Embracing AI with confidence – leveraging the benefits while mitigating the risks

CISO Melbourne 17 July 2024

Dr Greg Adamson

CISO, Department of Transport and Planning, Victoria

Speaking in a personal capacity

https://www.linkedin.com/in/gregadamson/

Cars and brakes

- Why do cars have brakes?
- So that they can go fast

Some starting principles

1. AI is an experimental technology in its infancy
2. AI makes the jobs of security professionals harder, not easier
3. Even the smartest AI experts can't explain exactly how AI works in a given situation
4. We can't give up

Abdullah Alsaibaie - Pexel

Handling AI with care

1. Key trends
2. AI on the attack
3. AI risk preparation
4. AI for the defence

Speaker bio

- Cyber security specialist since 1993
- First addressed this topic at the Melbourne Cyber Risk Meetup, 18 September 2018
- Chair of the international conference series on AI, IEEE Conference on Norbert Wiener in the 21st Century (Boston 2014, Melbourne 2016, Chennai 2021, West Lafayette 2023, Matsue (Japan) 2025)
- Researcher on explainable AI (honorary Associate Professor, University of Melbourne Faculty of Medicine and Dentistry)
- Technical Activities VP for IEEE Society on Social Implications of Technology

Key trends today

- Data loss
- Full attack surface threats
- Deep fakes
- Vendor promises (Threat intelligence, Security operations, Incident response, Risk management, Stakeholder engagement, … and every other task you can think of today ☺)
- This is not all new. GenAI could be described as a "machine … with a statistical preference for a certain sort of behavior" (Norbert Wiener, 1950)

Mathias Reding - Pexels

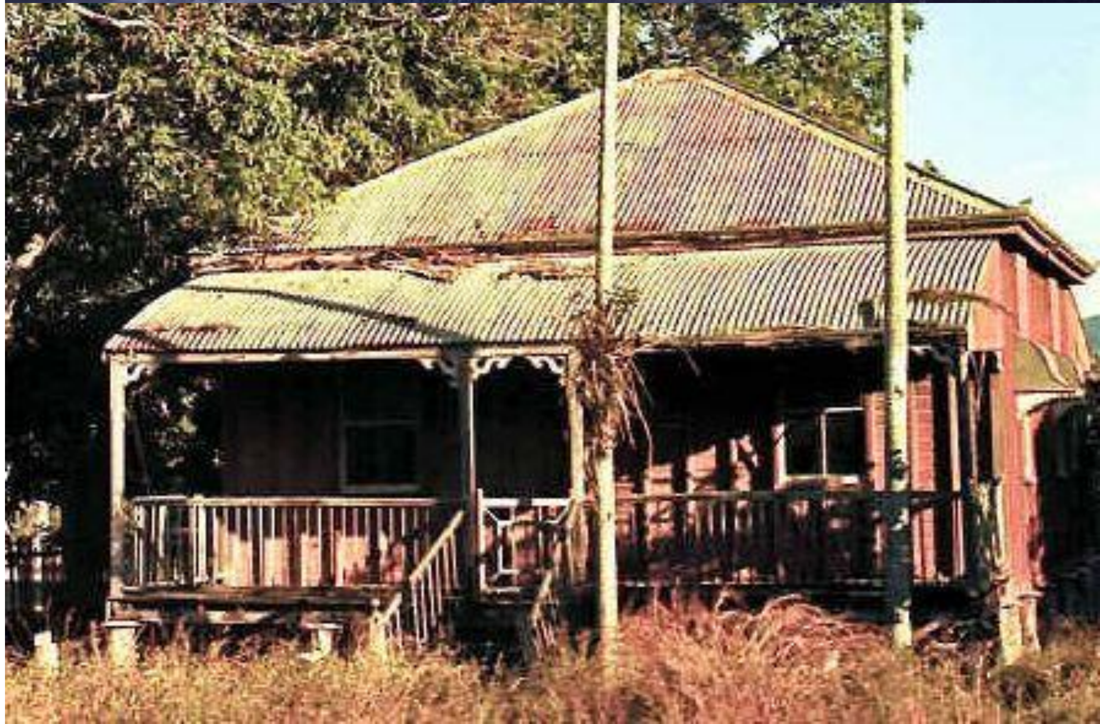
Key trend #1: Will an AI tool leak your data?

