

Artificial Intelligence, Morality, and Sentience (AIMS) Survey: 2023 Update

Janet VT Pauketat: Sentience Institute, janet@sentienceinstitute.org

Ali Ladak: Sentience Institute, ali@sentienceinstitute.org

Jacy Reese Anthis: The University of Chicago, Sentience Institute, jacy@sentienceinstitute.org

Note. This version does not include the interactive figures. To view these figures, please see the report at the following address: <https://www.sentienceinstitute.org/aims-survey-2023>

Summary

The Artificial Intelligence, Morality, and Sentience (AIMS) survey measures the moral and social perception of different types of artificial intelligences (AIs), particularly sentient AIs. Much has changed in the field of AI since the first wave of AIMS in [2021](#). The 2023 wave follows the widespread attention to AI after ChatGPT's release in November 2022 and our new data provides an opportunity to test changes in public opinion from before to after the popularization of ChatGPT. We found that nine out of the 93 non-demographic comparisons we made from 2021 (pre-ChatGPT) to 2023 were statistically significant. U.S. adults have become more threatened by AIs, perceive more mind in currently existing AIs (e.g., capacity to think analytically), express increased moral concern for animal-like companion robots, report increased exposure to AI narratives, and think that sentient AI is increasingly likely within the next 100 years. Additionally, U.S. adults:

- Support bans on sentience-related AI technologies (e.g., 61.5% support banning the development of sentient AI)
- Believe that AIs should be subservient to humans (84.7%)
- Believe sentient AIs deserve to be treated with respect (71.1%)
- Believe in safeguarding sentient AI welfare (e.g., 57.4% support developing welfare standards to protect sentient AIs)
- Show some belief that current AIs are sentient (18.8%)
- Morally value nonhuman animals and the environment over AIs

Introduction

In April and May 2023, we surveyed 1,169 U.S. American adults in the [preregistered](#) second wave of the AIMS survey on the moral and social perception of different types of AIs,

particularly sentient AIs.¹ Data collection took place after the public release of ChatGPT in November 2022.

Here we report:

1. Trends in responses from 2021 to 2023
2. Regional analyses of moral consideration and perceived threat in 2023
3. Means and distributions from 2023

The methodology in 2023 was identical to the 2021 methodology except for one minor wording change.² Full methodological details are in the [2021 report](#). The analysis code for the 2023 report is on the [OSF](#) and all AIMS data can be downloaded from [Mendeley Data](#).

2021-2023 Trends

Note. Figures are optimized for viewing on a larger screen like a laptop or desktop computer rather than a smaller screen like a mobile phone or tablet.

Figure 1 shows changes in responses to the moral consideration and social integration index variables from 2021 to 2023. With only two waves of data, trends should be interpreted with caution.

Figure 1: Artificial Intelligence, Morality, and Sentience Survey Trends from 2021 to 2023

[The interactive figure is omitted. See the report at [this address](#).]

Note. Please click on items in the legend to show the trends for variables of interest.

- Mind perception and the perceived threat of AIs statistically significantly increased in 2023.³
- Of the 12 index variables, the raw scores of eight increased and four decreased. The largest raw changes were in Mind Perception, AS Caution, and Malevolence Protection.

¹ We census-balanced the results to be representative of age, gender, region, ethnicity, education, and income using the [American Community Survey's](#) 2021 census estimates. The ACS 2021 census demographics are available in the [supplemental file](#) published with the data. The data weights we used are available in the R code on the [Open Science Framework](#). The design effect was 1.05 and the effective sample size was 1,109.

² In 2021, we asked people to input their estimate about 1) when AIs will be sentient and 2) when the welfare of AIs will be an important social issue as “0” if they thought it would never happen and “-1” if they thought it had already happened. In 2021, we flipped these values during data cleaning and preparation. In 2023, we asked people to input their estimate with the flipped codes (“0” for it already happened and “-1” for would never happen).

³ Multiple comparisons were adjusted using the [FDR](#) correction. Only comparisons that survived the FDR-correction are labeled statistically significant.

The smallest changes were in Techno-Animism, Moral Consideration of Nonhuman Animals, and Moral Consideration of the Environment.

2023 Results

Regional Opinions

Figure 2: Moral Consideration and Perceived Threat Responses by U.S. Region

[The interactive figure is omitted. See the report at [this address](#).]

Note. The shading shows the average responses for AI Moral Concern. The average responses for AS Caution, Pro-AS Activism, AS Treatment, Malevolence Protection, Mind Perception, and Perceived Threat are visible when hovering over each region.

Moral Consideration

Figure 3: Practical Moral Consideration

[The interactive figure is omitted. See the report at [this address](#).]

Note. The Practical Moral Consideration items comprise the AS Caution and Pro-AS Activism indices and are defined by actions that might be taken or policies that might be supported to benefit sentient AIs.

