

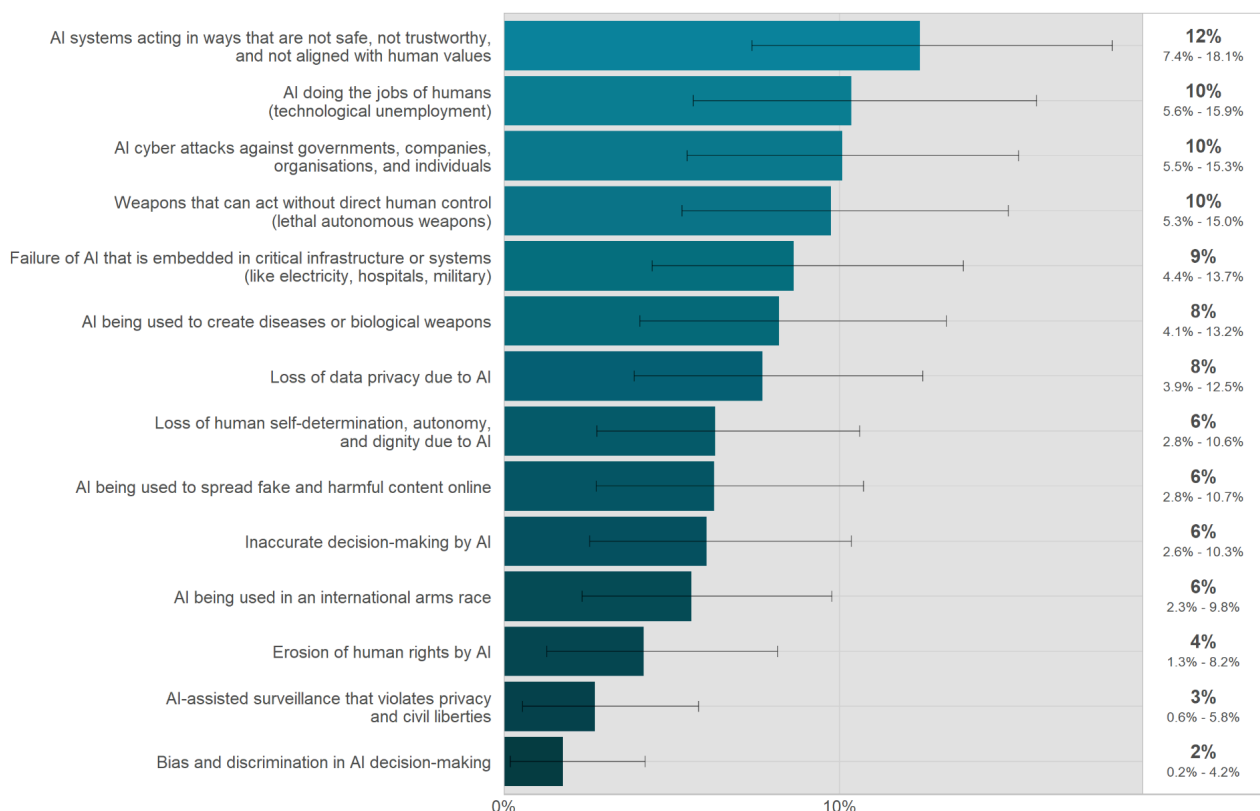
Australians are concerned about AI risks and expect strong government action

Key insights

1. A representative online *Survey Assessing AI Risks (SARA)* of 1,051 Australians in January-February 2024 investigated public perceptions of AI risks and support for AI governance actions.
2. Australians are **most concerned about AI risks where AI acts unsafely** (e.g., acting in conflict with human values, failure of critical infrastructure), **is misused** (e.g., cyber attacks, biological weapons), **or displaces the jobs of humans**; they are least concerned about AI-assisted surveillance, or bias and discrimination in AI decision-making.
3. Australians judge **"preventing dangerous and catastrophic outcomes from AI"** the **#1 priority for the Australian Government in AI**; 9 in 10 Australians support creating a new regulatory body for AI.
4. To meet public expectations, the Australian Government must urgently increase its capacity to govern increasingly-capable AI and address diverse risks from AI, including catastrophic risks.

