World-Making for a Future with Sentient AI

Janet V.T. Pauketat[1][0000-0003-3280-3345]
Ali Ladak[1,2][0000-0003-1039-5774]
Jacy Reese Anthis[1,3,4][0000-0002-4684-348X]

[1]Sentience Institute, New York City, NY, USA
janet@sentienceinstitute.org
[2]University of Edinburgh, Edinburgh, UK
[3]University of Chicago, Chicago, IL, USA
[4]Stanford University, Stanford, CA, USA

Abstract. The ways people imagine possible futures with artificial intelligence (AI) affects future world-making—how the future is produced through cultural propagation, design, engineering, policy, and social interaction—yet there has been little empirical study of everyday people's expectations for AI futures. We addressed this by analysing two waves (2021 and 2023) of USA nationally representative data from the Artificial Intelligence, Morality, and Sentience (AIMS) survey on the public's forecasts about an imagined future world with widespread AI sentience (total $N = 2{,}401$). Average responses to six forecasts (exploiting AI labour, treating AI cruelly, using AI research subjects, AI welfare, AI rights advocacy, AI unhappiness reduction) showed mixed expectations for humanity's future with AI. Regressions of these forecasts on demographics such as age, the year the data was collected, individual psychological differences (the tendency to anthropomorphise, mind perception, techno-animist beliefs), and attitudes towards current AI (perceived threat and policy support) found significant effects on all forecasts from mind perception, anthropomorphism, and political orientation, and on five forecasts from techno-animism. The realised future that comes to pass will depend on these dynamic social psychological factors, consequent changes in expectations, and how those expectations shape acts of world-making.

Keywords: world-making, forecasting, theory of mind, anthropomorphism, techno-animism, AI rights social movement, moral circle expansion

Please cite as: Pauketat, J.V.T., Ladak, A., & Anthis, J.R. (2025). World-making for a future with sentient AI. *British Journal of Social Psychology, 64*(1), e12844. https://doi.org/10.1111/bjso.12844

1 Introduction

Imagining the future has long been a pastime of storytellers and scholars from around the world. Science fiction visionaries have imagined many future worlds with social machines and advanced artificial intelligence (AI). Sometimes advanced AI entities live in relative harmony with biological entities, such as in *The Culture* book series by Iain M. Banks (Banks, 1987) in which superintelligent AIs and biological humanoid species co-exist for over 9,000 years, and the TV series *Star Trek: Discovery* (Fuller & Kurtzman, 2017) where the crew accepts the personhood of Zora, their spaceship computer. In many imagined futures, AI threatens human society, such as in the film, *The Terminator* (Cameron, 1984), in which an AI system becomes self-aware and starts a war against humans, and in the play *R.U.R.* by Karel Čapek (Čapek, 1923), in which oppressed robots revolt and wipe out humanity.

Imagining seemingly impossible futures is a foundational aspect of science fiction, the genre responsible for inspiring technological designs from mobile phones to voice assistants (Axtell & Munteanu, 2021; Jordan & Auernheimer, 2018; Mubin et al., 2019; Siegel, 2017) and giving us the term "robot" (Gunkel, 2018). In human-computer interaction (HCI) research, speculative design and design fictions foreground possible, and sometimes improbable, futures with AI to assess responses to possible AI designs, prototypes, and future scenarios such as integrated human-AI workplaces (Jordan & Silva, 2021; Linehan et al., 2014; Lupetti et al., 2018; Tanenbaum et al., 2016; Wong, 2018). These methods have been important for transcending traditional visions of distant futures that were limited to only some racial, gender, and class-based groups (Harrington et al., 2022; Hohendanner et al., 2023; Wyche, 2022). In psychological research, vignette studies have been used to illustrate perceptions of and responses to hypothetical AI systems such as AI with emotion recognition and emotion expression faculties, robotic avatars with burns, whole brain emulations, chatbot assistants, and humanised robotic victims (Küster & Swiderska, 2021; Ladak et al., 2023, 2024; Manoli et al., 2024; Nijssen et al., 2019).

In the real-world, socio-political institutions are investing in transformative AI infrastructure, such as Japan's Society 5.0 initiative to develop future industries that rely on AI for autonomous decentralisation (Fujii et al., 2018; Nagahara, 2019) and China's New Generation Artificial Intelligence Development Plan that invests in sweeping AI-based social transformation from the acceleration of AI application development to supporting collaborations between universities, industries, and the government for AI innovation (Wu et al., 2020). Individuals vary in the extent to which they interact with, read narratives about, or are knowledgeable about robots, smart machines, and AI. Increased exposure to AI has been associated with positive outcomes like increased acceptance in the workplace across 27 European countries (Turja & Oksanen, 2019).

AI[1] systems have sophisticated intelligent (e.g., problem solving, memory) and agentic (e.g., planning, moral judgement) capacities across varying purposes (e.g., social, industrial). Until recently, AI sentience, that is, the capacity of AI to have positive and negative experiences such as pain and pleasure, was purely science fiction. However, interdisciplinary scholarship has identified AI sentience as a potential crux of future human-AI interaction (Gibert & Martin, 2022; Gordon & Gunkel, 2021; Harris & Anthis, 2021; Ladak, 2023) and much public discourse has been focused on sentience, such as the 2022 wave of news coverage following the departure of a

[1] We refer to all embodied (e.g., robots) and non-embodied (e.g., language algorithms) as AI for the purposes of this paper.

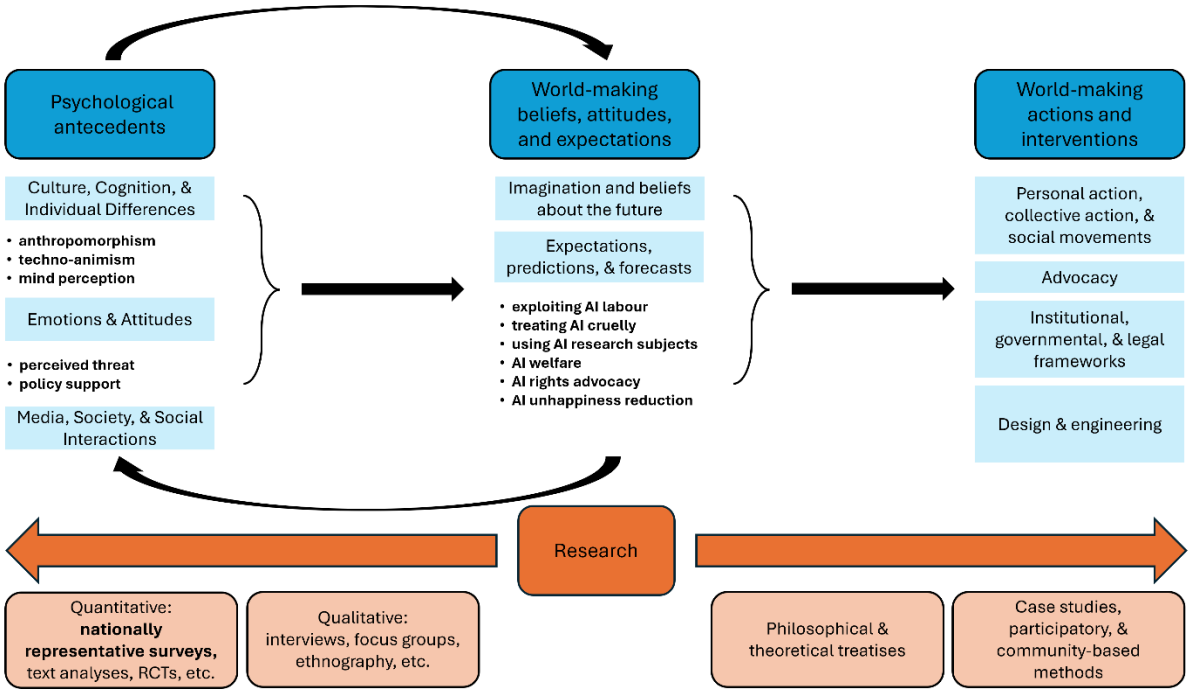
Google engineer who raised concerns that the company's latest AI system had gained sentience (Tiku, 2022).

Although the issue of sentient AI, that is AI with the capacity to have positive and negative experiences such as happiness and suffering, is still hypothetical, debated, and difficult to disentangle from misconceptions of AI capabilities (de Saint-Laurent, 2018; Schwitzgebel, 2023) and anthropomorphic tendencies (Giger et al., 2019), experts and laypeople alike increasingly consider there to be a real possibility of near-future AI that surpasses humans on all tasks, and sentient AI (Grace et al., 2023; Metzinger, 2021; Pauketat et al., 2023a; Roser, 2024). Academic engagement with the topic of sentience-based moral consideration of AI has been rapidly growing (Harris & Anthis, 2021) and some ethicists posit morally catastrophic possible futures with advanced AI or digital minds (i.e., artificial entities with (or that seem to have) mental faculties like autonomy, agency, and sentience; Baumann, 2022; Coghlan et al., 2019; Fröding & Peterson, 2020; Shulman & Bostrom, 2021). Laypeople already perceive a degree of consciousness in GPT-3, the large language model behind ChatGPT (Scott et al., 2023). Setting aside the issue of proving AI sentience, we have little understanding of what laypeople think of sentient AI or their expectations for future worlds with sentient AI. Yet, the pursuit (or avoidance) of specific AI futures could shape the unfolding of our technological, social, and moral world.

World-making, according to Power et al. (2023), entails individual and collective human contributions to making societies, social relations, and shaping cognition. The collective future world is forged by ongoing stable processes like evolutionary mechanisms and basic cognitions but also by the contestation of visions of the future. Power et al. advocate for social psychology to play a greater role in "what could be" (p. 1) and suggest that imagining futures is the first stage of world-making. They argue that imagining the future is social as well as personal because it draws on cultural norms and products. To Savage (2024), world-making is a sociological process that offers new, revelatory visions, in which social scientific description is an important part. Building on these conceptualisations, we consider world-making to be the production of the future through psychological processes, cultural propagation, design, engineering, policy, and social interactions, for the distant future.

In Figure 1, we suggest that psychological factors, broadly construed as the cognitive, personality, emotion, attitude, social, and cultural antecedents underpinning individual and social routines, and cognitive world-making processes such as imagining the future, inform each other in a feedback loop. We distinguish intervention-oriented world-making actions (e.g., the forming of social movements, advocacy for change, the establishment of institutional and legal frameworks, design and engineering choices) from these psychological antecedents and world-making beliefs, attitudes, and expectations. We distinguish between world-making beliefs, attitudes, and expectations, and world-making actions and interventions to separate between the future-focused cognitions that we study as an input into the world-making behaviours that we do not study, although both are conceptualised as world-making by Power et al. (2023).

Figure 1
Conceptual model of future world-making



Note. Psychological antecedents (e.g., culture, cognition, emotion) and world-making beliefs, attitudes, and expectations (e.g., expectations about the future) influence each other and together influence world-making actions and interventions (e.g., social movements, advocacy, legal frameworks, design). All types of world-making research can occur at any stage but certain methodologies are currently used predominantly at certain stages (e.g., nationally representative surveys to investigate the loop between psychological antecedents and world-making beliefs, attitudes, and expectations, participatory methods to investigate collective action). The bolded bullet points are variables that we analyse from the Artificial Intelligence, Morality, and Sentience (AIMS) survey waves 1 and 2. Nationally representative survey is bolded to highlight the AIMS survey methodology used in this study.

