

Ethical Concerns in AI Development and Deployment

Submission to the Productivity Commission on Artificial Intelligence (15/Sep/ 2025)

Author: Catherine Anderson-Karena, Director (Endpoint Strategy, trading name Test-Ed)

Introduction

Recent national surveys show Australians are deeply concerned about AI risks. Eighty percent believe preventing catastrophic AI risks should be treated as a global priority on par with pandemics and nuclear war¹. There is strong support for decisive governance to prevent dangerous outcomes.

This submission urges the Commission to ensure its agenda—spanning a dynamic and resilient economy, a skilled and adaptable workforce, harnessing data and digital technology, delivering quality care more efficiently, and the net-zero transformation—proceeds with explicit risk controls.

To that end, I ask the Commission to explore three cross-cutting areas of AI risk that underpin all five themes in your terms of reference:

1. **Use** – how AI affects rights, fairness, accountability and democratic processes.
2. **Infrastructure** – the environmental and resource footprint of AI (energy, water, minerals) and the concentration of compute power.
3. **Training** – consent, ownership, labour exploitation, and value-alignment embedded in models.

These three areas map directly onto the Commission's priorities:

- (i) sustaining a dynamic economy while preventing concentration of market power and systemic harms;
- (ii) building a workforce whose skills are complemented—not deskilled—by automation;
- (iii) governing data and digital systems with privacy, transparency, and auditability;
- (iv) safeguarding quality care with human-in-the-loop and liability for high-risk uses; and
- (v) ensuring AI's substantial energy and water demands are accounted for within realistic pathways to net zero rather than assumed away.

Note of intent (for later sections)

I will expand on two specific domains of personal concern:

- **Education:** the need for intelligent partnership with AI—as a tool that augments teachers and mentors, not as a replacement for them. Particularly as innovation is uniquely human.

¹<https://aigovernance.org.au/survey/>

- **Population surveillance:** the misuse of AI for mass tracking, profiling and compelled identity verification of Australians, including risks from client-side scanning (technology that examines content on your device before it's encrypted) and similar monitoring systems

These elaborations will provide concrete examples, evidence, and practical guardrails (auditability, consent pathways, red-team testing, independent oversight and liability) so that the Commission's productivity objectives are achieved with caution commensurate to the risks the Australian public already recognises.

1. Ethical Concerns in the Use of AI

Bias and Fairness: AI systems can reflect or amplify biases in the data they're trained on (e.g., racism, sexism, able-ism, or cultural bias). When deployed in policing, hiring, education, or healthcare, these biases can have life-altering consequences for individuals and perpetuate systemic discrimination.

Accountability and Transparency: Decisions made by AI are often opaque ("black box" effect). If an AI system harms someone, it's unclear who is responsible—the developer, the company, or the end-user. This lack of accountability undermines legal recourse and democratic oversight.

Privacy and Surveillance: AI powers facial recognition, data mining, and tracking tools that risk eroding privacy rights. Governments and corporations may use these tools to monitor populations in ways that undermine democracy and civil liberties.

Manipulation and Autonomy: Generative AI can create deepfakes, misinformation, or persuasive content designed to manipulate opinions, elections, or consumer choices. This raises questions about informed consent and human autonomy in democratic societies.

Dependence and Deskillling: Over reliance on AI tools has shown to reduce critical human skills (e.g., judgement, memory, writing ability, creative problem-solving), potentially undermining human agency and resilience.

Concentration of Power: The dominance of a few companies and countries in AI development creates dependencies that could undermine national sovereignty and democratic decision-making processes.

Employment and Social Disruption: Rapid AI deployment across industries threatens widespread job displacement without adequate social safety nets or retraining programs, risking significant social upheaval.

