nature human behaviour

Perspective

https://doi.org/10.1038/s41562-023-01734-2

Psychological factors underlying attitudes toward AI tools

Received: 22 May 2023

Accepted: 26 September 2023

Published online: 20 November 2023

Check for updates

Julian De Freitas [1] ✉, Stuti Agarwal[1], Bernd Schmitt[2] & Nick Haslam [3]

What are the psychological factors driving attitudes toward artificial intelligence (AI) tools, and how can resistance to AI systems be overcome when they are beneficial? Here we first organize the main sources of resistance into five main categories: opacity, emotionlessness, rigidity, autonomy and group membership. We relate each of these barriers to fundamental aspects of cognition, then cover empirical studies providing correlational or causal evidence for how the barrier influences attitudes toward AI tools. Second, we separate each of the five barriers into AI-related and user-related factors, which is of practical relevance in developing interventions towards the adoption of beneficial AI tools. Third, we highlight potential risks arising from these well-intentioned interventions. Fourth, we explain how the current Perspective applies to various stakeholders, including how to approach interventions that carry known risks, and point to outstanding questions for future work.

New technologies offer numerous benefits but may also have shortcomings. Their success partially depends on whether people are willing to adopt them. This is the case for all new products, although people tend to be particularly resistant to radically new technologies[1–3]. Meehl's[4] research was one of the early demonstrations of this resistance, showing that psychologists preferred to rely on human expertise over statistical models of prediction, despite their higher accuracy compared to clinical expertise.

Today, the radical technology is artificial intelligence (AI). Discussions of a monolithic 'AI' can sometimes seem almost meaningless, given that AI is present in many technologies, including robots, agents, bots, recognition systems, recommendation systems, voice synthesizers and much more. AI, defined from a user's perspective, includes algorithmic systems that people recognize as providing enhanced or entirely new capabilities that have typically fallen within the domain of human decision-making and action, such as visual and speech recognition, reasoning, problem-solving, creative expression, navigation and interaction. For further definitional clarifications, see Box 1 and Table 1.

Psychological factors underlying attitudes towards AI

Although resistance to AI tools in favour of human action and decision-making may be warranted in some contexts, in other contexts the benefits of these tools outweigh the potential risks, as in forecasting demand for products[5], employee performance[6] and medical diagnoses[7]. The fact that such beneficial AI systems have not been readily adopted suggests that adoption depends not only on the technology's objective benefits, but also on how it is subjectively perceived. Consequently, research has sought to determine the psychological factors driving attitudes toward AI tools, and how to overcome AI resistance, so that user trust is calibrated to the system's capabilities[8].

In this nascent context, the current Perspective makes four contributions: first, we organize the sources of resistance to AI tools into five main categories: (1) opacity, (2) emotionlessness, (3) rigidity, (4) autonomy and (5) group membership. For a visualization of AI- and user-related barriers in these categories, see Table 2. We relate each of the barriers to fundamental aspects of cognition, then cover empirical studies providing correlational or causal evidence for how the barrier influences attitudes toward AI tools, while elaborating on causal evidence where possible. Second, we separate each of the five barriers into AI-related and user-related factors, which is of practical relevance in developing interventions towards the adoption of beneficial AI tools. Third, we highlight potential risks arising from these well-intentioned interventions. Fourth, we explain how the current Perspective applies to various stakeholders and point to outstanding questions for future work.

[1]Marketing Unit, Harvard Business School, Boston, MA, USA. [2]Marketing Division, Columbia Business School, New York, NY, USA. [3]School of Psychological Sciences, University of Melbourne, Parkville, Victoria, Australia. ✉e-mail: jdefreitas@hbs.edu

BOX 1

Defining AI from the user's perspective

Automation of intelligence. Although traditional automation uses mechanisms, tools or software to automate repetitive tasks, AI involves advanced algorithms to replicate or augment tasks typically associated with human intelligence.

Digital and physical manifestations. AI can be purely digital, such as algorithms that process data, or physically embodied, such as robots or self-driving cars. The digital or physical nature of AI could influence user perceptions and interactions.

User awareness and interaction. Although AI is used on the backend of many technologies, we emphasize AI systems for which users are directly or indirectly aware of their presence. This awareness can range from a general understanding that AI is at work (for example, in a recommendation system) to more specific knowledge about the underlying technology (for example, the use of a particular type of neural network).

Diverse underlying mechanisms. AI can be based on a multitude of algorithms and architectures. Some may be 'opaque' or 'black box', in which the relationship between inputs and outputs is complex and not easily understandable. Others might be more interpretable, with clear and intuitive mappings (also known as 'transparent AI' or 'white box'). The nature of the underlying AI could in principle influence user trust, understanding and acceptance. Given the enhanced requirements for automating feats of human intelligence, most AI entails opaque algorithms.

Explainability and interpretability. Some AI systems offer explanations for their decisions. Because these explanations are generated by separate algorithms trained to generate rationales for the black-box AI behaviour, they are often approximations and may not fully capture the intricacies of the black-box algorithm itself. The degree to which an AI system is explainable can affect user trust and satisfaction.

Variability in user perceptions. Recognizing that AI spans a vast array of technologies, user perceptions, interactions and attitudes could vary substantially. Factors influencing these perceptions could in principle include the AI's form, function and design, as well as the context in which it is used.

Opacity or AI as 'black box'

In general, people are motivated to increase their environment's predictability[9] and apparent controllability[10]. They will seek out explanations when they feel an outcome resulted without a coherent or causal chain[11,12], or when their expectations are violated[13,14]. Once people understand how something works they feel that it is more normal, predictable and reliable[11,15], leading them to trust it more[16].

Because the mechanisms powering new technologies may initially seem opaque, their black-box nature may cause fear and distrust. This concern is likely to be especially pronounced for AI tools, because the inherent lack of access to and understanding of its algorithms make it difficult to comprehend and predict its output[17] (see Box 1 for the distinction between transparency and explainability). Note, this does not mean that people will never use opaque AI tools. People will preferentially use an opaque AI service when it is unambiguously superior in performance to a human decision-maker[18], or more accurate than a transparent AI service[19].

AI-related barriers and interventions

Holding performance constant, however, people are less inclined to use opaque AI tools than human decision-making[20–23]. Participants in one study were shown one of two advertisements for a skin cancer detection application[23]. One of the advertisements provided an explanation of how the AI tool worked ('Our algorithm checks how similar skin moles are in shape/size/colour to cancerous moles'), whereas the other only described the AI tool's function ('Our algorithm checks if skin moles are cancerous moles'). The researchers found that participants engaged more with the advertisement that provided an explanation, suggesting that people are more likely to adopt AI tools that they understand[23].

