Initial advice on Generative Artificial Intelligence in the public service

Joint guidance from data, digital, procurement, privacy and cyber security system leaders on responsible and trustworthy use of Generative Artificial Intelligence (GenAI) across the New Zealand Public Service.

July 2023.

UNCLASSIFIED

This advice provides interim guidance on using GenAI in the New Zealand Public Service

This guidance provides initial advice from the data, digital, privacy, procurement and security System Leaders about Public Service use of GenAI tools. This document and its attached A3 are intended to support agencies to make more informed decisions about using GenAI, balancing benefits and risks. Whilst we recognise that this guidance could have broader application and usefulness beyond the Public Service, it is intended for Public Service AI practitioners and decision-makers. This advice is the first collective effort of System Leaders to help agencies start to trial and use this new class of technology safely, ethically and in privacy-protecting ways.

This guide provides 'guardrails' supporting safe learning of GenAI tools and may be updated as the technology evolves and as the risks and their impacts are better understood. It is also an interim measure while longer-term plans for broader GenAI/AI and other emerging technologies issues are developed and progressed by System Leads. Work on economy wide opportunities and impacts, and NZ wide regulatory settings, may also be needed.

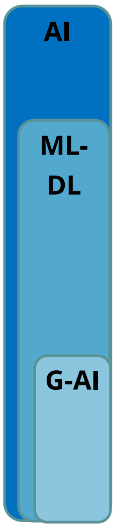
Who is this advice for?

This advice is intended for public service procurement, data, digital, privacy and security leaders. It is intended to help you to better understand the key risks of using GenAI within the New Zealand Public Service, and mitigations, to support you to develop your policy, standards and plans for responsibly using GenAI within your agency context.

> **We recommend that your organisation's data, digital, privacy, procurement and security leaders work together to create your organisation's policy and standards for trialling and using GenAI.**

What is GenAI?

Generative AI (GenAI) is incredibly popular across the world. We know that public servants are increasingly seeking to understand GenAI, and to use it to improve the services they provide to New Zealanders.



Artificial intelligence is the field of computer science that seeks to create engineered systems that can generate outputs for particular sets of objectives, without explicit programming.

Machine learning is a subset of AI that trains machines to learn from existing data and improve upon that data to make decisions or predictions. **Deep learning** is a more specialised machine learning technique in which more complex layers of data and neural networks are used to process data and make decisions.

GenAI can use prompts or questions to generate text or images that closely resemble human-created content. These tools work by matching user prompts to patterns in training data and probabilistically "filling in the blank." ChatGPT is the most well-known, free, example of a GenAI.

GenAI is being integrated into many tools we use

GenAI is something public servants will not be able to avoid, even if they wanted to. It is being integrated into many services across the cloud, security, creative and media sectors just to name a few. GenAI is here and, like any new potentially disruptive technology, we need to learn about it, understand it and work through how to use it safely.

We encourage agencies seeking to use GenAI to understand the necessity for it and use it cautiously

We think that GenAI could offer many benefits for the public service. These could include:

- **Efficiency and productivity** through process simplification and automation
- **Improved service design and delivery** through targeting and personalisation
- **Enhanced cyber monitoring and defence** through advanced predictive analysis, vulnerability assessment and threat detection
- **Innovation** from optimisation and access to "big-data" based insight
- **Improved policy development** through accessing fuller data and insight, and availability of nearer real-time analysis

We strongly recommend that you:

Don’t use GenAI tools for data classified at SENSITIVE[1] or above. The risks for security and potential impacts if SENSITIVE or above datasets were to be compromised could be catastrophic for our society and economy, and public services. Take all necessary steps to avoid inputting these types of at risk datasets into GenAI tools.

Don’t input personal information to GenAI tools if they are external to your environment. The risks to people, and their trust and confidence in their government, if their personal information is compromised, could be significant. Take all necessary steps to avoid inputting personal information into GenAI tools that are outside your network.

We also recommend that you:

Avoid inputting personal data, including client data, into GenAI in your network, and exercise extreme caution if personal information is involved. The government is held to a higher standard in respect of trust and confidence and so we recommend extreme caution where personal information is, or could be, involved. Don't include personal information (particularly client information) when using GenAI, unless (a) it isn’t possible to use non-personal or synthetic data; and (b) all potential harms have been addressed.