Finding: Australians are concerned about diverse risks from AI

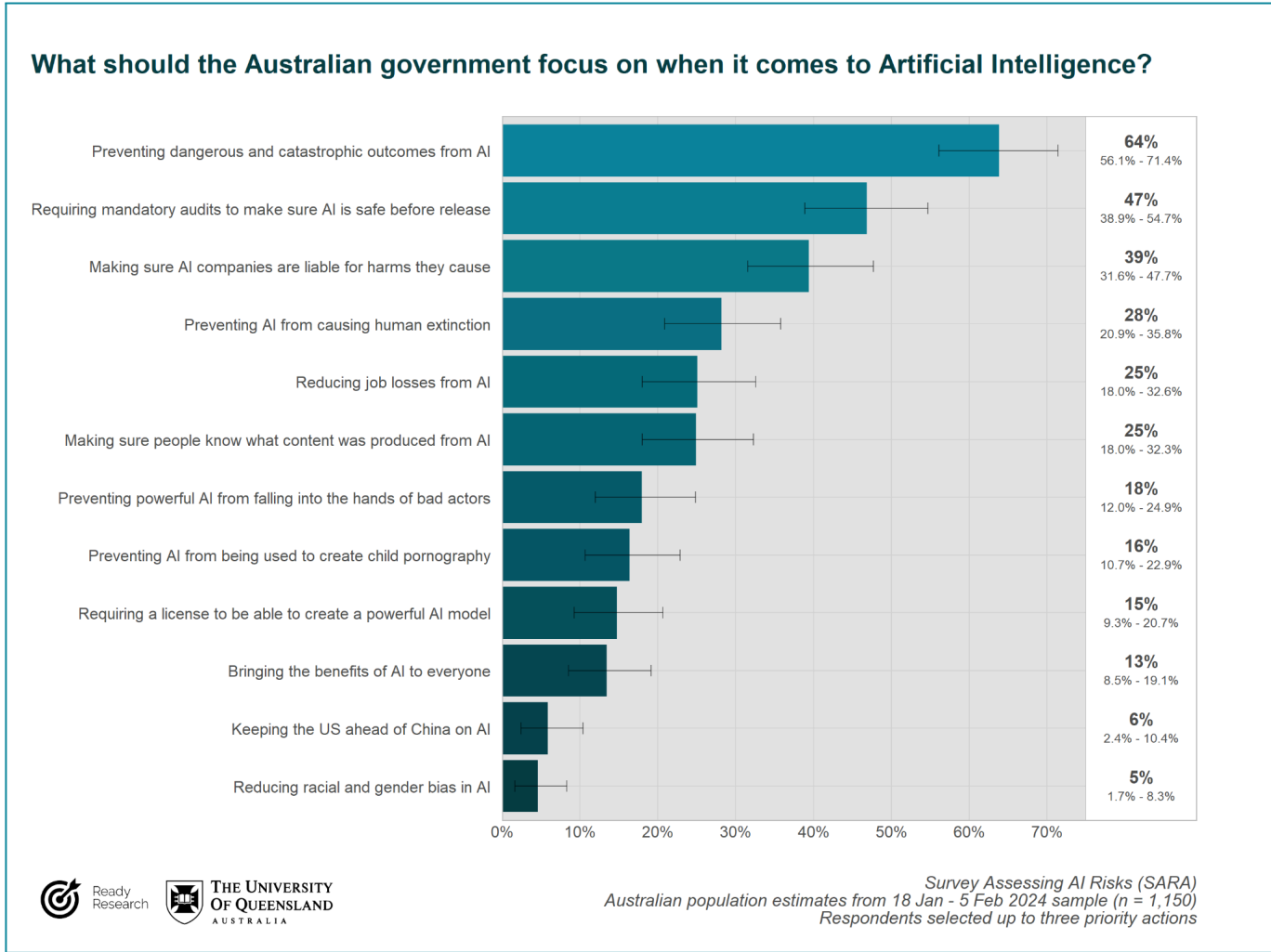
When it comes to the following potential risks from AI, which is most concerning to you?



