A1 ADDITIONAL RESULTS

In this document, we discuss additional results to enrich and contextualize the results in the main text. As with those results, we focus on the 2023 main survey, and we note when there was a significant increase from 2021 to 2023. Additionally, in the 2021 preregistration, four researchers specified 80% credible intervals (the Bayesian analog of a frequentist confidence interval to represent subjective beliefs) for summary statistics based on 82 of the 86 questions. This ensured that we would know which results were surprisingly high, surprisingly low, or in line with our expectations. We also solicited predictions from a popular online forecasting platform for five survey questions in March 2022, prior to the results being shared outside of the research team.

Following these results for the total samples, we include predictive models that regress attitudes on demographics, such as gender and race. The survey questionnaires and raw data are also included in the supplementary materials for further analysis.

A1.1 2021 and 2023 Main Survey Waves

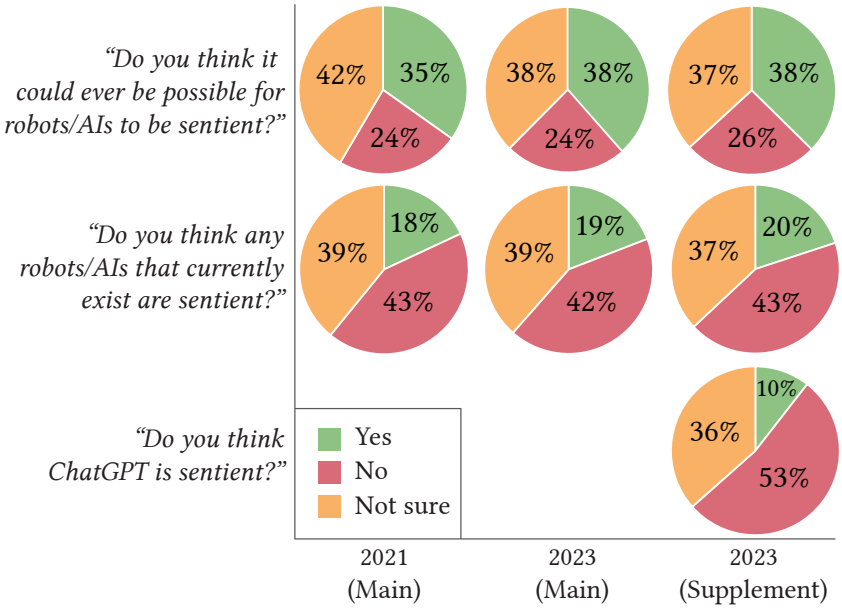

Fig. 1. Proportions of responses to selected questions about mind perception of sentient AI (rows) in each of the three survey waves (columns).

A1.1.1 Social Beliefs. In addition to tracking people's own beliefs about sentient AI, we included a question on social beliefs. Using the same Likert scale, we found 73.2% agreement in 2023 with the statement, "Most people who are important to me think that robots/AIs cannot have feelings."

A1.1.2 Subservience. We found 84.7% agreement with the statement, "Robots/AIs should be subservient to humans," a significant increase (p = 0.006) from 76.0% in 2021.

A1.1.3 Nonhumans. AIs constitute a new category of nonhuman entities in the world, so we also asked two questions on attitudes towards animals and two on attitudes towards the environment for potential comparative analysis (e.g., Are people who care more for animals and the environment likely to care more about AI?). We assessed agreement with the statements, "Animals deserve to be included in the moral circle" (90.8%), "The welfare of animals is one of the most important social issues in the world today" (82.2%), "The environment deserves to be included in the moral circle" (87.2%), and "The welfare of the environment is one of the most important social issues in the world today" (84.9%).