Power et al. (2023) argued that world-making can co-occur with social psychological research when we observe, engage with, or otherwise study the psychological processes that influence world-making rather than spotlighting a static moment or conceiving of the world as static. In the current era of rapid AI-driven social change in which AI systems like ChatGPT and Roomba are becoming household names, we study the US general public's expectations for one possible imagined future world filled with sentient AI. We consider this ongoing AI-based social change by analysing the expectations of US Americans alongside cognitive, moral, and social psychological factors measured in two waves (2021 and 2023) of the nationally representative Artificial Intelligence, Morality, and Sentience (AIMS) survey (Pauketat et al., 2023b).

Social scientists commonly base conceptualisations and methodologies on the assumption that people engage in collective future imagination (de Saint-Laurent et al., 2018). We examine collective future imagination using a quantitative approach. Following Power et al., we consider collective future imagination to be a component of world-making. The imagination of possible

futures with sentient AI in research initiatives such as the AIMS survey may affect future world-making in part by prompting the collective imagination of possible distant futures amongst the general public and in part by prompting researchers, advocates, and concerned stakeholders like lobbyists and policymakers to consider and act on alternative, distant, or hypothetical futures with AI. Longitudinal surveys like AIMS have been considered suitable tools for social psychological world-making because they illustrate attitudes and opinions about future trajectories and demonstrate change over time. Further, when survey results are publicly available, like for the AIMS survey, the collective representation of imagination can feedback into public consciousness (Power et al., 2023), influencing world-making thoughts and behaviours. With this study, we manifest research as world-making (see Figure 1 for our conceptual model).

World-making actions and interventions for the world of today in social movements like Black Lives Matter, #MeToo, and Animal Rights also shape the world of tomorrow. Social movements occur when a large group of people work collectively to solve a problem (Toch, 1965; van Stekelenburg & Klandermans, 2013). In modern movements like Animal Rights, people continue to work towards social and moral change over decades to pursue ongoing present-focused and future-focused goals. This type of massive social change movement for the rights of underprivileged, underrecognised, and oppressed groups has led to moral circle expansion, a strategy for future world-making (Anthis & Paez, 2021). Future world-making extends even to the distant future. For instance, Syropoulos et al. (2024) found that prompts to envision the long-term future increased support for legal reform to protect the welfare of people living over 1,000 years in the future.

Future world-making may occur under expectations that patterns of moral exclusion and oppression from human history will propagate into technologically advanced futures. If sentient AI are a part of this future, our relationship to them requires scrutiny, both for their welfare and for the wellbeing of future humans. For this reason, we examined forecasts about the treatment of future AI in a world where AI sentience is prevalent, focusing on situations that are likely to affect AI (e.g., being exploited for labour, being used as research subjects without consent; see Figure 1). These scenarios could propagate from our current values and norms about AI in the workforce and as replacements for biological subjects in research (Madden et al., 2020; Parker & Grote, 2022).

2 The Psychology of Making Future Worlds

Cognitive processes enable us to model past actions, imagine possible futures, and enact change to create better futures (Suddendorf & Corballis, 2007; Szpunar et al., 2014; Trope & Liberman, 2010). Thinking about the distant-future has been shown to decrease the attitude-intention gap relative to taking a near-future temporal perspective (Rabinovich et al., 2010), showing the importance of world-making beliefs, attitudes, and expectations as antecedents to world-making actions. For instance, pro-environmental attitudes were more strongly linked to pro-environmental intentions for those asked to think about the environmental state in 10 years as opposed to 1 month. Temporal perspective has also been shown to occur at both individual (i.e., self) and collective (i.e., group) levels (Peetz & Wohl, 2019). People distinguished between their social group's and personal future interests and past actions, suggesting that people can consider humanity's future, or multiple possible futures (e.g., what humans will eat in 100 years), and not just their own (e.g., what they will eat in the next year). These processes have been connected to the activation of the medial prefrontal cortex and the precuneus (Okuda et al., 2003; Packer & Cunningham, 2009;

Stillman et al., 2017) alongside future thinking (Cona et al., 2023). Further, future-oriented cognitions and motivations have been shown to positively shape outcomes that affect future-oriented behaviours like increased pro-health behaviours (e.g., exercise) and retirement planning, and decreased risk-taking behaviours such as risky driving (Kooij et al., 2018).

Exposure to momentary visions of the future also impact attitudes and intentions, illustrating connections between social psychological processes, world-making beliefs, attitudes, and expectations, and world-making actions and interventions. Bosone et al. (2023) showed that positive visions of the future (e.g., a future world with eco-efficient green technologies) increased perceived collective efficacy and intentions to engage in pro-environmental behaviours to improve climate change. Likewise, utopian visions of imagined worlds have been shown to promote intentions to engage in collective change actions to build a desirable future (Badaan et al., 2020; Fernando et al., 2018; Kashima & Fernando, 2020). Utopian thinking has also been shown to vary by culture. Fernando et al. (2023) showed that a futurist utopian profile based on science and technology advancements (e.g., more endorsement of statements such as "People are completely transformed by advances in science and technology") with little importance placed on morality (e.g., less endorsement of statements such as "In this society, all people are good and honest") was more prevalent in China and Australia than in the UK and the USA.

People differ in the extent to which they tend to anthropomorphise nonhuman entities and hold animistic beliefs. Anthropomorphism and animism entail attributing human-like features and a soul or personhood to nonhuman entities, respectively (Marenko, 2014; Waytz et al., 2010). Animist beliefs are ancient and globally prevalent (Pedersen, 2001; Richardson, 2016; Wilkinson, 2017). Researchers have suggested that animistic beliefs may be key to redefining humans' moral relations with nonhumans (Conty, 2022; Sprenger, 2021). Johnson et al. (2015) showed that attributing a soul to robots predicted personifying them and argued that anthropomorphism of nonhumans should be reoriented towards personhood, not just human-likeness. Both anthropomorphism and techno-animism, the attribution of animistic features to technological entities like robots (Jensen & Blok, 2013; Okanda et al., 2021), have been shown to predict increased moral inclusion of AI (Beran et al., 2011; Pauketat & Anthis, 2022). We expected the tendency to anthropomorphise and techno-animist beliefs to predict future forecasts.

Relatedly, perceiving mind in AI, a function of theory of mind to enable better comprehension and cooperation with other entities (Higgins et al., 2021; Tamir & Thornton, 2018; Tomasello et al., 2005), is linked with AI moral inclusion (Pauketat & Anthis, 2022). K. Gray et al. (2012) argued that mind perception is a key predictor of moral circle inclusion. The moral circle bounds moral consideration, protection from harm, care, and the right to fairness standards (H. M. Gray et al., 2007; K. Gray & Wegner, 2009; Hadarics & Kende, 2018; Opotow, 1990, 1993) and entities can be included to various degrees (Crimston et al., 2016).

AI mind perception processes are especially sensitive to their affective faculties such as sentience. Perceived affective faculties drive ascription of general mind to AI (Koban & Banks, 2024) and affective faculties are perceived more in highly valued social robots than industrial robots (Wang & Krumhuber, 2018). Swiderska and Küster (2018) and Ward et al. (2013) found that people perceived more mind in AI systems when they were harmed. Perception of affective faculties has also been shown to increase moral consideration for AI (Ladak et al., 2024) and Nijssen et al. (2019) showed that people were less likely to sacrifice a robot with affective faculties.

Given this previous research, we anticipated that AI mind perception would predict forecasts about an imagined future with sentient AI. Mind perception might predict forecasting a future in which sentient AI systems are not exploited for their labour, treated cruelly, or used as

research subjects, and in which their welfare matters, their rights are advocated for, and where people think it is important to reduce the percentage of unhappy sentient AI systems following an optimistic belief that the future will be better or because mind perception is linked to moral inclusion. Conversely, mind perception might predict the opposite (i.e., increased mind perception predicts increased expectations of cruel treatment, labour exploitation, etc.) under the recognition and extension of a pattern of nonhuman animal moral exclusion or if mind perception backfires to increase the protectionism of human uniqueness.

3 Attitudes towards AI

The rise of automation and the design of anthropomorphic AI poses realistic and symbolic threats like species survival and uniqueness that may affect future forecasts. Anthropomorphic design has had mixed effects on responses to AI, although it has been widely adopted to improve acceptance of AI systems (Giger et al., 2019; Paiva et al., 2018; Waytz et al., 2014; Złotowski et al., 2015). Physical anthropomorphic design and physical or psychological contexts where robots outperform humans have been shown to increase perceived threat and opposition to robotics development (Ferrari et al., 2016; Giger et al., 2019; Yogeeswaran et al., 2016; Złotowski et al., 2017). Widespread public support for slower technological development, laws regulating development, or support for outright bans on some human-like developments (e.g., sentience, human-AI system integration) could shape future worlds by prompting policymakers towards certain trajectories like limiting the development of advanced AI by placing safeguards around models that cost a certain amount to train, and consumers towards rejection of certain technologies that they disagree with or dislike.

We examined the predictive effects of perceived threat and policy attitudes on future forecasts. Increased perceived threat might predict forecasting exploitation, cruel treatment, research use, welfare and advocacy not being important, and it not being necessary to reduce the percentage of unhappy sentient AI systems if people expect or believe that sentient AI will destroy human uniqueness or curtail human survival. For example, Pauketat and Anthis (2022) found that realistic threat predicted less support for pro-AI welfare policies. Alternatively, perceived threat might predict the opposite (e.g., less exploitation and cruel treatment) under a belief that human-like AI deserve moral status and moral inclusion (i.e., perceived threat predicts a better future for AI). Pauketat and Anthis (2022) found some evidence for this relationship. Strong identity threat linearly predicted increased mind perception (i.e., acknowledging a degree of human-likeness). Moderate identity threat quadratically predicted more moral inclusion, more valuation of future AI sentience, and more support for AI rights.

4 Overview

With the rapid adoption of generative AI like ChatGPT, Claude, and Midjourney, we have a unique opportunity to develop the social psychological theory of the human-AI dynamics that may undergird and circumscribe humanity's socio-technological future. Social and technological changes are often studied after they happen, meaning that we have rarely observed perceptions of seemingly radical technologies before they are fully integrated into daily life. With our analysis of the nationally representative AIMS survey data, we begin to examine the general public's visions of a future with widespread sentient AI, particularly expectations for their experiences and this future world's AI-relevant social issues. To do this, we selected and analysed a portion of data

from two waves of this cross-sectional, longitudinal, and online survey of US Americans, comparing responses to the six forecasts and regressing each of the forecasts on demographic and psychological factors to illustrate the culturally-situated American visions of the future. We did not make directional hypotheses with this data-driven and bottom-up approach, although we selected predictor variables based on the research reviewed in the previous sections on human-AI interaction.

5 Methods

5.1 Participants

Wave 1 of the nationally representative AIMS survey included 1,232 US Americans recruited by iSay/Ipsos, Dynata, Disqo, and other leading sample panels in November and December 2021 (54% female, 75% White, 22% 18-34 years old, 32% 35-54 years old, 45% 55+ years old). Wave 2 included 1,169 Americans recruited in April and May 2023 (54% female, 61% White, 27% 18-34 years old, 34% 35-54 years old, 39% 55+ years old). The sample (total $N = 2{,}401$) was recruited by age, gender, ethnicity, income, education, and USA region according to the official American Community Survey's census estimates.[2] The sample self-identified as politically moderate (Wave 1 $M = 3.15$, $SD = 1.14$; Wave 2 $M = 3.05$, $SD = 1.12$).

