Attitudes towards Artificial Intelligence

A Artificial intelligence (AI) can already predict the future. Police forces are using it to map when and where crime is likely to occur. Doctors can use it to predict when a patient is most likely to have a heart attack or stroke. Researchers are even trying to give AI imagination so it can plan for unexpected consequences.

Many decisions in our lives require a good forecast, and AI is almost always better at forecasting than we are. Yet for all these technological advances, we still seem to deeply lack confidence in AI predictions. Recent cases show that people don't like relying on AI and prefer to trust human experts, even if these experts are wrong.

If we want AI to really benefit people, we need to find a way to get people to trust it. To do that, we need to understand why people are so reluctant to trust AI in the first place.

B Take the case of Watson for Oncology, one of technology giant IBM's supercomputer programs. Their attempt to promote this program to cancer doctors was a PR disaster. The AI promised to deliver top-quality recommendations on the treatment of 12 cancers that accounted for 80% of the world's cases. But when doctors first interacted with Watson, they found themselves in a rather difficult situation. On the one hand, if Watson provided guidance about a treatment that coincided with their own opinions, physicians did not see much point in Watson's recommendations. The supercomputer was simply telling them what they already knew, and these recommendations did not change the actual treatment.

On the other hand, if Watson generated a recommendation that contradicted the experts' opinion, doctors would typically conclude that Watson wasn't competent. And the machine wouldn't be able to explain why its treatment was plausible because its machine-learning algorithms were simply too complex to be fully understood by humans. Consequently, this has caused even more suspicion and disbelief, leading many doctors to ignore the seemingly outlandish AI recommendations and stick to their own expertise.

C This is just one example of people's lack of confidence in AI and their reluctance to accept what AI has to offer. Trust in other people is often based on our understanding of how others think and having experience of their reliability. This helps create a psychological feeling of safety. AI, on the other hand, is still fairly new and unfamiliar to most people. Even if it can be technically explained (and that's not always the case), AI's decision-making process is usually too difficult for most people to comprehend. And interacting with something we don't understand can cause anxiety and give us a sense that we're losing control.

Many people are also simply not familiar with many instances of AI actually working, because it often happens in the background. Instead, they are acutely aware of instances where AI goes wrong. Embarrassing AI failures receive a disproportionate amount of media attention, emphasising the message that we cannot rely on technology. Machine learning is not foolproof, in part because the humans who design it aren't.

D   Feelings about AI run deep. In a recent experiment, people from a range of backgrounds were given various sci-fi films about AI to watch and then asked questions about automation in everyday life. It was found that, regardless of whether the film they watched depicted AI in a positive or negative light, simply watching a cinematic vision of our technological future polarised the participants' attitudes. Optimists became more extreme in their enthusiasm for AI and sceptics became even more guarded.

This suggests people use relevant evidence about AI in a biased manner to support their existing attitudes, a deep-rooted human tendency known as "confirmation bias". As AI is represented more and more in media and entertainment, it could lead to a society split between those who benefit from AI and those who reject it. More pertinently, refusing to accept the advantages offered by AI could place a large group of people at a serious disadvantage.

E   Fortunately, we already have some ideas about how to improve trust in AI. Simply having previous experience with AI can significantly improve people's opinions about the technology, as was found in the study mentioned above. Evidence also suggests the more you use other technologies such as the internet, the more you trust them.

Another solution may be to reveal more about the algorithms which AI uses and the purposes they serve. Several high-profile social media companies and online marketplaces already release transparency reports about government requests and surveillance disclosures. A similar practice for AI could help people have a better understanding of the way algorithmic decisions are made.

F   Research suggests that allowing people some control over AI decision-making could also improve trust and enable AI to learn from human experience. For example, one study showed that when people were allowed the freedom to slightly modify an algorithm, they felt more satisfied with its decisions, more likely to believe it was superior and more likely to use it in the future.

We don't need to understand the intricate inner workings of AI systems, but if people are given a degree of responsibility for how they are implemented, they will be more willing to accept AI into their lives.