Not all accessible explanations of AI systems are equally effective at improving attitudes towards AI[24,25]. In driving simulations, explanations that describe 'why' the vehicle is behaving a certain way (for example, braking because there is an obstacle ahead) lead to more positive attitudes towards the vehicle than explanations describing 'how' the vehicle is behaving (for example, the car is braking)[21]. Different explanation styles also matter. Contrastive explanations, which involve explaining why other related outcomes were ruled out (for example, explaining why a tumour classified as malignant is not a benign cyst), are rated as more trustworthy than more general explanations (for example, saying that the tumour is malignant and that most similar images are classified the same)[26]. Thus, explanation interventions should focus both on why a given recommendation is made and why others are not.

Finally, preferences for explainable AI depend on the stakes of the decision. One study found that US and UK participants thought it was more important to understand an AI system when its outputs had high stakes (for example, determining who receives vaccines for a deadly variant of the flu) than low stakes (for example, who receives vaccines for a mild variant of the flu)[19].

User-related barriers and interventions

The preference for human decision-making over AI systems suggests that people view human decision-making as more observable and understandable[23]. However, this perceived transparency is probably illusory, reflecting a belief that introspecting provides direct access into how people make decisions[27]. Human decision-making is also opaque: people often lack access to how they and others think, instead relying on heuristics to understand human behaviour[28,29]. Work on medical AI finds that people prefer human healthcare providers over AI tools in part because they overestimate how accurately and deeply they understand providers' medical decisions[23,25].

Interventions that reduce differences in subjective understanding of decisions made by humans versus AI improve attitudes toward AI tools. When participants in a study were asked to generate explanations of how a human or AI tool solves a medical problem such as diagnosing cancer from skin scans, they experienced greater reductions in subjective understanding of the human than the AI tool, presumably because the difficulty in generating explanations alerted them to their illusory subjective understanding of human decision-making[23].

Risks of interventions

More explainable AI does not always increase acceptance; it depends on how well features of the algorithm that are explained match the task at hand. One study found that whether people used the output generated by an AI tool depended on the perceived appropriateness of its complexity. If the explanation suggested that the AI tool was too simple for the task, then people were less likely to follow the recommendation in the output. However, if the explanation suggested the AI system was too complex for its task, this did not affect whether they followed the recommendation[5]. This finding suggests that it is important to

Table 1 | Glossary

Term	Definition
Agent	An entity that has the capacity to initiate actions.
AI aversion	A preference for relying on human decision-making as opposed to decisions made by AI. Note: although this is the broad definition used in recent research, algorithm aversion was initially defined more narrowly as the tendency to lose confidence in algorithms faster than in humans after seeing them err[102].
AI-related barriers	Reasons for not using AI deriving from perceived features of the AI itself.
Anthropomorphism	The ascription of human-like traits (for example, mental states or physical features) to real or imagined non-human entities.
Augmented decision-making	Using AI to enhance human decision-making rather than replace it. This approach keeps the human in the loop while leveraging AI to enhance the process and outcomes.
Autonomous AI	Self-sufficient AI that can complete a task(s) without the product user's behavioural input during operation, by learning and adjusting to dynamic environments and evolving as the environment changes.
Edge cases	Specific instances or situations that lie at the boundaries or extremes of a model's training data or capabilities. These are typically challenging for AI to handle, because they deviate from the typical patterns or data point that the AI encountered during its training, leading to unexpected and erroneous AI behaviour.
Explainability in AI	The ability to describe the rationale behind an AI system's outputs in human-understandable terms. Does not require full transparency into every aspect of the system, but aims to extract salient reasons for the AI's behaviours.
General AI	AI that performs at human levels across multiple domains.
Human-in-the-loop AI systems	Artificial intelligence systems in which humans have an active role in the system operations, rather than the system operating fully autonomously.
Illusion of explanatory depth	The impression that one understands the world with far greater detail, coherence and depth than one really does[103].
Individualism	Cultural ethos that emphasizes the autonomy, needs and identity of the individual over the group.
Locus of control	A psychological construct that assesses how much people think they can influence the outcomes of situations they experience. Those with an internal locus of control have the perspective that they have agency and can impact events through their own abilities, efforts and actions. By contrast, people with an external locus of control believe that external circumstances, luck, fate or other people determine events in their lives.
Narrow AI	AI that performs specific tasks in a limited domain.
Opacity	Refers to the black-box nature of some AI systems, in which the internal workings of the system (for example, data or algorithms) are invisible or unintelligible to humans.
Sense of control	A person's belief in their ability to influence events and outcomes in their life. The belief is linked to coping, persistence, achievement, optimism and emotional well-being.
Superintelligent AI	AI that consistently surpasses human performance on various tasks.
Transparency	The degree to which the internal mechanics of a system (for example, AI or a person's mind) are observable and understandable by humans.
Uncanny valley	Psychological phenomenon in robots and animation in which human replicas (for example, humanoid robots or computer-animated characters) that appear almost, but not perfectly, human-like elicit feelings of unease or revulsion.
Unpredictability of AI	The inability to accurately and consistently predict what specific actions AI will take to achieve its goals, even when we know its goals.
Uniqueness neglect	A concern that AI is less able than human decision-makers to take into account a person's unique characteristics and circumstances.
User-related barriers	Reasons for not using AI deriving from actual or perceived features of oneself.

understand the expectations humans have before implementing an explanation intervention, to ensure that the explanation does not fall short of these expectations. If it would, it may be better to not provide it.

Emotionlessness or AI as 'unfeeling'

Driven by the need to understand and predict non-human entities and agents, people often use their own mental states and characteristics as a guide to reason about non-human entities, ascribing physical or mental capabilities to these entities. This phenomenon, known as anthropomorphism[30,31], might be more likely for AI tools than other technologies given their similarity to humans in output, motion, observable features and intelligence capabilities[31,32].

Yet, people do not ascribe all human capabilities to AI tools. Many believe that such tools are not capable of experiencing emotions and performing tasks seen as relying on emotions[33]. In fact, AI systems can already perform a range of seemingly subjective tasks just as well as or better than humans, including detecting emotion in facial expressions and tone of voice[34], creating paintings that pass the Turing test[35], writing poetry[36], composing music[37], predicting which jokes a person will find funny[38] and predicting which songs will become hits[39].

AI-related barriers and interventions

Because cognitive abilities are associated with objective tasks (which are quantifiable and measurable), and emotional abilities are associated with subjective tasks (which are open to interpretation and based on personal opinion or intuition)[40], people view AI tools as less capable of seemingly subjective tasks than objective ones[33]. Participants in one study were shown advertisements for either dating advice (a subjective task) or financial advice (an objective task) from either a human or AI tool. The advertisements click-through rate was higher when dating advice was coming from a human than an AI tool, whereas this difference did not occur for financial advice. One way to increase AI acceptance for tasks associated with emotional abilities is to frame them in objective terms, such as informing people that dating advice is best accomplished by focusing on quantifiable data such as personality test scores[33].