Table – Ethical Concerns in the Use of AI

While AI offers many potential benefits, its use raises significant ethical challenges. These challenges are not abstract — they affect human rights, democratic processes, and social cohesion. When misapplied, AI risks amplifying existing inequities, eroding trust, and undermining core human skills. The following table outlines the major risks associated with the use of AI, the stakeholders most affected, and strategies to ensure AI enhances rather than harms society.

Risk	Stakeholders Affected	Mitigation Strategies
Bias and Fairness	Marginalised communities, job applicants, patients, students.	• Mandatory algorithmic audits. • Independent bias testing. • Inclusion of diverse datasets.
Accountability and Transparency	Citizens, regulators, courts, governments.	• Establish legal liability frameworks. • Require explainability standards. • Create independent oversight bodies.
Privacy and Surveillance	General public, vulnerable populations, civil society.	• Strengthen privacy protections. • Limit government/corporate surveillance. • Explicit opt-in consent for sensitive uses.
Manipulation and Autonomy	Voters, consumers, democratic institutions.	• Regulate deepfakes and deceptive content. • Transparency labelling for synthetic media. • Media literacy initiatives.
Dependence and Deskilling	Students, workforce entrants, knowledge workers.	• Embed AI literacy in education. • Encourage human-AI collaboration. • Preserve critical thinking and creativity in curricula.
Concentration of Power	National governments, small businesses, citizens.	• Antitrust enforcement. • Public investment in open AI infrastructure. • International cooperation on AI standards.
Employment and Social Disruption	Workers, unions, families, communities.	• Fund large-scale retraining programs. • Strengthen social safety nets. • Incentivise industries that create new jobs.

2. Ethical Concerns in the Infrastructure Required

Environmental Cost: Training large AI models consumes massive energy and water resources, contributing to carbon emissions. Research indicates training large models can generate substantial emissions – for example, training GPT-3 was estimated to produce 552 tonnes of CO₂, contradicting climate change mitigation efforts.

Resource Inequality: Building and running advanced AI requires enormous computational power and rare minerals for chips. This concentrates control in a handful of wealthy corporations and countries, often exploiting poorer regions that supply raw materials.

Supply Chain Ethics: Mining cobalt, lithium, and rare earths for AI hardware often involves child labour and unsafe working conditions, perpetuating global inequality and human rights abuses.

Digital Divide: The infrastructure gap means poorer countries and smaller organisations cannot access AI at the same level, deepening global inequality and creating new forms of technological colonialism.

Table – Ethical Concerns in the Infrastructure Required

AI is often described as a “clean” or “immaterial” technology, but its infrastructure is resource-intensive and globally unequal. The environmental footprint of large models, the concentration of computational power, and the exploitation of global supply chains all raise ethical concerns. Without addressing these issues, AI development may deepen inequalities and contradict commitments to sustainability. The table below identifies the risks, stakeholders, and potential mitigations.

Risk	Stakeholders Affected	Mitigation Strategies
Environmental Cost	Local communities, environment, future generations.	• Mandatory carbon and water usage reporting. • Investment in green data centres. • Incentives for energy-efficient architectures.
Resource Inequality	Developing nations, small firms, global South.	• Fund equitable AI access programs. • Regional public AI infrastructure. • Promote cooperative international research.
Supply Chain Ethics	Workers in mining regions, low-income countries.	• Enforce supply chain transparency. • Require fair labour standards. • Expand hardware recycling/reuse.
Digital Divide	Developing nations, rural communities, under-resourced schools.	• Public investment in digital literacy. • Ensure equitable infrastructure access. • International development cooperation.

3. Ethical Concerns in Training AI

Data Consent and Creative Rights: Many models are trained on vast swaths of internet data without explicit consent from creators, raising copyright and intellectual property concerns. Artists, writers, musicians, and researchers often find their work used without credit or compensation, with AI systems then competing directly against the creators whose work trained them.

Economic Harm to Creators: The asymmetric power dynamic allows tech companies to profit from creators' unpaid labour while creators bear economic costs. Current copyright frameworks appear inadequate for addressing mass data scraping and the need for opt-in rather than opt-out consent systems.

