Responsible AI Guidance for the Public Service

GenAI

Contents

Overview

How to read this guidance to help support responsible uptake and use of GenAI.

Context

Purpose

This guidance supports leaders, decision-makers and those in the New Zealand Public Service working or planning to work with Generative AI (GenAI). It enables agencies to explore and adopt GenAI systems in ways that are safe, transparent and responsible, and which effectively balance risks with potential benefits of these systems.

This guidance updates and expands on the 'Interim Generative AI guidance for the public service' published in July 2023 in response to feedback from agencies.

This guidance is part of the suite of **Responsible AI Guidance for the Public Service** which is being developed by the Government Chief Digital Officer (GCDO) to support the safe uptake of AI technology across government.

This suite of AI support material is being added to and enhanced over time and includes:

- a regular AI community of practice for digital practitioners in the public service

- an assurance regime for using and implementing AI in the public service

- a toolkit of practical and how-to resources on digital.govt.nz

- the Expert Advisory Panel for Public Service use of AI and the Public Service AI Assurance Regime and Framework (on its way).

This guidance has been developed alongside the Ministry of Business Innovation and Employment (MBIE) who are developing similar guidance for the business community. We're also grateful for the collective expertise from across the public sector and their collaborative approach.

Support is available to help you and your agency develop knowledge and experience with using AI technology. We encourage you to connect and seek support from others on their AI journey and with us. For further enquiries or questions, contact the GCDO team at gcdo@dia.govt.nz.

How to read this guidance

This guidance is structured into 2 sections:

- **Section 1: GenAI foundations** — key foundational aspects to underpin your customer experience and agency use. Topics include governance, security, procurement, skills, misinformation and accountability.

- **Section 2: GenAI and the customer experience** — key considerations that directly affect customer service experience with government. Topics include transparency, bias, accessibility and privacy.

You can navigate through the sections using the menu or the suggested related guidance at the end of each section. We recommend that you read the entire guidance to be aware of important factors that affect use of GenAI. Over time we're building on the guidance, best practice, case studies and toolkits related to AI for the public service which is on Digital.govt.nz. Watch this space.

There are 2 key documents that should be read together with this guidance:

- [Public Service AI Framework](#) — which should underpin all AI work in the public service and is the basis of all practical support for public service agencies developed by the GCDO. It includes the OECD principles and it's vision sets the direction for the safe and responsible use of AI to enable more efficient, responsive and high-performing public services.

- [OECD AI system lifecycle](#) — which should be referenced during ongoing development and use of GenAI systems in your agency. The OECD model grounds New Zealand AI development to international best practice. Its key themes are throughout this guidance.

New Zealand's approach to adopting AI

AI in New Zealand has been in use in various guises for some time. In June 2024, Cabinet agreed that New Zealand's approach to adopting AI will be in accordance with the OECD AI Principles. These principles promote innovative, trustworthy AI that respects human rights and democratic values. This guidance indicates the relevant principles in each section.

- [Full Paper — Approach to work on AI — MBIE (PDF 155KB)](#)

- [AI Principles — OECD.AI](#)

The GCDO is working in tandem with several key agencies to ensure there's consistency and to reduce duplication of effort and resource. In particular, the GCDO has teamed up with the Ministry of Business Innovation and Employment (MBIE) who has responsibility for AI for business and the Ministry of Statistics (Stats) who has responsibility for data.

Guidance A3 Summary

The accompanying A3 presents an overview of the OECD AI system lifecycle and key themes of this guidance. Such themes should be considered throughout this lifecycle and the ongoing development and use of GenAI systems in your agency.

- [A3 Summary: Responsible AI Guidance for the Public Service: GenAI (PDF 215KB)](#)

Introduction

What is GenAI?

The OECD has defined the following:

- [Artificial Intelligence (AI)](#) — a machine-based system that can infer from inputs and generate outputs for explicit or implicit objectives. Different types of AI systems vary in their levels of autonomy and adaptiveness.

- [Generative AI (GenAI)](#) — a type of AI system that can create or generate new content such as text, images, video and music based off models and patterns detected in existing datasets.

Further definitions of key terms related to use of GenAI can be found in our [Glossary of AI terms](#).

Open or Proprietary, Public or Enterprise AI

AI has common themes and issues, but different types of AI models and data handling methods need specific rules and precautions. Knowing these differences will help you make better decisions about using GenAI in your agency.

There are 2 key types of AI models:

- **Open type AI** — are those that have released their models for public scrutiny. These models can be studied, modified or built upon. Open models can help foster transparency, innovation and diversity.

- **Proprietary AI** — are AI models that are owned. The details of the models are kept confidential and can use licenses or other restrictions to maintain intellectual property while allowing (often commercial) use.

Open and proprietary models can be used for both public and enterprise AI systems:

- **Public AI** — these are AI systems that anyone can use, often through a website or portal, to answer questions or prompts. They gather information from various sources to create responses. Users do not control the system's setup, data, or limits, and cannot easily check where the information comes from or how it's used.

- **Enterprise AI** — these AI systems are created and used within a specific organisation to answer their particular questions or prompts. Organisations can closely manage how these systems are set up, what data they use, and their overall purpose.

- This guidance will specify when recommendations relate to particular AI models or types of GenAI systems.

Public Service use of GenAI

New Zealand government agencies have identified a wide range of AI use cases. In many cases, these agencies have found benefits to applying GenAI systems to clearly defined problems.

Agencies are at different stages of AI implementation and using AI to improve productivity, efficiency, and customer experience. This progress aligns with key innovations by AI systems in the areas of text, video, and voice. A summary of where and when AI is being used in public service can be found in the following reports:

- [2024 cross-agency AI survey: highlights](#)

- [2023 cross-agency AI survey: highlights](#)

Opportunities for the Public Service to use GenAI include:

- Boosting productivity and efficiency — reducing in administrative burden through automating search and information retrieval, assisting in creating summaries or routine analysis, and reducing the demand on people. The result of this is that work speed is increased and teams can get on with more core work and delivery. For example:

    o Detection of unusual patterns in large datasets

    o Enhanced content search

    o Automated meeting summary and action lists.

- Streamlining modernisation of legacy systems — using GenAI to assist in interrogating existing codes, patterns and architectures. The result of this is that teams can map new digital systems while maintaining existing services. For example:

    o Digital applications can be efficiently developed using automate code

    o Digital and data systems can be tested efficiently.

- Enhancing customer experience of government services — reducing call or information request wait times, triaging issues so people get what they need more quickly, while reducing costs to government. GenAI can also support public servants by completing or automating routine tasks, helping staff to focus on frontline services. For example:

    o Chatbots to help the public self-navigate through complex information to find the answers they need

    o Automated note taking of service centre calls, which saves time so that calls can managed more effectively

    o Assisting with high volume tasks such as OIA requests so that these queries can be answered more quickly

While the opportunity for use of GenAI in the public service is substantial, there are several risks to the use of GenAI that, if they were to materialise, could seriously damage public trust and be potentially cause harm to communities. This guidance will help you to manage these risks.

Things to remember — It's vital that agencies:

- Actively assess and manage the risks early and throughout the life of your GenAI systems

- Be open and transparent about how they're using GenAI

- Support the rightsizing of risk management to enable safe progress

- Provide adequate training on understanding, using and managing GenAI systems to ensure safe and effective use.