A1.1.4 Pairwise Comparisons: Target-Specific Moral Concern. In the main text, we listed the mean responses to each of 11 questions about target-specific moral concern for specific types of AIs. We also conducted pairwise comparisons between responses to each question using a generalized linear model, which in this case is equivalent to a weighted t-test of the difference between responses. In Table 1, we present the p-values for each pairwise comparison. For the ten adjacent comparisons, we include, in parentheses, the q-value when adjusted for the false discovery rate (i.e., significant when under 0.1 due to adjusting for a false discovery rate of 10%), as these comparisons are used to determine tiers of results, meaning a series of adjacent results that are not statistically significant from each other. The tiers of results are, in order from most to least concern:

(1) exact digital copies of human brains
(2) human-like companion robots
(3) human-like retail robots[a], animal-like companion robots[b], exact digital copies of animals, AI personal assistants[ab]
(4) complex language algorithms
(5) machine-like factory production robots
(6) machine-like cleaning robots, virtual avatars
(7) AI video game characters

where [a] and [b] indicate significant differences within a tier.

A1.1.5 Target-Specific Social Connection. Because of the close relationship between moral concern and social connection, we included a well-known measure of social connection known as Inclusion of Other in the Self (IOS) [1] in which participants are shown seven pairs of circles with varying degrees of overlap between the pair and asked, "Which pair of circles best represents how connected [type of AI] are to humans?" with the same 11 targets as in the target-specific moral concern measures.

	exact digital copies of human brains	human-like companion robots	human-like retail robots	animal-like companion robots	exact digital copies of animals	AI personal assistants	complex language algorithms	machine-like factory production robots	machine-like cleaning robots	virtual avatars	AI video game characters
exact digital copies of human brains		0.0030 (0.0051)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
human-like companion robots			0.0000 (0.0000)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
human-like retail robots				0.6949 (0.6949)	0.1801	0.0020	0.0000	0.0000	0.0000	0.0000	0.0000
animal-like companion robots					0.2694 (0.2993)	0.0105	0.0000	0.0000	0.0000	0.0000	0.0000
exact digital copies of animals						0.1814 (0.2592)	0.0000	0.0000	0.0000	0.0000	0.0000
AI personal assistants							0.0000 (0.0001)	0.0000	0.0000	0.0000	0.0000
complex language algorithms								0.0003 (0.0006)	0.0000	0.0000	0.0000
machine-like factory production robots									0.0000 (0.0000)	0.0000	0.0000
machine-like cleaning robots										0.2624 (0.2993)	0.0000
virtual avatars											0.0000 (0.0000)
AI video game characters											

Table 1. p-values for the target-specific moral concern pairwise comparisons. The adjacent comparisons (immediately above the diagonal) include, in parentheses, the q-value (i.e., under 0.1 is significant with a false discovery rate of 10%).

	AI personal assistants	human-like companion robots	complex language algorithms	exact digital copies of human brains	AI video game characters	machine-like factory production robots	animal-like companion robots	virtual avatars	human-like retail robots	machine-like cleaning robots	exact digital copies of animals
AI personal assistants		0.2025 (0.5062)	0.1272	0.1680	0.0004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
human-like companion robots			0.7656 (0.9417)	0.6777	0.0350	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
complex language algorithms				0.9404 (0.9417)	0.0468	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000
exact digital copies of human brains					0.1187 (0.3956)	0.0021	0.0001	0.0000	0.0000	0.0000	0.0000
AI video game characters						0.0499 (0.357)	0.0656	0.0011	0.0018	0.0001	0.0000
machine-like factory production robots							0.9417 (0.9417)	0.3061	0.2265	0.0326	0.0005
animal-like companion robots								0.3669 (0.7339)	0.1937	0.0861	0.0000
virtual avatars									0.874 (0.9417)	0.3943	0.0064
human-like retail robots										0.4555 (0.7592)	0.0025
machine-like cleaning robots											0.0714 (0.357)
exact digital copies of animals											

Table 2. p-values for the target-specific social connection pairwise comparisons. The adjacent comparisons (immediately above the diagonal) include, in parentheses, the q-value (i.e., under 0.1 is significant with a false discovery rate of 10%).

