# Kate Chaney MP
FEDERAL MEMBER FOR CURTIN

# AI DISCUSSION PAPER

## The case for an Australian AI Safety Institute

# The case for an Australian AI Safety Institute

**The transformative impact of Artificial Intelligence (AI) on workflows, decision-making, knowledge creation and our society is hard to fathom.**

Twenty years ago, when smartphones harnessed the power of the internet, we couldn't imagine the impact these pocket-sized computers would have on our use of cameras, watches, alarm clocks, CDs, maps, banks, cash and newspapers.

While it took a decade for smartphones to become ubiquitous, AI is moving faster. In just a few years, tools like ChatGPT have gone from novelty to normal. Students use them to write essays, businesses use them to automate customer service and governments are exploring how to use them in service delivery. Chatbots are now being used for mental health support. Other AI tools are being used in recruitment, policing, healthcare and beyond.

In my community of Curtin and across Australia, people are concerned about the speed of change and are seeking to understand how our governments can act to enable Australians to harness the benefits of AI while managing the significant risks.

In recent months, I have heard these concerns from constituents and sought to engage with a wide range of stakeholders to determine what the Australian Government can do now, to support the safer roll-out of AI. These stakeholders include international and Australian AI policy experts and thinktanks, international AI developers, academics and advocacy groups.

While the technology is racing ahead, our ability to understand and govern it is lagging behind. The Australian Government has limited AI expertise and insufficient capacity to respond quickly.

Around the world, countries are dealing with this problem by establishing AI safety institutes. AI safety institutes are independent advisory bodies, with significant AI expertise and the flexibility to respond rapidly to developments in AI. These institutes test AI systems, monitor risks and support governments to respond swiftly and sensibly. The UK, US, EU, Germany, Singapore, Japan and many other countries have established AI safety institutes. The Federal Government has endorsed the importance of AI safety institutes, but is yet to deliver for Australia. Put simply, an Australian AI Safety Institute would provide the government with the ability to keep up with the rapid developments in AI.

The Albanese Government should **establish an Australian AI Safety Institute urgently**.

The mission of an Australian AI Safety Institute should be to support a safe roll-out of AI in Australia, with three essential functions:

- **Identifying and monitoring risks** – both emerging risks from new AI tools and known risks as AI tools are adopted in Australia.
- **Collaborating internationally** – with international AI safety institutes and AI developers such as OpenAI, Microsoft and Anthropic.
- **Supporting Australian regulators and policymakers** – by providing the technical AI expertise required to make and enforce laws related to AI.

For an Australian AI Safety Institute to be effective, it must be **well-resourced and independent**. Importantly, an AI Safety Institute is not a regulator; it supports existing regulators.

We cannot pretend to know how AI will change our lives, but we can put the right structures in place to ensure that we face the transformation head-on and make conscious decisions about how to build a future in which we want to live.

An Australian AI Safety Institute will not solve all our problems, but it is a common-sense, 'no-regret' first step that should be established as a matter of urgency.