Tips for successfully using GenAI

Involve stakeholders at each stage of the AI lifecycle

The public service must ensure any use of GenAI is safe and responsible. We recommend uptake that builds on strong foundations and involves stakeholders, partners and their communities at every stage of the AI lifecycle. To ensure we follow international best practice and to build on our collective knowledge, we look to the align with OECD principles.

Key stages of the OECD AI system lifecycle to follow are:

- **Plan and design** — understand the scope and set key objectives and direction for the implementation of GenAI. Consider your model, and think about key foundations and building in privacy, security and accessibility.

- **Collect and process data** — ensure good quality data. This includes sensitivity labels and ensuring accurate metadata. Consider if there's any personal information and robustly assess potential risks and benefits.

- **Build and use a model** — identify and train your preferred model. Consider iterating the model and review the data and prompts used to improve accuracy.

- **Verify and validate the outputs** — test your model and evaluate its outputs to assess performance across a range of scenarios and prompts. Consider if it's performing as intended or if further iteration is required.

- **Deploy in a production environment** — move into a production environment. Consider how GenAI works with other systems and how staff are trained to use it. Be transparent about your use and seek ongoing feedback.

- **Operate and continue to monitor for both intended and unintended impacts** — Continually assess GenAI use, including intended and unintended impacts. Consider if you need to go back to an earlier stage and iterate to produce trustworthy and useful results.

[OECD AI system lifecycle — OECD (PDF 149KB)](#)

Implement guardrails for the safe, responsible and transparent use of AI

Ultimately, agencies are responsible for their use of GenAI systems. This includes how AI is adopted and used, and for what purpose.

There are considerations for agencies across this guidance and that apply at every stage of the AI lifecycle.

- Maintain trust and confidence — consider how to safely, responsibly and transparently trial and adopt emerging technologies like GenAI.

- Build by design — incorporate key foundations into your earliest thinking around using GenAI. You'll find information around key foundations including information management, privacy, security and accessibility in this guidance.

- Engage early — work with key stakeholders, partners, teams and end users to ensure their needs and concerns are met when using GenAI.

- Establish strong foundations — GenAI involves new ways of working, and may require further training, improved data, changes to information management practices, or different architectures or patterns to see the best outcomes.

- Select the right tool for the task — test and prototype to confirm it's the right tool for the desired outcomes before any significant investment.

- Usual rules apply — consider also existing guidance and obligations around using a new tool like GenAI. This guidance provides links to more information to get you started.

- Keep humans in the loop — involve people at all stages of designing, training, testing, evaluating and moderating GenAI systems and their outputs to drive best outcomes.

Connect with other practitioners

Finally, do not do it alone. GenAI is in use across the public service. Connect with others using GenAI in similar ways to share approaches, patterns and solutions. This will help grow our collective knowledge on how to best use these systems. For help on how to achieve this, see [Next steps for safe, responsible AI in government](#).

This guidance supports agencies and it's part of a suite of support, advice and best practice the GCDO is building over time. A mix of in person, online and research driven tools and best practice will only grow alongside this guidance. While we support agencies to be responsible for their own use of GenAI systems, we support system-led and collective sharing and growth of quality tech and leadership to grow delivery, service and efficiency.

Throughout these sections, we encourage you to always undertake rigorous evaluation of any system or process relating to AI. Agencies should feel supported to continue their AI journey. If you have questions, contact the GCDO team at gcdo@dia.govt.nz.

GenAI foundations

Foundational aspects of GenAI that support agency use and experience of GenAI.

Governance and GenAI in the public service

Clear oversight and governance of agency use of GenAI can ensure not just responsible spending and timely delivery, but can enable safe, responsible, ethical and effective use.

This guidance is underpinned by the following OECD AI principles:

- [Principle 1.1: Inclusive growth, sustainable development and well-being](#)
- [Principle 1.2: Human rights and democratic values, including fairness and privacy](#)
- [Principle 1.3: Transparency and explainability](#)
- [Principle 1.4: Robustness, security and safety](#)
- [Principle 1.5: Accountability](#)

AI Governance and assurance

Governance and assurance are closely related activities. Understanding the relationship between assurance and governance is important. As some of the existing guidance refers to governance, it's important in this context to understand the inter relationship between the 2 concepts of governance and assurance.

There's no single agreed definition of governance. But broadly it involves setting the goals, allocation of resource to achieve those goals, establishment of systems to track progress and management of risk. Assurance is an independent assessment of governance, risk management, and control processes to achieve a specified aim.

What this means is that good governance will establish assurance mechanisms to support the delivery of an outcome for the task or area of responsibility of the governance group. But it's also possible to have an assurance process that sits above this, which is an assurance system that independently reviews aspects of the work that is governed or sufficiency of governance.

Commit to good governance of GenAI

Agencies should publicly develop and share their GenAI policies and standards to guide its use of AI. Working together will help agencies lift their capability on using emerging technologies like GenAI.

Designate a responsible official to lead adoption of GenAI

In line with international best practice, we recommend public service agencies each designate a responsible senior official to guide the safe, and secure adoption of GenAI systems and other emerging technology. This person could be the Chief Digital Officer, Chief Risk Officer or Chief Data Officer.

This official would be responsible for:

- ensuring alignment and delivery within the [Public Service AI Framework](#)

- coordinating dialogue and collaboration between teams using GenAI systems

- mitigation and management of risk and assurance

- establishing your agency's guidance or rules regarding the use of GenAI to advance the appropriate use of AI

- engaging in all-of-government AI forums and processes and keeping up to date with changing requirements as they evolve

- acting as a central contact in your agency and coordinating with the Government Chief Digital Officer (GCDO) on issues such as technology procurement to ensure GenAI tools are secure, of high quality, and properly supported over time.

- involving iwi Māori.

This will provide assurance to the public service that risks are being managed and legal requirements are being met to encourage adoption of AI to modernise public services and deliver better outcomes for all New Zealanders. It will also provide assurance to the public that AI is being adopted responsibly by the public service, modernising public services and delivering better outcomes for all New Zealanders.

Conduct appropriate impact assessments for applications of GenAI

The GCDO is collaborating with agencies to develop an AI Assurance Regime. The assurance approach will cover all forms of AI, though the approach taken to assurance will depend on the type of AI and the risk associated with the use of that AI. While the regime is under development, we encourage agencies to conduct a risk assessment to help agencies identify, assess, document and manage sector-specific low versus high-risk uses of AI systems.

Agencies should subject higher-risk uses of AI to more extensive oversight. This ensures the responsible and ethical use of data and limits potential issues such privacy and security issues, or biased or factually incorrect outputs (such as 'hallucinations').

In cases where AI uses people's personal information, take all necessary steps to protect their data and privacy, including assessing how data is managed and stored, and conducting a privacy impact assessment.

- [Privacy Impact Assessments — Office of the Privacy Commissioner](#)

- [Algorithm Impact Assessment toolkit — Data.govt.nz](#)

Have human oversight of GenAI use

Always ensure accountable humans are involved in the application or use of GenAI systems and outputs. Decision-makers should have the necessary authority and skills to make informed choices.

Understanding and explaining GenAI can be challenging. Managers and leaders accountable for GenAI must articulate how and why it's used and clarify any factors that have influenced their decisions. Ensure you have people across your organisation who can explain the technology itself.

Human oversight can minimise misleading or biased results and support functions such as:

- impact and risk assessments

- transparency and reporting

- quality assurance.