- The structure you need to prevent leaks (every user’s access perfectly configured)

Key trend #1: Will an AI tool leak your data?

- The structure you need to prevent leaks (every user's access perfectly configured)

- Your current structure



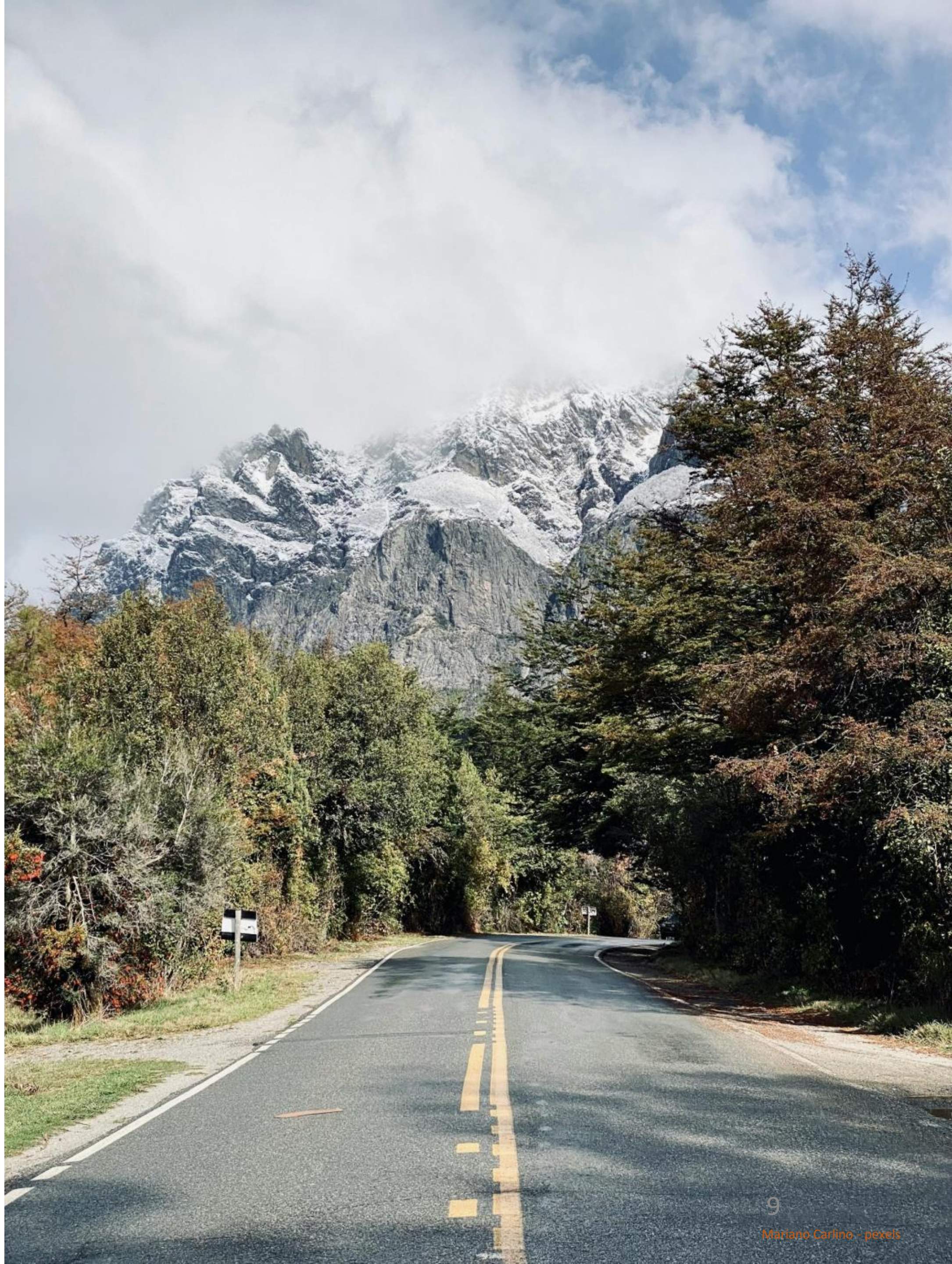
AI on the attack

- Deep fake puts Business Email Compromise (BEC) on steroids: $40m HK loss
- Writing and executing social engineering just got a whole lot easier
- Because AI does things in so many ways, anomaly detection (deviation from patterns of expected behavior) applies less today
- There is no downside for bad actors using AI (they don’t have to worry about bias, breaching IP, hallucination)
- In 1950 Norbert Wiener warned of faulty model training, and we can expect to see malicious actors take this path.