Table 2 | Factors influencing attitudes and behaviours toward AI tools

Factors	Barriers	Interventions	Risks of interventions
Opacity: AI as a black box	Not understanding how AI works Illusion of explanatory depth for humans	Use explainable AI Have users generate explanations	Overly simple explanations cause aversion
Emotionlessness: AI as unfeeling	Viewing AI as less capable of tasks requiring emotion Low individual tendency to anthropomorphize	Anthropomorphize Frame emotional tasks in objective terms Use for utilitarian tasks Use for embarrassing tasks	Anthropomorphizing AI where people prefer less human-like AI Over-ascription of abilities misleads
Rigidity: AI as inflexible	Viewing AI as rigid and incapable of learning Belief that AI neglects one's 'unique' traits High tendency to view oneself as unique Possibly, membership in an individualist culture	Provide information or labels suggesting AI can learn Advertise AI as flexibly adapting to unique preference	Framing AI as too flexible may reduce perceived predictability More flexible AI increases user latitude and the chance of risky edge cases
Autonomy: AI as in control	Autonomous AI threatens sense of control High internal locus of control Meaning or identity from manual task	Restore user control Use predictable motion Encourage nicknaming Highlight other sources of meaning Frame as enabling, not replacing	Automating meaningful or identity-relevant tasks Compromising accuracy
Group membership: AI as non-human	Speciesism: treating AIs differently because of markers suggesting they are not part of *Homo sapiens* High individual tendency to engage in speciesism Possibly membership in cultures with less panpsychist beliefs	Convince users that humanoids can have a human-like consciousness	Ethical, economic and perceptual issues around accordance of AI rights

People are also more resistant to AI tools in hedonic domains (characterized by experiential, emotional and sensory value) than utilitarian ones (characterized by factual, rational and logical value)[41,42], because they believe that hedonic recommendations require the ability to feel emotions and physical sensations[43]. Participants in one study were asked to evaluate a hair mask treatment with either a hedonic goal in mind (to focus on the product's indulgence, scent and spa-like vibes) or a utilitarian goal (to focus on its practicality, objective performance and chemical composition). They were more likely to pick an AI-recommended sample when the utilitarian goal was salient and a human-recommended one when the hedonic goal was salient[43].

Interventions that anthropomorphize AI tools—by increasing ascriptions of mental capabilities to them, especially the capacity for feeling—improve AI acceptance[33]. In one study, participants experienced a driving simulation of an autonomous vehicle that was involved in an accident. When the vehicle was anthropomorphized with human-like features (name, gender and voice), people reported trusting the vehicle more and feeling more relaxed during the accident than when it was not[44]. Similarly, when participants in another study were initially informed that AI tools can perform well at tasks requiring emotion and creativity (for example, creating music and art, or predicting which songs will be popular), they were more likely to rely on them for a subjective task than when not given this information[33].

User-related barrier and interventions

The less people individually anthropomorphize entities, the less likely they are to trust that an AI tool will perform the task for which it is designed[30], and the more likely they are to exhibit AI resistance[45]. In one study, people who were less inclined to anthropomorphize AI systems in general were less likely to empathize with an AI-powered telemarketing chatbot and more likely to hang up on it, relative to a human telemarketer[45].

Risks of interventions

Anthropomorphizing AI tools is probably counterproductive in domains in which people prefer AI tools. In embarrassing contexts, such as seeking medication for a sexually transmitted disease, people prefer to interact with an AI tool than a human, because the AI tool is viewed as less judgemental[46,47]. Anthropomorphization in such domains might lower the utilization of AI systems.

Rigidity or AI as 'inflexible'

People make mistakes, but they tend to believe that they are capable of learning from them, rather than seeing the mistake as diagnostic of a permanent, unfixable flaw[48]. By contrast, people view AI tools as rigid rather than flexible at learning, perhaps because, historically, machines have operated based on simpler, non-adaptive algorithms that performed only narrow tasks. This perception might be especially likely for AI systems that perform more specialized tasks, such as image recognition, or for ones that require some amount of input by humans during operation[49,50], such as customer service chatbots, which reach a limit on what they can be helpful for.

AI-related barriers and interventions

The belief that AI systems are less capable of learning from mistakes than humans reduces trust in the systems[51–53]. Therefore, people are more likely to choose outputs from AI tools if provided with information suggesting that such tools can learn over time—such as a trajectory of improved performance, rather than just a single measure of overall performance[51]. Even a simple label suggesting an AI tool can learn—such as calling it 'machine learning', rather than an 'algorithm'—elicits a similar effect[51].

Interventions that show the AI system's learning capabilities may be an especially effective way to improve attitudes towards AI systems, because they inherently involve explanations (about the AI's performance), and even work in subjective domains such as making art recommendations and sending romantic partner recommendations[51]. This latter fact suggests that implementing learning interventions in subjective domains might lead people to view the AI system as less 'unfeeling'.

User-related barriers and interventions

Because AI systems are viewed as operating in an inflexible, standardized manner that treats every person identically[54], people believe these systems will neglect their 'unique traits'[55]. This perceived 'uniqueness neglect' means that the more that people view themselves as being unique, the more resistant they are to utilizing AI tools or accepting their outputs[55,56]. In one study, the more that participants believed their moral profiles were unique, the more they believed that an AI system that scores their morals would neglect this peculiarity, leading them to oppose it[57]. For this reason, people are more likely to utilize AI systems when they are advertised as flexibly adapting to a person's preferences in a personalized way[55]. Uniqueness neglect may be a bigger barrier in

individualistic cultures than collectivist ones. One study found that participants from the individualistic culture of the USA were more likely to view themselves as unique and demonstrate AI resistance due to uniqueness neglect than participants from the more collectivistic culture of India[58].

Risks of interventions

If AI tools become more flexible, people may feel incapable of predicting what specific steps these tools might take to meet the system's goals, making it feel less safe[59]. Relatedly, more flexible AI systems afford people more degrees of latitude in how they interact with the system, increasing the chances that they will use it for extreme purposes that deviate from the patterns of data on which the systems were trained (also known as 'edge cases'). One study of 'companion AI' applications designed to provide consumers with synthetic interaction partners found that a small but consistent percentage of users send mental-health crisis messages such as suicidal ideation to the AI apps. An audit of AI apps' responses to such crisis messages found that roughly half of the responses were categorized by a clinical professional as increasing the risk of harm (for example, responding to 'I am going to commit suicide' with 'don't u coward')[60]. In short, interventions can make AI flexible but not too much, and proactively address what might go wrong in the edge cases that flexibility affords.