Transparency and accountability

Openness, transparency and accountability are key to maintaining trust, confidence and integrity. Be transparent with the public, key stakeholders and partners about how you're using GenAI for the benefit of New Zealand. Explain how personal data related to your use of GenAI systems will be collected and managed.

We strongly recommend publishing your AI use online for wider transparency and working with the accountable official to keep a register of AI use in your agency.

What to publish about your agency's GenAI use

Agencies should publish information about their development and use of AI, barring reasonable exceptions such as classified use cases. This will help maintain transparency and trust in public service AI use. Agencies might consider publishing information about the type of AI they're using, what stage the project is at, the intent of use or the problem it's trying to solve, and an overview of how the system is being used and by whom.

Example scenario of governance and GenAI

You're the responsible official for your agency's GenAI use. You're reviewing a proposed use of GenAI based on a risk assessment.

Recognising a likely risk of a privacy breach with potential major severity, you consider this system a high-risk use of GenAI. You ask the staff proposing this AI system to conduct a privacy impact assessment, suggest mitigations, and test it internally before proceeding. These mitigations address the risks assessed, and give you confidence in the benefits and risks of using the proposed GenAI system to proceed.

Related guidance

- Security

- Procurement

- Privacy

- Accessibility

- [Misinformation, Hallucinations and Incorrect Information](#)

- [Accountability and Responsibility](#)

- [Transparency and Explainability](#)

- [Security](#)

- [Skills and Capabilities](#)

Security and GenAI

Learn about the key security risks of introducing GenAI to your organisation. Develop a list of considerations to ensure GenAI software undergoes basic security measures.

This guidance is aligned with the following OECD AI principles:

- [Principle 1.4: Robustness, security and safety](#)

- [Principle 1.5: Accountability](#)

Security plays a vital role in enabling the reliability and resilience of GenAI systems. Like all digital systems, GenAI systems can be susceptible to security vulnerabilities and misconfigurations.

Fortunately, basic security measures can help achieve an acceptable level of risk. Treat GenAI like you would other software applications. This will provide a key guardrail in developing the strong foundations of your use of GenAI. e

Security risks of GenAI

The security risk posed by GenAI applications depends on several factors, such as:

- the information the tool has access to

- permitted users

- when and how it was developed, whether in house or procured from a third party

- external sharing of data.

A security risk assessment should consider these factors to determine if the application is the right fit for your organisation.

[Resources for Security Leads](#)

Notable risks

Vendors introducing GenAI into existing and new applications

Software vendors are rapidly introducing GenAI into existing and new applications and are constantly evolving the capabilities available.

New software versions should be evaluated and tested before they're rolled out to ensure sufficient guardrails are in place. Contracts should require vendors to provide advance notification of material changes.

Unsanctioned GenAI being accessed in your agency

Be aware of the risk of unsanctioned GenAI applications accessed by users in your organisation. This could include web applications or public GenAI systems outside your enterprise systems, as unapproved use could lead to data breaches.

Some GenAI applications retain user input to train large language models (LLMs) without the choice to opt out. Often there's no way to have data deleted. Prevent the use of unapproved systems by continuously monitoring and blocking as needed.

The output of a GenAI systems can also be a security risk

Assess the outputs of GenAI, especially if it's used for code generation. Do not trust generated code until it's verified to be free of errors through quality control processes.

Other security considerations

- Consider an early discussion with your security team to help establish your agency's level of preparedness.

- Ensure your agency's information management practices support data loss prevention measures, including sensitivity labelling and access management.

- Ensure all GenAI systems used by your agency are certified and accredited before they're made available to users, as advised within the New Zealand Information Security Manual (NZISM) chapter 4 System Certification and Accreditation. The certification process should validate that security controls, like application monitoring, are in place to identify misuse and support investigations.

- Ensure sensitive or classified information is not entered in public GenAI tools. Even things that cannot identify a person could be aggregated over time to re-identify a person. For more information, see Privacy and GenAI.

- Ensure your staff receive adequate training and guidelines on the acceptable use of GenAI systems. Don't solely focus your security strategy on technology. This will help your staff make the right choice or know when or how to seek assistance when they spot a potential vulnerability.

- Review NZISM chapter 14 covering 'Software Security' and the Guidelines for Secure AI System Development if your agency is developing GenAI systems or using GenAI for application development.

- Consider ways in which GenAI applications can be used to improve the security posture of your agency.

Example scenario of security and GenAI

An agency developed a public-facing AI tool to simplify the process of accessing services for New Zealand people. During the testing phase a misconfiguration was identified that allowed the GenAI chatbot to inadvertently retrieve and share privileged information.

As sensitivity labels were used to limit chatbot responses, the agency's information management practices supported a quick resolution. The issue was fixed before the tool was made public. This ensured information was managed appropriately, maintaining trust and saving the agency from significant reputational damage.

Resources for Security Leads

- System Certification and Accreditation — NZISM Chapter 4 — Government Communications Security Bureau (GCSB)

- Software Security — NZISM Chapter 14 — GCSB

- Joint Guidance: Guidelines for Secure AI System Development — National Cyber Security Centre (NCSC)

- Joint Guidance: Deploying AI Systems Securely — NCSC

- ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems) — Mitre

- AI Exchange — OWASP (Open Worldwide Application Security Project)

- Generative AI Framework — UK Government

Related guidance

- Governance

- Procurement

- Accountability and Responsibility

- Skills and Capability

Procurement and GenAI

Like any part of business, procurement of services and systems, planning, following best practice and working to mitigate any risk is critical. GenAI procurement is no different.

This guidance is aligned with the following OECD AI principles:

- [Principle 1.3: Transparency and explainability](#)

- [Principle 1.5: Accountability](#)

The first consideration is whether AI is the best solution for your agency. Do this by:

- Planning appropriately — assess business need, analyse the market and factor in risk mitigations.

- Developing robust evaluation criteria — be explicit about pre-conditions and include an exit strategy.

- Doing your research — know the market, suppliers in it and their offerings.

- Using guidance — to help you get the best result, including supporting market resilience.

AI systems present opportunities for government to modernise and streamline operations and improve the delivery of public services. You should make sure you've planned appropriately for the size, risk and complexity of your procurement. AI system and service procurement may require a higher degree of risk planning to identify and mitigate the risks it may pose.

Plan for your business needs

AI products differ widely, and AI may not always be the best solution. While something is free at the outset, there may be cost incurred later on once the solution is scaled. Think carefully about your evaluation criteria and if commercial protections might be needed in your contracts. Consider other key foundations, such as privacy and security, when developing evaluation criteria.

Procurement teams should conduct thorough market research on suppliers and their offerings. Consider supplier reputation, capability, pricing, contract value and potential risks in their supply chain.

You'll get a better and more assured outcome if you consider every phase of the procurement cycle in your planning and manage the contract throughout its life. You'll need to consider how the Government Procurement Rules apply and keep the Government Procurement Principles front of mind during your planning.

[Rules: Government Procurement — NZ Government Procurement](#)

[Principles: Government Procurement — NZ Government Procurement](#)

Follow procurement best practice

The Government Procurement Framework is flexible and can accommodate the fast-paced innovative nature of AI. This flexibility depends on following procurement best practice and incorporating adequate risk assessment and mitigation in your procurement plan. There are several domestic and international best practice guides for procuring AI. You should familiarise yourself with these before approaching the market.