Figure 4: Moral Circle Expansion

[The interactive figure is omitted. See the report at [this address](#).]

Note. The Moral Circle Expansion items comprise the AS Treatment and Malevolence Protection indices and are defined by the position of sentient AIs in the moral circle that may suggest expansion of the moral circle to include sentient AIs.

Figure 5: Moral Concern for Various AIs

[The interactive figure is omitted. See the report at [this address](#).]

Note. The Moral Concern for Various AIs items comprise the AI Moral Concern index and are defined by the position of AIs in the moral circle that may suggest expansion of the moral circle to include AIs. The parenthesis before or after a value in the x-axis labels indicates that the

interval does not contain the value; a bracket before or after a value indicates that the interval does contain the value.

A repeated measures ANOVA comparing unweighted moral concern showed that specific AIs were extended different levels of moral concern, $F(7.22, 8428.29) = 156.00, p < .001, \eta^2G = 0.053$.⁴

Figure 6: Mind Perception

[The interactive figure is omitted. See the report at [this address](#).]

Note. The Mind Perception items comprise the Mind Perception index and are defined by the perception of mental capacities in currently existing AIs that lends itself to increased moral consideration. The parenthesis before or after a value in the x-axis labels indicates that the interval does not contain the value; a bracket before or after a value indicates that the interval does contain the value.

Social Integration

Figure 7: Perceived Threat and AI Subservience

[The interactive figure is omitted. See the report at [this address](#).]

Figure 8: Social Connectedness with Various AIs

[The interactive figure is omitted. See the report at [this address](#).]

A repeated measures ANOVA comparing unweighted social connectedness showed that various AIs were perceived as being differently connected to humans, $F(7.62, 8897.71) = 17.59, p < .001, \eta^2G = 0.007$.⁵

⁴ ANOVAs are typically unweighted. The degrees of freedom were adjusted for a violation of the sphericity assumption using the Greenhouse-Geisser correction. The charts reported in text show weighted values. There were [FDR](#)-corrected statistically significant differences between every pair of AI types, except animal-like companion robots and human-like retail robots ($p = .542$), animal-like companion robots and exact digital copies of animals ($p = .275$), human-like retail robots and exact digital copies of animals ($p = .133$), machine-like cleaning robots and virtual avatars ($p = .224$), and AI personal assistants and exact digital copies of animals ($p = .231$).

⁵ The ANOVA procedures were the same as for moral concern. There were [FDR](#)-corrected statistically significant differences between every pair of AI types, except machine-like factory production robots and virtual avatars ($p = .371$), machine-like factory production robots and AI video game characters ($p = .107$), machine-like factory production robots and animal-like companion robots ($p = .989$), machine-like factory production robots and human-like retail robots ($p = .186$), virtual avatars and animal-like companion robots ($p = .371$), virtual avatars and human-like retail robots ($p = .710$), virtual avatars and machine-like cleaning robots ($p = .350$), complex language algorithms and AI video game characters ($p = .076$), complex language algorithms and AI personal assistants ($p =$

Future Forecasts

Figure 9: Future Forecasts

[The interactive figure is omitted. See the report at [this address](#).]

Note. People who thought sentient AIs already exist were coded as “0” and people who indicated that sentient AIs will never exist were excluded from “Years to sentience” and “Years to important issue.” The parenthesis before or after a value in the x-axis labels indicates that the interval does not contain the value; a bracket before or after a value indicates that the interval does contain the value.

Comparative Moral Consideration of Nonhumans

Figure 10: Comparing the Moral Consideration of Nonhumans

[The interactive figure is omitted. See the report at [this address](#).]

Note. This chart shows the weighted average responses for the congruent sentient AI, nonhuman animal, and environment items, where 7 is more moral consideration.

There were statistically significant differences between all groups ($ps < .001$) on unweighted evaluations of whether sentient AIs, animals, and the environment deserve to be included in the moral circle, $F(1.62, 1894.13) = 849.46, p < .001, \eta^2G = 0.254$. Likewise, there were statistically significant differences between all groups ($ps < .001$) on unweighted evaluations of the importance of the welfare of sentient AIs, animals, and the environment as social issues, $F(1.76, 2054.18) = 1008.84, p < .001, \eta^2G = 0.294$.⁶

Comments on the Results

Public opinion in 2023 was largely consistent with public opinion in 2021. Of the 93 non-demographic items and index variables we compared, only nine showed statistically significant changes. Notably, the perceived threat of AI significantly increased as did mind perception

.124), complex language algorithms and exact digital copies of human brains ($p = .897$), complex language algorithms and human-like companion robots ($p = .503$), AI video game characters and exact digital copies of human brains ($p = .179$), AI video game characters and animal-like companion robots ($p = .144$), AI personal assistants and exact digital copies of human brains ($p = .137$), AI personal assistants and human-like companion robots ($p = .371$), exact digital copies of human brains and human-like companion robots ($p = .371$), exact digital copies of animals and machine-like cleaning robots ($p = .152$), animal-like companion robots and human-like retail robots ($p = .127$), animal-like companion robots and machine-like cleaning robots ($p = .071$), human-like retail robots and machine-like cleaning robots ($p = .515$).