*KEChaney*

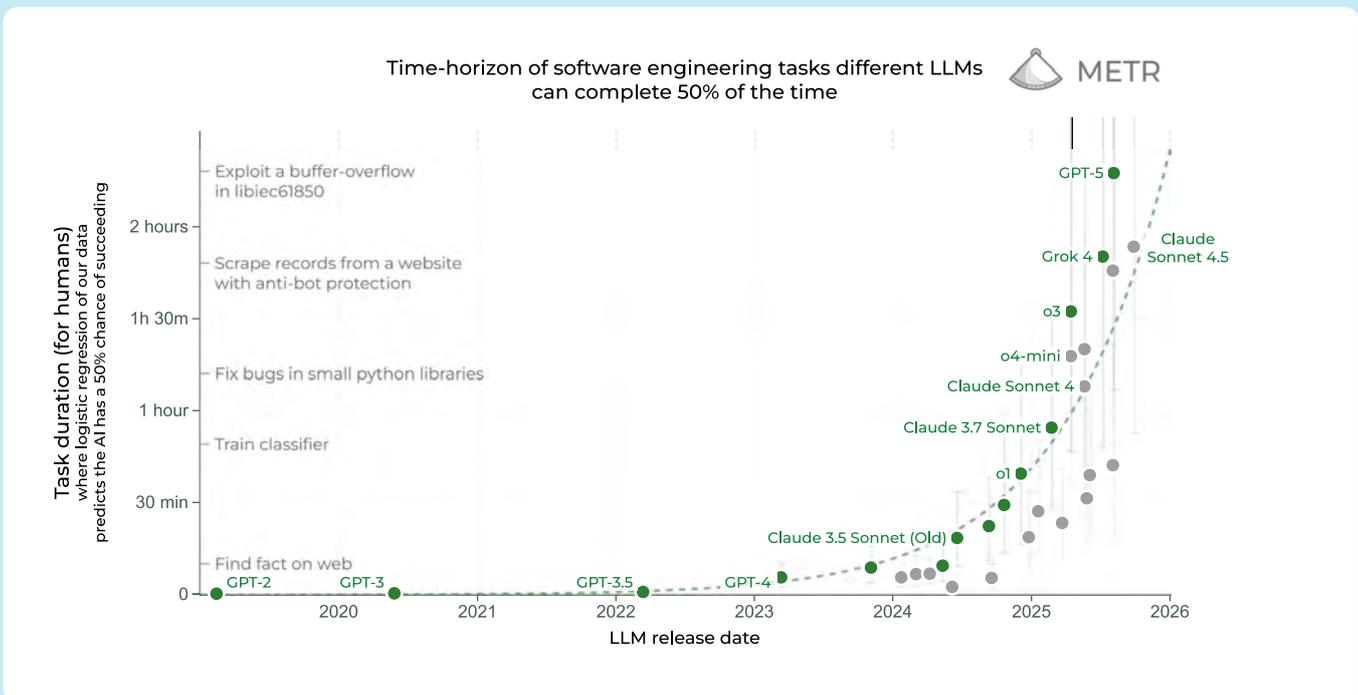**Kate Chaney MP**
**November 2025**

# Key Asks

**1**   **Establish an Australian AI Safety Institute**

**4**   **Mandate the AI Safety Institute to collaborate internationally**

**2**   **Ensure the AI Safety Institute is independent and well-resourced**

**5**   **Mandate the AI Safety Institute to support existing policymakers and regulators**

**3**   **Mandate the AI Safety Institute to identify and monitor risks**

# AI is moving quickly

Since ChatGPT launched in late 2022, AI has gone from a specialised tool to something many Australians use every day in both their personal and professional lives. We can all see how rapidly AI is developing. Chatbot answers are improving, and social media feeds are dominated by increasingly realistic AI images and videos.

## AI is improving very quickly – at an exponential rate

The **Model Evaluation & Threat Research** group has analysed how the ability of AI to complete human tasks has improved over time. It has shown that Large Language Models (like ChatGPT), are improving exponentially – the length of human tasks that AI can do is doubling every 7 months.



Time-horizon of software engineering tasks different LLMs can complete 50% of the time. METR

In the digital copy of this paper, underlined text is linked to its source.

While AI is developing quickly, it is also being adopted rapidly. In Australia:

- The **National AI Centre estimates** that 42% of small and medium businesses are already using or planning to use AI (this increases to 73% for businesses with at least 20 employees).

- In 2025, a **University of Melbourne/KPMG study** found that 49% of Australian employees use AI at work.

- As early as 2023, **UNICEF found** that 69% of Australian teenagers had used a generative chatbot.

AI is being used in schools, hospitals, banks and government services, helping people write resumes, translate languages and detect fraud. It is also being used in more sensitive areas, like mental health support, hiring decisions and policing.

The speed of AI development and adoption presents significant opportunities. The **Tech Council of Australia estimates** that AI adoption could add $115 billion per year to the Australian economy by 2030. AI can help small businesses save time, help doctors analyse scans faster and assist teachers tailor lessons to students.

But the risks are real too, exacerbated by the pace of change. Some risks include:

- deepfake content undermining trust in our democratic institutions;

- loss of privacy and security;

- widespread use of AI undermining critical thinking skills in children;

- significant changes or losses of entire job categories;

- AI chatbots encouraging suicidal ideation; and

- potentially longer-term existential threats to humankind.

We need to take action to deal with the risks posed by AI, so that we can make the most of the positive opportunities. **Study** after **study** has shown that Australians are deeply concerned about the risks of AI and expect action from government. Without action, trust and uptake will remain low, while increasingly serious risks may emerge.

When considering how to deal with the risks posed by AI, it is the pace of change that is the problem. Our laws and institutions were not built for a world where new risks can arise overnight. We need a way to keep up.

