Acceptable Use of Generative AI

VUW AI and Society Group
March 2025

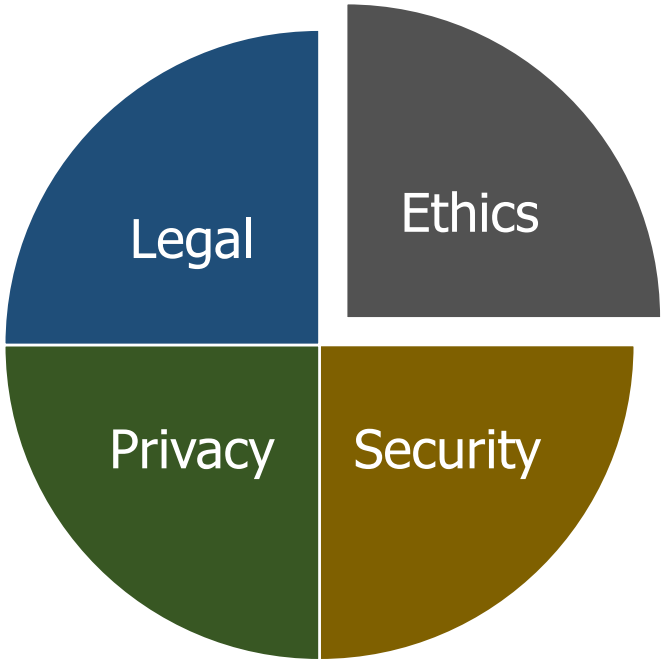
Dr. Andrew Chen
Chief Advisor: Technology Assurance, New Zealand Police

NEW ZEALAND POLICE
Ngā Pirihimana o Aotearoa

Technology Assurance

Strategy - maintaining public trust and confidence in Police

- Managing technology-related risks for our people, for NZ Police, and for the public
- Trial or Adoption of New Policing Technology policy
- Internal governance processes
  + External Expert Panel on Emergent Technologies
- Aiming to improve transparency where we can
  - Publicly-available Technology Capabilities List



NEW ZEALAND POLICE
Ngā Pirihimana o Aotearoa

AI in Law Enforcement

- Balancing pressure to adopt AI to help fight crime and keep communities safe against concerns about potential harms from poor use of these tools

- Many, many frameworks in general and law enforcement contexts
  - No commonly accepted definition of artificial intelligence
  - Two literature review reports written by EBPC, based on open-source info
- No AI-specific legal framework in New Zealand (yet – GCDO is working on it)
  - Patchwork of other legislation (e.g. Privacy Act 2020)
  - Voluntary Commitments (e.g. Algorithm Charter)

- Some (non-generative) AI tools have been in use by NZP for many years
- Blanket ban on the use of generative AI since May 2023

Existing Uses of (Non-Generative) AI at NZP

- Triaging online forms (105)
- Controlling RPAS (drones)
- ANPR (historical and real-time) and Facial Recognition (retrospective only)
- Triaging drug samples
- Risk scoring for youth offenders, domestic violence
- Some “algorithms” that some people might call “AI”

Generative AI Policy Considerations

- 15,000 person organisation, somewhat decentralised structure with some autonomy in Districts, a wide variance in tech capability

- High interest internally – seeing what is being achieved overseas + AI hype vs those skeptical of genAI's capabilities + those fearful of automation
- High interest externally – Ministerial comments to adopt technology to improve delivery of the public service vs concern about Police use of Technology
  - Consequences of NZP misuse/errors are high for the public

- Need to consider operational vs corporate use cases (i.e. front/back office)
- Need to consider speed of adoption vs constrained fiscal environment
- Hard to identify a business owner for a broadly enabling technology

NEW ZEALAND POLICE
Ngā Pirihimana o Aotearoa

Acceptable Use of Generative AI

- Policy approved on 10 March 2025 – becomes a chapter of Police Instructions
  - Made publicly available to support transparency (12pm today!)
- Need to mitigate bias, inaccuracies, privacy, infosec, automation bias risks

- Each tool to be evaluated separately with a limited approval scope
- All users must go through online training before getting access to genAI tools
- People remain accountable for using outputs of genAI

- Fully automated use of genAI tools not currently permitted
- **Use of genAI outputs in any court context not currently permitted**
- Six-monthly audits of use, and six-month review period of policy