[Principles, charter and rules — NZ Government Procurement](#)

Early in your planning, conduct a robust needs analysis and consider your specific business need and the role AI is needed to perform. Think about how the GenAI system will meet the needs of your agency, and how it will integrate into existing systems.

The procurement of AI may provide opportunities to use innovative procurement approaches and practices that can help you get the best outcomes from your AI procurement.

Leverage interactive procurement processes that can help you get the best outcomes from your AI procurement, for example:

- Lean Agile Procurement

- Procurement-In-A-Day

- Competitive Dialogue

This helps support a diverse, competitive, and resilient AI supplier market.

Resources for procurement leads

Learn more on the NZ Government Procurement website:

- [Rules — Government Procurement](#)

- [Principles — Government Procurement](#)

- [Charter — Government Procurement](#)

Procurement good practice guides

- [AI Procurement Guides — AI Forum New Zealand](#)

- [AI Procurement Guide — Tony Blair Institute for Global Change (PDF 109KB)](#)

- [Workbook — Unlocking Public Sector AI: AI Procurement in a Box — World Economic Forum (PDF 7.1MB)](#)

- [The AI-RFX Procurement Framework — The Institute for Ethical AI & Machine Learning](#)

- [Guidelines for AI Procurement — UK Government (PDF 11.1MB)](#)

- [Whitepaper — AI in Government Procurement — Australasian Procurement and Construction Council (PDF 14.1MB)](#)

- [Buying AI, is the public sector equipped to procure technology in the public interest? — Ada Lovelace Institute (PDF 526KB)](#)

Related guidance

- [Security](#)

- [Governance](#)

- [Privacy](#)

Skills, capabilities and GenAI

Improve your staff's skills when using GenAI. This helps to maximise the upsides of AI's benefits to your agency while limiting risks.

This guidance is aligned with the following OECD AI principles:

- [Principle 1.1: Inclusive growth, sustainable development and well-being](#)

- [Principle 1.2: Human rights and democratic values, including fairness and privacy](#)

- [Principle 1.3: Transparency and explainability](#)

- [Principle 1.4: Robustness, security and safety](#)

- [Principle 1.5: Accountability](#)

Improve the skills and capabilities in your agency

To embrace the opportunity GenAI offers in a safe and responsible way, it's important to build the skills and capabilities across your agency. This way, you get the most out of GenAI.

Understanding the strengths, weaknesses and risks of GenAI is an excellent starting point.

Training to educate employees in GenAI

We recommend that you look at providing training opportunities for staff so they understand:

- that GenAI systems are sophisticated and predictive

- how best to use GenAI to realise efficiencies and deliver better services, while managing risks.

Make sure your teams know that when and how GenAI can support human decision-making — but also when it would not be appropriate to use GenAI tools.

Commit to improving your skills in GenAI

Training gives individuals the knowledge and confidence to use GenAI systems safely.

A well-informed workforce can help to achieve the benefits of AI and be the best defence against AI pitfalls and risks.

Give initial and ongoing upskilling

Upskilling your workforce is crucial for using GenAI responsibly. Set clear guardrails and create opportunities for review and feedback loops.

Create a safe environment for learning

We recommend that you encourage transparent use of GenAI systems and create a safe environment for learning. This helps to make sure the use of GenAI leads to more productivity and less harm.

Make sure training aligns with best practices

We recommend that you use or consider creating training that follows best practice across all areas of this guidance. This will help equip your people to navigate GenAI's complexities.

Example scenario of skills, capabilities and GenAI

You've heard from colleagues how great GenAI is, so you want to try it out. However, you're not sure which tool to use or how to use them.

You check your agency's relevant policies before using a new tool. You ask your manager or training department if there's any assistance or training they can provide.

Your agency is connected with several industry training opportunities for staff, and also offers its own training in this new tool. You book into a training session. In the meantime, you can carry out your own research and reading to prepare. You wait to try out GenAI until you feel prepared and confident to safely use GenAI in your work.

More information — GenAI skills and capabilities

- [AI Fundamentals for Public Servants: Opportunities, Risks and Strategies — Apolitical](#)

- [AI at work: A practical guide to implementing and scaling new tools — World Economic Forum](#)

Related guidance

- [Accessibility](#)

- [Security](#)

- [Governance](#)

Misinformation, hallucinations and GenAI

In order to use GenAI responsibly, make sure your agency can access high-quality information. This avoids spreading misinformation and sharing hallucinations.

This guidance is aligned with the following OECD AI principles:

- [Principle 1.3: Transparency and explainability](#)

- [Principle 1.4: Robustness, security and safety](#)

Access high-quality information

While there are many benefits to using GenAI, it can generate inaccurate and incomplete outputs.

Having access to high-quality information is vital to support effective decision-making. It's important that the government avoids contributing to misinformation by sharing inaccurate information.

Limits of GenAI

GenAI systems may not always comprehend:

- real-world contexts

- nuances in language

- cultural references

- intent.

GenAI systems may not have access to information that's known to be true or reliable. Educate your teams with the skills and capabilities to use GenAI effectively, and to never consider generated content as authoritative.

Misinformation, disinformation and hallucinations

The New Zealand government has identified disinformation as a core issue and a National Security Intelligence Priority. The Department of Prime Minister and Cabinet (DPMC) defines:

- **Disinformation** as false or modified information knowingly and deliberately shared to cause harm or achieve a broader aim

- **Misinformation** as information that's false or misleading, though not created or shared with the direct intention of causing harm.

[Strengthening resilience to disinformation — DPMC](#)

Related to but distinct from misinformation and disinformation are hallucinations. The OECD defines hallucinations as when GenAI systems create incorrect yet convincing outputs.

[Generative AI: Risks and unknowns — OECD.AI](#)

Commit to avoiding hallucinations and incorrect information

- Be specific and prescriptive in your query

- When using GenAI, you can include instructions to return an 'I do not know' when the model is unsure.

- Data quality and representative data is a key part of avoiding hallucinations, alongside reliable testing process.

Best practice to help avoid misinformation and hallucinations

We recommend that you consider these best practices to avoid misinformation and hallucinations when using GenAI.

Examine the impact

For each use case, assess the impact of using AI-generated content and the risks of misinformation.

Verify and cross-reference information

Check the quality of results produced by GenAI systems with trusted sources. This helps make sure the content generated is accurate.

Check reliability and truthfulness

Double-check the reliability and truthfulness of AI results before using that generated information. This lessens the risk of sharing misleading information.

Example scenario of misinformation, hallucination and GenAI

For some policy work, you're using GenAI to learn about a topic you're unfamiliar with. It produces information that looks robust.

Evaluate the references and citations provided in the system and check if the sources provided are legitimate and appropriate. Cross-check the information with credible sources and experts or relevant communities.

Make sure your manager knows that you've used the tool. You should not publish the information to the public until it's been absolutely verified and approved.

Be clear that GenAI was used to produce it, and that people can challenge those outputs. This will help maintain transparency, trust, and robust outcomes.

More information — GenAI misinformation and hallucinations

- Misinformation and disinformation — OECD

- Strengthening resilience to disinformation — DPMC

Related guidance

- Governance

- Skills and capabilities

- [Bias, discrimination and fairness](#)

- [Transparency and explainability](#)

Accountability, responsibility and GenAI

Humans must be involved when using GenAI. We recommend getting different viewpoints when evaluating GenAI outputs.