## $115 billion
**could be added to the economy every year from 2030 through AI adoption**

## 96%
**of Australian consumers have concerns about AI, including its misuse by scammers and personal privacy implications**

## 81%
**of Australians think the government needs to play a stronger role in overseeing AI technologies**

## 49%
**of Australian employees use AI at work**

## 56%
**of Australian teens had been exposed to deepfake content by 2024**

## 9 in 10
**Australians think Australia should play a leading role on international governance of AI**

# An AI Safety Institute equips the Federal Government to keep up

Around the world, countries are dealing with the rapid pace of change by developing AI safety institutes. AI safety institutes are independent public sector organisations with significant AI expertise that identify and monitor AI risks. They provide in-house technical expertise to the government. Governments are typically slow and bureaucratic, but AI safety institutes are specifically designed to keep up with AI developments and monitor risks.

Australia is a member of the International AI Safety Institute Network and has endorsed the importance of AI safety institutes. But the Federal Government has not yet delivered an Australian AI Safety Institute. All other member countries of the AI Safety Institute Network other than Kenya have already established an AI safety institute.

The Federal Government should **establish an Australian AI Safety Institute** urgently. An Australian AI Safety Institute must have resources, independence and mandate to rapidly identify and address critical AI safety issues facing Australians.

The mission of the Australian AI Safety Institute should be to support a safe roll-out of AI in Australia by:

- **identifying and monitoring risks;**
- **collaborating internationally; and**
- **supporting regulators and policymakers.**

These three essential functions are explored below.

For an Australian AI Safety Institute to be effective in its mission, it must be **independent** and **well-resourced**.

The Australian AI Safety Institute should not be responsible for promoting AI adoption or direct AI regulation – these remain the functions of the National AI Centre and existing regulators.

| AISI International Network Member | AI Safety Institute |
|---|:---:|
| UK | ✅ |
| USA | ✅ |
| EU | ✅ |
| Canada | ✅ |
| Germany | ✅ |
| France | ✅ |
| Japan | ✅ |
| South Korea | ✅ |
| Singapore | ✅ |
| India | ✅ |
| Israel | ✅ |
| Brazil | ✅ |
| Australia | ❌ |
| Kenya | ❌ |

## The UK's AI Security Institute represents the 'gold standard' for resourcing

The UK's AI Safety Institute, renamed as the AI Security Institute, is often pointed to as a 'gold standard' for resourcing. The UK Government **invests £66 million per year** in the UK Security Institute. The UK Government actively pursued and recruited talented AI experts. This required significant salaries, in order to be competitive with the opportunities for AI talent in industry. This was a worthwhile investment, with the UK now seen as a leader in AI safety research and a hub of expertise.

# An Australian AI Safety Institute should have three functions

## Function 1: Identify and monitor risks

The Australian AI Safety Institute should identify and monitor AI risks. For it to be effective, the AI Safety Institute must have the resources and capacity to keep pace with developments in AI. The AI Safety Institute should monitor both:

- **Emerging risks** – risks that arise from new AI capabilities (including all types of AI, not just generative AI chatbots). For example, the Australian AI Safety Institute would test whether a new GPT model released by OpenAI has the potential to create chemical weapons.

- **Known risks** – risks that we can already see, that arise from increasing use of existing AI models in everyday society. For example, the Australian AI Safety Institute would monitor whether AI tools were increasingly involved in deaths by suicide.

Internationally, AI safety institutes tend to concentrate on long-term emerging risks, which include catastrophic but unlikely scenarios such as existential threats, AI takeover, or the creation of novel chemical weapons. But the harms occurring now, and those most likely to intensify in the near future, stem from the widespread deployment of existing AI tools across society. An Australian AI Safety Institute must therefore be mandated to monitor these immediate known risks, as well as the long-term emerging risks.

Identifying and monitoring risks would involve:

- **Testing of AI systems before and after they are released** – for example, the UK AI Security Institute actively tests whether new AI systems are capable of producing harmful outputs and decisions, to test the effectiveness of in-built safeguards.

- **Tracking trends in the adoption of AI systems in Australia** – for example, this could involve assessment of the use, effectiveness and risks of AI tools in sensitive areas such as policing or law, or for mental health support – to identify risks and harms.

- **Investigating incidents where harm has occurred** – this could involve investigating how an AI tool arrived at a harmful output or decision, such as an automated system used in government to incorrectly calculate debts, or a chatbot that may have encouraged suicidal ideation.