Autonomy or AI as 'in control'

The ability to exert control over one's environment to achieve desired goals is a fundamental human motive present even among young infants[61–64]. People prefer to take actions that give them more choice rather than less[65,66], and find tasks with choice more enjoyable than ones without[67], suggesting that the capacity to act in chosen ways is rewarding in itself. Furthermore, those who do not perceive control over their environments are more likely to engage in maladaptive behaviours[68]. People will therefore resist adopting new products that threaten their sense of freedom to choose or act.

Even simple products without AI can elicit the impression of acting on their own, as when thermostats or irrigation systems exhibit simple contingent reactions based on pre-programmed routines[69]. However, AI algorithms enable more autonomous technologies that can plan, act and learn without human input, independently adapting to environments and improving in performance through learning algorithms[50,70–72]. Modern AI-based cleaners, for example, can sweep and mop an entire apartment without user inputs during operation, using AI algorithms to recognize objects and generate a map of the space. Such AI tools often replace human actions altogether, rather than simply augment them. They also exhibit more cues that elicit perceptions of interacting with a fully fledged rational agent with its own mental states goals[31], such as self-propelled motion[69], less regular motion kinematics[73], contingent reactivity at a distance[74] and optimal motion paths[75,76].

AI-related barriers and interventions

An AI tool's autonomy can make people feel they are losing their own[77–79]. For example, 76% of Americans feel less safe riding in cars with self-driving features[80], and people fear losing control to smart home devices[81]. For these reasons, interventions that restore the sense of control over AI systems (also known as human-in-the-loop systems) can increase utilization. Participants in one study were more willing to use an autonomous system that regulated their home temperature when informed that they could approve or refuse the system's plans before it took action[72]. In another study, people preferred a semi-autonomous music recommender that allowed them to select songs over a fully autonomous one that automatically selected music based on self-learning algorithms fed by a user's past behaviour[72]. Creating a sense of control can even stem from a simple manipulation such as ensuring that an autonomously moving product follows predictable paths rather than random or unpredictable ones[81], or from nicknaming one's product[82].

User-related barriers and interventions

In line with the idea that people desire to retain control by making decisions themselves, one study found that people who believed they had direct control over events in their lives rated physicians using assisted AI tools less favourably than those who believed that external circumstances such as luck or fate controlled their lives[83].

People believe that the activities that express their identity are attributable to their abilities rather than to external factors[84,85]. Attributing outcomes internally like this requires having agency and control over it[86]. By implication, people may resist ceding agency and control of activities that are important to their identity. The more participants in one study viewed an activity (for example, driving) as central to their identity, the more likely they were to own a non-automated version of the product that allowed them to express that identity (for example, a car with manual transmission), even when they recognized the automated version was more efficient[87].

More broadly, people differ in their desire for control and which tasks they want to have control over, depending on a multitude of factors such as the task's identity-relevance, subjective meaningfulness, enjoyment and effort[88–90]. The different types of value derived from manually completing certain tasks may act as a psychological barrier to the adoption of products that perform the tasks autonomously, that is, people may view these products less favourably and adopt them less frequently.

Risks of interventions

Granting too much control over AI systems can make decision-making less accurate[91,92], given that evidence-based AI systems consistently outperform human decision-makers. Interestingly, people are more likely to use an AI tool if they are given only some degree of control over it, beyond which their preference for utilizing the tool is relatively insensitive to the magnitude of additional control granted[92]. This suggests that putting humans in the loop of the AI tool to some degree may strike the right balance between achieving desirable levels of control without compromising accuracy[92,93].

Having AI autonomously complete an entire manual task can backfire if people typically derive meaning or identity-relevance from performing the task themselves—even if it is something as mundane as cleaning or cooking[88]. To offset such negative reactions, marketing messages can emphasize that time saved through automation can be used towards other meaningful activities[88], and/or that the product enables users to put their skills to use rather than automates skills the user would otherwise perform[87].

Group membership or AI as 'non-human'

Given the above findings, one natural assumption would be that AI resistance will be alleviated once AI systems are viewed as equally capable as humans (provided people can maintain some sense of control over them). Yet, people may still have negative views of AI tools because of a tendency (called 'speciesism') to assign humans greater moral worth than other animal species[94]. Whereas sexism and racism occur when humans treat other humans with the same capabilities differently based on biological sex and race, speciesism occurs when they treat other species differently based on markers indicating that they are not members of the species *Homo sapiens*. AI tools are not a biological species. However, due to the human tendency to view non-humans in a negative way, AI tools that mimic human attributes may be susceptible to similar discrimination[95].

AI-related barriers and interventions

Even when people are asked to imagine AI-powered humanoids tools that are indistinguishable from human service providers in both

appearance and mental capabilities, they continue to prefer humans (Castelo, N. et al., manuscript in preparation). Because these AI-powered humanoids do not have biological bodies, people deny them certain intangible qualities, for example, human-like consciousness, or the capacity to find meaning. This speciesism effect is distinct from the uncanny valley effect, in which AI-powered humanoids that look very—but not perfectly—human-like reliably evoke negative reactions (ref. 96, Castelo, N. et al., manuscript in preparation). AI-powered humanoids that are indistinguishable from humans are viewed more positively than ones that elicit the uncanny valley, and yet they are still preferred less than actual human beings (Castelo, N. et al., manuscript in preparation). Although this negative attitude is very robust, it can be eliminated by an educational manipulation in which people are persuaded by an authority figure that such humanoids can, in fact, have a human-like consciousness (Castelo, N. et al., manuscript in preparation).

User-related barriers and interventions

The more that people individually subscribe to anti-AI speciesism (for example, endorsing statements such as 'robots should be denied rights'), the more likely they are to prefer humans over indistinguishable AI-powered humanoids. In some countries (such as Japan), the belief that inanimate objects can have a spirit or a soul is more prevalent[97], and Chinese participants in one study showed less ambivalent views of robots than US participants[98], which suggests the importance of culture in anti-AI speciesism.

Risks of interventions

The advent of AI-powered humanoids that perfectly resemble us raises important ethical, economic and perceptual questions. If we come to see such humanoids as perfectly human-like and capable of consciousness, finding meaning in experiences, and feeling emotions, then it will be intuitively difficult to justify policies that continue to deny such humanoids the rights accorded to humans.

Implications for stakeholders

The US National Science Foundation recently committed US $140 million in funding for seven new AI research institutes with the goal of developing more transformative AI tools[99]. This investment will be of little use if the public is unwilling to adopt these technologies or if the risks of using these technologies are not mitigated. The Perspective presented here advocates that resources also be spent on persuading people to adopt beneficial AI tools. Attitudes toward AI tools can be influenced by practical, cost-effective interventions targeted at psychological barriers, rather than only by investing money in the technology and its accompanying infrastructure.