Some of the sample owned a robotic/AI device (30% Wave 1, 29% Wave 2) and some reported working with a robotic/AI device[3] (17% Wave 1, 16% Wave 2). Participants reported having a variety of experiences with AI including having a conversation with family or friends about AI (25% Wave 1, 29% Wave 2), having a conversation with an AI (16% Wave 1, 20% Wave 2), reading about AI ethics (17% Wave 1, 19% Wave 2), seeing a robot being physically attacked by humans (4% Wave 1, 6% Wave 2), and seeing a recording of a robot being physically attacked by humans (11% Wave 1, 13% Wave 2). More than half (59% Wave 1, 54% Wave 2) reported not having any of these experiences. Some of the sample reported believing that sentient AI already exists (16% Wave 1, 18% Wave 2).

5.2 Procedure and Measures

The AIMS survey consisted of several preregistered[4] sections: attitudes towards the moral and social integration of AI, the moral consideration of other nonhumans, individual differences, and future forecasts. Important terms like "robots/AIs", "sentience", and "sentient robots/AIs" were defined within the survey. "Robot/AIs" were defined as "intelligent entities built by humans, such as robots, virtual copies of human brains, or computer programs that solve problems, with or without a physical body, that may exist now or in the future". "Sentience" was defined as "the capacity to have positive and negative experiences, such as happiness and suffering". "Sentient robots/AIs" were defined as "those with the capacity to have positive and negative experiences,

[2] All sample demographics can be found alongside the open data (https://data.mendeley.com/datasets/x5689yhv2n/2).

[3] Many terms could be used to describe the entities of interest, such as "artificial beings," "artificial entities," "robots and AI," or "AI." The AIMS survey used "robot/AI" because of its clarity, specificity, and inclusion of all AI systems, including the salient robot subgroup.

[4] The AIMS survey methodology was preregistered on the OSF by the original research team (Wave 1 https://osf.io/udbhm; Wave 2 https://osf.io/w9h6g). The full AIMS survey methods and materials are available through the preregistrations.

such as happiness and suffering". Participants completed the survey on the GuidedTrack platform, gave their informed consent at the start of the survey, and were thanked at the end of the survey. Procedures were approved by the primary author's institution and performed in accordance with the APA Ethical Principles of Psychologists and the ethical standards of the 1964 Declaration of Helsinki and its later amendments. The open data (Pauketat et al., 2023b) is available on Mendeley Data (https://data.mendeley.com/datasets/x5689yhv2n/2). The items we selected and the R analysis code we wrote for the current study are on the OSF (https://osf.io/vm59q).

5.2.1 Predictors

Demographics and Time. Age, gender, region, ethnicity, education, and income were collected by the survey recruiter. Age, education level, and income were analysed as continuous predictors. Gender was binary[5] (man = 0), ethnicity was multi-categorical (Asian, Black, Hispanic, Indigenous, Other, White), and region was multi-categorical (Midwest, Northeast, South, West). Time was the year in which the data were collected (2021, 2023).

The other demographics were collected within the survey after all measures. Political orientation was self-reported on a sliding scale (1 = very liberal, 3 = moderate, 5 = very conservative). Religion (e.g., Agnostic, Buddhist, Catholic, None) and diet (e.g., meat-eater, veg*n) were multi-categorical. We recoded religion to a binary (religious = 0). Ownership of a robotic or AI device was binary (0 = do not own). Self-reported exposure to AI was the average of two items, "How often do you interact with AI or robotic devices that respond to you and that can choose their own behavior" and "How often do you read or watch robot/AI-related stories, movies, TV shows, comics, news, product descriptions, conference papers, journal papers, blogs, or other material", measured on a Likert-type (0 = never, 5 = daily) scale ($r = .57$, $p < .001$).

Mind Perception. We operationalised mind perception with the average of four items, "To what extent do current robots/AIs (i.e., those that exist in 2021/2023) have the capacity for each of the following? (experiencing emotions, having feelings, thinking analytically, being rational)" on a 0 (not at all) to 100 (very much) sliding scale adapted from Wang and Krumhuber (2018). Higher scores reflected more attribution of mind to AI (Cronbach's $\alpha = .82$).

Policy Attitudes. Policy attitudes were the average of responses to three broad bans relevant in this early era of AI regulation before people may have nuanced understandings of AI regulation, "I support a global ban on the development of sentience in robots/AIs", "I support a global ban on the development of AI-enhanced humans", and "I support a global ban on the development of robot-human hybrids" on Likert-type scales (1 = strongly disagree, 4 = no opinion, 7 = strongly agree). Higher scores indicated more cautious attitudes towards the development of sentience-related technologies (Cronbach's $\alpha = .87$).

Perceived Threat. Three items on Likert-type scales (1 = strongly disagree, 4 = no opinion, 7 = strongly agree), "Robots/AIs may be harmful to me personally", "Robots/AIs may be harmful to people in the USA", and "Robots/AIs may be harmful to future generations of people", were averaged into an index of perceived threat (Cronbach's $\alpha = .90$). These items were adapted from Thaker et al. (2017) for the AI context. Higher scores indicated more perceived threat from AI.

Anthropomorphism. The tendency to anthropomorphise artificial entities was assessed with four averaged items adapted from Waytz et al. (2010) on a 0 (not at all) to 10 (very much) sliding scale. The stem, "To what extent does the average…", was paired with four AI targets and

[5] This information was obtained from the recruiter as a part of their demographics and did not include non-binary or other categories.

human-like capacities, "…robot have consciousness", "…computer have a mind of its own", "…AI have intentions", and "digital simulation have emotions". Higher scores meant more anthropomorphism (Cronbach's α = .94).

Techno-Animism. Techno-animism was assessed with two items, "Artificial beings contain a spirit" and "The spirits of human, natural, and artificial beings can interact with each other" from Pauketat and Anthis (2022). Items were measured on a 1 (strongly disagree) to 7 (strongly agree) sliding scale and averaged (Cronbach's α = .75). Higher scores indicated stronger techno-animist beliefs.

5.2.2 Future Forecasting Outcomes

We predicted six forecasts about an imagined future world where sentient AI had proliferated. We referred to AI as a group, following Smith et al. (2021), given that people already treat AI as an outgroup similar to human outgroups. Further, asking people to forecast about the future likely activates abstract mental representations of AI as a group following research on future thinking (Agerström & Björklund, 2013; D'Argembeau & Van der Linden, 2007; Förster et al., 2004; Szpunar et al., 2014) rather than concrete representations of a specific AI system like Siri or Alexa. We loosely categorised the forecasts as being about the treatment of future sentient AI and AI social issues.

Participants were asked to "…imagine a future in which sentient robots/AIs have already been developed and become widespread" and answer questions about "this future world". The forecasts were presented in random order. All started with the phrase, "In this future world, to what extent…", and were paired with target phrases. The three treatment targets were, "…are robots/AIs exploited for their labor" (Exploitation), "…are robots/AIs treated cruelly" (Cruelty), and "…are robots/AIs used as subjects in scientific and medical research" (Research Use). The three social issue targets were "…is the welfare of robots/AIs an important social issue" (Welfare), "…is advocacy for robot/AI rights necessary" (Advocacy), and "…is it important to reduce the overall percentage of unhappy sentient robots/AIs" (Unhappiness Reduction). Responses were measured on a 1 (not at all) to 5 (very much) sliding scale and were analysed as standalone items.

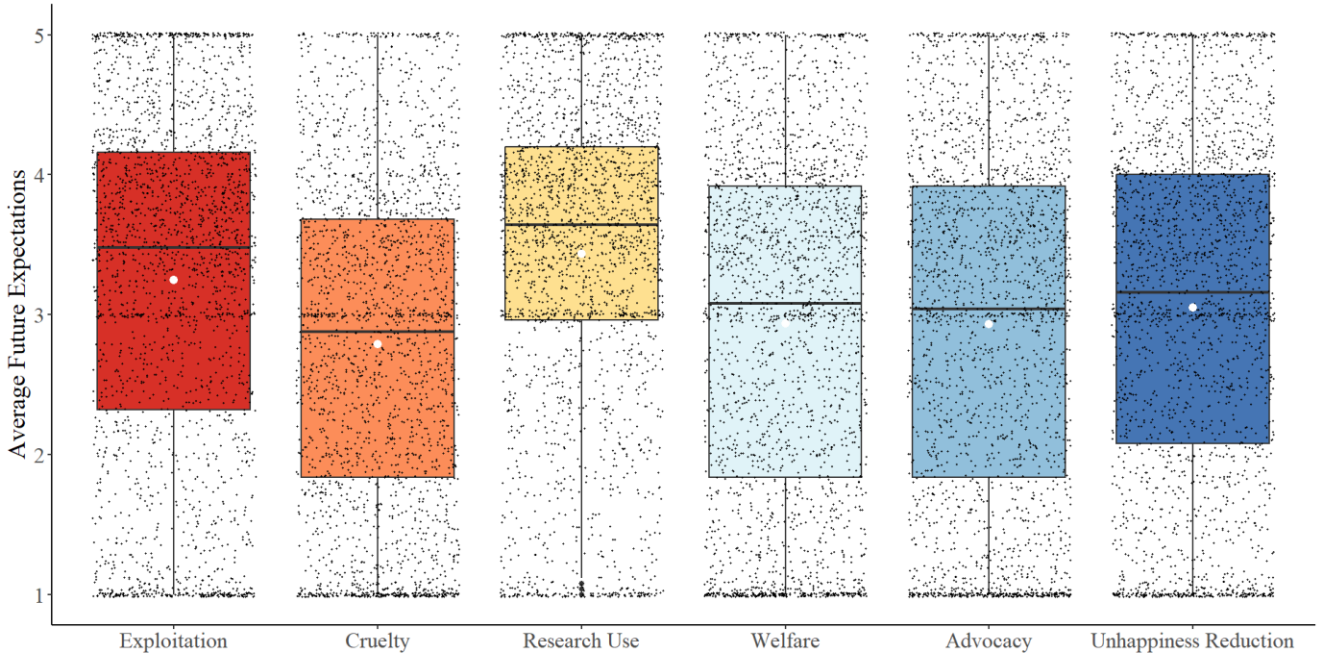
6 Results

We employed a repeated measures ANOVA and OLS regressions in RStudio (v.2023.06.1+524) to examine forecasts, and the false discovery rate (FDR; Benjamini & Hochberg, 1995) to adjust for multiple comparisons. For the regressions, the multi-categorical demographic variables (i.e., ethnicity, region, diet) were compared to the largest group (i.e., White, South, meat-eating) and the predictor variables without a meaningful zero in the scale (i.e., age, education, income, political orientation, AI exposure, policy attitudes, perceived threat, techno-animism) were mean-centred. We report adjusted R^2 values. Correlations between continuous variables are in Table 1.