Avoid permitting GenAI to be used as Shadow IT. There could be risk of unsanctioned use of GenAI by teams operating within your environment. This may create risk of technologies being used in ways that could compromise security, increase risk of data or privacy breaches, and/or add complexity to your technology environment in ways that could lead to service disruption or conflict.

Also, be aware that free GenAI tools could carry higher risk, and paid GenAI also carries risk. Free GenAI tools may not have the same robust privacy and security controls as paid GenAI tools, and there could be potential for inputted data to be used for unauthorised purposes or in non-transparent ways. Quality and reliability may also be an issue. Free and paid GenAI tools come with their own sets of risks, and we recommend that you consider factors such as cost, availability, functionality, privacy and security and maintenance/support when making choices about free vs. paid GenAI tools

We recommend blocking access to GenAI tools through your network until this guidance has been applied.

Avoid inputting into GenAI tools any information that would be withheld under the Official Information Act. The risks for the integrity of the Public Service, and potential impacts if redacted information were to be accessed and/or inappropriately used, could be

[1] NZ PSR Classification System

extremely damaging for public trust and confidence. Take care when using GenAI for data where that data would satisfy OIA withholding grounds.

Avoid using GenAI for business-critical information, systems or public-facing channels. GenAI can generate inaccurate and incomplete outputs and has the potential to perpetuate bias and mis/dis information. Further, AI systems can be complex to understand, causing issues where understanding decision-making processes is important. Avoid using GenAI for your agency's essential systems and services.

Understanding and actively managing risks is key to maintaining the integrity of the Public Service

Whilst the benefit potential is substantial, there are several risks to Public Service use of GenAI that, if they were to materialise, could seriously damage public trust be potentially harmful for the public. **We recommend fully assessing and actively managing for the risks,** to support informed decisions on when and how your agency uses GenAI.

Robustly govern the use of GenAI

Consider your governance system and obtain senior approval of GenAI decisions relating to the use and application of GenAI in your agency context. Consider also appropriate involvement of your Te Tiriti partners.

We recommend that your agency develops a GenAI/AI policy and standards to guide your agency's use of GenAI/AI and share this policy with the Government Chief Privacy Officer (email: gcpo@dia.govt.nz).

Because GenAI has the potential to cause significant harm, any input into it and use of its outputs should be robustly governed by accountable humans. Decisions based on GenAI can be significant for society, the economy and environment; and it is important that GenAI outputs, and human-made decisions based on them, are fair, transparent and unbiased.

Governance of use and application of GenAI should be based on the principles of safe and ethical use, security and privacy by design, be Te Tiriti-based and aligned to Government's Procurement Rules, and grounded in openness, transparency and accountability. Ongoing dialogue and collaboration between teams using GenAI tools and your agency's functional leads[2] will be important to effective governance of the use and application of GenAI tools.

Assess and manage for privacy risks

Take all necessary steps to protect privacy. This includes doing and getting senior approval of a robust privacy impact assessment for any testing or use of GenAI.

[2] For procurement, data, digital, privacy and security

Applying privacy by design principles can help to build trust in GenAI tools and systems through ensuring they are privacy-compliant, transparent and respect people's privacy and reduce risks of privacy breaches.

Consider how Te Tiriti in relation to Māori privacy might apply to any use or application of GenAI. Engage Iwi Māori about the impacts of GenAI for Māori data and/or outcomes.

We recommend identifying privacy risks and minimising how GenAI tools that contain personal data are used. Data anonymisation is recommended to reduce risk of identification. Access, encryption and minimisation should also be considered, to ensure that outputs are only accessible to those who are authorised to view or use them.

Openness, transparency and accountability are key to maintaining trust, confidence and integrity. Be transparent with your stakeholders and the public more broadly about how you are using GenAI for the benefit of the public, and how personal data related to GenAI tools used will be collected, managed and used.

A privacy impact assessment (PIA) will help you to identify and manage privacy risks. Obtain senior approval of your PIA so as to ensure your senior leaders are properly informed. Also copy your PIA to the GCPO to enable the GCPO to identify further support that agencies may need. Actively govern and manage for the identified risks and seek support from the GCPO if you are unsure of what to do, or if critical risks materialise.