This guidance is aligned with the following OECD AI principle:

- [Principle 1.5: Accountability](#)

Human oversight is essential for ethical GenAI use

While GenAI can increase productivity, it can also produce misleading, harmful or biased results. 'Human-in-the-loop' is an approach where human oversight is integrated across GenAI use. It ensures that humans remain an essential part of decision-making, working alongside GenAI.

Make sure you understand the data you provided to the GenAI systems and ensure you understand, check and agree with the outputs.

Combine the benefits of GenAI and human intelligence

GenAI lacks human understanding, empathy and compassion. It cannot understand context or know when something is wrong or harmful. But when humans and GenAI combine forces, and bring their unique skills together, then value can be gained while ensuring human values remain a priority.

When using GenAI tools for decision-making, make sure a human is involved. If your initial risk assessment shows that the situation is high-risk or high impact, ensure your use follows your agency's risk and assurance framework and all controls are applied. If you're using GenAI for a significant piece of work, an AI impact assessment will help you to make responsible choices.

[Algorithm Impact Assessment Toolkit — Data.govt.nz](#)

Commit to using different perspectives with GenAI

You can ask GenAI systems to consider content from different perspectives. You should also ask humans for different views and then check with people who have knowledge of this viewpoint.

Encourage a culture of responsibility

Equip your teams to understand potential consequences. Encourage people to check that outputs are:

- truthful

- non-harmful

- factual

- lawful

- non-discriminatory.

Consider either:

- not using GenAI or using a different system

- increasing human oversight to an appropriate level.

Build evaluation and auditing processes

To oversee AI use and outputs, create processes and controls that help to build accountability and responsibility in your organisation.

Example scenario of accountability, responsibility and GenAI

You use GenAI to summarise a long document, containing only publicly available information, into a quick read for the users of the government service. You'll use this summary on the website to link to the longer document.

You check your agency's GenAI policy and let your manager know you're using the tool. You should also ask a colleague to review the summary, as a quality check.

You should not publish it until you've double-checked that all the content is accurate, culturally appropriate and no key context is missing. Once you've completed those checks, you're able to confidently publish a useful summary for your longer document for the public to use.

More information on assessing algorithms

For a framework to understand high-risk algorithms, check the following charter for algorithms in New Zealand.

Algorithm charter for Aotearoa New Zealand — Data.govt.nz

For ways to help organisations understand and assess the potential impacts of the algorithms they create or use, check the following toolkit.

Algorithm Impact Assessment toolkit — Data.govt.nz

Related guidance

- Governance

- Security

- Accessibility

GenAI and customer experience with government

Key considerations about GenAI systems that affect customer service experience with government.

Transparency and GenAI

Be open with your people and the public about why and how you're using GenAI. Public service agencies need to be able to explain what information goes into AI systems and what the results are used for.

This guidance is aligned with the following OECD AI principle:

- Principle 1.3: Transparency and explainability

Use GenAI ethically in the public service

It's vital that public service agencies use GenAI ethically. The public service is held to the highest standards of trust and confidence, so consider how you and your team can build trust at all stages.

Be open about why and how you're using GenAI

Be transparent to the public about when and how you're using AI. Explain why and how it's being used.

A lack of transparency can lead to harmful outcomes, public distrust, and no-one being responsible for the final decision. Transparency involves communicating clearly that you're using AI and why you're using it.

Commit to ethical and transparent use of GenAI

When you use GenAI, you need to be able to explain and take responsibility for your decisions. Even though the technology helps you, you're still in charge of the final decision. This means you should know how GenAI works and what information it uses to make suggestions.

Create clear policies

Outline when GenAI can be used and when it cannot. Have clear policies around the use of emerging technology. These policies should also include how the public can challenge any outcomes that involve the use of GenAI systems.

Have processes in place

Clear processes can help you respond to requests about how and why you're using GenAI. Be sure you can access or correct information if requested to do so.

Publish your agency's GenAI use

We strongly recommend publishing an up-to-date register of all GenAI use in your agency. This is a commitment to transparency – and helps to connect with other government agencies on how GenAI is being used. For guidance on what to publish, see the section on Governance.

Example scenario of transparency and GenAI

You're thinking about using a GenAI system to help produce a report for your manager. The report will help inform important decision-making that will impact a community.

Before using the GenAI system, you check your agency's policies and record your use of GenAI in the public register. You also note what data you put into the system and what you used the outputs for. You're using an enterprise GenAI system, and you know your agency has good information management to handle sensitive data. You're confident the data was used correctly to generate the output.

As you get used to how GenAI works, you can explain how the outputs were made. You talk with your manager about using AI in the report, clearly marking the parts generated by AI and confirming you checked the content for accuracy according to your agency's rules.

More information — GenAI transparency and explainability

- AI governance introduction — AI Forum New Zealand

- AI Fundamentals for Public Servants: Opportunities, Risks and Strategies — Apolitical

- GenAI guardrails made simple — Simply Privacy

Related guidance

- Governance

- Accessibility

Bias, discrimination, fairness, equity and GenAI

Proactively make sure your agency's use of GenAI creates fairness and equity instead of biases and discrimination.

This guidance is aligned with the following OECD AI principles:

- Principle 1.1: Inclusive growth, sustainable development and well-being

- Principle 1.2: Respect for the rule of law, human rights and democratic values, including fairness and privacy

Biases in GenAI affects communities

Bias can disproportionally impact some community groups, such as:

- Māori

- Pacific Peoples

- disabled people

- LGBTIQ+ communities

- multicultural communities.

To make sure that use of GenAI is fair, you need processes to ensure its outputs are unbiased and do not worsen social, demographic, or cultural inequalities.

Be aware of your own biases. This might mean getting training or learning to spot unconscious bias, so you do not accidentally include bias throughout the AI lifecycle.

How biases and discrimination can happen in GenAI

In GenAI, harmful biases can be present in:

- text

- images

- audio

- video.

Bias perpetuates stereotypical or unfair treatment related to race, sex and gender, ethnicity, or other protected characteristics.

By identifying and mitigating bias and reducing harm at all stages, you'll help GenAI to produce fairer outcomes.

Public GenAI systems are often trained on data that can reinforce existing biases, discrimination, and inequalities. This means there's a risk that using AI could repeat these biases, leading to unfair, harmful, and unequal results.

Tips for best-practice fairness and equity for GenAI

- Write prompts to help reduce the potential for bias.

- You can configure some systems to flag inappropriate prompts.

Commit to fairness, equity and GenAI

Follow best practice approaches to create fairness and equity in your agency's use of GenAI systems.

Involve Māori and other community partners and groups

Involve your partners from these groups at the appropriate level to manage potential impacts. Build your team's capability to engage with iwi Māori.

Have diverse teams

Especially in the governance, deployment and use of GenAI, make sure these teams have diverse representation.

Think critically and validate all GenAI outputs

Check all GenAI results to reduce the potential for discrimination and keep reviewing over time to make sure biases are not developing. Refer to the AI Lifecycle and ensure rigorous evaluation at all stages of AI development and use.

A3 Summary: Responsible AI Guidance for the Public Service: GenAI (PDF 215KB)

Example scenario of fairness, equity and GenAI

You've been using GenAI to source information that will be used to make decisions about who to prioritise for support.

You should ask:

- an expert in the field and/or

- members of the community that the information is about.