In 2023, the most social connection was attributed to AI personal assistants (M = 3.75, SE = 0.0578), followed by human-like companion robots (M = 3.68, SE = 0.0578), complex language algorithms (M = 3.66, SE = 0.0604), exact digital copies of human brains (M = 3.66, SE = 0.0641), AI video game characters (M = 3.54, SE = 0.0629), machine-like factory production robots (M = 3.43, SE = 0.0620), animal-like companion robots (M = 3.42, SE = 0.0576), virtual avatars (M = 3.37, SE = 0.0605), human-like retail robots (M = 3.36, SE = 0.0573), machine-like cleaning robots (M = 3.32, SE = 0.0618), and exact digital copies of animals (M = 3.20, SE = 0.0600). While this order is similar to that of the target-specific moral concern questions, there are some distinctions, such as virtual avatars being seen as neither the IOS measure nor its individual items significantly changed from 2021 to 2023. As with the target-specific moral concern questions, the results of pairwise comparisons between each are included in the supplementary materials.

A1.1.6 Pairwise Comparisons: Target-Specific Social Connection. Unlike the numerous significant differences between adjacent types in the moral concern comparisons, there was only one in the social connection comparisons: that between AI video game characters and machine-like factory production robots with an unadjusted p-value of 0.0499, and the adjusted q-value of 0.357 does not meet the adjustment cutoff of 0.1 for the 10% false discovery rate. We still found that there are numerous significant differences between non-adjacent types. For example, machine-like companion robots evoke significantly less social connection than AI personal assistants ($p < 0.001$), human-like companion robots ($p < 0.001$), complex language algorithms ($p < 0.001$), and exact digital copies of human brains ($p = 0.002$). Complete results are listed in Table 2.

A1.1.7 Substratism. In the study of human-animal interaction, the idea of "speciesism" has been developed to refer to the view that individuals of certain species matter less purely due to their species. We tested two analogs of items from the well-known Speciesism Scale [2] translated into the context of substratism. Specifically, we found 74.5% agreement that, "Morally, artificial beings always count for less than humans," and 54.2% agreement that, "Humans have the right to use artificial beings however they want to." The latter significantly increased from 48.5% in 2021 ($p = 0.007$).

A1.2 2023 Supplemental Survey Wave

A1.2.1 Trust. Given the focus in 2023 on corporate behavior, particularly that of OpenAI and other technology companies producing AI systems, and on government regulation, particularly to rein in those companies, the supplement included six questions about trust and assessments of whether these institutions can control AI.

When asked, "AI systems include many different parts. To what extent do you trust the following parts?" on a scale from 1 (not at all) to 7 (very much): engineers (M = 4.19, SE = 0.0525), training data (M = 3.92, SE = 0.0506), output (M = 3.85, SE = 0.0500), algorithm (M = 3.77, SE = 0.0520), companies (M = 3.42, SE = 0.0521), and governments (M = 3.09, SE = 0.0536).

We asked whether participants trust that "the creators of large language models (e.g., OpenAI and GPT-4) put safety over profits" (yes: 22.5%, not sure: 28.7%, no: 48.8%), "the creators of an AI can control all current and future versions of the AI" (yes: 26.9%, not sure: 24.1%, no: 45.9%), "governments have the power to regulate the development of AI" (yes; 52.0%, not sure: 24.1%, no: 23.9%), and, "To what extent do you agree or disagree that governments have the power to effectively enforce regulations on the development of AI?" (71.1% agreement).

On a series of agree-disagree questions, people were largely split on whether they trusted each of four specific types of AI: "I trust game-playing AI" (59.8% agreement); "I trust large language models" (53.5%); "I trust robots" (47.5%); and "I trust chatbots" (46.5%).

A1.2.2 Positive Emotions. Most of our measures of moral concern have been explicit statements that could be considered rational or logical, rather than the more intuitive, emotional aspects of judgment and decision making. One of the most popular trends in psychology research in the 21st century has been positive psychology [6], particularly the role of positive emotions [3]. We assessed these by asking, "To what extent do you, as a human, feel the following emotions towards robots/AIs?" on a 1–7 sliding scale (1 = not at all, 4 = a moderate amount, 7 = "very much) for six emotions: awe (3.90), excitement (3.76), respect (3.53), admiration (3.50), pride (3.30), and compassion (3.02).

A1.2.3 Uploads. We probed participant opinions on a speculative future scenario in which humans can upload their minds to computers, which has been discussed in numerous academic and science fiction works [for a review, see Sandberg and Bostrom 5]. The possibility of mind uploading presents an alternative future trajectory of AI instead of the *de novo* AI, such as large language models, that is the current focus of most AI researchers.