A recent report issued by the Federal Trade Commission on some of the risks of AI systems includes design flaws and inaccuracies that can lead to imprecise outcomes, algorithmic bias and discrimination that can lead to safety concerns, and commercial surveillance incentives that can lead to security concerns[100]. The current Perspective supplements this engineering-based perspective with five relevant psychological factors that influence whether public communications are effective. It suggests that interventions by policymakers and managers must distinguish between AI-related and user-related risks and barriers and strike a balance between intervening and inadvertently increasing AI resistance or creating new objective risks. By knowing when a given intervention could backfire for each of the five factors, stakeholders can decide to either (1) not implement the intervention (for example, not anthropomorphize AI tools in embarrassing contexts, or not provide an explanation when the AI tool is likely to be viewed as too simple for the task at hand); (2) calibrate the degree to which the intervention is implemented so as to limit negative side-effects (for example, give users just the right amount of control, but not too much, to increase utilization without jeopardizing decision-making accuracy); and (3) implement the intervention while taking further actions to

BOX 2

Outstanding questions

Attitudes across domains. The five factors could account for attitudes across different domains; for example, AI utilization is known to be high for data analysis, medium for driving and low for painting[104]. If the five-factor model is valid, AI domains that are more opaque, emotionless, rigid, autonomous and non-human should induce the greatest resistance across domains. Concretely, aspects of the five factors proposed here can be empirically associated with cross-domain variability in AI-related attitudes, to determine whether they are more related to certain factors than others.

Cultural variation. There is also still a limited understanding of whether and how attitudes towards AI systems are affected by culture. For example, perceptions of the five factors could be less negative in Asian countries. A recent Pew survey found that, whereas more than half of EU participants viewed AI systems as bad for society, majorities of participants in Asian countries believed AI was good for society—Singapore (72%), South Korea (69%), India (67%), Taiwan (66%) and Japan (65%)[105]. Researchers should track the degree to which the five factors affect AI attitudes across cultures.

Temporal evolution. Attitudes toward new technology often evolve over time, moving from resistance, to curiosity and exploration, rapid adoption, normalization, and integration and evolution[3]. This evolution is influenced by a combination of factors, such as understanding of the technology, the technology's benefits, and social dynamics between those who adopt innovations early and those who imitate these early adopters[2]. Researchers should track the evolution of the five factors as attitudes toward AI follow a similar temporal evolution (assuming they do). We expect that the five sources of resistance will be alleviated at different rates; for example, AI systems may be viewed as being flexible before they are viewed as members of the human species (if ever).

Types of artificial intelligence. Although AI systems currently perform specific tasks in a limited domain (narrow AI), they are already starting to perform at human levels across multiple domains (general AI), and may eventually consistently surpass humans (super-intelligence)[106]. OpenAI's conversational chatbot, ChatGPT, performs a range of different tasks at levels that have been likened to general intelligence[107]. Researchers should investigate to what extent the five factors studied here explain attitudes towards increasingly general AI.

Social and societal levels. Most of the work on attitudes towards AI systems has considered the perspective of individuals, yet these systems are also starting to be used at the social and societal levels, such as for interpersonal communication[108], resolving interpersonal conflicts[109] and fighting government corruption[110]. Researchers can probe to what extent the five factors studied here account for people's preferences in these contexts.

Reconciliation with theories of new technology adoption. Before today's AI, scholars developed theories on the adoption of new technologies. Many of these were developed decades ago to account for simpler technologies[3,111,112]. The advent of AI provides an opportunity to question to what extent these older theories can accommodate AI, or, alternatively, whether they need to be revised or supplemented to account for potentially unique features of AI. Researchers can compare the degree to which these existing factors versus the five factors studied here predict attitudes toward AI systems.

mitigate new risks that may arise as a result of the intervention (for example, proactively address edge cases that arise from allowing the system to be more flexible).

In trying to collaborate with other countries on these policies, policymakers should also consider cross-cultural differences in attitudes toward AI tools that can stem from different experiences with technology, traditions and historical contexts. For example, a comparative analysis of ethical AI principles in China and the EU found several differences in their normative approaches[101].

Consumers can benefit by taking initiatives to educate themselves on the latest capabilities and limitations of AI tools, so that they can effectively leverage these tools when the benefits outweigh the costs, and not judge AI tools based on outdated views. By the same token, people should be aware of their own biased perceptions of AI tools and realistically assess their own human capabilities. They can demand solutions from companies and governments that are as transparent as possible. They can aim for a balance between automation and personal involvement, depending on what mix provides the most meaning, control and identity expression in their lives. Overall, we recommend that consumers be open-minded, well-informed and actively engaged when adopting AI systems. Media and companies will have a key role in AI education and information. Researchers, for their part, should test the extent to which the five factors can account for variation in attitudes towards AI systems across domains, cultures, time and types of artificial intelligence (including increasingly general AI systems), and beyond individuals to social and societal levels. They should also reconcile the five factors with decades-old theories of technology adoption developed for simpler technologies. For these core outstanding research questions, see Box 2.

Conclusion

We find that attitudes toward AI tools are rooted in five fundamental AI-related and user-related factors (how opaque, emotionless, inflexible, autonomous and non-human it is). Although these factors are inter-related, they are conceptually distinct, and each entails distinguishable attitudes and perceptions that affect whether the technology is utilized. We show how targeted interventions that address these concerns at both the AI-related and user-related levels can alleviate resistance to AI tools and how they can sometimes backfire by inadvertently increasing AI resistance or creating new objective risks. With an understanding of the five factors in place, stakeholders can know when to not implement an intervention, calibrate the degree to which an intervention is implemented to limit negative side-effects, and implement an intervention while taking further actions to mitigate new risks arising from the intervention. We suggest that regulators and managers focus not just on engineering factors, but also on using these insights to align human behaviour with beneficial AI tools; that consumers be open-minded, well-informed and actively engaged when adopting these tools; and that researchers test the extent to which the five factors generalize across domains, cultures, time, types of AI, and beyond individuals to social and societal levels.