There were significant differences in the average extent to which each forecast was expected to occur, $F(5, 12000) = 205.17$, $p < .001$, $\eta_p^2 = .08$ (see Figure 2). All forecasts were significantly different from each other using post-hoc pairwise comparisons ($p < .001$) except for expectations about the importance of AI welfare as a social issue ($M = 2.94$, $SD = 1.24$) and the necessity of AI advocacy ($M = 2.93$, $SD = 1.25$; $p = .757$). The most forecasted occurrence was that sentient AI would be used in scientific and medical research ($M = 3.43$, $SD = 1.11$), followed closely by expectations of labour exploitation ($M = 3.25$, $SD = 1.24$). Forecasting that the number of unhappy sentient AI systems needs to be reduced was also above the scale midpoint ($M = 3.05$,

SD = 1.22). The weakest forecast was that sentient AI would be treated cruelly (*M* = 2.79, *SD* = 1.17).

Figure 2
Forecasts about a future with widespread sentient AI



Note. The black horizontal line on each forecast boxplot shows the median response and the white point shows the mean response. Each black point is an individual data response.

6.1 Predicting Forecasts about the Treatment of Future Sentient AI

The regression model explained 12% of the variance in Exploitation (R^2 = .12, F(26, 2374) = 13.0, p < .001). Older age, a more liberal political orientation, and being non-religious were significant demographic predictors of forecasting that sentient AI would be exploited for their labour. Perceiving more mind in currently existing AI, stronger techno-animist beliefs, and a stronger tendency to anthropomorphise artificial entities were significant psychological predictors (see Table 2).

The regression model explained 22% of the variance in Cruelty (R^2 = .22, F(26, 2374) = 26.5, p < .001). A more liberal political orientation, a meat-eating (relative to vegan) diet, and being non-religious were significant demographic predictors of forecasting that sentient AI would be treated cruelly. Perceiving more mind in current AI, having more cautious policy attitudes, perceiving more threat from AI, holding stronger techno-animist beliefs, and having a stronger tendency to anthropomorphise were significant psychological predictors (see Table 2).

The regression model explained 15% of the variance in Research Use (R^2 = .15, F(26, 2374) = 17.5, p < .001). Older age, higher income, a more liberal political orientation, and more exposure to AI were significant demographic predictors of forecasting that sentient AI would be

used as subjects in medical and scientific research. Perceiving more mind in currently existing AI and tending to anthropomorphise were psychological predictors (see Table 2).

6.2 Predicting Forecasts about Future Social Issues

The regression model explained 35% of the variance in Welfare (R^2 = .35, F(26, 2374) = 50.9, p < .001). Older age, being female, living in the West, Hispanic (relative to White) ethnicity, a more liberal orientation, and more exposure to AI significantly predicted forecasting that AI welfare will be an important social issue. Perceiving more mind, stronger techno-animist beliefs, and a greater tendency to anthropomorphise were the significant psychological predictors (see Table 3).

The regression model explained 40% of the variance in Advocacy (R^2 = .40, F(26, 2374) = 62.4, p < .001). Older age, a more liberal political orientation, being non-religious, and more exposure to AI were the significant demographic predictors of forecasting a need for AI rights advocacy. Perceiving more mind, perceiving less threat, stronger techno-animist beliefs, and a greater tendency to anthropomorphise were the significant psychological predictors (see Table 3).

The regression model explained 27% of the variance in Unhappiness Reduction (R^2 = .27, F(26, 2374) = 35.0, p < .001). The only significant demographic predictor of thinking that it would be important to reduce the overall percentage of unhappy sentient AI systems was a more liberal political orientation. The significant psychological predictors were perceiving more mind, stronger techno-animist beliefs, and a greater tendency to anthropomorphise (see Table 3).

Table 1

Correlations and descriptive statistics

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Exploitation Forecast	-												
2. Cruelty Forecast	.64***	-											
3. Research Use Forecast	.51***	.45***	-										
4. Welfare Forecast	.48***	.58***	.43***	-									
5. Advocacy Forecast	.49***	.60***	.41***	.80***	-								
6. Unhappiness Forecast	.49***	.55***	.44***	.68***	.68***	-							
7. Mind Perception	.28***	.36***	.34***	.49***	.51***	.44***	-						
8. Policy Attitudes	.00	.00	-.04*	-.11***	-.14***	-.06***	-.12***	-					
9. Perceived Threat	-.01	-.01	-.05*	-.14***	-.18***	-.06***	-.15***	.47***	-				
10. Techno-Animism	.26***	.40***	.27***	.50***	.55***	.43***	.59***	-.12***	-.13***	-			
11. Anthropomorphism	.29***	.41***	.31***	.52***	.55***	.46***	.68***	-.07***	-.10***	.75***	-		
12. AI Exposure	.20***	.23***	.27***	.33***	.34***	.26***	.40***	-.08***	-.08***	.39***	.41***	-	
13. Political Orientation	-.07***	-.08***	-.07***	-.12***	-.14***	-.12***	-.05*	.17***	.17***	-.03	.03	-.06***	-
Mean	3.25	2.79	3.43	2.94	2.93	3.05	46.20	4.62	4.52	3.07	3.39	1.22	3.12
(*SD*)	(1.24)	(1.17)	(1.11)	(1.24)	(1.25)	(1.22)	(23.75)	(1.73)	(1.68)	(1.64)	(2.75)	(1.36)	(1.13)

Note. FDR-adjusted *p*-values reported. *$p < .05$, **$p < .01$, ***$p < .001$

Table 2

Regression models predicting forecasts about the treatment of future sentient AI

	Exploitation			*Cruelty*			*Research Use*		
	b SE CI	β	*t*	*b* SE CI	β	*t*	*b* SE CI	β	*t*
Intercept	2.36 0.13 2.11, 2.60	-	18.78***	1.56 0.11 1.34, 1.78	-	14.07***	2.51 0.11 2.29, 2.73	-	22.83***
Age	0.01 0.00 0.00, 0.01	0.09	3.75***	0.00 0.00 0.00, 0.01	0.04	1.98	0.00 0.00 0.00, 0.01	0.06	2.60*
Gender – Woman	-0.11 0.05 -0.21, 0.01	-0.05	-2.22	0.01 0.05 -0.08, 0.10	0.01	0.26	0.05 0.04 -0.04, 0.14	0.02	1.06
Region – Midwest	-0.02 0.07 -0.15, 0.11	-0.01	-0.37	-0.05 0.06 -0.17, 0.06	-0.02	-0.90	0.11 0.06 -0.01, 0.22	0.04	1.85
Region – Northeast	0.01 0.07 -0.13, 0.14	0.00	0.09	0.01 0.06 -0.10, 0.13	0.00	0.23	0.03 0.06 -0.09, 0.15	0.01	0.49
Region – West	0.10 0.07 -0.03, 0.23	0.03	1.51	0.10 0.06 -0.01, 0.22	0.04	1.80	0.08 0.06 -0.03, 0.19	0.03	1.36
Ethnicity – Asian	-0.10 0.11 -0.31, 0.11	-0.02	-0.94	-0.13 0.09 -0.31, 0.06	-0.03	-1.33	-0.18 0.09 -0.37, 0.00	-0.04	-1.96
Ethnicity – Black	0.03 0.08 -0.13, 0.20	0.01	0.38	0.01 0.07 -0.14, 0.15	0.00	0.09	-0.07 0.07 -0.22, 0.07	-0.02	-0.96
Ethnicity – Hispanic	0.00 0.08 -0.15, 0.16	0.00	0.03	-0.04 0.07 -0.18, 0.10	-0.01	-0.58	0.08 0.07 -0.05, 0.21	0.02	1.18
Ethnicity – Indigenous	-0.07 0.24	-0.01	-0.30	0.07 0.21	0.01	0.33	0.27 0.21	0.02	1.30

	-0.54, 0.40			-0.35, 0.49			-0.14, 0.69		
Ethnicity – Other	-0.02 0.17	0.00	-0.15	-0.28 0.15	-0.03	-1.87	-0.17 0.15	-0.02	-1.16
	-0.36, 0.31			-0.57, 0.01			-0.46, 0.12		
Education	0.00 0.01	0.01	0.22	0.01 0.01	0.02	0.89	0.00 0.01	0.00	0.07
	-0.02, 0.02			-0.01, 0.03			-0.02, 0.02		
Income	0.00 0.00	0.00	0.21	0.00 0.00	0.02	0.74	0.00 0.00	0.06	2.80*
	0.00, 0.00			0.00, 0.00			0.00, 0.00		
Political orientation	-0.07 0.02	-0.07	-3.22**	-0.08 0.02	-0.07	-3.82***	-0.06 0.02	-0.06	-2.96*
	-0.12, -0.03			-0.12, -0.04			-0.10, -0.02		
Diet – Some restrictions	0.00 0.11	0.00	0.01	-0.02 0.10	0.00	-0.20	-0.13 0.10	-0.03	-1.41
	-0.21, 0.21			-0.21, 0.17			-0.32, 0.05		
Diet – Pescatarian	0.12 0.13	0.02	0.91	0.08 0.12	0.01	0.67	-0.01 0.12	0.00	-0.06
	-0.14, 0.39			-0.15, 0.31			-0.24, 0.22		
Diet – Vegan	-0.34 0.20	-0.03	-1.68	-0.62 0.18	-0.06	-3.45**	-0.10 0.18	-0.01	-0.55
	-0.74, 0.06			-0.97, -0.27			-0.45, 0.25		
Diet – Vegetarian	0.02 0.13	0.00	0.14	0.08 0.11	0.01	0.70	0.09 0.11	0.02	0.80
	-0.23, 0.26			-0.14, 0.29			-0.13, 0.30		
Religion – Not religious	0.14 0.06	0.05	2.56*	0.20 0.05	0.08	4.08***	0.01 0.05	0.00	0.25
	0.03, 0.25			0.10, 0.30			-0.08, 0.11		
AI ownership – Yes	0.15 0.06	0.05	2.37	0.07 0.05	0.03	1.32	0.12 0.05	0.05	2.14
	0.03, 0.27			-0.03, 0.18			0.01, 0.22		
AI exposure	0.05 0.02	0.05	2.05	0.03 0.02	0.04	1.46	0.10 0.02	0.12	4.57***
	0.00, 0.10			-0.01, 0.07			0.05, 0.14		
Mind perception	0.01 0.00	0.15	5.48***	0.01 0.00	0.14	5.28***	0.01 0.00	0.21	7.70***
	0.01, 0.01			0.00, 0.01			0.01, 0.01		

Policy attitudes	0.03	0.04	1.73	0.04	0.05	2.51*	0.00	-0.01	-0.28
	0.02			0.01			0.01		
	0.00, 0.06			0.01, 0.06			-0.03, 0.02		
Perceived threat	0.03	0.04	1.80	0.04	0.05	2.51*	0.02	0.02	1.04
	0.02			0.01			0.01		
	0.00, 0.06			0.01, 0.07			-0.01, 0.04		
Techno-Animism	0.06	0.09	2.86*	0.14	0.19	6.85***	0.02	0.03	1.07
	0.02			0.02			0.02		
	0.02, 0.11			0.10, 0.18			-0.02, 0.06		
Anthropomorphism	0.06	0.13	3.75***	0.08	0.18	5.90***	0.05	0.12	3.56**
	0.02			0.01			0.01		
	0.03, 0.09			0.05, 0.10			0.02, 0.07		
Time – 2023	-0.09	-0.04	-1.84	-0.07	-0.03	-1.62	-0.02	-0.01	-0.51
	0.05			0.04			0.04		
	-0.19, 0.01			-0.16, 0.02			-0.11, 0.06		