When asked about a diverse set of 14 possible negative outcomes from AI, Australians were most concerned about *AI systems acting in ways that are not safe, not trustworthy, and not aligned with human values*. Other high-priority risks include AI replacing human jobs, enabling cyber attacks, operating lethal autonomous weapons, and malfunctioning within critical infrastructure.

Australians are skeptical of the promise of artificial intelligence: 4 in 10 support the development of AI, 3 in 10 oppose it, and opinions are divided about whether AI will be a net good (4 in 10) or harm (4 in 10).

Finding: Australians support regulatory and non-regulatory action to address risks from AI



When asked to choose the top 3 AI priorities for the Australian Government, the #1 selected priority was preventing dangerous and catastrophic outcomes from AI. Other actions prioritised by at least 1 in 4 Australians included (1) requiring audits of AI models to make sure they are safe before being released, (2) making sure that AI companies are liable for harms, (3) preventing AI from causing human extinction, (4) reducing job losses from AI, and (5) making sure that people know when content is produced using AI.

Almost all (9 in 10) Australians think that AI should be regulated by a national government body, similar to how the Therapeutic Goods Administration acts as a national regulator for drugs and medical devices. 8 in 10 Australians think that Australia should lead the international development and governance of AI.

Finding: Australians take catastrophic and extinction risks from AI seriously

Australians consider the prevention of dangerous and catastrophic outcomes from AI the #1 priority for the Australian Government. In addition, a clear majority (8 in 10) of Australians agree with AI experts, technology leaders, and world political leaders that preventing the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war ¹.

Artificial Intelligence was judged as the third most likely cause of human extinction, after nuclear war and climate change. AI was judged as more likely than a pandemic or an asteroid impact. About 1 in 3 Australians think it's at least 'moderately likely' AI will cause human extinction in the next 50 years.

Implications and actions supported by the research

Findings from SARA show that Australians are concerned about diverse risks from AI, especially catastrophic risks, and expect the Australian Government to address these through strong governance action.

Australians' ambivalence about AI and expectation of strong governance action to address risks is a consistent theme of public opinion research in this area ²⁻⁴

Australians are concerned about more diverse risks from AI, compared to Government

The Australian Government published an interim response to its Safe and Responsible AI consultation ⁵. As part of its interim response, the Government plans to address known risks and harms from AI by strengthening existing laws, especially in areas of privacy, online safety, and mis/disinformation.

Findings from SARA show that some Australians are concerned about privacy, online safety, and mis/disinformation risks, so government action in these areas is a positive step. However, the risks that Australians are most concerned about are not a focus of the Government's interim response. These priority risks include AI systems being misused or accidentally acting in ways that harm people, AI-enabled cyber attacks, and job loss due to AI. The Government must broaden its consideration of AI risks to include those identified as high priority by Australians.

Australians want Government to establish a national regulator for AI, require pre-release safety audits, and make companies liable for AI harms

The Government plans to develop a voluntary AI Safety Standard and voluntary watermarking of AI-generated materials. Findings from SARA show that Australians support stronger Government action, including mandatory audits to make sure AI is safe before release ⁶, and making AI companies liable for harms caused by AI ⁷. Australians show strong support for a national regulatory authority for AI; this support has been consistently high since at least 2020 ⁴. To meet expectations, Government should establish a national regulation for AI, and implement strong action to limit harms from AI.

Australians want Government action to prevent dangerous and catastrophic outcomes from frontier and general-purpose models

In its interim response, the Government described plans to establish mandatory safeguards for 'legitimate, high-risk settings' to 'ensure AI systems are safe when harms are difficult or impossible to reverse', as well as for 'development, deployment and use of frontier or general-purpose models'.

Findings from SARA indicate that Australians want the government, as a #1 priority action, to prevent dangerous and catastrophic outcomes from AI. Frontier and general-purpose models carry the greatest risk for catastrophic outcomes ⁸, and are also advancing in capabilities without clear safety measures. Australians believe preventing the risk of extinction from AI should be a global priority. To meet Australians' expectations, Government must ensure it can understand and respond to emergent and novel risks from these AI models.