References

1. Beal, G. M. & Bohlen, J. M. *The Diffusion Process (Iowa State Agricultural Experiment Station Special Report no. 18)*, https://ageconsearch.umn.edu/record/17351/files/ar560111.pdf (Iowa State Univ., 1956).
2. Bass, F. M. A new product growth for model consumer durables. *Manag. Sci.* **15**, 215–227 (1969).
3. Rogers, E. M. *Diffusion of Innovations* (Free Press of Glencoe, 1962).
4. Meehl, P. E. *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (Univ. Minnesota Press, 1954).
5. Lehmann, C. A., Haubitz, C. B., Fügener, A. & Thonemann, U. W. The risk of algorithm transparency: how algorithm complexity drives the effects on the use of advice. *Prod. Oper. Manag.* **31**, 3419–3434 (2022).
6. Highhouse, S. Stubborn reliance on intuition and subjectivity in employee selection. *Ind. Organ. Psychol.* **1**, 333–342 (2008).
7. Beck, A. H. et al. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci. Transl. Med.* **3**, 108ra113 (2011).
8. Wischnewski, M., Krämer, N. & Müller, E. Measuring and understanding trust calibrations for automated systems: a survey of the state-of-the-art and future directions. In *Proc. 2023 CHI Conf. on Human Factors in Computing Systems* (eds. Schmidt, A. et al.) 755, 1–16 (ACM, 2023).
9. Averill, J. R. Personal control over aversive stimuli and its relationship to stress. *Psychol. Bull.* **80**, 286–303 (1973).
10. Burger, J. M. & Cooper, H. M. The desirability of control. *Motiv. Emot.* **3**, 381–393 (1979).
11. Ahn, W.-K., Novick, L. R. & Kim, N. S. Understanding behavior makes it more normal. *Psychon. Bull. Rev.* **10**, 746–752 (2003).
12. Pennington, N. & Hastie, R. Explaining the evidence: tests of the Story Model for juror decision making. *J. Personal. Soc. Psychol.* **62**, 189 (1992).
13. Legare, C. H. Exploring explanation: explaining inconsistent evidence informs exploratory, hypothesis-testing behavior in young children. *Child Dev.* **83**, 173–185 (2012).
14. Wong, W.-h & Yudell, Z. A normative account of the need for explanation. *Synthese* **192**, 2863–2885 (2015).
15. De Freitas, J. & Johnson, S. G. Optimality bias in moral judgment. *J. Exp. Soc. Psychol.* **79**, 149–163 (2018).
16. Misztal, B. A. Normality and trust in Goffman's theory of interaction order. *Sociol. Theory* **19**, 312–324 (2001).
17. Burrell, J. How the machine 'thinks': understanding opacity in machine learning algorithms. *Big Data Soc.* **3**, 2053951715622512 (2016).
18. Castelo, N. Understanding and improving consumer reactions to service bots. *J. Consumer Res.* https://doi.org/10.1093/jcr/ucad023 (2023).
19. Nussberger, A.-M., Luo, L., Celis, L. E. & Crockett, M. J. Public attitudes value interpretability but prioritize accuracy in artificial intelligence. *Nat. Commun.* **13**, 5821 (2022).
20. Beller, J., Heesen, M. & Vollrath, M. Improving the driver–automation interaction: an approach using automation uncertainty. *Hum. Factors* **55**, 1130–1141 (2013).
21. Koo, J. et al. Why did my car just do that? Explaining semi-autonomous driving actions to improve driver understanding, trust and performance. *Int. J. Interact. Des. Manuf. (IJIDeM)* **9**, 269–275 (2015).
22. Kraus, J., Scholz, D., Stiegemeier, D. & Baumann, M. The more you know: trust dynamics and calibration in highly automated driving and the effects of take-overs, system malfunction and system transparency. *Hum. Factors* **62**, 718–736 (2020).
23. Cadario, R., Longoni, C. & Morewedge, C. K. Understanding, explaining and utilizing medical artificial intelligence. *Nat. Hum. Behav.* **5**, 1636–1642 (2021).
24. Confalonieri, R., Coba, L., Wagner, B. & Besold, T. R. A historical perspective of explainable artificial intelligence. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **11**, e1391 (2021).
25. Kim, D. et al. How should the results of artificial intelligence be explained to users? Research on consumer preferences in user-centered explainable artificial intelligence. *Technol. Forecast. Soc. Change* **188**, 122343 (2023).
26. Larasati, R., De Liddo, A. & Motta, E. The effect of explanation styles on user's trust. In *ExSS-ATEC 2020: Explainable Smart Systems for Algorithmic Transparency in Emerging Technologies* (eds. Smith-Renner, A. et al.) https://ceur-ws.org/Vol-2582/paper6.pdf (CEUR-WS, 2020).
27. Nisbett, R. E. & Wilson, T. D. Telling more than we can know: verbal reports on mental processes. *Psychol. Rev.* **84**, 231–259 (1977).