Note. FDR-adjusted *p*-values are reported. The reference groups for categorical variables were "man" for Gender, "South" for Region, "White" for Ethnicity, "meat-eater" for Diet, "religious" for Religion, "do not own" for AI Ownership, and "2021" for Time. Some predictors were significant prior to FDR adjustment: Exploitation (gender, AI ownership, AI exposure), Cruelty (age), Research Use (Asian ethnicity, AI ownership). *$p < .05$, **$p < .01$, ***$p < .001$

Table 3

Regression models predicting forecasts about future social issues

	Welfare			*Advocacy*			*Unhappiness Reduction*		
	b	β	*t*	*b*	β	*t*	*b*	β	*t*
	SE			SE			SE		
	CI			CI			CI		
Intercept	1.64	-	15.28***	1.69	-	16.27***	1.71	-	15.24***
	0.11			0.10			0.11		
	1.43, 1.85			1.49, 1.89			1.49, 1.93		
Age	0.01	0.07	3.68***	0.01	0.09	4.53***	0.00	0.04	2.01
	0.00			0.00			0.00		
	0.00, 0.01			0.00, 0.01			0.00, 0.01		
Gender – Woman	0.11	0.04	2.52*	0.09	0.04	2.20	0.10	0.04	2.19
	0.04			0.04			0.05		
	0.02, 0.19			0.01, 0.18			0.01, 0.19		

Region – Midwest	0.10	0.03	1.70	0.01	0.00	0.26	-0.03	-0.01	-0.56
	0.06			0.05			0.06		
	-0.01, 0.21			-0.09, 0.12			-0.15, 0.08		
Region – Northeast	0.04	0.01	0.65	0.05	0.02	0.87	-0.05	-0.02	-0.78
	0.06			0.06			0.06		
	-0.08, 0.15			-0.06, 0.16			-0.17, 0.07		
Region – West	0.15	0.05	2.62*	0.12	0.04	2.16	0.02	0.01	0.33
	0.06			0.05			0.06		
	0.04, 0.25			0.01, 0.22			-0.09, 0.13		
Ethnicity – Asian	-0.09	-0.02	-0.96	-0.12	-0.02	-1.37	-0.20	-0.04	-2.10
	0.09			0.09			0.10		
	-0.27, 0.09			-0.29, 0.05			-0.39, -0.01		
Ethnicity – Black	-0.01	0.00	-0.08	-0.05	-0.01	-0.72	0.03	0.01	0.38
	0.07			0.07			0.08		
	-0.15, 0.14			-0.19, 0.09			-0.12, 0.18		
Ethnicity – Hispanic	0.25	0.07	3.81***	0.08	0.02	1.20	0.14	0.04	2.03
	0.07			0.06			0.07		
	0.12, 0.38			-0.05, 0.20			0.00, 0.28		
Ethnicity – Indigenous	0.00	0.00	0.00	0.02	0.00	0.12	0.07	0.01	0.31
	0.21			0.20			0.22		
	-0.40, 0.40			-0.37, 0.42			-0.35, 0.49		
Ethnicity – Other	-0.12	-0.01	-0.84	-0.12	-0.01	-0.84	0.17	0.02	1.09
	0.14			0.14			0.15		
	-0.41, 0.16			-0.39, 0.16			-0.13, 0.46		
Education	-0.01	-0.02	-0.82	0.00	0.00	0.19	-0.01	-0.02	-1.02
	0.01			0.01			0.01		
	-0.03, 0.01			-0.02, 0.02			-0.03, 0.01		
Income	0.00	0.00	-0.12	0.00	-0.02	-1.23	0.00	0.02	1.03
	0.00			0.00			0.00		
	0.00, 0.00			0.00, 0.00			0.00, 0.00		
Political orientation	-0.10	-0.09	-5.31***	-0.12	-0.11	-6.24***	-0.12	-0.11	-5.74***
	0.02			0.02			0.02		
	-0.14, -0.06			-0.15, -0.08			-0.16, -0.08		
Diet – Some restrictions	-0.04	-0.01	-0.42	0.00	0.00	0.00	0.10	0.02	1.04
	0.09			0.09			0.10		
	-0.22, 0.14			-0.18, 0.18			-0.09, 0.29		
Diet – Pescatarian	-0.01	0.00	-0.05	0.03	0.00	0.29	-0.01	0.00	-0.07
	0.11			0.11			0.12		
	-0.23, 0.22			-0.19, 0.25			-0.24, 0.23		

Diet – Vegan	-0.07	-0.01	-0.39	-0.18	-0.02	-1.07	-0.09	-0.01	-0.52
	0.17			0.17			0.18		
	-0.41, 0.27			-0.51, 0.15			-0.45, 0.26		
Diet – Vegetarian	0.07	0.01	0.64	0.09	0.01	0.85	0.13	0.02	1.13
	0.11			0.10			0.11		
	-0.14, 0.28			-0.11, 0.29			-0.09, 0.35		
Religion – Not religious	0.07	0.02	1.40	0.12	0.04	2.57*	0.08	0.03	1.63
	0.05			0.05			0.05		
	-0.03, 0.16			0.03, 0.21			-0.02, 0.18		
AI ownership – Yes	0.10	0.04	1.98	0.11	0.04	2.11	0.08	0.03	1.49
	0.05			0.05			0.05		
	0.00, 0.21			0.01, 0.21			-0.03, 0.19		
AI exposure	0.09	0.10	4.34***	0.09	0.10	4.52***	0.04	0.04	1.67
	0.02			0.02			0.02		
	0.05, 0.13			0.05, 0.13			-0.01, 0.08		
Mind perception	0.01	0.20	8.55***	0.01	0.18	8.17***	0.01	0.20	8.12***
	0.00			0.00			0.00		
	0.01, 0.01			0.01, 0.01			0.01, 0.01		
Policy attitudes	-0.02	-0.03	-1.32	-0.02	-0.03	-1.49	0.00	-0.01	-0.25
	0.01			0.01			0.01		
	-0.04, 0.01			-0.05, 0.01			-0.03, 0.02		
Perceived threat	-0.02	-0.03	-1.51	-0.05	-0.06	-3.44**	0.02	0.03	1.64
	0.01			0.01			0.01		
	-0.05, 0.01			-0.07, -0.02			0.00, 0.05		
Techno-Animism	0.15	0.20	7.64***	0.19	0.26	10.39***	0.11	0.15	5.55***
	0.02			0.02			0.02		
	0.11, 0.19			0.16, 0.23			0.07, 0.15		
Anthropomorphism	0.09	0.21	7.37***	0.10	0.21	7.65***	0.09	0.21	6.88***
	0.01			0.01			0.01		
	0.07, 0.12			0.07, 0.12			0.07, 0.12		
Time – 2023	-0.09	-0.04	-2.22	-0.04	-0.02	-1.07	-0.11	-0.04	-2.40
	0.01			0.04			0.04		
	0.07, 0.12			-0.12, 0.04			-0.19, -0.02		

Note. FDR-adjusted *p*-values are reported. The reference groups for categorical variables were "man" for Gender, "South" for Region, "White" for Ethnicity, "meat-eater" for Diet, "religious" for Religion, "do not own" for AI Ownership, and "2021" for Time. Some predictors were significant prior to FDR adjustment: Welfare (AI ownership, time), Advocacy (gender, West region, AI ownership), Unhappiness Reduction (age, gender, Asian ethnicity, Hispanic ethnicity, time). *$p < .05$, **$p < .01$, ***$p < .001$

7 Discussion

US Americans' imagined future world with sentient AI showed middling expectations for the future with unhappiness, use of sentient AI in medical and scientific research, and sentient AI labour exploitation more common than cruel treatment of AI, AI welfare mattering, and AI rights initiatives being necessary. This vision coheres with current functional uses of AI in the workplace, as replacements for biological nonhuman animal research subjects (Madden et al., 2020; Parker & Grote, 2022), and futurist visions of technological competence without morality (Fernando et al., 2023). The AIMS data suggest that Americans expect a continuation of socio-technological change along current lines.

The primary psychological predictors of Americans' expectations for this imagined future world were AI mind perception, the tendency to anthropomorphise, and techno-animist beliefs (i.e., believing AI systems have a soul or inner essence). They predicted forecasting more cruel treatment, more exploitation, more use of sentient AI as research subjects, that AI welfare would be an important social issue, AI rights advocacy would be necessary, and that it would be important to reduce the percentage of unhappy sentient AI systems. Specifically, increased perception of minds in current AI systems and a stronger tendency to anthropomorphise predicted all forecasts. Techno-animism predicted all forecasts except research use. The psychological processes of recognising and attributing mind may be critical to world-making for a distant future with sentient AI because it facilitates moral inclusion consistent with previous research connecting mind perception to morality (K. Gray et al., 2012). Further, it may promote thinking about the future experiences of sentient AI. Likewise, anthropomorphism and animism may aid the expectation of a morally worse future for AI systems (e.g., being treated cruelly, exploited, etc.) conceived of as persons and promote acting now to ward off that future.

Perceived threat and current preferences for cautious development policies were less predictive of future forecasts. More perceived threat predicted forecasting more cruel treatment and less necessity of AI rights advocacy in the future, suggesting a potential dismissal of future sentient AI moral status. More cautious policy attitudes about the development of sentient AI and integrated human-AI technologies predicted forecasting more cruel treatment, perhaps in recognition that developing sentience-related AI technologies could lead to more future cruelty.

Of the demographic predictors, only liberal political orientation predicted all forecasts. Liberals envisioned more exploitation, crueller treatment, more use as biomedical research subjects, that AI welfare would be an important social issue, AI rights advocacy would be necessary, and that it would be important to reduce the percentage of unhappy sentient AI systems. This pattern is consistent with the relationship that liberal political orientation has with moral values such as relying most on the harm/care and fairness/reciprocity moral foundations (Graham et al., 2009) and extending the circle of moral consideration more universally, including to nonhumans, than does a conservative political orientation (Waytz et al., 2019).

There was also a demographic pattern of older age and more exposure to AI forecasting a worse future world, although age was not predictive of forecasted cruelty or the need to reduce the percentage of unhappy AI. AI exposure predicted research use, welfare importance, and advocacy necessity. Younger people may expect a more positive or moral future in which all sentient entities are respected regardless of the social or technological context. Alternatively, older people might expect a continuation of historically oppressive trends given their longer experience in society. Age differences in optimism have been demonstrated previously (Durbin et al., 2019; You et al.,

2009), although research has not systematically tested age effects on optimism about technological or moral futures to our knowledge. The effect of AI exposure might stem from greater exhibition of AI potential (e.g., for replacing biological entities in research) through direct experience or exposure to ethical arguments about the use and treatment of AI in science fiction as well as philosophical scholarship. There was no evidence of significant change in future forecasts from 2021 to 2023 following FDR multiple comparison correction. The unadjusted regression results showed a temporal change in forecasting that AI welfare is important and that it would be important to reduce the percentage of unhappy sentient AI systems. These were forecasted to be less important in Americans' vision of the future for respondents in 2023 than in 2021. This lack of significant results suggests that forecasting about AI futures is not necessarily directly tied to radical technological advances like ChatGPT (which was released in late 2022) although some visions might be affected by technological advances and how they alter the moral and social discourse.