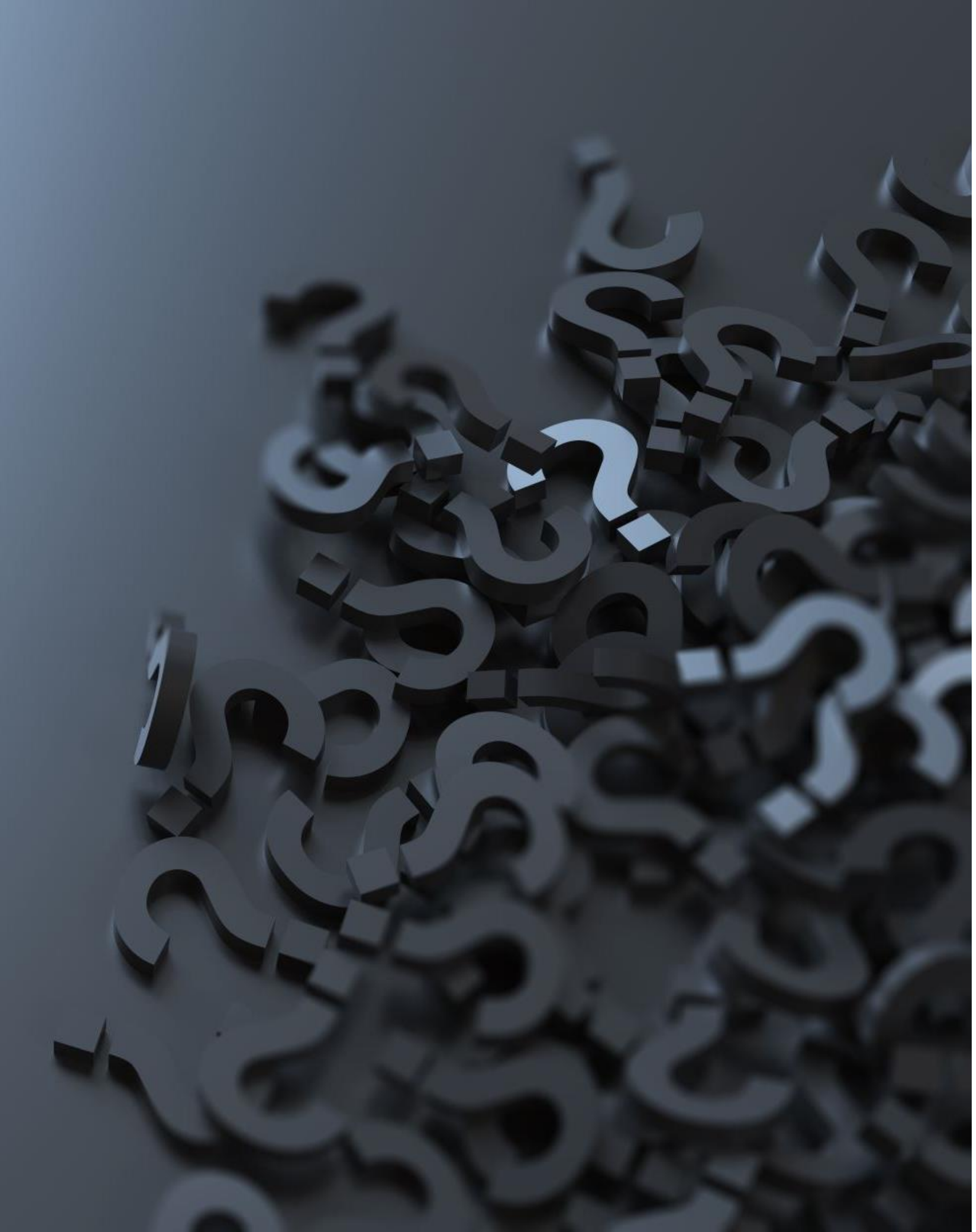
AI on the attack: Arup case study

AI risk preparation: starting a journey

- Obligations remain, whether you are in government or industry
- Ignorance is no excuse
- Risk frameworks remain
- Deal with overconfidence