Assess and control for security risks

GenAI can increase risk of security breaches. Attackers can use these tools to manipulate or socially engineer your staff and potentially make it easier to produce malware. Managing these risks requires strong cybersecurity practices – e.g., robust security culture around emails.

Be cautious of using GenAI to generate code for your agency, as this could result in insecure code.

Data spills: third party GenAI services are run by independent organisations that will likely have visibility of the queries or prompts that you use to generate its answers. You have limited visibility of where prompt information is stored and how it is used. This creates risks relating to data spills, reverse-engineering datasets and potential disclosure of sensitive information.

You have limited control of how they are run, limited configuration control and how they store or use data. This is a similar risk to many software as a service cloud services that you might consume. However, the difference is that GenAI is still a new field, there are not yet well understood transparency and assurance tools (e.g., SOC2 or ISO27001 reports) available to help you understand and manage these risks.

Insecure code generation: GenAIs often create insecure code that have vulnerabilities and common mistakes. Agencies should not be using code generated from GenAI to put into their production systems without robust review.

You should always conduct testing and assurance processes to ensure that any outputs your staff use are checked and confirmed as correct and safe to use.

Through applying your agency's normal security principles and processes, your developers and users of GenAI can help to ensure that AI systems are secure, trustworthy and resilient to potential threats.

Several GenAI tools have "opt out" functionality that provides choices for retaining your data for further training the AI. Opt out of the GenAI tools retaining and using your data for training purposes, if possible.

Consider Te Tiriti o Waitangi (the Treaty of Waitangi)

Work with Iwi Māori where GenAI may use Māori data and/or its use may impact Māori including services to Māori.

Māori representatives have expressed varying views, some strongly held, in respect of Government use of GenAI tools. There is heightened concern among Indigenous groups, in particular about discrimination at the hands of GenAI.[3] We strongly recommend working with your Te Tiriti partners, where Māori data is involved and/or where Māori interests or outcomes could be affected through using GenAI.

We recommend understanding important context for Māori and the Crown, including why GenAI is being considered, how it could impact Māori including services to Māori, what Māori data might be involved and its status across "tapu" or "noa," and how Māori Data Governance might apply.

Where Māori data is involved, we recommend aligning to your agency's existing Māori-Crown relationship approach and leveraging existing engagement models. As part of this you may wish to consider working more actively with your Tiriti partners on the opportunities for sharing of decisions, and/or exploring of more Māori-led approaches.

Considering how your teams work with Māori for mutual benefit is recommended. You may wish to consider how engagements can be simple, equitable, safe and value-adding for Māori participants. We also recommend considering how to ensure your team has the capability needed to engage with Māori confidently and successfully.

Use AI ethically and ensure accuracy

[3] See e.g., "Indigenous groups in NZ, US fear colonisation as AI learns their languages", Reuters, (3 April 2023)

GenAI can perpetuate bias and mis/dis information. Understand the limitations and take active steps to check for accuracy when using GenAI outputs, to avoid harm.

Ethics should be at the centre of how the Public Service uses GenAI systems and tools. AI can perpetuate existing biases and discrimination if they are not properly designed and used. GenAI systems should be designed to avoid discrimination, with outputs regularly check for bias and other possible harms.

Māori communities could be at higher risk of bias and discrimination in the results and application of GenAI outputs, so work with your Tiriti partners to understand and actively manage for the impacts of GenAI for Māori.

We recommend validating and scrutinising GenAI outputs, to reduce the potential for discrimination against minority peoples/groups, including for example women, ethnic, older and/or disabled communities.

Validate all outputs before using them in practice. There are many examples of GenAI 'making up' information or returning misleading results ("hallucinations").

Ensuring the accuracy of GenAI outputs is critical to trustworthy use of GenAI. It is essential that data used to train AI tools is of high-quality, for quality outputs. Cleansing, validating and quality-assuring data can help to ensure accuracy and reliability of outputs. Good algorithms also play a key role; if algorithms are poorly designed the results, they return could be erroneous, inaccurate or inappropriate. Iterative testing, adjustment, validation and monitoring are key to tuning and optimising algorithmic performance.