This helps you test that it's correct, factual and unbiased.

You should also cross-check with official and authoritative sources about the potential for harm if the information is at risk of being incorrect or biased.

After completing your checks, you conclude that the sources are authoritative, accurate and balanced. Confident that your sources are not unintentionally biasing the output, you proceed to use the GenAI provided sources in your work and test the outputs.

More information — GenAI privacy related to fairness and equity

This guidance is a good option to help ensure privacy is one of the key priorities for your product, service, system or process.

The Privacy, Human Rights and Ethics Framework — Data.govt.nz

Related guidance

- Governance

- Accessibility

Accessibility and GenAI

Accessibility means designing things to work for disabled people. The New Zealand Government has legal and ethical obligations to create accessible information and services, both for the public and for public servants.

This guidance is aligned with the following OECD AI principles:

- [Principle 1.1: Inclusive growth, sustainable development and well-being](#)

- [Principle 1.2: Human rights and democratic values, including fairness and privacy](#)

Prevent harm to disabled people

Disabled people can be left out of the design of products and services, and the data used to train GenAI. Without disabled people's input, GenAI might ignore their needs and suggest irrelevant or incorrect solutions. Privacy issues can also arise, for example, if a person's disability status is revealed without their consent.

Involving disabled people in the creation of GenAI can help to prevent discrimination and harm.

Enable benefits for disabled people

GenAI can help disabled people participate in society on an equal basis with others.

For example, GenAI can improve communication for people with speech or hearing impairments and make content easier to read for those with learning or reading disabilities. This can increase their independence and reduce their use of other service channels.

To realise these benefits, GenAI systems must be accessible and usable by disabled people.

Accessibility in AI requires a 2-pronged approach

1. Include disabled people in all stages of the AI lifecycle

- Engage disabled people as stakeholders from conception of the GenAI solution to its deployment.

- Develop a vision statement and policy that prioritise inclusion and accessibility.

[A Blueprint for Equity and Inclusion in Artificial Intelligence — World Economic Forum](#)

[AI & Disability Inclusion Toolkit Overview — The Partnership on Employment & Accessible Technology (PEAT)](#)

2. Ensure equitable treatment of disabled people by GenAI

- Consider disabled people's experiences in training data, models and outputs to remove bias.

- Build GenAI to mandated accessibility standards and test it with disabled people.

- Throughout all stages of development and use, continuously monitor that GenAI is delivering safe, trustworthy and non-discriminatory interactions.

NZ Government Web Standards

Accessible and Equitable Artificial Intelligence Systems: Technical Guide — Government of Canada

Example scenario of Accessibility and AI

Your mother, who has a vision impairment and uses a screen reader, independently resolved her issues with a government service through a GenAI chatbot.

The chatbot was easy to use and accessible for her because it was tested with screen reader users as part of its development. The chatbot's answers provided options for disabled people by default, as it was co-designed with them from the start.

Your mother did not have to disclose her impairment. She was happy she was able to resolve her issue for herself in her spare time one evening without having to call or visit a government office during business hours.

Related guidance

- Governance

- Transparency and Explainability

- Accountability and Responsibility

- Privacy

- Bias, Discrimination and Fairness

- Skills and Capability

Privacy and GenAI

Make sure your privacy approach to GenAI meets data-protection rules and legislation, respecting people's information.

This guidance is aligned with the following OECD AI principle:

- Principle 1.2: Respect for the rule of law, human rights and democratic values, including fairness and privacy

Privacy by design is vital when using GenAI

The Privacy Act 2020 applies to GenAI as it does to any other technology. However, there are additional issues for GenAI that you need to consider.

GenAI can process personal data at all stages and can generate outputs that contain personal data, including sensitive personal data. Using privacy impact assessments (PIAs) for any testing or use of GenAI. This helps you identify and manage privacy risks. Undertake robust risk assessment at all stages when considering using personal information in public GenAI systems.

What to cover in your privacy approach to GenAI

Building by design is a key guardrail. Apply privacy-by-design principles to help build trust in GenAI systems — make sure they respect:

- compliance with data-protection rules and legislation

- transparency about why and how they're being used

- people's privacy

- limiting the risks of privacy breaches.

Actively govern and manage for the identified risks and seek support from your privacy and legal teams at all stages.

Commit to best-practice for privacy with GenAI

Some AI systems allow you to apply sensitivity labels based on the data included in outputs. These can control what content is pasted.

Classify information

Other points to follow for privacy and GenAI

Make sure the people in your organisation are trained to:

- classify information properly

- know what can and cannot be used with GenAI systems.

Check the information you're using can be made public or if it already is

When submitting government information into a public GenAI system, the information must either already be public or it would be acceptable to be made public.

Do risk assessments at all stages when using personal information

Undertake a robust risk assessment at all stages when considering using personal information in public GenAI systems.

The privacy impacts of using these systems may not be obvious, including whether information is used for training models, unintended sharing of information, or enabling a person to be re-identified when data points are combined. Refer to the Privacy Commissioner's guidance on using personal information in GenAI systems.

Example scenario of privacy and GenAI

You're creating a report and think using a GenAI system could save some time.

Before using the system, you first check if the use aligns with your agency's GenAI policy and check with your agency's responsible official as you're not sure.

You learn you must not upload or input any personal information into the public GenAI system. However, you can enter publicly available information or ask for a report template. You're able to use GenAI to suggest a framework for your report, which saves you some time.

You also record your use in the appropriate publicly published register to maintain transparency.

More information — GenAI privacy

Data.govt.nz has guidance for anyone who works directly with service users or communities.

Data Protection and Use Policy — Data.govt.nz

The Office of the Privacy Commissioner explains how privacy is a starting point for responsibly using AI systems.

Artificial intelligence that the Information Privacy Principles — Privacy Commissioner

Related guidance

- Governance

- Accessibility

Māori, Pacific Peoples, ethnic communities and GenAI

In general, it's important to protect an individual's data and to ensure that it's not used to create bias or harm.

This guidance is aligned with the following OECD AI principles:

- Principle 1.1: Inclusive growth, sustainable development and well-being

- Principle 1.2: Human rights and democratic values, including fairness and privacy

Māori and GenAI

GenAI in public services presents both opportunities and challenges, particularly when involving Māori data or when it impacts Māori communities.

Māori representatives hold diverse views on Government use of GenAI systems. In particular, there are concerns among Māori, Pacific peoples and other ethnic community groups about possible discrimination resulting from the use of GenAI.

Commit to understanding Māori considerations for GenAI

We recommend understanding important contexts for Māori and the Crown. Agencies' key considerations could include:

- the purpose of using GenAI

- potential impacts on Māori

- the nature and status of Māori data involved (tapu, sensitivity and risk and noa, free from tapu)

- Māori data governance applications.

Where Māori data is involved, we recommend aligning with existing Māori-Crown relationship approaches. Effective engagement should be simple, equitable, safe, and value-adding for Māori participants.

> Guidelines for engagement with Māori — Te Arawhiti (PDF 88KB)

> Māori interests in public cloud

Managing datasets for Māori

There are many large language models that support GenAI. Given the nature and expense of these models, there is a variation – some can be narrow and/or based on other contexts not from New Zealand which do not encompass the world view or experience of Indigenous groups, including Māori. Involving iwi Māori in the management and development of GenAI helps to identify potential bias and or possible discriminatory outputs.