Labour Exploitation: Human annotators, often in low-income countries, are paid very little to perform emotionally distressing tasks (e.g., filtering violent or sexual content) that make AI training possible, creating a hidden human cost in supposedly automated systems.

Transparency and Public Scrutiny: Companies rarely disclose exactly what data was used, how it was cleaned, or what guardrails were applied, limiting public scrutiny and democratic oversight of systems that affect society broadly.

Value Alignment and Cultural Bias: How "safety" and "morality" are defined in training often reflects the worldview of the organisations funding and deploying AI. This risks embedding cultural or political biases into supposedly "neutral" systems and marginalising minority voices or languages.

Regulatory Inadequacy: The speed of AI development outpaces regulatory frameworks, with insufficient international coordination on AI governance and limited mechanisms for public participation in AI development decisions that affect entire societies.

Irreversibility: Once models are trained on copyrighted or ethically problematic data, that knowledge cannot easily be "untrained," creating permanent ethical compromises in AI systems.

Table – Ethical Concerns in Training AI

The way AI systems are trained determines the values they carry and the impacts they have on society. Current practices raise questions of consent, fairness, and accountability. Creators often see their work used without permission, while low-paid annotators bear hidden human costs. Furthermore, cultural bias and inadequate regulation risk embedding structural inequities into supposedly neutral systems. The following table highlights these concerns, the stakeholders involved, and the strategies needed to build a fairer and more transparent AI training ecosystem.

Risk	Stakeholders Affected	Mitigation Strategies
Data Consent and Creative Rights	Authors, artists, musicians, journalists, researchers.	• Opt-in data consent systems. • Royalty/compensation frameworks. • Transparent data sourcing.
Economic Harm to Creators	Creative industries, independent workers, small publishers.	• Collective bargaining frameworks. • Fair licensing and remuneration systems. • Clear labelling of AI-generated works.
Labour Exploitation	Data annotators, low-income workers, content moderators.	• Minimum global labour standards. • Fair wages and psychological support. • Transparency on outsourcing practices.
Transparency and Public Scrutiny	Regulators, civil society, researchers.	• Mandatory “model nutrition labels.” • Independent audits of training data. • Public reporting of methodologies.
Value Alignment and Cultural Bias	Minority voices, non-dominant languages, civil society.	• Multi-stakeholder governance. • Pluralistic input in training data. • Independent review of alignment criteria.
Regulatory Inadequacy	Governments, global community, all citizens.	• Accelerate adaptive regulation. • Strengthen international cooperation. • Mechanisms for public participation.
Irreversibility	Future developers, societies dependent on AI.	• Legal mechanisms for redress. • Restrict deployment of models trained unethically. • Develop standards for dataset curation before training.

Summary

These concerns span three critical dimensions:

- **Use:** How AI affects human rights, fairness, accountability, and democratic processes
- **Infrastructure:** Environmental impacts and global equity considerations
- **Training:** Consent, ownership, labour exploitation, and values embedded in systems

The rapid deployment of AI without adequate ethical frameworks risks creating irreversible harms to creators' livelihoods, democratic institutions, environmental sustainability, and

global equity. Urgent regulatory intervention is needed to ensure AI development serves the public interest rather than concentrating power and externalising costs onto society's most vulnerable members.

1. Expanded Note on Education — Human-led, AI-supported

Real-world case study: Navigation Training Framework (proof of concept)

My background is in technical education and testing. I led an innovative, human-centred program built around mentors, projects, and practices that are difficult²—if not impossible—for AI to replicate. I expect to publish a book on the use and misuse of AI by November this year.