Be accountable

Always ensure accountable humans are making decisions in respect of the application or use of GenAI outputs, and that the decision-makers have the necessary authority and skills.

Human oversight and control are essential for ensuring GenAI is used ethically and responsibly. GenAI can produce misleading or biased results, so we strongly recommend human oversight of validating, verifying and interpreting AI outputs. Human governance over the data provided to AI tools and how the outputs are applied is also recommended.

Human decision-making of the overall risk assessment, taken by decision-makers with appropriate authority and capability, is also highly recommended. It is important that the people in these decision-making roles have the necessary authority and capability to make these decisions on behalf of the agency.

Be transparent, including to the public

Be open and transparent in terms of what GenAI is being used for and why. Ensure processes are in place to respond to citizen requests to access/correct information.

Citizens are concerned about ethical use of GenAI, and the public has expectations about how GenAI is being used, particularly for the Public Service. The Public Service is held to a higher standard; so, consider your social licence and how to assure transparency, accountability, and fairness in how your agency is using and applying GenAI, whether directly or as part of a wider technology solution.

Exercise caution when using publicly available AI

Be aware of the potential security, quality, intellectual property and supply chain risks of publicly available AI and mitigate risks where possible before using AI tools. Publicly available AI tools sit outside an agency's own environment. They are third party AI platforms, tools or software that have not necessarily been risk assessed to the standard expected for NZ government agencies or are not part of a commercial contract established through government assurance and procurement processes.

For example, some publicly available AI software may be developed by communities of developers, without assurance that the software is secure and free from vulnerabilities. This could lead to security breaches or other issues if the software is not properly tested and maintained within a controlled environment.

Further, contributing public AI software developers could have differing levels of experience, so quality and absence of errors cannot always be assured.

Lastly, support and distribution licenses for publicly available AI software vary, attracting some risk around intellectual property protection, performance, levels of support and maintenance over time.

We recommend taking steps to assess the testing, maintenance, and governance of any type of AI software you are using to ensure it is secure, appropriate, of high quality, and properly supported over time.

Apply the Government procurement principles

There may be increased risk of vendor lock-in, and exposure if providers are using GenAI in the services they provide you. Use the procurement rules and consider the mitigations needed for these when sourcing GenAI tools.

Some of your providers could be, or could be planning to, use GenAI for the services they provide you. It is essential to have visibility, and ideally control, over how your vendor is using and/or integrating GenAI into the solutions/services they provide you.

AI systems are often proprietary, and vendors can be reluctant to support integration and / or interoperability with other systems. The technology is also swiftly evolving, and as a result, capability can become quickly obsolescent.

Procurement teams should conduct market research on vendors and their offerings, considering vendor reputation, capability, pricing and the supply chain. We recommend a coordinated approach across agency functional leads[4] to support this evaluation. For trustworthy use of AI, and ultimately value for taxpayer dollar, teams are encouraged to also consider including specific commercial protections in contracts with your vendors for: privacy, security and ethical risks, technology obsolescence, vendor lock-in, and reliance on third-party provided services/AI.

Testing safely

Create guardrails and dedicated testing spaces, like sandboxes[5], for your teams to safely trial and learn to use GenAI.

Safely trialling GenAI is important to ensuring that AI systems and their outputs are as expected, and do not cause unintended harm to people, communities, society, the economy and/or the environment. We suggest selecting appropriate “safe” and “lower-risk” datasets to learn and trial safe use AI tools, gauge data quality and test outputs before they are deployed. This should include testing under various conditions to identify issues, and to validate that models and training data are appropriate for use in New Zealand and will work in ways that are accurate and fair for New Zealanders.

Agencies could learn about these new tools and set up appropriate guardrails for their own contexts

GenAI represents significant benefit potential, and many public servants are keen to use this modern technology.

As system leaders we think agencies should start learning about GenAI to understand how to safely realise the benefits in a secure and privacy-protecting way. While we cannot predict how this new area of AI/ML will be used in the future, we are working on the assumption that it will be part of the lives of knowledge workers across the country.

Rather than suggest an interim ban, or a halt on using GenAI, we encourage agencies to safely trial and learn about it. But you should do so cautiously.