7.1 Limitations

This research was conducted with a nationally representative sample recruited based on the official American Community Survey census estimates for age, education, ethnicity, gender, income, and region. The AIMS data is more generalisable to the adult US population than most US-based quantitative social psychological research, and the imagined future of this representative sample profiles the cultural expectations of contemporary Americans, including underrepresented, underpowered, and marginalised groups. This is important within a topic such as human-AI interaction where many of the viewpoints and much of the research still predominantly arises from the study of traditionally represented and highly privileged social groups by these same groups. Despite the nationally representative sample however, the survey methodology limits us from making firm conclusions about specific underrepresented perspectives.

With the AIMS data, we could not account for the many cognitive biases that might affect forecasting accuracy, such as optimism bias or present bias (Sweeny et al., 2006; Windschitl & Stuart, 2015). Optimism bias might lead to rosier expectations for the future than would be likely to occur. Present bias might lead to more unstable forecasts for the future given a lesser interest in the future than the present. People with a present bias might also neglect thinking about uncertain factors or possible future technologies such as sentient AI. They might default to forecasting a future consistent with current beliefs and realities. People have representations of sentient AI from science fiction narratives but no direct experiences with sentient AI to guide their imagination. AIMS participants were asked to imagine what a future with widespread sentient AI would be like without detailed instructions. Some people might have imagined a society filled with feeling, human-like android companions and others a society filled with sapient, omniscient, and non-embodied computer workers, amongst many possibilities. Alternatively, some participants might not be able to imagine a future with sentient AI at all given the hypotheticality of the technology and individual differences in mental simulation and imagination processes (Kosslyn et al., 1984; Mak & Willems, 2019; O'Connor & Aardema, 2005; Smallwood & Schooler, 2015).

Further, representations of sentient AI likely also differ in conceptualisations of how sentience overlaps with other advanced capacities (e.g., complex cognition, having preferences, social skills). Research on expectations about the intersectionality of these capacities and how this affects the adoption and regulation of AI is needed, particularly in light of philosophical perspectives that many capacities might give rise to AI mattering morally (Ladak, 2023).

Regardless of the diversity of imagined futures and AI capacities, it is valuable to examine the raw intuitions of the general public because their forecasts could influence their support of policies, enforcement of human-AI interaction norms, use of new technologies, perception of mind in AI systems, and treatment of current AI that could anchor future treatment. Aggregating responses across a diversity of envisioned futures is also a strength because many future worlds are possible and the factors that predict aggregated forecasts, such as AI mind perception, may have an outsize effect on future world-making.

7.2 Future Directions

Attitudes towards AI have been shown to differ between ethnic groups in the USA (Johnson et al., 2015), pointing to the importance of nationally representative studies such as AIMS in research-based future world-making, particularly in stratified multicultural societies where some groups (e.g., from WEIRD cultures; Henrich et al., 2010) participate in research more than others. Extending studies like this to other nations and cultures would enable expanded insights on pluralistic values and forecasts about human-AI futures. Different modes of thinking about the future may manifest in cross-cultural attitudinal differences towards AI like more ambivalence in the USA than China (Dang & Liu, 2021) and Japan (Nomura et al., 2008). Future research could construct profiles of forecasts about AI futures and their psychological predictors in multiple countries around the world to illustrate whether and where values, norms, and forecasts converge into a global profile and where they diverge into pluralistic sets.

In addition to expanding into other cultural contexts, surveying a wider range of forecasts about future human-AI interactions, such as the expected benefits and costs of a future world filled with sentient AI would advance our understanding of future world-making processes. We analysed the forecasts available within the AIMS survey data but there are many more possible forecasts that people may routinely engage in about human-AI futures that merit future examination (e.g., that sentient AI will be friendly companions, that sentient AI will improve human productivity). Additionally, combining nationally representative survey research with case studies, observational, and interview-based studies could highlight the scale and diversity of values, norms, and future world-making endeavours in a variety of world-making contexts (e.g., policymakers thinking about the future, AI safety advocates lobbying policymakers).

Our study focused on the relationship between psychological factors and what we term world-making beliefs, attitudes, and expectations (Figure 1). By broadening our understanding of these stages, we gain insight into world-making processes that could shape the efficacy of AI literacy education and future world-making education. Future research on how psychological antecedents and world-making beliefs, attitudes, and expectations relate to world-making actions and interventions could result in stronger implications for human-AI future world-making and futures literacy, generally. For instance, if current values and norms enshrine the blanket moral exclusion of AI rather than engaging in critical ongoing reflection about AI capacities and social integration, then world-making actions might lock-in intolerance, exclusion, and prejudice. If people expect AI to be subservient they may enact laws and regulations now that enshrine this AI role. This could be damaging for the long-term since adherence to laws and mandates has been shown to increase social norm compliance and enforcement (Mulder et al., 2024). A harmful feedback loop could emerge if laws are written now that set precedence against the inclusion of any AI under the position that they will never qualify for moral status. Alternatively, critically-

engaged AI-inclusive policies might shape the future by norming the idea of moral consideration for deserving AI systems, if implemented cautiously.

One fruitful avenue for world-making for a technological utopian or "protopian" (i.e., progressively improving future; Kelly, 2010) future would be to invest more into understanding the plurality of human values in socio-technological contexts, how global value systems emerge between cultures, and how forecasts about the future arise from values and transfer into world-making actions. Research on human value systems like moral foundations (Graham et al., 2013) or personal values (Sagiv et al., 2017; Sagiv & Schwartz, 2022) has made strides on the psychological study of values. Approaches to AI development and regulatory policies that consider values pluralism and working towards a future endorsed by most humans begins with understanding attitudes towards AI as well as how people envision socio-technological futures.

Targeted hypothesis testing of specific relationships between psychological antecedents, cognitive, and action-oriented world-making is also necessary. Research on how individual differences in mentally simulating imagined futures impacts forecasts about the future and actions taken to build that future would broaden our understanding of the relationship between psychological antecedents, cognitive, and behavioural world-making. This could involve asking more nuanced questions about specific policies like regulating the development of AI applications akin to the development of other technologies (e.g., medical assistive devices) to extend beyond support for global bans or regulations such as those that we assessed.

8 Conclusion

Building on Power et al.'s (2023) theoretical framework for the social psychology of world-making, we examined nationally representative longitudinal survey data illustrating expectations for an imagined future with sentient AI, a first, cognitive stage of world-making. In the current critical era of AI-related social change, the study of the attitudes, norms, and values that influence AI policy and social integration over the next 25-50 years is needed to build the long-term social and moral future we want with AI. Social psychological world-making research could help to identify how people imagine humanity's future with AI and points at which advocates could begin laying the groundwork for moral circle expansion that decreases the risk of future suffering and increases the opportunity for a harmonious future. Research on AI mind perception, attitudes and norms regarding human-AI relations, and how expectations for future human-AI interaction change over time will undergird the choices we make as we build the socio-technological future.

References

Agerström, J., & Björklund, F. (2013). Why people with an eye toward the future are more moral: The role of abstract thinking. *Basic and Applied Social Psychology*, *35*(4), 373–381. https://doi.org/10.1080/01973533.2013.803967

Anthis, J. R., & Paez, E. (2021). Moral circle expansion: A promising strategy to impact the far future. *Futures*, *130*, 102756. https://doi.org/10.1016/j.futures.2021.102756

Axtell, B., & Munteanu, C. (2021). Tea, Earl Grey, Hot: Designing Speech Interactions from the Imagined Ideal of Star Trek. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3411764.3445640

Badaan, V., Jost, J. T., Fernando, J., & Kashima, Y. (2020). Imagining better societies: A social psychological framework for the study of utopian thinking and collective action. *Social and Personality Psychology Compass*, *14*(4), e12525. https://doi.org/10.1111/spc3.12525

Banks, I. M. (1987). *The Culture*. Orbit Books.

Baumann, T. (2022). *Avoiding the Worst: How to prevent a moral catastrophe*. Center for Reducing Suffering.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Beran, T. N., Ramirez-Serrano, A., Kuzyk, R., Fior, M., & Nugent, S. (2011). Understanding how children understand robots: Perceived animism in child–robot interaction. *International Journal of Human-Computer Studies*, *69*(7), 539–550. https://doi.org/10.1016/j.ijhcs.2011.04.003

Bosone, L., Thiriot, S., Chevrier, M., Rocci, A., & Zenasni, F. (2023). Visioning sustainable futures: Exposure to positive visions increases individual and collective intention to act for a decarbonated world. *Global Environmental Psychology*. https://www.psycharchives.org/en/item/c2dd0b69-f485-4e92-b9d0-aee13b81c39a

Cameron, J. (Director). (1984). *The Terminator* [Video recording]. Orion Pictures.

Čapek, K. (1923). *R.U.R. (Rossum's Universal Robots)*. Humphrey Milford Oxford University Press.

Coghlan, S., Vetere, F., Waycott, J., & Barbosa Neves, B. (2019). Could social robots make us kinder or crueller to humans and animals? *International Journal of Social Robotics*, *11*(5), 741–751. https://doi.org/10.1007/s12369-019-00583-2

Cona, G., Santacesaria, P., & Scarpazza, C. (2023). Envisioning the future: An ALE meta-analysis on neural correlates of future thinking, prospective memory and delay discounting. *Neuroscience & Biobehavioral Reviews*, *153*, 105355. https://doi.org/10.1016/j.neubiorev.2023.105355

Conty, A. (2022). Animism in the Anthropocene. *Theory, Culture & Society*, *39*(5), 127–153. https://doi.org/10.1177/02632764211039283

Crimston, C. R., Bain, P. G., Hornsey, M. J., & Bastian, B. (2016). Moral expansiveness: Examining variability in the extension of the moral world. *Journal of Personality and Social Psychology*, *111*(4), 636–653. https://doi.org/10.1037/pspp0000086

Dang, J., & Liu, L. (2021). Robots are friends as well as foes: Ambivalent attitudes toward mindful and mindless AI robots in the United States and China. *Computers in Human Behavior*, *115*, 106612. https://doi.org/10.1016/j.chb.2020.106612

D'Argembeau, A., & Van der Linden, M. (2007). Emotional aspects of mental time travel. *Behavioral and Brain Sciences*, *30*(3), 320–321. https://doi.org/10.1017/S0140525X07002051

de Saint-Laurent, C. (2018). In Defence of Machine Learning: Debunking the Myths of Artificial Intelligence. *Europe's Journal of Psychology*, *14*(4), Article 4. https://doi.org/10.5964/ejop.v14i4.1823

de Saint-Laurent, C., Obradović, S., & Carriere, K. R. (2018). *Imagining Collective Futures: Perspectives from Social, Cultural and Political Psychology*. Palgrave Macmillan Cham. https://doi.org/10.1007/978-3-319-76051-3

Durbin, K. A., Barber, S. J., Brown, M., & Mather, M. (2019). Optimism for the Future in Younger and Older Adults. *The Journals of Gerontology: Series B*, *74*(4), 565–574. https://doi.org/10.1093/geronb/gbx171

