is a range of technologies and applications used by actors with different strengths, priorities and values. In turn, a single 'finish line' implies a single 'winner' of a zero sum game in which those that did not win must have lost. In doing so, the race metaphor emphasises competition over cooperation, sharing and mutual benefits. It may compel us to follow the direction and pace of those we consider to be in front, rather than following our own path. A more specific version of this metaphor invokes an 'AI arms race', a framing which has been criticised for closing down debate and transforming investments in militarised AI from options into necessities.

Finally, metaphors are also used to refer to positions in the AI debate. For example, the 'terminator' metaphor often serves to frame concerns about AI development as unjustified fears of fictional technologies that reveal a misunderstanding of its capabilities. However, studies of Europeans' attitudes towards robotics and AI show that citizens overwhelmingly associate robots with production-line machines, and not humanoid forms like the terminator. Indeed, respondents were broadly supportive of AI while expressing some concrete concerns about today's applications, notably their impacts on employment. While the terminator metaphor misrepresents how people make sense of AI, it serves as a powerful metaphor for public perspectives that may undermine questions about the concrete impacts of today's AI.

Anticipatory policy-making

The definition of AI as 'systems that display intelligent behaviour' – as used in the 2018 European Commission communication AI for Europe – would be too ambiguous for a legal text. Notably, the recent AI Act was more precise, defining AI not by apparent intelligence, but with reference to specific techniques such as machine learning, expert systems, and statistics. That such a diverse range of tools came to be bundled together in a 'tech-specific' legislative proposal is testament to the power of the AI metaphor in policy.

In debates about AI, we could follow the approach of many developers in referring to specific techniques and application contexts. For example, if we target messages towards 'machine learning diagnostic support' or 'biometric identification in public spaces' rather than just 'AI', we might reveal more common ground between our positions. Furthermore, by reducing our reliance on core metaphors such as 'intelligence' and 'trust', which allude to human qualities and capabilities, we could create space for new metaphors that describe AI in its own terms. In doing so, we would be better placed to articulate visions for the future and benchmarks for success, which focus on complementing rather than competing with humanity. Likewise, when talking about competition in global AI development, we could move on from the 'race' metaphor to speak of an 'AI Olympics' which celebrates a plurality of global achievements, or of 'moonshots' that invoke the Apollo project to inspire grand ambitions for the benefit of all humanity. While still capturing the notion of competition in global development, they would articulate a role for cooperation, sharing and mutual benefits. Ultimately, whichever metaphors we use, it is important that they articulate agency for Europe to define the direction and pace of its development, and the criteria for evaluating its success.