For information purposes, Appendix One attaches some examples of free vs. paid GenAI. It should be noted that list is point in time and will evolve as AI technologies evolve. We are considering the options for maintaining an accurate and up to date-list of available GenAI products, and assurance options for these.

Some example uses of GenAI include:

- HR advisors reviewing job ads for bias or overly complex language
- Policy analysts providing impact analysis during policy development

[4] digital, data, procurement, security, privacy

[5] A **sandbox** is a controlled testing environment created to conduct extensive testing of an application without affecting the actual software system

- Data or security analysts transferring code / scripts (e.g., KSL to SQL or R to SAS)
- Developers reviewing code for vulnerabilities
- Developers seeking explanations of error messages in code.
- Security analysts writing hunting queries when trying to understand the scope of an incident.

These are all small scale examples of how you can start to use GenAI tools, understand and learn how to prompt and speed up some of the tasks that public servants do daily.

What next?

This guide is interim advice that will be reviewed as AI/GenAI evolves, and as risks are better understood.

There is a need for further work to consider the longer-term issues and opportunities, and any further actions required for public service, wider society and economy, and our regulatory settings.

Whilst this initial advice is focussed on the public service; longer term plans for a broader look at AI issues and opportunities for New Zealand are being developed and progressed by System and Policy Leads.

For further inquiries or questions, please contact:

Digital Gcdo@dia.govt.nz	**Data** datalead@stats.govt.nz	**Information Security** info@ncsc.govt.nz
Procurement Gcdo@dia.govt.nz	**Privacy: Public Service** gcpo@dia.govt.nz	**Privacy: All** enquiries@privacy.org.nz

Further resources

Case studies

Ministry of Education Policy for Using AI

Further reading

For Data leads

Data Protection and Use Policy,

Algorithm Charter

Ngā Tikanga Paihere

Privacy, Human Rights and Ethics Framework.

For privacy leads

Office of the Privacy Commissioner | Generative Artificial Intelligence

For procurement leads

Government Procurement Rules

Government procurement principles

Government Procurement Charter

For security leads

Cloud Security Alliance white paper:
https://cloudsecurityalliance.org/artifacts/security-implications-of-chatgpt/

Trail of Bits: AI risk assessment paper

https://blog.trailofbits.com/2023/03/14/ai-security-safety-audit-assurance-heidy-khlaaf-odd/

NCSC UK on ChatGPT and large language model AI: https://www.ncsc.gov.uk/blog-post/chatgpt-and-large-language-models-whats-the-risk

MITRE ALTAS catalogue:

https://atlas.mitre.org/

Appendix one: examples of free vs. paid GenAI tools:

For information purposes, this Appendix includes some examples of free vs. paid GenAI. It should be noted that list is point in time and will evolve as AI technologies evolve. We are considering the options for maintaining an accurate and up to date-list of available GenAI products, and assurance options for these.

Examples of free GenAI tools	Examples of paid GenAI tools
• ChatGPT – conversational AI trained on a large set of text from 2021. • Bing and The New Bing – ChatGPT-based discussions and search results inside Microsoft Bing. • Google Bard is designed to function similarly to Chat GPT with the biggest difference being that Google pulls its data from the web. • Stable Diffusion – free stable artwork generator; no login required. • PicsArt AI Writer – generates marketing material (e.g., slogans, LinkedIn) for free. • Eightify summarises YouTube Video with AI ChatGPT – creates YouTube summaries. • Gimme Summary AI – Chrome extension that summarises web articles. • Whisper – provides online transcription of audio files using LLM from OpenAI. • Runway – is an AI photo and video editor. • Riffusion – generates music from text descriptions using Stable Diffusion.	• Microsoft Copilot (not yet available in NZ) – a GenAI tool that can automate tasks within the Office365 tools (e.g., Outlook, Word, Excel, PowerPoint and Teams). • Duet AI for Google Workspace – embeds the power of GenAI across all Google Workspace apps, helping to write, organise and visualise. • Zoom IQ – Zoom IQ is a smart companion to Zoom tools that use technology from OpenAI to help provide meeting information. • Slack GPT – is bringing the power of GenAI, through having natively embedded summarisation and text generation, within Slack's existing features for its core messaging services. • GrammarlyGO – is expanding beyond grammar and spell-checking to writing.