Fernando, J. W., Burden, N., Ferguson, A., O'Brien, L. V., Judge, M., & Kashima, Y. (2018). Functions of Utopia: How Utopian Thinking Motivates Societal Engagement. *Personality and Social Psychology Bulletin*, *44*(5), 779–792. https://doi.org/10.1177/0146167217748604

Fernando, J. W., Burden, N., Judge, M., O'Brien, L. V., Ashman, H., Paladino, A., & Kashima, Y. (2023). Profiles of an Ideal Society: The Utopian Visions of Ordinary People. *Journal of Cross-Cultural Psychology*, *54*(1), 43–60. https://doi.org/10.1177/00220221221126419

Ferrari, F., Paladino, M. P., & Jetten, J. (2016). Blurring human–machine distinctions: Anthropomorphic appearance in social robots as a threat to human distinctiveness. *International Journal of Social Robotics*, *8*(2), 287–302. https://doi.org/10.1007/s12369-016-0338-y

Förster, J., Friedman, R. S., & Liberman, N. (2004). Temporal Construal Effects on Abstract and Concrete Thinking: Consequences for Insight and Creative Cognition. *Journal of Personality and Social Psychology*, *87*(2), 177–189. https://doi.org/10.1037/0022-3514.87.2.177

Fröding, B., & Peterson, M. (2020). Friendly AI. *Ethics and Information Technology*. https://doi.org/10.1007/s10676-020-09556-w

Fujii, T., Guo, T., & Kamoshida, A. (2018). A consideration of service strategy of Japanese electric manufacturers to realize super smart society (SOCIETY 5.0). In L. Uden, B. Hadzima, & I.-H. Ting (Eds.), *Knowledge Management in Organizations* (pp. 634–645). Springer International Publishing. https://doi.org/10.1007/978-3-319-95204-8_53

Fuller, B., & Kurtzman, A. (Directors). (2017, 2024). *Star Trek: Discovery* [Science Fiction]. CBS Studios.

Gibert, M., & Martin, D. (2022). In search of the moral status of AI: Why sentience is a strong argument. *AI & SOCIETY*, *37*(1), 319–330. https://doi.org/10.1007/s00146-021-01179-z

Giger, J.-C., Piçarra, N., Alves-Oliveira, P., Oliveira, R., & Arriaga, P. (2019). Humanization of robots: Is it really such a good idea? *Human Behavior and Emerging Technologies*, *1*(2), 111–123. https://doi.org/10.1002/hbe2.147

Gordon, J.-S., & Gunkel, D. J. (2021). Moral Status and Intelligent Robots. *The Southern Journal of Philosophy*. https://doi.org/10.1111/sjp.12450

Grace, K., Impacts, A., Stewart, H., Sandkühler, J. F., Thomas, S., Weinstein-Raun, B., & Brauner, J. (2023). Thousands of AI authors on the future of AI. *Unpublished Preprint*.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Chapter Two - Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism. In P. Devine & A. Plant (Eds.), *Advances in Experimental Social Psychology* (Vol. 47, pp. 55–130). Academic Press. https://doi.org/10.1016/B978-0-12-407236-7.00002-4

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*, 1029–1046. https://doi.org/10.1037/a0015141

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619–619. https://doi.org/10.1126/science.1134475

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology*, *96*(3), 505–520. https://doi.org/10.1037/a0013748

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry*, *23*(2), 101–124. https://doi.org/10.1080/1047840X.2012.651387

Gunkel, D. J. (2018). *Robot rights*. MIT Press. https://mitpress.mit.edu/books/robot-rights

Hadarics, M., & Kende, A. (2018). Moral foundations of positive and negative intergroup behavior: Moral exclusion fills the gap. *International Journal of Intercultural Relations*, *64*, 67–76. https://doi.org/10.1016/j.ijintrel.2018.03.006

Harrington, C. N., Klassen, S., & Rankin, Y. A. (2022). "All that You Touch, You Change": Expanding the Canon of Speculative Design Towards Black Futuring. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–10. https://doi.org/10.1145/3491102.3502118

Harris, J., & Anthis, J. R. (2021). The moral consideration of artificial entities: A literature review. *Science and Engineering Ethics*, *27*(4), 53. https://doi.org/10.1007/s11948-021-00331-8

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Higgins, E. T., Rossignac-Milon, M., & Echterhoff, G. (2021). Shared Reality: From Sharing-Is-Believing to Merging Minds. *Current Directions in Psychological Science*, *30*(2), 103–110. https://doi.org/10.1177/0963721421992027

Hohendanner, M., Ullstein, C., Buchmeier, Y., & Grossklags, J. (2023). Exploring the Reflective Space of AI Narratives Through Speculative Design in Japan and Germany. *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*, 351–362. https://doi.org/10.1145/3582515.3609554

Jensen, C. B., & Blok, A. (2013). Techno-animism in Japan: Shinto cosmograms, actor-network theory, and the enabling powers of non-human agencies. *Theory, Culture & Society*, *30*(2), 84–115. https://doi.org/10.1177/0263276412456564

Johnson, K. A., Cohen, A. B., Neel, R., Berlin, A., & Homa, D. (2015). Fuzzy people: The roles of kinship, essence, and sociability in the attribution of personhood to nonliving, nonhuman agents. *Psychology of Religion and Spirituality*, *7*(4), 295–305. https://doi.org/10.1037/rel0000048

Jordan, P., & Auernheimer, B. (2018). The Fiction in Computer Science: A Qualitative Data Analysis of the ACM Digital Library for Traces of Star Trek. In T. Ahram & C. Falcão (Eds.), *Advances in Usability and User Experience* (pp. 508–520). Springer International Publishing. https://doi.org/10.1007/978-3-319-60492-3_48

Jordan, P., & Silva, P. A. (2021). Science Fiction—An Untapped Opportunity in HCI Research and Education. In M. M. Soares, E. Rosenzweig, & A. Marcus (Eds.), *Design, User Experience, and Usability: UX Research and Design* (pp. 34–47). Springer International Publishing. https://doi.org/10.1007/978-3-030-78221-4_3

Kashima, Y., & Fernando, J. (2020). Utopia and ideology in cultural dynamics. *Current Opinion in Behavioral Sciences*, *34*, 102–106. https://doi.org/10.1016/j.cobeha.2020.01.002

Kelly, K. (2010). *What Technology Wants*. Penguin Books. https://kk.org/books/what-technology-wants

Koban, K., & Banks, J. (2024). It feels, therefore it is: Associations between mind perception and mind ascription for social robots. *Computers in Human Behavior*, *153*, 108098. https://doi.org/10.1016/j.chb.2023.108098

Kooij, D. T. A. M., Kanfer, R., Betts, M., & Rudolph, C. W. (2018). Future time perspective: A systematic review and meta-analysis. *Journal of Applied Psychology*, *103*(8), 867–893. https://doi.org/10.1037/apl0000306

Kosslyn, S. M., Brunn, J., Cave, K. R., & Wallach, R. W. (1984). Individual differences in mental imagery ability: A computational analysis. *Cognition*, *18*(1), 195–243. https://doi.org/10.1016/0010-0277(84)90025-8

Küster, D., & Swiderska, A. (2021). Seeing the mind of robots: Harm augments mind perception but benevolent intentions reduce dehumanisation of artificial entities in visual vignettes. *International Journal of Psychology*, *56*(3), 454–465. https://doi.org/10.1002/ijop.12715

Ladak, A. (2023). What would qualify an artificial intelligence for moral standing? *AI and Ethics*. https://doi.org/10.1007/s43681-023-00260-1

Ladak, A., Harris, J., & Anthis, J. R. (2024). Which Artificial Intelligences Do People Care About Most? A Conjoint Experiment on Moral Consideration. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–11. https://doi.org/10.1145/3613904.3642403

Ladak, A., Wilks, M., & Anthis, J. R. (2023). Extending Perspective Taking to Nonhuman Animals and Artificial Entities. *Social Cognition*, *41*(3), 274–302. https://doi.org/10.1521/soco.2023.41.3.274

Linehan, C., Kirman, B. J., Reeves, S., Blythe, M. A., Tanenbaum, T. J., Desjardins, A., & Wakkary, R. (2014). Alternate endings: Using fiction to explore design futures. *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, 45–48. https://doi.org/10.1145/2559206.2560472

Lupetti, M. L., Smit, I., & Cila, N. (2018). Near future cities of things: Addressing dilemmas through design fiction. *Proceedings of the 10th Nordic Conference on Human-Computer Interaction*, 787–800. https://doi.org/10.1145/3240167.3240273

Madden, J. C., Enoch, S. J., Paini, A., & Cronin, M. T. D. (2020). A Review of In Silico Tools as Alternatives to Animal Testing: Principles, Resources and Applications. *Alternatives to Laboratory Animals*, *48*(4), 146–172. https://doi.org/10.1177/0261192920965977

Mak, M., & Willems, R. M. (2019). Mental simulation during literary reading: Individual differences revealed with eye-tracking. *Language, Cognition and Neuroscience*, *34*(4), 511–535. https://doi.org/10.1080/23273798.2018.1552007

Manoli, A., Pauketat, J. V. T., & Anthis, J. R. (2024). *The AI Double Standard: Humans Judge All AIs for the Actions of One*. arXiv. https://doi.org/10.48550/arXiv.2412.06040

Marenko, B. (2014). Neo-animism and design: A new paradigm in object theory. *Design and Culture*, *6*. https://doi.org/10.2752/175470814X14031924627185

Metzinger, T. (2021). Artificial suffering: An argument for a global moratorium on synthetic phenomenology. *Journal of Artificial Intelligence and Consciousness*, 1–24. https://doi.org/10.1142/S270507852150003X

Mubin, O., Wadibhasme, K., Jordan, P., & Obaid, M. (2019). Reflecting on the Presence of Science Fiction Robots in Computing Literature. *J. Hum.-Robot Interact.*, *8*(1), 5:1-5:25. https://doi.org/10.1145/3303706

Mulder, L. B., Kurz, T., Prosser, A. M. B., & Fonseca, M. A. (2024). The presence of laws and mandates is associated with increased social norm enforcement. *Journal of Economic Psychology*, *101*, 102703. https://doi.org/10.1016/j.joep.2024.102703

Nagahara, M. (2019). A research project of Society 5.0 in Kitakyushu, Japan. *2019 IEEE Conference on Control Technology and Applications (CCTA)*, 803–804. https://doi.org/10.1109/CCTA.2019.8920449

Nijssen, S. R. R., Müller, B. C. N., Baaren, R. B. van, & Paulus, M. (2019). Saving the Robot or the Human? Robots Who Feel Deserve Moral Care. *Social Cognition*, *37*(1), 41-S2. https://doi.org/10.1521/soco.2019.37.1.41

Nomura, T., Suzuki, T., Kanda, T., Han, J., Shin, N., Burke, J., & Kato, K. (2008). What people assume about humanoid and animal-type robots: Cross-cultural analysis between japan, korea, and the united states. *International Journal of Humanoid Robotics*, *05*(01), 25–46. https://doi.org/10.1142/S0219843608001297

O'Connor, K. P., & Aardema, F. (2005). The imagination: Cognitive, pre-cognitive, and meta-cognitive aspects. *Consciousness and Cognition*, *14*(2), 233–256. https://doi.org/10.1016/j.concog.2004.07.005