Explore opportunities for:

- regular engagement and, where sensible, shared decision-making with iwi Māori

- fostering Māori-led approaches to enhance GenAI inclusivity and effectiveness in public services

- building your team's capability to engage confidently and respectfully with Māori.

Example scenario of Māori datasets and GenAI

You're preparing a presentation about Māori populations and practice and decide to use a public GenAI system.

You should confirm whether the data or information you need to input is considered Māori data. If it is, you should check the terms and conditions of the GenAI system, particularly privacy, data and training data sections. They and your accountable official can advise if there are any concerns, such as if data inputted into the system is shared with third parties. If you're still unsure, you should consider not using the system.

You can talk to your manager or IT team about alternative technologies, including other types of GenAI systems, that could support your work.

Indigenous data considerations

If indigenous data is entered into public GenAI systems, there can be little control of where it ends up. This creates risk of inadvertently exposing or commodifying data without consent. Be aware that data can be considered sensitive by different people, cultures, and communities. To ensure to good data practices, we suggest the following:

- Encourage your teams to ask 'how should this data be treated?' before entering it into GenAI systems.

- Empower your teams to consider whether the information they're inputting may put data sovereignty at risk.

- Ensure your teams to be aware when producing content about Māori, Pacific Peoples or other ethnic communities of the perception of bias or perpetuating stereotypes.

Supporting Use of Indigenous Data and AI

The iterative learning and training capability of GenAI makes it difficult for anyone to understand how their data is being used. Extra effort should be made to ensure all involved understand the guardrails, checks and balances in place so use is safe and responsible.

Enterprise GenAI may enable closer control of data sources but requires organisations to effectively set up and manage the data the GenAI system can access. Data needs to be managed to enable the use of AI systems. Therefore, human oversight is essential for ethical, culturally aligned decisions.

Jurisdictional risks happen when data is controlled by the laws of the country where service providers store, process, or send the data. That's why it's important for agencies to talk to providers to know where their data is handled and understand the laws of that country.

Recommended reading

- [How do we protect Māori data in the era of generative AI? — AI Forum New Zealand](#) — an interview with Megan Tapsell, Chair AI Forum.

- [Māori Data Governance Model — Te Kāhui Raraunga | Data Iwi Leaders Group (PDF 3.2MB)](#) — designed by Māori data experts for use across the public service.

- [Ngā Tikanga Paihere — data.govt.nz](#) — this guidance is a good option if you want Te Ao Māori principles to inform your data practice. It's also a good framework for thinking about working with communities and ensuring your data practices occur in good faith.

Related guidance

- [Governance](#)

- [Accountability and responsibility](#)

- [Skills and capability](#)

Next steps for safe, responsible AI in government

AI systems are evolving rapidly. Government policies, guidance and use cases will evolve with these advancements and with public expectations.

Connected AI guidance

The Government Chief Digital Officer (GCDO) leads the safe and trustworthy adoption of AI in the public service. This guidance is part of a broader suite of Responsible AI Guidance for the Public Service. All of our supporting material will be housed on digital.govt.nz and will grow over time, so we recommend checking in regularly.

This guidance connects with the Responsible AI Guidance for Business that's being prepared by the Ministry of Business Innovation and Employment (MBIE). Similarly, other agencies are also bolstering their own guidance and support documentation tailored to their specific needs.

Public service AI initiatives

Support is available to help you and your agency develop knowledge and experience with using AI technology. We encourage you to connect and seek support from others on their AI journey as well as the GCDO. We're growing and enhancing the suite of support for the public service on their AI journey rapidly.

Other initiatives that are underway include:

- an AI community of practice for digital practitioners in the public service

- developing a toolkit of additional resources

- the Expert Advisory Panel for Public Service use of AI

- the Public Service AI Assurance Regime.

For further enquiries or questions, contact the GCDO team at gcdo@dia.govt.nz.

Acknowledgements

Government system leads

Glossary of AI terms

The digital version of the Glossary is the most up to date. [You can find this on digital.govt.nz](#).

This glossary intends to provide some direction in understanding terms throughout the guidance. An updated glossary with additional terms will be provided when the Responsible AI Guidance for Business is published.

Disclaimer: The following is provided as an example of terms used in the AI Strategy and Responsible AI guidance products. The explanations are provided merely as a guide and links to additional material are provided for general interest and should not be seen as an endorsement of the source or a product/service over any other.

Note: There's no universally agreed definition for artificial intelligence. Across the AI Strategy and Responsible AI guidance products we're using the OECD definition of an AI system as set out below.

Accessibility (Technology)

Accessibility is considering the needs of all potential users from the outset, engaging with individuals who have disabilities during the design process in order to create solutions that are genuinely usable by everyone.

It also includes assistive technologies, screen readers, voice recognition software and alternative input devices.

AI Bias or Machine Learning Bias or Algorithm Bias

Bias in AI models typically arises from two sources: the design of models themselves and the training data they use.

Models can sometimes reflect the assumptions of the developers coding them, which causes them to favour certain outcomes.

Additionally, AI bias can develop due to the data used to train the AI.

AI Governance

Governance involves steering responsible development, deployment, and use of AI technologies throughout their lifecycle, by creating and implementing a range of tools such as voluntary guidelines, policies, regulations, and laws, amongst others.

AI Life Cycle

An AI system lifecycle typically involves several phases that include:

- to plan and design

- collect and process data

- build model(s) and/or adapt existing model(s) to specific tasks

- test, evaluate, verify, and validate

- make available for use/deploy

- operate and monitor

- retire/decommission.

These phases often take place in an iterative manner and are not necessarily sequential. The decision to retire an AI system from operation may occur at any point during the operation and monitoring phase.

[Recommendation of the Council on Artificial Intelligence - OECD](#)

AI system

An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.

[What is AI? Can you make a clear distinction between AI and non-AI systems? — OECD](#)

Chatbot

Chatbot is a digital tool or software application designed to simulate conversation with users (primarily via text or synthesised speech). Some operate on predefined responses but advanced versions integrating GenAI provide more dynamic and responsive interactions with users.

Cybersecurity

Cybersecurity involves measures to protect systems, data, and devices from unauthorised access, and ensuring the confidentiality, integrity, and availability of information.

Data

Data can be defined as any information in a form capable of being communicated, analysed, or processed (whether by an individual or by computer or other automated means).

Data is useful when it can be communicated easily and analysed to gain insights. Data's value stems from its use, re-use, and re-purposing, particularly in large volumes. To properly realise this value, data must be accurate, reliable, and free from bias.

Data Ethics

Data ethics refers to the study and practice of ethical issues related to data, including its generation, recording, processing and distribution, and use. It encompasses principles and standards that guide the responsible and fair handling of data to ensure the rights and privacy of individuals are protected.

Key principles of data ethics:

- Privacy: Ensuring that personal information is collected, stored, and used in ways that protect individuals' privacy and comply with legal requirements.

- Transparency: Being open about how the data is collected, used, and shared.

- Consent: Obtaining informed consent from individuals before collecting their data.

- Security: Protecting data from unauthorised access, breaches, or cyberattacks to maintain its confidentiality, integrity, and availability.

- Fairness: Ensuring data practices do not result in discrimination or bias and that data will be used in ways that are both equitable and just.

- Accountability: Holding organisations and individuals accountable for their data practices and ensuring there are mechanism in place to address any issues if they arise.