Mariano Carlino - pexels

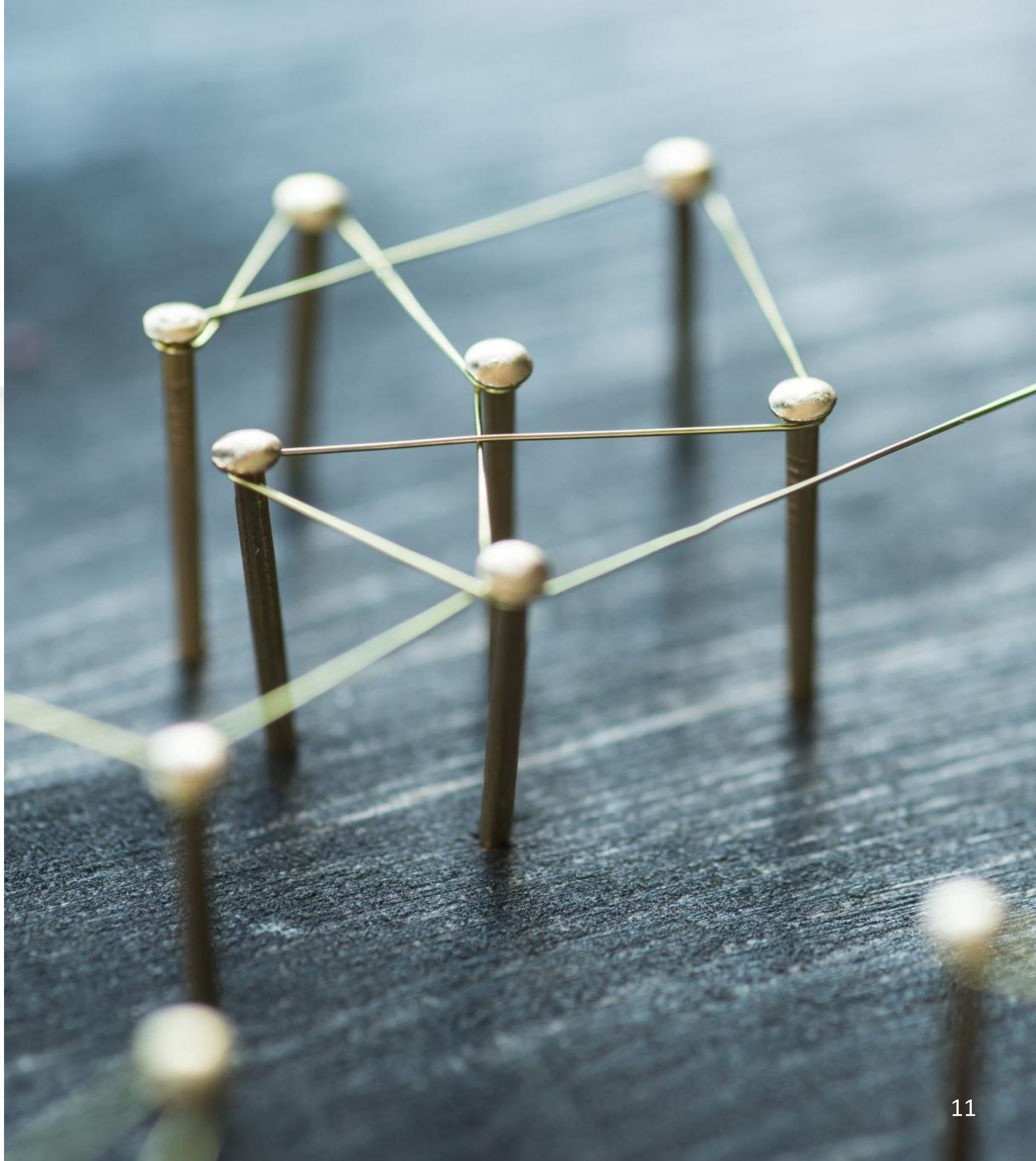


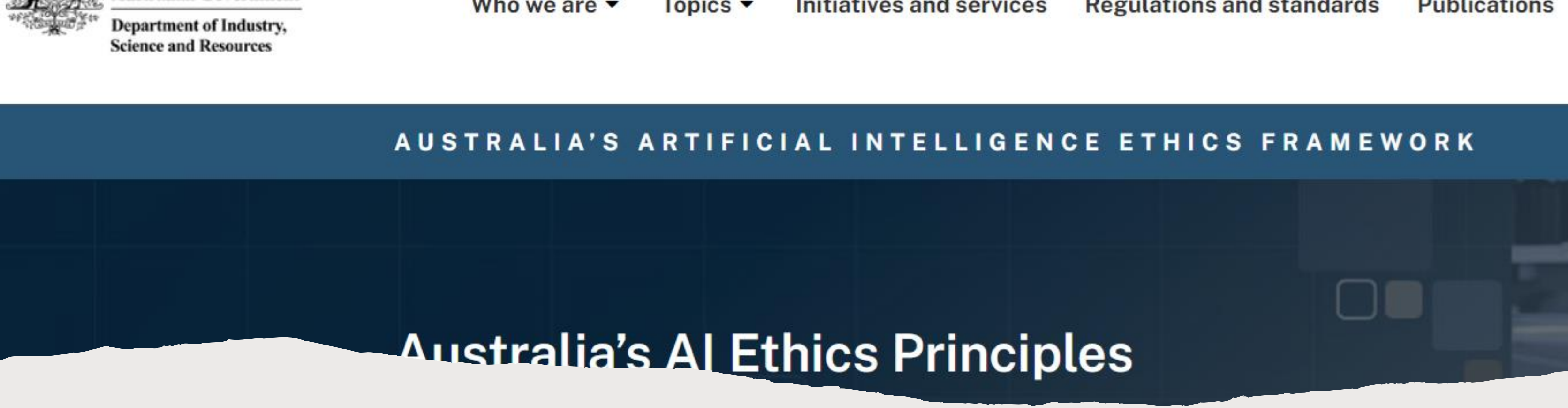
AI risk preparation: Questions for a PoC

- There are many checklists of AI risks
- These can inform Proofs of Concept
- To begin, the users should be trained to observe
- Don’t ask “How great was your experience? How much time did you save?”
- Use your identified AI risks to ask:
    - How many cases of bias did you discover?
    - What were the hallucinations you saw?
    - Could you pick any copyright infringements?
- Otherwise it isn’t a PoC, it’s a snow job.

AI for the defence: what can be done

- Understanding our attack surface
- Identifying gaps in data storage and classification
- Minimising false positives
- Making connections that are beyond human capacity





AUSTRALIA'S ARTIFICIAL INTELLIGENCE ETHICS FRAMEWORK

Australia's AI Ethics Principles

AI for the defence: IEEE CertifAIed

The first two Australian AI Ethics Principles state:

1. **Human, societal and environmental wellbeing:** AI systems should benefit individuals, society and the environment.
2. **Human-centred values:** AI systems should respect human rights, diversity, and the autonomy of individuals.

These were developed in collaboration with IEEE (Institute of Electrical and Electronics Engineers, with 470,000 members in 161 countries), which provides a certification program for AI compliance, IEEE CertifAIed.

Greg Adamson

Questions?

Key takeaways

- AI is in its infancy
- Cyber jobs are getting harder
- No one can explain what AI does
- We can't give up

https://www.linkedin.com/in/gregadamson/