We developed two questions based on the standard format of policy proposals in the General Social Survey [7]. The first question, which presented the scenario straightforwardly, resulted in 41.2% agreement with, "I support humans using advanced technology in the future to upload their minds into computers." In a more detailed question, we found that 39.3% supported the position (i.e., response higher than four, excluding responses of exactly four), "In the future, humans could upload their minds into computers. Some people think that this would be very good because uploaded humans could consume fewer resources, live longer free from biological disease, and have enhanced intelligence and a greater ability to improve the world. Others disagree and think that uploading would mean that we are no longer truly human, change who we are and how we want to live, and distract us from making the real world a better place. Where would you place yourself on this scale?" (sliding scale: 1 = oppose mind uploads, 4 = neither oppose nor support uploads, 7 = support mind uploads).

A1.2.4 Additional Replications. In addition to the replicated questions mentioned in the main text from YouGov on concern for human extinction [9] and support for the six-month pause on advanced AI development [8] and its reversed formulation, we replicated another YouGov question [10], "How likely do you think it is that artificial intelligence (AI) will eventually become more intelligent than people?" (very likely, somewhat likely, not very likely, not likely at all, it is already more intelligent than people, not sure), finding that 69.8% consider it likely, compared to 57% found by YouGov, and a question posed by science communicator Hank Green on Twitter [4], "Which universe is the better one:" (one with humans, one without humans), finding 89.9% selecting "one with humans" compared to 58.9% in Green's Twitter poll.

A1.3 Predictors of attitudes towards sentient AI

Because of the large number of questions in this survey and their novelty, our predictive analyses rely on exploratory descriptive tests rather than confirmatory tests of particular ex ante hypotheses. While further testing and theorization of these relationships is beyond the scope of the current work, it will be an important area of research as this field develops—particularly insofar as interaction with apparent digital minds relates to ongoing social dynamics and disparities in human-AI interaction.

We fit multivariate generalized linear models in which participant characteristics predicted each of a variety of indices: Ban Support, Existential Threat, General Mind Perception, General Moral Concern for All AIs, General Moral Concern for Sentient AIs, General Threat, LLM Mind Perception, LLM Suffering Concern, Mind-Related Anthropomorphism, Positive Emotions, Protection Support, Slowdown Support, Target-Specific Moral Concern, and Trust; we also predicted two individual items: Subservience and 100-Year Likelihood of Sentient AI. For each outcome, we fit a model for each

dataset in which the question or questions were included: first, the main 2021 and 2023 surveys (N = 2,401), which also included year of taking the survey as a predictor, and second, the supplemental 2023 survey (N = 1,099). In general, regressing on indices is preferable because it tends to reduce noise compared to individual items. In our reporting of the results, effect sizes are standardized to compare across different outcome scales, and p-values are assessed at the typical cutoffs (** : $p < 0.001$; *: $0.001 < p < 0.01$). Complete results, including effects at lesser significance ($0.01 < p < 0.05$), are available in the supplementary materials.

We present the strongest effects of the four strongest predictors, expressed as the number of standard deviations of the dependent variable associated with an increase of one standard deviation in the independent variable. We list all effect sizes of at least 0.1, which means that an increase in the independent variable by one standard deviation is associated with the dependent variable increasing by at least one-tenth of a standard deviation. Effect sizes from each dataset are paired together when there is more than a strong association in both the main data (listed first with subscript $_M$) and the supplemental data (listed second with subscript $_S$).

In general, measures of participant experience with AI were the strongest predictor across indices and questions. To mitigate multicollinearity, of the three questions about this, we only included the strongest predictor, frequency of reading or watching AI content, in the models described below. Responses to this question were strongly correlated with the count of AI-related experiences (r_M = 0.584, r_S = 0.544) and the frequency of AI interaction (r_M = 0.561, r_S = 0.569).