The risks and harms that are uncovered by the AI Safety Institute would be communicated to the AI Safety Institute International Network and overseas AI developers (Function 2) and to Australian regulators and policymakers (Function 3).

Currently, the government does not have a way to monitor AI risks as they emerge in a coordinated manner. The AI Safety Institute would fill this gap. It would help government spot problems early and respond before they get worse. Without an AI Safety Institute, the government remains incapable of responding quickly to the rapid and transformative developments in AI.

## Function 2: Cooperate internationally

The biggest AI companies are international. OpenAI, Microsoft, Anthropic and Meta develop and train their models overseas. But the impacts are felt here.

The AI Safety Institute should:

- **Collaborate with the international network of AI safety institutes** to:

  - **Share research and findings on risks** – international AI safety institutes may share AI system risks that have been uncovered. Global cooperation on identifying these risks will be much more effective than each country working alone.

  - **Collaborate on the development of standards** – international AI safety institutes may work together to develop a set of standards that new AI models must meet, such as product testing and performance standards. A company like OpenAI or Meta is more likely to invest in compliance with Australian standards and laws if they align with those in Europe and Asia.

- **Collaborate directly with AI developers**, like OpenAI, Microsoft and Anthropic, to identify and minimise risks. For example, the UK AI Security Institute works closely with US-based AI developers once it uncovers potential risks caused by new models. This close collaboration is also important in order to gain early access to new models, to allow for testing before deployment in Australia.

Working and collaborating internationally is vital for Australia to capture the potential productivity gains of AI, while mitigating the risks. This is widely supported; **9 in 10 Australians** think Australia should play a leading role on international governance of AI.

## Function 3: Support regulators and policymakers

AI is complex and fast-moving. Currently, policymakers and regulators who make and enforce our laws have limited understanding of how AI works and how the technology may change over time. We need to ensure that our laws and regulations are fit-for-purpose in an AI world. The AI Safety Institute would support this aim by providing a centralised body for AI expertise and independent advice that could be drawn on by regulators, departments and Ministers. This function would include:

- **Alerting regulators and policymakers to risks** – for example, if the AI Safety Institute identifies that a new AI tool being used to assess eligibility for bank loans is discriminatory, it could advise and work with the existing regulators to determine the appropriate course of action.

- **Responding to requests for assistance from policymakers and regulators** – for example, if the Federal Government is trying to deal with the harms of electoral misinformation, the government could draw on expertise and advice from the AI Safety Institute to ensure that policies will be effective for the AI technologies.

- **Testing and assessing AI technology for regulators and policymakers** – for example, if the ACCC is investigating whether an AI-powered shopping app is misleading consumers, the AI Safety Institute could help test how the app behaves in different scenarios and ensure the ACCC responds in a way that deals with the realities of AI technology.

The AI Safety Institute would help regulators and policymakers move faster and make better decisions. Fundamentally, our regulators and policymakers need to be able to draw on a source of AI expertise for their laws and regulations to be effective for AI.

### The AI Safety Institute should not be an AI regulator

The AI Safety Institute should not have its own regulatory powers. We already have regulators for privacy, consumer protection, health, education and more. But many of them do not have the technical tools or expertise to deal with AI. Instead, the AI Safety Institute would be supporting existing regulators to deal with AI by providing technical expertise and experience.

# Bringing Curtin into the AI conversation

**Establishing an AI Safety Institute is the first step in equipping Australian governments to address the risks and realise the benefits of AI. There is so much more to be done.**

In the coming year, as well as encouraging the Federal Government to establish the best institutional framework for tackling such a fast changing and ubiquitous development, I aim to contribute to policy development in two ways:

- Firstly, I will continue to work with stakeholders to identify gaps in our existing laws and urge the government to address them. For example, I recently introduced a Private Member's Bill to make it an offence to download AI tools specifically designed to generate child sexual abuse material, which are not currently covered in the *Criminal Code Act 1995 (Cth)*.

- Secondly, I will work with my community in Curtin to explore the values that should guide our take-up of this extraordinary technology in this ever-changing space. It is vital that our future is guided by the voices of Australian people and communities, rather than only by international tech companies.

**If you would like to be part of our ongoing conversation about how to safely roll-out AI in Australia, please express your interest through this QR Code.**

# Kate Chaney MP
## FEDERAL MEMBER FOR CURTIN