Research context and motivation

The development and use of AI technologies is accelerating. Across 2022 and 2023, new large-scale models have been announced monthly, and are achieving increasingly complex and general tasks⁹; this trend continues in 2024 with Google DeepMind Gemini, OpenAI Sora, and others. Experts in AI forecast that development of powerful AI models could lead to radical changes in wealth, health, and power on a scale comparable to the nuclear and industrial revolutions ^{10,11}.

Addressing the risks and harms from these changes requires effective AI governance: forming robust norms, policies, laws, processes and institutions to guide good decision-making about AI development, deployment and use¹². Effective governance is especially crucial for managing extreme or catastrophic risks from AI that are high impact and uncertain, such as harm from misuse, accident or loss of control.

Understanding public beliefs and expectations about AI risks and their possible responses is important for ensuring that the ethical, legal, and social implications of AI are addressed through effective governance. We conducted the Survey Assessing Risks from AI (SARA) to generate 'evidence for action', to help public and private actors make the decisions needed for safer AI development and use.

About the Survey Assessing Risks from AI

Ready Research and The University of Queensland collaborated to design and conduct the *Survey Assessing Risks from AI* (SARA). This briefing presents topline findings. Contact Dr Alexander Saeri (a.saeri@uq.edu.au) for a detailed briefing, and access to the forthcoming technical report.

Between 18 January and 5 February 2024, The University of Queensland surveyed 1,151 adults living in Australia, online using Qualtrics survey platform. Participants were recruited through the Online Research Unit's panel, with nationally representative quota sampling by gender, age group, and state/territory. Multilevel regression with poststratification (MRP) was used to create Australian population estimates and confidence intervals, using 2021 Census information about sex, age, state/territory, and education. The research project was reviewed and approved by UQ Research Ethics (Project 2023/HE002257).

This project was funded by the Effective Altruism Infrastructure Fund.

References

1. Center for AI Safety. Statement on AI risk. <https://www.safe.ai/statement-on-ai-risk> (2023).
2. Lockey, S., Gillespie, N. & Curtis, C. *Trust in artificial intelligence: Australian insights*. <https://espace.library.uq.edu.au/view/UQ:b32f129> (2020) doi:10.14264/b32f129.
3. Gillespie, N., Lockey, S., Curtis, C., Pool, J. & Akbari, A. *Trust in Artificial Intelligence: A global study*. <https://espace.library.uq.edu.au/view/UQ:00d3c94> (2023) doi:10.14264/00d3c94.
4. Selwyn, N., Cordoba, B. G., Andrejevic, M. & Campbell, L. *AI for social good - Australian attitudes toward AI and society*. https://bridges.monash.edu/articles/report/AI_for_Social_Good_-_Australian_Attitudes_Toward_AI_and_Society_Report_pdf/13159781/1 (2020) doi:10.26180/13159781.V1.
5. *Safe and responsible AI in Australia consultation: Australian Government's interim response*. <https://consult.industry.gov.au/supporting-responsible-ai> (2024).
6. Shevlane, T. *et al.* Model evaluation for extreme risks. *arXiv [cs.AI]* (2023) doi:10.48550/arXiv.2305.15324.
7. Weil, G. Tort Law as a Tool for Mitigating Catastrophic Risk from Artificial Intelligence. (2024) doi:10.2139/ssrn.4694006.
8. Anderljung, M. *et al.* Frontier AI Regulation: Managing Emerging Risks to Public Safety. *arXiv [cs.CY]* (2023) doi:10.48550/arXiv.2307.03718.
9. Maslej, N. *et al.* Artificial Intelligence Index Report 2023. *arXiv [cs.AI]* (2023).
10. Grace, K., Stein-Perlman, Z., Weinstein-Raun, B. & Salvatier, J. *2022 Expert Survey on Progress in AI*. <https://aiimpacts.org/2022-expert-survey-on-progress-in-ai/> (2022).
11. Davidson, T. *What a compute-centric framework says about takeoff speeds*. <https://www.openphilanthropy.org/research/what-a-compute-centric-framework-says-about-takeoff-speeds/> (2023).
12. Dafoe, A. AI governance: Opportunity and theory of impact. <https://www.allandafoc.com/opportunity> (2018).