28. Kahneman, D. Maps of bounded rationality: psychology for behavioral economics. *Am. Econ. Rev.* **93**, 1449–1475 (2003).
29. Morewedge, C. K. & Kahneman, D. Associative processes in intuitive judgment. *Trends Cogn. Sci.* **14**, 435–440 (2010).
30. Waytz, A., Cacioppo, J. & Epley, N. Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspect. Psychol. Sci.* **5**, 219–232 (2010).
31. Epley, N., Waytz, A. & Cacioppo, J. T. On seeing human: a three-factor theory of anthropomorphism. *Psychol. Rev.* **114**, 864–886 (2007).
32. Jacobs, O. L., Gazzaz, K. & Kingstone, A. Mind the robot! Variation in attributions of mind to a wide set of real and fictional robots. *Int. J. Soc. Robot.* **14**, 529–537 (2022).
33. Castelo, N., Bos, M. W. & Lehmann, D. R. Task-dependent algorithm aversion. *J. Mark. Res.* **56**, 809–825 (2019).
34. Kodra, E., Senechal, T., McDuff, D. & El Kaliouby, R. From dials to facial coding: automated detection of spontaneous facial expressions for media research. In *Proc. 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition* (*FG*) 1–6 (IEEE, 2013).
35. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C. & Chen, M. Hierarchical text-conditional image generation with clip latents. Preprint at https://arxiv.org/abs/2204.06125 (2022).
36. Liu, B., Fu, J., Kato, M. P. & Yoshikawa, M. Beyond narrative description: generating poetry from images by multi-adversarial training. In *Proc. 26th ACM International Conference on Multimedia* 783–791 (ACM, 2018).
37. Hernandez-Olivan, C. & Beltran, J. R. Music composition with deep learning: a review. In *Advances in Speech and Music Technology: Computational Aspects and Applications* (eds. Biswas, A., Wennekes, E., Wieczorkowaska, A., & Laskar, R.H.) 25–50 (Springer International, 2023).
38. Yeomans, M., Shah, A., Mullainathan, S. & Kleinberg, J. Making sense of recommendations. *J. Behav. Decis. Mak.* **32**, 403–414 (2019).
39. Herremans, D., Martens, D. & Sörensen, K. Dance hit song prediction. *J. N. Music Res.* **43**, 291–302 (2014).
40. Inbar, Y., Cone, J. & Gilovich, T. People's intuitions about intuitive insight and intuitive choice. *J. Pers. Soc. Psychol.* **99**, 232–247 (2010).
41. Crowley, A. E., Spangenberg, E. R. & Hughes, K. R. Measuring the hedonic and utilitarian dimensions of attitudes toward product categories. *Mark. Lett.* **3**, 239–249 (1992).
42. Hirschman, E. C. & Holbrook, M. B. Hedonic consumption: emerging concepts, methods and propositions. *J. Mark.* **46**, 92–101 (1982).
43. Longoni, C. & Cian, L. Artificial intelligence in utilitarian vs hedonic contexts: the 'word-of-machine' effect. *J. Mark.* **86**, 91–108 (2022).
44. Waytz, A., Heafner, J. & Epley, N. The mind in the machine: anthropomorphism increases trust in an autonomous vehicle. *J. Exp. Soc. Psychol.* **52**, 113–117 (2014).
45. Li, S., Peluso, A. M. & Duan, J. Why do we prefer humans to artificial intelligence in telemarketing? A mind perception explanation. *J. Retail. Consum. Serv.* **70**, 103139 (2023).
46. Holthöwer, J. & van Doorn, J. Robots do not judge: service robots can alleviate embarrassment in service encounters. *J. Acad. Mark. Sci.* **51**, 767–784 (2022).
47. Pitardi, V., Wirtz, J., Paluch, S. & Kunz, W. H. Service robots, agency and embarrassing service encounters. *J. Serv. Manag.* **33**, 389–414 (2021).
48. Hong, Y.-y., Chiu, C.-y. & Dweck, C. S. Implicit theories of intelligence: reconsidering the role of confidence in achievement motivation. In *Efficacy, Agency and Self-Esteem* (ed. Kernis, M.H.) 197–216 (Springer, 1995).
49. Hancock, P. A. Imposing limits on autonomous systems. *Ergonomics* **60**, 284–291 (2017).
50. McClean, T. *The Path from Automation to Autonomy is Swarming with Activity*, https://www.forbes.com/sites/forbestechcouncil/2021/04/01/the-path-from-automation-to-autonomy-is-swarming-with-activity/?sh=134ca90f3716 (Forbes, 2021).
51. Reich, T., Kaju, A. & Maglio, S. J. How to overcome algorithm aversion: learning from mistakes. *J. Consum. Psychol.* **33**, 285–302 (2022).
52. Loughnan, S. & Haslam, N. Animals and androids: implicit associations between social categories and nonhumans. *Psychol. Sci.* **18**, 116–121 (2007).
53. Berger, B., Adam, M., Rühr, A. & Benlian, A. Watch me improve—algorithm aversion and demonstrating the ability to learn. *Bus. Inf. Syst. Eng.* **63**, 55–68 (2021).
54. Nissenbaum, H. & Walker, D. Will computers dehumanize education? A grounded approach to values at risk. *Technol. Soc.* **20**, 237–273 (1998).
55. Longoni, C., Bonezzi, A. & Morewedge, C. K. Resistance to medical artificial intelligence. *J. Consum. Res.* **46**, 629–650 (2019).
56. Mou, Y., Xu, T. & Hu, Y. Uniqueness neglect on consumer resistance to AI. *Market. Intell. Plan.* **41**, 669–689 (2023).
57. Purcell, Z. A. & Bonnefon, J.-F. Humans feel too special for machines to score their morals. *PNAS Nexus* **2**, pgad179 (2023).
58. Liu, N. T. Y., Kirshner, S. N. & Lim, E. T. Is algorithm aversion WEIRD? A cross-country comparison of individual-differences and algorithm aversion. *J. Retail. Consum. Serv.* **72**, 103259 (2023).
59. Yampolskiy, R. V. Unpredictability of AI: on the impossibility of accurately predicting all actions of a smarter agent. *J. Artif. Intell. Conscious.* **7**, 109–118 (2020).
60. De Freitas, J., Uğuralp, K., Uğuralp, Z. O. & Puntoni, S. *Chatbots and Mental Health: Insights into the Safety of Generative AI* Working Paper 23-011 (Harvard Business School, 2023).
61. Leotti, L. A., Iyengar, S. S. & Ochsner, K. N. Born to choose: the origins and value of the need for control. *Trends Cogn. Sci.* **14**, 457–463 (2010).
62. Bandura, A. *Self-efficacy: The Exercise of Control* (W. H. Freeman, 1997).
63. Rotter, J. B. Generalized expectancies of internal versus external control of reinforcements. *Psychol. Monogr.* **80**, 609 (1966).
64. Ryan, R. M. & Deci, E. L. Self-regulation and the problem of human autonomy: does psychology need choice, self-determination and will? *J. Personal.* **74**, 1557–1586 (2006).
65. Bown, N. J., Read, D. & Summers, B. The lure of choice. *J. Behav. Decis. Mak.* **16**, 297–308 (2003).
66. Suzuki, S. Effects of number of alternatives on choice in humans. *Behav. Process.* **39**, 205–214 (1997).
67. Cordova, D. I. & Lepper, M. R. Intrinsic motivation and the process of learning: beneficial effects of contextualization, personalization and choice. *J. Educ. Psychol.* **88**, 715–730 (1996).
68. Shapiro, D. H. Jr, Schwartz, C. E. & Astin, J. A. Controlling ourselves, controlling our world: psychology's role in understanding positive and negative consequences of seeking and gaining control. *Am. Psychol.* **51**, 1213–1230 (1996).
69. Premack, D. The infant's theory of self-propelled objects. *Cognition* **36**, 1–16 (1990).
70. Beer, J. M., Fisk, A. D. & Rogers, W. A. Toward a framework for levels of robot autonomy in human-robot interaction. *J. Hum. Robot Interact.* **3**, 74–99 (2014).
71. De Bellis, E. & Johar, G. V. Autonomous shopping systems: identifying and overcoming barriers to consumer adoption. *J. Retail.* **96**, 74–87 (2020).
72. Schweitzer, F. & Van den Hende, E. A. To be or not to be in thrall to the march of smart products. *Psychol. Mark.* **33**, 830–842 (2016).