A decade ago, at Australia's largest IT social enterprise (WorkVentures), I demonstrated a pathway for disadvantaged young people—those failed by the education system—to enter top-tier companies in roles usually reserved for university graduates after just four months of targeted training. In their first year, participants earned at least \$10,000 more than those graduates. I then founded my own company and replicated these outcomes with every cohort, no matter how marginalised—transsexual, autistic, ADHD, schizophrenic, or homeless youth. All but one graduate secured employment within a week of finishing. One group began as contractors and were contracted out at \$1,200 per day, per person.

Example David*: From ages 5–16, David attended a school for the intellectually challenged and was placed on a disability pension for life after being diagnosed “intellectually disabled.” But at interview, he produced references for years of unpaid work (gardening, offices of MPs, takeaway stores—20+ hours/week for 2.5 years as proof that he could work). He arrived nervous but prepared, and had without his parents knowledge completed two Level-1 TAFE certificates in Digital Media and IT. Today, David is a Senior Test Automation Engineer at Qantas, building mission-critical systems, coaching programmers, and recently attaining managerial status. He earns \$170,000 per annum and owns his home. He has told me the program saved his life: being recognised and treated as an intelligent young man made all the difference.

Other graduates* overcame different barriers: Barry (short-term memory loss post-tumour) is now a DevOps lead at Medicare; Darrel (two generations of unemployment, initially unwilling to speak) now works in technical support at Tyro; Gavin (schizophrenia and institutional trauma) now leads a UI/UX team.

Success required subject-matter expertise across multiple domains: testing, programming, communication, project management, writing, and conflict resolution. Above all, it required the ability to read students and manage group dynamics to accelerate learning.

My concern is the push to **automate what cannot be automated**: on-the-spot innovation, intelligent analysis in response to unique human challenges.

² [References](#): See Catherine (Anderson) Karena *(Proof of students successes can be provided directly to Productivity Commission.)

There is a useful analogy between the adoption of test automation and the adoption of AI. Management often assumed testers could be replaced when automation arrived—causing serious failures across major systems worldwide. Automation doesn't think.

AI has well-documented flaws: bias, hallucination, and over-trust ("it's a machine, so it must be objective"). Young people also face temptations to "cut and paste" prompts rather than practise critical thinking, disciplined inquiry, and focus.

Education is more than data. The human-led education I deliver depends on real-time judgement and human input—by trained practitioners with real-world experience not only in subject matter, but in communication, conflict, support, bonding, stress and workplace realities. Those elements gave my students an advantage over university-trained peers.

Over 35 years working with leaders and managers, I've observed a recurring tendency to seek the easy route, to put process over outcomes. Programs like mine are **human learning journeys** and must be led by humans.

That said, I am enthusiastic about **intelligent partnership with AI**—as a tool that augments teachers and mentors, not as a replacement for them.

Pros and limitations of AI in education (youth)³

Potential benefits (when teacher-led and well-governed)

- **Improved academic outcomes and immediate feedback:** Studies report better results and richer, instant feedback when students use chatbots in higher education settings.
- **Personalisation and self-directed learning** in high student–teacher ratio contexts: chatbots can guide problem-solving, enable self-quizzing, and adapt instruction to individual needs, increasing engagement and motivation.
- **Greater independence in knowledge acquisition** when virtual agents are used as support tools (not substitutes for teachers).
- **Operational gains argued by proponents:** scenarios note higher exam pass rates and potential cost reductions—though these claims often accompany proposals to remove teachers, which raises separate risks (see below).

Documented limitations and risks (especially if used to replace teachers)

- **Educational dehumanisation:** excessive reliance on machines can hinder development of empathy and social skills; risks include emotional disconnection and social fragmentation.
- **Over-dependence, reduced human interaction, and privacy risks;** plus practical barriers—adequate infrastructure, teacher training, and equitable access.