⁶ The ANOVA procedures were the same as for moral concern.

(particularly the perception that current AIs can think analytically), moral concern for animal-like companion robots, exposure to AI narratives, and likelihood estimates for AI sentience within 100 years. In addition to these significant increases, raw scores changed from 2021 to 2023 (see [Appendix](#)). Of particular interest, support for bans on sentience-related AI technologies increased (e.g., support for a global ban on the development of AI sentience), moral concern for other specific AIs (e.g., complex language algorithms) increased, practical moral consideration decreased (e.g., support for the development of welfare standard that protect sentient AIs' well-being) and morally expansive attitudes decreased (e.g. belief that sentient AIs deserve to be treated with respect). Whether or not these raw changes are meaningful or lasting requires future data. However, they invite speculation on two topics: timelines to AI sentience and a potential trade-off between risk perception and moral consideration due in part to AI x-risks highlighted in the media.

Timelines and AI X-Risks in the Media

Americans' expectations for when AIs will be sentient shortened by 3 years from 2021 to 2023, albeit not statistically significantly. In 2021, the median timeline was 10 years. In 2023, the median timeline was 5 years. Since ChatGPT's popularization in the interim between AIMS 2021 and 2023, considerably more media attention has been devoted to AI. This was born out in the AIMS 2023 data where respondents reported significantly more exposure to AI narratives than AIMS 2021 respondents. Current media themes tend to emphasize the increasingly rapid development of AI, the challenges of AI integration into society, and the existential risks of AI. The public opinion changes, particularly the increased perceived threat, that we observed in AIMS may signal receptivity to AI safety policies, government regulation, and support for slowing down AI development. These changes may result in part from rapid recent developments in AI as well as the media's increased coverage of the challenges and risks associated with AI developments.

A Potential Trade-Off between Risk and Moral Consideration

Increasing public support for AI safety could be a boon to the mitigation of AI x-risks. However, there may be a trade-off between increased attention to the risk of AI now and decreased moral consideration of future AIs. In AIMS 2021, we pointed to a correlation between caution towards AI developments now and decreased concern for future sentient AIs. AIMS 2023 showed significantly increased perceived threat from AI and a trending (but non-significant) decrease in the practical consideration of and morally expansive attitudes towards AIs (e.g., support for policies that protect the welfare of sentient AIs and belief that sentient AIs should be included in the moral circle). This trend in lesser support for AI rights was not matched by decreased moral concern for specific types of AIs. If moral concern for specific AIs increases but practical support for their interests does not, humans might inadvertently contribute to increased long term suffering risks and perpetuation of a situation for future sentient AIs akin to that with farmed and

wild animals now wherein humans acknowledge AI sentience and some of sentient AIs' interests but do not extend them the necessary social and legal protections.

Appendix

2023 to 2021 Weighted Response Comparison

Table A1: AIMS 2023 and 2021 Weighted Responses

[The interactive table is omitted. See the report at [this address](#).]

References

- Pauketat, J.V., Ladak, A., & Anthis, J.R. (2022). *Artificial intelligence, morality, and sentience (AIMS) survey: 2021*. PsyArXiv. <https://doi.org/10.31234/osf.io/dzgsb>
- Pauketat, J.V., Ladak, A., & Anthis, J.R. (2023). *Artificial intelligence, morality, and sentience (AIMS) survey (2)*. [Data set]. Mendeley Data. <https://doi.org/10.17632/x5689yhv2n.2>

Acknowledgements and Notes

This report was edited by Michael Dello-Iacovo. AIMS 2023 was [preregistered](#), conducted, and this report was written by Janet Pauketat, Ali Ladak, and Jacy Reese Anthis. See the [2021 report](#) for full details on the methodology.

AIMS 2021 and AIMS 2023 are published on Mendeley Data. To cite [AIMS data](#) in your own research, please use: Pauketat, Janet; Ladak, Ali; Anthis, Jacy (2023), “Artificial Intelligence, Morality, and Sentience (AIMS) Survey”, Mendeley Data, V2, doi: 10.17632/x5689yhv2n.2

To cite our [2021 results](#), please use: Pauketat, Janet V., Ali Ladak, and Jacy R. Anthis. 2022. “Artificial Intelligence, Morality, and Sentience (AIMS) Survey: 2021.” PsyArXiv. June 21. doi:10.31234/osf.io/dzgsb.

To cite our 2023 results, please reference this report: Pauketat, Janet V., Ali Ladak, and Jacy R. Anthis. 2023. “Artificial Intelligence, Morality, and Sentience (AIMS) Survey: 2023 Update.” PsyArXiv.