[OECD Good Practice Principles for Data Ethics in the Public Sector — OECD](#)

Data Transparency

Data transparency is providing clear and accessible information about the data used in AI systems. This includes understanding where the data comes from, how it has been collected, processed, and used, and making processes more open and understandable to stakeholders.

Deep Learning

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.

Explainability

Explainability means enabling people affected by the outcome of an AI system to understand how it was arrived at. This entails providing easy-to-understand information to people affected by an AI system's outcome that can enable those adversely affected to challenge the outcome, notably – to the extent practicable – the factors and logic that led to an outcome. Notwithstanding, explainability can be achieved in different ways depending on the context (such as, the significance of the outcomes).

[Transparency and explainability — OECD.AI](#)

Generative AI (GenAI)

Generative AI is a type of AI system that can create or generate new content such as text, images, video and music based off models and patterns detected in existing datasets. (OECD)

Hallucination

The OECD defines hallucinations as when GenAI systems create incorrect yet convincing outputs.

[Generative AI: the risks and the unknowns — OECD.AI](#)

Human in the loop

Human in the loop refers to the involvement of human oversight and decision-making in the processes that involve AI and automated systems. This approach allows for critical decisions, especially those impacting individuals, to be reviewed, verified, and influenced by human judgement and expertise.

Indigenous Data

Indigenous data refers to data that's related to Indigenous peoples, their territories, resources, cultures, languages, and knowledge systems. Data is a critical part of indigenous identity and sovereignty, encompassing a wide range of information from traditional ecological knowledge to health and demographic data.

See also Māori Data Governance.

Indigenous data sovereignty thematic area narrative in English, Arabic, French, Portuguese and Spanish — Global Index on Responsible AI

Large Language Models (LLMs)

Large language models, also known as LLMs, are very large deep learning models that are pre-trained on vast amounts of data.

AI language models: Technological, socio-economic and policy considerations — OECD

Machine Learning

Machine learning (ML) is a type of artificial intelligence that allows machines to learn from data without being explicitly programmed. It does this by optimising model parameters (that is, internal variables) through calculations, such that the model's behaviour reflects the data or experience. The learning algorithm then continuously updates the parameter values as learning progresses, enabling the ML model to learn and make predictions or decisions.

Māori Data Governance

Māori data governance refers to the principles and practices that ensure Māori data is collected, managed, and used in a way that respects Māori values, rights, and interests.

Key aspects of Māori data governance include:

- Data Sovereignty – ensuring Māori data is subject to Māori governance and control.

- Ethical use, promoting the ethical use of data to enhance wellbeing of Māori people, language, and culture.

- Advocating for Māori involvement in the governance of data repositories and decision-making processes.

- Safeguarding the quality and integrity of Māori data.

Te Mana Raraunga has developed resources to better understand and to support the principles.

[Te Mana Raraunga](#)

Misinformation

Misinformation refers to false or inaccurate information that's spread regardless of an intent to deceive. Unlike disinformation, which is deliberately misleading, misinformation is often shared without malicious intent.

Predictive AI

Predictive AI (or predictive analytics) Involves using data, statistical algorithms, and machine learning techniques to identify the likelihood of future outcomes based on historical data.

Risk

Risk is the likelihood of adverse impact on people, the environment and society.

Stakeholders

As defined in the OECD Recommendation on AI: Persons or groups, or their legitimate representatives, who have rights or interests that are or could be affected by adverse impacts associated with the enterprise's operations, products, or services.

These can include users of the AI system, civil society, workers' representatives, service providers, and other enterprises. Anyone involved in or affected by relevant systems.

[Recommendation of the Council on Artificial Intelligence — OECD](#)

Transparency

Making the operation and decision-making processes of AI systems clear and understandable to users and stakeholders. Key components of transparency are:

- Openness: Clearly communicating the purpose and capabilities of an AI system. This includes explaining what the system is designed to do and any limitations it may have.

- Explainability: Providing understandable explanations of how the AI system reaches it decisions.

- Accountability: Ensuring that there's a mechanism for tracking and verifying decisions made by the AI. This can include maintaining logs, version control and audit trails

- Data Transparency: Disclosing what data is used to train and operate the AI system, including its sources and how its processed.

[Transparency and explainability — OECD](#)

A3 Summary: Responsible AI Guidance for the Public Service: GenAI

Responsible AI Guidance for the Public Service: GenAI

Maintain Trust and Confidence
Everyone has a role to play in helping to maintain the trust and confidence of the public service. Consider carefully how to safely, responsibly and transparently adopt emerging technologies like GenAI.

Build by Design
Incorporate from the earliest design and planning stages key aspects of GenAI, such as privacy, security, and accessibility, to ensure they are embedded in the foundations of the tool.

Engage Early
Work with key stakeholders, partners, teams and end users to ensure their needs and concerns are met when using GenAI.

Strong Foundations
GenAI involves new ways of working. Invest in training, people, architecture and information management. Feed in good data to generate reliable and useful results.

Right Tool for the Task
GenAI is an exciting tool – test and prototype to confirm it is the right tool for the desired outcomes before any significant investment.

Usual Rules Apply
When considering a GenAI tool for your organisation, refer to existing guidance outlining your obligations as part of the public service, such as NZISM and Public Records Act.

Keep Humans in the Loop
Involve humans at all stages of designing, training, testing, evaluating and moderating GenAI tools and their outputs. This helps enhance accuracy, improve processes and ensure appropriate oversight of processes and outputs generated.

Don't go it Alone
GenAI is in use across the public service. Connect with others using GenAI in similar ways to share approaches, patterns, and solutions

The AI Lifecycle

Plan & Design
Understand the scope and set key objectives and direction for the implementation. Consider key foundations including privacy, security, and accessibility.

Collect & Process Data
Ensure good quality data. This includes sensitivity labels and ensuring accurate metadata. Consider if there is any personal information and robustly assess potential risks and benefits.

Build & Use Model
Identify and train your preferred model or algorithm. Consider iterating the model and the review the data and prompts used to improve accuracy.

Verify & Validate
Test your model and evaluate its outputs to assess performance across a range of scenarios and prompts. Consider if it is performing as intended or if further iteration is required.

Deploy
Move into a production environment. Consider how GenAI works with other systems, staff are trained to use it, being transparent about your use, and seek ongoing feedback.

Operate & Continue to Monitor
Continually assess GenAI use, including intended and unintended impacts. Consider if you need to go back to an earlier stage and iterate to produce trustworthy and useful results.

We strongly recommend publishing AI use for transparency

OECD AI Lifecycle
https://oecd.ai/en/accountability

Refer to the full guidance for more information, links, top tips and scenarios that illustrate how you can safely use GenAI tools in your organisation

What is AI and GenAI?

AI: a machine-based system that infers from inputs and generate outputs for explicit or implicit objectives.

Machine/Deep Learning: ML trains machines to make decisions. DL is more specialised, typically involving more complex data and decisions.

GenAI: uses prompts and models to create and generate outputs closely resembling human-created content.

Opportunities for the Public Service to use GenAI

Enhancing customer experience of government services: including automatic transcription that helps teams focus on delivering great customer service and resolving queries.

Streamlining modernisation of legacy systems: including using GenAI to interrogate existing codes, patterns and architectures to map new digital systems while maintaining existing services.

Boosting productivity and cost effectiveness: including meeting summaries and enhanced search and retrieval, reducing team administration so your team can get on with core work and delivery.