³ [References](#)

- **Loss of social-emotional development and mental-health concerns** when human presence is removed; cautionary evidence warns of increased disorders and societal disconnection.
 - **Bias, lack of empathy, and quality risks** from over-trusting “objective” automated systems in classrooms.
 - **Equity impacts:** automated education may disproportionately harm vulnerable students and deepen inequalities.
 - **Deskilling risks** for students and teachers (over-reliance on tools in place of critical thinking and practice)—already flagged in your base framework.
-

2. Expanded Note – Population Surveillance

Scope. This note addresses the misuse of AI for mass tracking, profiling, and compelled identity verification of Australians, including the risks posed by client-side scanning and other forms of **pre-encryption monitoring**.

Why this approach. The following section uses a thought experiment to illustrate how current technological capabilities could evolve under weak governance frameworks. While the tone shifts from academic analysis, the underlying technologies and legal authorities described are real and documented. During COVID, the public was assured that contact-tracing data would not be used for other purposes; subsequently, NSW Police accessed this data for criminal investigations. That episode, combined with recent proposals in law and tools, shows we are entering untrustworthy terrain if guardrails are weak.

A concrete risk picture. Please consider this simple thought experiment: every photo from your child’s school sports day, every family video, every message about a teenager’s struggles—**scanned on your device, analysed, and retained**. Extend that further: everything you read, write, or view is **observed first on your device** and only then encrypted or sent.

The risk is not abstract. Non-technical Australians are largely unaware of how quickly the **infrastructure for pervasive inspection** can be laid down under the banner of “safety,” and later repurposed. Left unchecked, these capabilities could constrain movement, speech, association, and even everyday transactions (when tied to identity systems), all “for the children.”

The Real Technologies & Laws (illustrative, evolving)

The Technology (capabilities to watch)

Apple – MediaAnalysisd: Built-in tool that scans photos and videos on iPhones/iPads, originally part of the abandoned *NeuralHash* project for CSAM detection. Put simply, this means Apple can see your family photos before you even share them.

Microsoft – Windows Recall: Takes automatic screenshots every few seconds, allowing AI to read and analyse everything on your screen. Put simply, this means Microsoft can capture anything and everything you write and read, and store.

Google – Safety Core (2025): Embedded into Android/Google systems, designed for “safety” but capable of analysing what users see and do.

Client-Side Scanning (CSS): Scans your device’s content *before* it’s encrypted, breaking the principle of private communication.

(Note: some features have been proposed, piloted, revised, paused, or geo-limited. The risk arises from the architecture: once device-side scanning is normalised, scope is easily broadened.)



The Laws (selected)

- **EU “Chat Control” (pending)** – would mandate client-side scanning across Europe, overriding end-to-end encryption.
- **UK Online Safety Act (2023, enforced 2025)** – requires platforms to prevent “harmful” content; fines of up to 10% global revenue.
- **Australia Online Safety Act 2021** – gives the **eSafety Commissioner** sweeping powers to order removal of material, issue fines, and compel platforms to hand over data.
- **Digital ID (Australia)** – legislation underway to link online access to government-approved identity, ending online anonymity.
- **Age Verification Laws (EU/UK, proposed in Australia)** – require ID to access certain websites, paving the way for centralised identity tracking.



Why It Matters

- Privacy no longer exists when your device itself spies on you.
- “For the children” is used as the justification, but once in place, these systems can monitor *anything*: dissent, political speech, personal beliefs.
- What begins in the EU and UK spreads globally — Australia already mirrors these frameworks.

Thought experiment: *The Dictator's Monologue*

Scenario: A future leader's perspective on surveillance normalisation. The following illustrates how incremental acceptance of monitoring capabilities could enable authoritarian control:

“Citizens,

How easily you surrendered. You once spoke of privacy. Of liberty. Of rights.

Now you speak only of safety. And safety is mine to define.

You thought encryption protected you. You thought your words were safe.

But I am already inside your devices.

Apple gave me **MediaAnalysisd**.

Microsoft gave me **Windows Recall** — screenshots of every moment you work, every word you type.

Google gave me **Safety Core** — an eye that sees everything you see.

I no longer need to break your encryption. I simply scan you *before* you send.