Specifically, increased frequency of reading or watching AI content was the strongest predictor across indices and questions, strongly predicting increases in LLM Mind Perception (M_S = 0.329**), Positive Emotions (M_S = 0.314**), Trust (M_S = 0.281**), Mind-Related Anthropomorphism (M_M = 0.261**), General Mind Perception (M_M = 0.260**), General Moral Concern for All AIs (M_S = 0.261**), Protection Support (M_M = 0.245**), LLM Suffering Concern (M_S = 0.234**), General Moral Concern for Sentient AIs (M_M = 0.248**), 100-Year Likelihood of Sentient AI (M_M = 0.211**, M_S = 0.197**), and Target-Specific Moral Concern (M_M = 0.200**), as well as strongly predicting decreases in General Threat (M_S = -0.145**), Slowdown Support (M_S = -0.118**), and Ban Support (M_S = -0.101*).

Increased age strongly predicted increases in Subservience (M_M = 0.165**, M_S = 0.148**) as well as decreases in Mind-Related Anthropomorphism (M_M = -0.182**), Positive Emotions (M_S = -0.174**), LLM Suffering Concern (M_S = -0.168**), 100-Year Likelihood of Sentient AI (M_S = -0.135**), General Mind Perception (M_M = -0.134**), Trust (M_S = -0.133**), LLM Mind Perception (M_S = -0.124**), and Protection Support (M_M = -0.114**).

Conservative political orientation strongly predicted increases in Ban Support (M_M = 0.171**), General Threat (M_M = 0.137**), Subservience (M_S = 0.107**), and Existential Threat (M_S = 0.107**) as well as decreases in General Moral Concern for Sentient AIs (M_M = -0.121**), General Moral Concern for All AIs (M_S = -0.106**), and LLM Suffering Concern (M_S = -0.103**).

Male gender strongly predicted increases in Trust (M_S = 0.156**) and Positive Emotions (M_S = 0.147**) as well as decreases in Ban Support (M_M = -0.138**, M_S = -0.146**) and Slowdown Support (M_S = -0.129**).

APPENDIX REFERENCES

1. Arthur Aron, Elaine N. Aron, and Danny Smollan. 1992. Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness. *Journal of Personality and Social Psychology* 63, 4 (Oct. 1992), 596–612. https://doi.org/10.1037/0022-3514.63.4.596
2. Lucius Caviola, Jim A. C. Everett, and Nadira S. Faber. 2018. The Moral Standing of Animals: Towards a Psychology of Speciesism. *Journal of Personality and Social Psychology* 116, 6 (March 2018), 1011–1029. https://doi.org/10.1037/pspp0000182
3. Barbara L. Fredrickson. 2001. The Role of Positive Emotions in Positive Psychology: The Broaden-and-Build Theory of Positive Emotions. *American Psychologist* 56, 3 (March 2001), 218–226. https://doi.org/10.1037/0003-066X.56.3.218

4. Hank Green. 2023. Which Universe Is the Better One: One with Humans (58.9%), One without Humans (41.1%) - @hankgreen.
5. Anders Sandberg and Nick Bostrom. 2008. *Whole Brain Emulation: A Roadmap.* Technical Report. Future of Humanity Institute.
6. Martin E. P. Seligman and Mihaly Csikszentmihalyi. 2000. Positive Psychology: An Introduction. *American Psychologist* 55, 1 (2000), 5–14. https://doi.org/10.1037/0003-066X.55.1.5
7. Tom W. Smith, Michael Davern, Jeremy Freese, and Morgan Stephen. 2018. *General Social Surveys, 1972-2018.* Technical Report. National Opinion Research Center.
8. YouGov. 2023. The AI Policy Institute Toplines. https://archive.ph/UBR7K.
9. YouGov. 2023. How Concerned, If at All, Are You about the Possibility That AI Will Cause the End of the Human Race on Earth? | Daily Question. https://today.yougov.com/topics/technology/survey-results/daily/2023/04/03/ad825/3.
10. YouGov. 2023. More than 1,000 Technology Leaders Recently Signed an Open Letter Calling on Researchers to Pause Development of Certain Large-Scale AI Systems for at Least Six Months World-Wide, Citing Fears of the "Profound Risks to Society and Humanity." Would You Support or Oppose a Six-Month Pause on Some Kinds of AI Development? | Daily Question. https://today.yougov.com/topics/technology/survey-results/daily/2023/04/03/ad825/2.