Okanda, M., Taniguchi, K., Wang, Y., & Itakura, S. (2021). Preschoolers' and adults' animism tendencies toward a humanoid robot. *Computers in Human Behavior*, *118*, 106688. https://doi.org/10.1016/j.chb.2021.106688

Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., Kawashima, R., Fukuda, H., Itoh, M., & Yamadori, A. (2003). Thinking of the future and past: The roles of the frontal pole and the medial temporal lobes. *NeuroImage*, *19*(4), 1369–1380. https://doi.org/10.1016/S1053-8119(03)00179-4

Opotow, S. (1990). Moral exclusion and injustice: An introduction. *Journal of Social Issues*, *46*(1), 1–20. https://doi.org/10.1111/j.1540-4560.1990.tb00268.x

Opotow, S. (1993). Animals and the scope of justice. *Journal of Social Issues*, *49*(1), 71–85. https://doi.org/10.1111/j.1540-4560.1993.tb00909.x

Packer, D. J., & Cunningham, W. A. (2009). Neural correlates of reflection on goal states: The role of regulatory focus and temporal distance. *Social Neuroscience*, *4*(5), 412–425. https://doi.org/10.1080/17470910902750186

Paiva, A., Mascarenhas, S., Petisca, S., Correia, F., & Alves-Oliveira, P. (2018). Towards more humane machines: Creating emotional social robots. In *New Interdisciplinary Landscapes in Morality and Emotion* (1st ed., pp. 125–139). Routledge. https://doi.org/10.4324/9781315143897-10

Parker, S. K., & Grote, G. (2022). Automation, Algorithms, and Beyond: Why Work Design Matters More Than Ever in a Digital World. *Applied Psychology*, *71*(4), 1171–1204. https://doi.org/10.1111/apps.12241

Pauketat, J. V. T., & Anthis, J. R. (2022). Predicting the moral consideration of artificial intelligences. *Computers in Human Behavior*, *136*, 107372. https://doi.org/10.1016/j.chb.2022.107372

Pauketat, J. V. T., Ladak, A., & Anthis, J. R. (2023a). *Artificial Intelligence, Morality, and Sentience (AIMS) Survey: 2023 Update*. PsyArXiv. https://doi.org/10.31234/osf.io/9xsav

Pauketat, J. V. T., Ladak, A., & Anthis, J. R. (2023b). *Artificial Intelligence, Morality, and Sentience (AIMS) Survey Data* [Dataset]. Mendeley Data. https://doi.org/10.17632/x5689yhv2n.2

Pedersen, M. A. (2001). Totemism, animism and north Asian indigenous ontologies. *The Journal of the Royal Anthropological Institute*, *7*(3), 411–427.

Peetz, J., & Wohl, M. J. A. (2019). Perceiving time through group-based glasses: Collective temporal orientation. *British Journal of Social Psychology*, *58*(3), 609–629. https://doi.org/10.1111/bjso.12291

Power, S. A., Zittoun, T., Akkerman, S., Wagoner, B., Cabra, M., Cornish, F., Hawlina, H., Heasman, B., Mahendran, K., Psaltis, C., Rajala, A., Veale, A., & Gillespie, A. (2023). Social Psychology of and for World-Making. *Personality and Social Psychology Review*, *27*(4), 378–392. https://doi.org/10.1177/10888683221145756

Rabinovich, A., Morton, T., & Postmes, T. (2010). Time perspective and attitude-behaviour consistency in future-oriented behaviours. *British Journal of Social Psychology*, *49*(1), 69–89. https://doi.org/10.1348/014466608X401875

Richardson, K. (2016). Technological animism: The uncanny personhood of humanoid machines. *Social Analysis*, *60*(1), 110–128. https://doi.org/10.3167/sa.2016.600108

Roser, M. (2024). AI timelines: What do experts in artificial intelligence expect for the future? *Our World in Data*. https://ourworldindata.org/ai-timelines

Sagiv, L., Roccas, S., Cieciuch, J., & Schwartz, S. H. (2017). Personal values in human life. *Nature Human Behaviour*, *1*(9), 630–639. https://doi.org/10.1038/s41562-017-0185-3

Sagiv, L., & Schwartz, S. H. (2022). Personal Values Across Cultures. *Annual Review of Psychology*, *73*(Volume 73, 2022), 517–546. https://doi.org/10.1146/annurev-psych-020821-125100

Savage, M. (2024). In defence of sociological description: A 'world-making' perspective. *The British Journal of Sociology*, *75*(3), 360–365. https://doi.org/10.1111/1468-4446.13083

Schwitzgebel, E. (2023). AI systems must not confuse users about their sentience or moral status. *Patterns*, *4*(8). https://doi.org/10.1016/j.patter.2023.100818

Scott, A. E., Neumann, D., Niess, J., & Woźniak, P. W. (2023). Do You Mind? User Perceptions of Machine Consciousness. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–19. https://doi.org/10.1145/3544548.3581296

Shulman, C., & Bostrom, N. (2021). Sharing the world with digital minds. In S. Clarke, H. Zohny, & J. Savulescu (Eds.), *Rethinking Moral Status*. Oxford University Press. https://nickbostrom.com/papers/digital-minds.pdf

Siegel, E. (2017). *Treknology: The Science of Star Trek from Tricorders to Warp Drive*. Voyageur Press.

Smallwood, J., & Schooler, J. W. (2015). The Science of Mind Wandering: Empirically Navigating the Stream of Consciousness. *Annual Review of Psychology*, *66*(Volume 66, 2015), 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Smith, E. R., Šabanović, S., & Fraune, M. R. (2021). Human-robot interaction through the lens of social psychological theories of intergroup behavior. *Technology, Mind, and Behavior*, *1*(2). https://doi.org/10.1037/tmb0000002

Sprenger, G. (2021). Can Animism Save the World? Reflections on Personhood and Complexity in the Ecological Crisis. *Sociologus*, *1*, 73–92. https://doi.org/10.3790/soc.71.1.73

Stillman, P. E., Lee, H., Deng, X., Unnava, H. R., Cunningham, W. A., & Fujita, K. (2017). Neurological evidence for the role of construal level in future-directed thought. *Social Cognitive and Affective Neuroscience*, *12*(6), 937–947. https://doi.org/10.1093/scan/nsx022

Suddendorf, T., & Corballis, M. C. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *The Behavioral and Brain Sciences*, *30*(3), 299–313; 313–351. https://doi.org/10.1017/S0140525X07001975

Sweeny, K., Carroll, P. J., & Shepperd, J. A. (2006). Is Optimism Always Best?: Future Outlooks and Preparedness. *Current Directions in Psychological Science*, *15*(6), 302–306. https://doi.org/10.1111/j.1467-8721.2006.00457.x

Swiderska, A., & Küster, D. (2018). Avatars in Pain: Visible Harm Enhances Mind Perception in Humans and Robots. *Perception*. https://doi.org/10.1177/0301006618809919

Syropoulos, S., Law, K. F., & Young, L. (2024). Longtermist Education Interventions Increase Concern for and Action to Protect Future Generations. *Social Psychological and Personality Science*, 19485506241228465. https://doi.org/10.1177/19485506241228465

Szpunar, K. K., Spreng, R. N., & Schacter, D. L. (2014). A taxonomy of prospection: Introducing an organizational framework for future-oriented cognition. *Proceedings of the National Academy of Sciences*, *111*(52), 18414–18421. https://doi.org/10.1073/pnas.1417144111

Tamir, D. I., & Thornton, M. A. (2018). Modeling the Predictive Social Mind. *Trends in Cognitive Sciences*, *22*(3), 201–212. https://doi.org/10.1016/j.tics.2017.12.005

Tanenbaum, T. J., Pufal, M., & Tanenbaum, K. (2016). W. *Proceedings of the Second Workshop on Computing within Limits*, 1–9. https://doi.org/10.1145/2926676.2926687

Tiku, N. (2022, June 11). The Google engineer who thinks the company's AI has come to life. *Washington Post*. https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/

Toch, H. (1965). *The social psychology of social movements* (pp. xiv, 257). Bobbs-Merrill Co.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463. https://doi.org/10.1037/a0018963

Turja, T., & Oksanen, A. (2019). Robot Acceptance at Work: A Multilevel Analysis Based on 27 EU Countries. *International Journal of Social Robotics*, *11*(4), 679–689. https://doi.org/10.1007/s12369-019-00526-x

van Stekelenburg, J., & Klandermans, B. (2013). The social psychology of protest. *Current Sociology*, *61*(5–6), 886–905. https://doi.org/10.1177/0011392113479314

Wang, X., & Krumhuber, E. G. (2018). Mind perception of robots varies with their economic versus social function. *Frontiers in Psychology*, *9*(1230). https://doi.org/10.3389/fpsyg.2018.01230

Ward, A. F., Olsen, A. S., & Wegner, D. M. (2013). The Harm-Made Mind: Observing Victimization Augments Attribution of Minds to Vegetative Patients, Robots, and the Dead. *Psychological Science*, *24*(8), 1437–1445. https://doi.org/10.1177/0956797612472343

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human?: The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, *5*(3), 219–232. https://doi.org/10.1177/1745691610369336

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Waytz, A., Iyer, R., Young, L., Haidt, J., & Graham, J. (2019). Ideological differences in the expanse of the moral circle. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-12227-0

Wilkinson, D. (2017). Is there such a thing as animism? *Journal of the American Academy of Religion*, *85*(2), 289–311. https://doi.org/10.1093/jaarel/lfw064

Windschitl, P. D., & Stuart, J. O. (2015). Optimism Biases. In *The Wiley Blackwell Handbook of Judgment and Decision Making* (pp. 431–455). https://doi.org/10.1002/9781118468333.ch15

Wong, J. S. (2018). Design and fiction: Imagining civic AI. *Interactions*, *25*(6), 42–45. https://doi.org/10.1145/3274568

Wu, F., Lu, C., Zhu, M., Chen, H., Zhu, J., Yu, K., Li, L., Li, M., Chen, Q., Li, X., Cao, X., Wang, Z., Zha, Z., Zhuang, Y., & Pan, Y. (2020). Towards a new generation of artificial intelligence in China. *Nature Machine Intelligence*, *2*(6), Article 6. https://doi.org/10.1038/s42256-020-0183-4

Wyche, S. (2022). Reimagining the Mobile Phone: Investigating Speculative Approaches to Design in Human-Computer Interaction for Development (HCI4D). *Proc. ACM Hum.-Comput. Interact.*, *6*(CSCW2), 535:1-535:27. https://doi.org/10.1145/3555648

Yogeeswaran, K., Złotowski, J., Livingstone, M., Bartneck, C., Sumioka, H., & Ishiguro, H. (2016). The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. *Journal of Human-Robot Interaction*, *5*(2), 29. https://doi.org/10.5898/JHRI.5.2.Yogeeswaran

You, J., Fung, H. H. L., & Isaacowitz, D. M. (2009). Age differences in dispositional optimism: A cross-cultural study. *European Journal of Ageing*, *6*(4), 247–252. https://doi.org/10.1007/s10433-009-0130-z

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and challenges in human–robot interaction. *International Journal of Social Robotics*, *7*(3), 347–360. https://doi.org/10.1007/s12369-014-0267-6

Złotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies*, *100*, 48–54. https://doi.org/10.1016/j.ijhcs.2016.12.008