NEW ZEALAND POLICE
Ngā Pirihimana o Aotearoa

	Very Low Risk	Low Risk	Medium Risk	High Risk	Unacceptable Risk
Necessity	In alignment with Policing values				Not in alignment with Policing values
	One-off or limited uses	Built into processes or SOPs with ongoing evaluation of effectiveness and necessity			Repeatedly used without evaluation
	Internal impacts only, no interaction with the public, some organisational reliance on the outputs		Internal or External impacts, with limited public interaction	Internal or External impacts, including interacting with any member of the public	
Effectiveness	No evaluation required due to low risk		Performance of the system is monitored and evaluated		No evaluation or monitoring planned
Lawfulness	Complies with all legal and regulatory requirements and obligations, including consideration for disclosure and discoverability				Significant uncertainty over legality of use
Partnership	No engagement required due to low risk and no interaction with the public		Engagement with affected communities to incorporate their perspectives		No engagement with affected communities
Fairness	System bias is not relevant to use case	System bias is well understood and managed		System bias is uncontrolled or unknown	
	Context and edge cases are comprehensively understood		Trials conducted to ensure effectiveness and safety of genAI system, with some uncertainty		No planning to identify edge cases or mitigate potential harms
	There are no genAI-enabled decisions	There is a process for people to contest genAI-enabled decisions			There is no process for people to contest genAI-enabled decisions
Privacy	No use of personal information	Minimal use of personal information, with no privacy harm	Use of personal information, with potential privacy harm(s) for individuals or groups		Use of personal information that will likely result in significant privacy harm
Security	No use of Police information	High quality Police information exposed to the genAI system		Uncontrolled or low-quality Police information exposed to the genAI system	
	UNCLASSIFIED information only	Information up to IN-CONFIDENCE		Information up to SENSITIVE	Information above SENSITIVE
	Used within the Police Enterprise Network only			Used through external cloud-based services	
	No risk (or controlled risk) of data leakage				Some risk of data leakage to third parties
Proportionality	Errors have negligible impact	Low to no harm from system malfunction	Some or uncertain harm from system malfunction	High or uncertain harm from system malfunction	Extreme harm from system malfunction (e.g. incorrect arrest or imprisonment)
		Errors can be detected and corrected without causing any harm		Errors detectable but harm may not be mitigated	No mechanism to detect or correct errors, causing harm
	The benefit of using the genAI tool outweighs any potential harm			The benefit of using the genAI tool is uncertain or unjustified	
Oversight and Accountability	A human is fully accountable for outcomes		Some process automation is hidden from humans, with ability for humans to override or intervene		Full automation (no human-in-the-loop)
	Users are trained and understand the limitations of the tools				Users undergo no training before use of the tools
Transparency	Use of genAI acknowledged and labelled		Use of genAI acknowledged publicly but not labelled for operational reasons	Use of genAI not acknowledged publicly for operational reasons	

- Used when evaluating genAI tools
- EU AI Act-esque approach (but in the EU any law enforcement use is High Risk or above)
- Aligns with existing Technology Assurance Framework and process
- Approval scopes may be limited based on specific use cases, work groups, authorisation requirements, risk levels, or a combination
- Allows a balance to let people use the tools where it is safe, rather than a blanket open/closed approach

NEW ZEALAND POLICE
Ngā Pirihimana o Aotearoa

Acceptable Use of Generative AI

- Other policy features:
  - Labelling – transparency and accountability
  - Authority for Urgent Use – with high authorisation threshold
  - Use of Generative AI by Vendors – closing off loopholes
  - Governance model – business owner at Superintendent or higher
  - Supporting the ANZPAA AI *Principles* and *Framework*
  - Encouraging people to ask for help when unsure about risk
  - Free online genAI tools continue to be blocked – data leakage unacceptable
  - Future real-time monitoring of prompts and assessing risk

Current Use of Generative AI at NZP

- None approved for general/operational use at this time

- TRIAL: M365 Copilot Chat (250 participants)
  - “General Purpose” genAI tool, very hard to limit by use case
  - Very easy for people to use, mostly targeted towards corporate use cases
  - Technically very limited (e.g.1MB file size limit) but cheap

Future Use of (gen)AI

- Broader use of transcription and translation
- Supporting natural language information retrieval
- Further opportunities in:
  - Process automation (improving data quality and process speed)
  - Anomaly detection and saliency in text, audio, video
  - Summarisation of very large amounts of content
- “Can we do it?” versus “Should we do it?”
  - Trust and Confidence are critical to our organisation
  - Not doing things can be just as important – as long as we are transparent
  - Data quality limited for some high-value applications
  - Natural justice demands human accountability

Being a critical friend

tech.assurance@police.govt.nz

NEW ZEALAND POLICE
Ngā Pirihimana o Aotearoa