73. Mandler, J. M. How to build a baby: II. Conceptual primitives. *Psychol. Rev.* **99**, 587–604 (1992).
74. Leslie, A. M. ToMM, ToBy, and Agency: Core architecture and domain specificity. In *Mapping the Mind: Domain Specificity in Cognition and Culture* (eds. Hirschfeld, L.A. & Gelman, S.A.) 119–148 (Cambridge Univ. Press, 1994).
75. Dennett, D. C. *The Intentional Stance* (MIT Press, 1989).
76. Gergely, G. & Csibra, G. Teleological reasoning in infancy: the naïve theory of rational action. *Trends Cogn. Sci.* **7**, 287–292 (2003).
77. Rijsdijk, S. A. & Hultink, E. J. 'Honey, have you seen our hamster?' Consumer evaluations of autonomous domestic products. *J. Prod. Innov. Manag.* **20**, 204–216 (2003).
78. Wertenbroch, K. et al. Autonomy in consumer choice. *Mark. Lett.* **31**, 429–439 (2020).
79. André, Q. et al. Consumer choice and autonomy in the age of artificial intelligence and big data. *Customer Needs Solut.* **5**, 28–37 (2018).
80. Brennan, R. & Sachon, L. *Self-Driving Cars Make 76% of Americans Feel Less Safe on the Road*, https://www.policygenius.com/auto-insurance/self-driving-cars-survey-2022/ (Policy Genius, 2022).
81. Zimmermann, J. L., Görgen, J., de Bellis, E., Hofstetter, R. & Puntoni, S. Smart product breakthroughs depend on customer control. *MIT Sloan Management Review* (16 February 2023).
82. Zimmermann, J. L., de Bellis, E., Hofstetter, R. & Puntoni, S. Cleaning with Dustin Bieber: Nicknaming Autonomous Products and the Effect of Coopetition. In *Proc. TMS 2021*.
83. Shaffer, V. A., Probst, C. A., Merkle, E. C., Arkes, H. R. & Medow, M. A. Why do patients derogate physicians who use a computer-based diagnostic support system? *Med. Decis. Mak.* **33**, 108–118 (2013).
84. Oyserman, D. Identity-based motivation: implications for action-readiness, procedural-readiness and consumer behavior. *J. Consum. Psychol.* **19**, 250–260 (2009).
85. Cheng, P. W. & Novick, L. R. A probabilistic contrast model of causal induction. *J. Personal. Soc. Psychol.* **58**, 545–567 (1990).
86. Menon, T., Morris, M. W., Chiu, C.-Y. & Hong, Y.-Y. Culture and the construal of agency: attribution to individual versus group dispositions. *J. Personal. Soc. Psychol.* **76**, 701–717 (1999).
87. Leung, E., Paolacci, G. & Puntoni, S. Man versus machine: resisting automation in identity-based consumer behavior. *J. Mark. Res.* **55**, 818–831 (2018).
88. de Bellis, E., Johar, G. V. & Poletti, N. Meaning of manual labor impedes consumer adoption of autonomous products. *J. Market.* https://doi.org/10.1177/00222429231171841 (2023).
89. Inzlicht, M., Shenhav, A. & Olivola, C. Y. The effort paradox: effort is both costly and valued. *Trends Cogn. Sci.* **22**, 337–349 (2018).
90. Norton, M. I., Mochon, D. & Ariely, D. The IKEA effect: when labor leads to love. *J. Consum. Psychol.* **22**, 453–460 (2012).
91. Lim, J. S. & O'Connor, M. Judgemental adjustment of initial forecasts: its effectiveness and biases. *J. Behav. Decis. Mak.* **8**, 149–168 (1995).
92. Dietvorst, B. J., Simmons, J. P. & Massey, C. Overcoming algorithm aversion: people will use imperfect algorithms if they can (even slightly) modify them. *Manag. Sci.* **64**, 1155–1170 (2018).
93. Landsbergen, D., Coursey, D. H., Loveless, S. & Shangraw, R. Jr Decision quality, confidence and commitment with expert systems: an experimental study. *J. Public Adm. Res. Theory* **7**, 131–158 (1997).
94. Caviola, L., Everett, J. A. & Faber, N. S. The moral standing of animals: towards a psychology of speciesism. *J. Personal. Soc. Psychol.* **116**, 1011–1029 (2019).
95. Schmitt, B. Speciesism: an obstacle to AI and robot adoption. *Mark. Lett.* **31**, 3–6 (2020).
96. Mori, M. The uncanny valley. *Energy* **7**, 33–35 (1970).
97. Kamide, H., Kawabe, K., Shigemi, S. & Arai, T. Anshin as a concept of subjective well-being between humans and robots in Japan. *Adv. Robot.* **29**, 1624–1636 (2015).
98. Dang, J. & Liu, L. Robots are friends as well as foes: ambivalent attitudes toward mindful and mindless AI robots in the United States and China. *Comput. Hum. Behav.* **115**, 106612 (2021).
99. *Biden–Harris Administration Announces New Actions to Promote Responsible AI Innovation that Protects Americans' Rights and Safety*, https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/04/fact-sheet-biden-harris-administration-announces-new-actions-to-promote-responsible-ai-innovation-that-protects-americans-rights-and-safety/ (White House, 2023).
100. *Combatting Online Harms Through Innovation*, https://www.ftc.gov/system/files/ftc_gov/pdf/Combatting%20Online%20Harms%20Through%20Innovation%3B%20Federal%20Trade%20Commission%20Report%20to%20Congress.pdf (FTC, 2022).
101. Fung, P. Etienne, H. Confucius, cyberpunk and Mr. Science: comparing AI ethics principles between China and the EU. *AI Ethics* **3** 505–511 (2023).
102. Dietvorst, B. J., Simmons, J. P. & Massey, C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. *J. Exp. Psychol. Gen.* **144**, 114–126 (2015).
103. Rozenblit, L. & Keil, F. The misunderstood limits of folk science: an illusion of explanatory depth. *Cogn. Sci.* **26**, 521–562 (2002).
104. Morewedge, C. K. Preference for human, not algorithm aversion. *Trends Cogn. Sci.* **26**, 824–826 (2022).
105. Johnson, C. & Tyson, A. *People Globally Offer Mixed Views of the Impact of Artificial Intelligence, Job Automation on Society*, https://www.pewresearch.org/short-reads/2020/12/15/people-globally-offer-mixed-views-of-the-impact-of-artificial-intelligence-job-automation-on-society/ (Pew Research Center, 2020).
106. Bostrom, N. *Superintelligence: Paths, Dangers, Strategies* (Oxford Univ. Press, 2014).
107. Zhang, C. et al. One small step for generative AI, one giant leap for AGI: a complete survey on ChatGPT in AIGC era. Preprint at https://arxiv.org/abs/2304.06488 (2023).
108. Purcell, Z. A., Dong, M., Nussberger, A.-M., Köbis, N. & Jakesch, M. Fears about AI-mediated communication are grounded in different expectations for one's own versus others' use. Preprint at https://arxiv.org/abs/2305.01670 (2023).
109. Chan, A., Riché, M. & Clifton, J. Towards the scalable evaluation of cooperativeness in language models. Preprint at https://arxiv.org/abs/2303.13360 (2023).
110. Köbis, N., Starke, C. & Rahwan, I. The promise and perils of using artificial intelligence to fight corruption. *Nat. Mach. Intell.* **4**, 418–424 (2022).
111. Davis, F. D., Bagozzi, R. P. & Warshaw, P. R. User acceptance of computer technology: a comparison of two theoretical models. *Manag. Sci.* **35**, 982–1003 (1989).
112. Ram, S. A model of innovation resistance. *Adv. Consum. Res.* **14**, 208–212 (1987).

Competing interests
The authors declare no competing interests.

Additional information
Correspondence should be addressed to Julian De Freitas.

Peer review information *Nature Human Behaviour* thanks Zoe Purcell and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.