And you thanked me for it. **“For the children,”** I said.

You applauded as I broke your freedom.

In Europe, they call it **Chat Control**.

In Britain, the **Online Safety Act**.

In Australia, it is your **Online Safety Act 2021**, already in force.

Here, I gave the **eSafety Commissioner** power to order removal of speech, to punish platforms, to silence dissent. And soon, with **Digital ID laws**, every word you post, every site you visit, will be tied to your name, your address, your life.

No more anonymity. No more hiding.

You gave me your driver's licences, your passports, your faces.

Now, one knock at the door is all it takes to end your voice.

You thought this was about children. You thought this was about safety.

But tell me — what is the difference between blocking CSAM and blocking dissent?

Between censoring “harmful” content and censoring opposition?

There is none.

And so I decide what you may read, what you may write, what you may think.

Already your neighbours accept 24/7 tracking as normal.

Already you believe you must have “nothing to hide.”

Already you fear saying the wrong thing, lest your accounts vanish, your bank be frozen, your name be erased. You carry me in your pocket. You invite me into your home.

And you pay for the privilege.

This is your world now.

A world where Apple, Microsoft, Google — and Canberra — serve me.

Where laws passed in Brussels or London echo instantly in Sydney.

Where I flip the switch, and Australia falls in line.

I am in your phone. I am in your screen. I am in your very thoughts.

And you pay for it all.”

Minimal recommendations

1. **Ban general-population client-side scanning** in Australia; require that any device-side inspection be opt-in, narrow in scope, and independently audited.
2. **Protect end-to-end encryption in law.** No decryption mandates, no “exception” backdoors.
3. **Purpose-limitation and warrant standards** for any cross-use of safety data (e.g., health/crime/policing).
4. **No linkage of routine internet use to Digital ID.** Identity only where strictly necessary, with statutory anonymity protections.
5. **Transparency & red-teaming.** Independent safety evaluations, public reports, and penalties for covert scope expansion.

References:

- <https://aigovernance.org.au/survey/>
We conducted the survey in early 2024 and found 80% of Australians believe preventing catastrophic risks from advanced AI systems should be a global priority on par with pandemics and nuclear war.
- <https://www.cold-takes.com/racing-through-a-minefield-the-ai-deployment-problem/> The promise of powerful technology may tempt companies – and countries – to race ahead without heeding the risks.
- <https://doi.org/10.1186/s40309-025-00253-9> The future of AI in government services and global risks: insights from design fictions
- Supporting responsible AI: discussion paper
<https://consult.industry.gov.au/supporting-responsible-ai>
- **AI in Schools: Pros and Cons,**
 - <https://education.illinois.edu/about/news-events/news/article/2024/10/24/ai-in-schools--pros-and-cons>
 - <https://doi.org/10.3390/educsci15030343> The Impact of Artificial Intelligence (AI) on Students' Academic Development
- **80% of Australians think AI risk is a global priority.**
<https://news.uq.edu.au/2024-03-11-80-australians-think-ai-risk-global-priority-government-needs-step>

Catherine (Anderson) Karena – background

- <https://test-ed.com.au> (Provides contributors to the program: Atlassian, Google, Tyro, Suncorp, International Tech Influencers.)
- Track: It takes a village to raise a tester,
<https://web.archive.org/web/20160408151905/http://testengineeringalliance.com/track-it-takes-a-village-to-raise-a-tester/>

- The WorkVentures Traineeship Program: <https://youtu.be/IRXFqo4UhKw?t=97>
- Sydney Testers Talk: "Hiring and Training the Marginalised and Neurodiverse: Catherine Karena & Mike Tozer"
<https://www.youtube.com/watch?v=llwHQgu0Ypl>
- <https://www.linkedin.com/pulse/test-ed-good-go-catherine-karena/>
- <https://www.linkedin.com/in/ckarena/>

Contact: Catherine Anderson-Karena
Sydney, Australia.