New Zealand Government
Te Kāwanatanga o Aotearoa

Responsible AI Guidance for Businesses

Investing with confidence

Accelerating Private Sector AI Adoption and Innovation

JULY 2025

GOING FOR GROWTH

MINISTRY OF BUSINESS, INNOVATION & EMPLOYMENT
HĪKINA WHAKATUTUKI

Ministry of Business, Innovation and Employment (MBIE) Hīkina Whakatutuki – Lifting to make successful

MBIE develops and delivers policy, services, advice and regulation to support economic growth and the prosperity and wellbeing of New Zealanders.

More information

Information, examples and answers to your questions about the topics covered here can be found on our website: **www.mbie.govt.nz.**

Disclaimer

This document is a guide only. It should not be used as a substitute for legislation or legal advice. The Ministry of Business, Innovation and Employment is not responsible for the results of any actions taken on the basis of information in this document, or for any errors or omissions.

Use of Artificial Intelligence in this material

MBIE has used AI to develop aspects of this Guidance in line with the Government Chief Digital Officer's *Responsible AI Guidance for the Public Service: GenAI*, including with human oversight of any generated material. MBIE is a signatory of the Algorithm Charter for Aotearoa; a commitment to use algorithms in a fair, ethical, and transparent way.

Print: 978-1-99-106955-9 Online: ISBN 978-1-99-106999-3

July 2025

This guidance has been produced for a digital-first format. Please refer to mbie.govt.nz for current guidance and further information.

©Crown Copyright
The material contained in this report is subject to Crown copyright protection unless otherwise indicated. The Crown copyright protected material may be reproduced free of charge in any format or media without requiring specific permission. This is subject to the material being reproduced accurately and not being used in a derogatory manner or in a misleading context. Where the material is being published or issued to others, the source and copyright status should be acknowledged. The permission to reproduce Crown copyright protected material does not extend to any material in this report that is identified as being the copyright of a third party. Authorisation to reproduce such material should be obtained from the copyright holders.

Contents

About **3**

What is this Guidance? **3**

Who is this Guidance for? **4**

Why use this Guidance? **4**

How should it be used? **5**

Navigating the Guidance **6**

Understanding your 'why' for AI **7**

Good business foundations for responsible AI **8**

Governance and accountability **8**

Assembling a team 8

Compliance and legal obligations 9

Risk management 11

Recordkeeping 13

Supporting capabilities **14**

Procurement 14

IT + Cybersecurity 14

Privacy 16

Skills and knowledge building 16

Stakeholder interactions **17**

Engagement and consultation 17

Transparency 18

Feedback and complaints 19

AI system specific considerations **20**

Data & Modelling **20**

Fit-for-purpose data 20

Legal and ethical data 22

Model efficacy 25

Use & Outputs **26**

GenAI user inputs 26

Use of GenAI outputs 27

Human-in-the-loop decision-making 29

Continuing the conversation **32**

Appendix One: Glossary **33**

Appendix Two: Checklist resources **37**

Appendix Three: Options to ethically source datasets including copyright works **40**

Appendix Four: Other Guidance and resources **41**

About

What is this Guidance?

New AI developments and applications are advancing at a rapid pace and are helping support business productivity, competitiveness, and innovation. There is strong potential for AI to help lift New Zealand's economic performance and the Government has set out its ambition in New Zealand's AI Strategy.

The Responsible AI Guidance for Businesses (the Guidance) is a voluntary resource to help businesses (including sole traders, non-profits, and individual professionals) to realise AI's benefits through using and developing **AI systems** in a trustworthy way. The Organisation of Economic Cooperation and Development (OECD) AI principles provide a broad direction for this that highlights:

- engaging in responsible stewardship of AI and pursuing beneficial outcome for people and the environment
- designing systems that respect the rule of law, human rights, and democratic values
- building transparency and responsible disclosure regarding AI systems
- prioritising robustness, security, and safety in AI systems
- establishing accountability and a systematic risk management approach across an AI system lifecycle.

"AI" is an umbrella term of technologies with many actual and potential applications. These include, for example, fraud detection, inventory management, and targeted ads as well as autonomous vehicles and disease diagnosis.

According to the OECD:

- **Artificial Intelligence** refers to a machine-based system's ability to 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** 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.

Recently, there has been increased awareness around **Large Language Models** (such as Open AI's 'ChatGPT,' Anthropic's 'Claude,' Google's 'Gemini,' Meta's 'Llama,' or the Chinese model 'Deepseek') and **Generative AI** (GenAI) more broadly. But these are just a portion of AI systems available today. More 'traditional' rule and logic-based AI systems have been around for decades, with more modern AI systems including **machine- and deep-learning.** These support applications such as facial recognition, speech detection, and automated cybersecurity systems.

This Guidance reflects Government and wider expectations around how businesses might assess and understand the implications of any **AI system** that they are using, deploying, designing, or developing. It is in line with New Zealand's proportionate risk-based approach to AI (agreed by Cabinet), commonly seen in other countries and international initiatives advancing AI, where potential risks are treated in proportion to their likelihood, magnitude, and context.

The Guidance outlines various types of considerations that businesses can take into account when using or developing AI systems. These include potential risks to: cybersecurity; privacy; human rights; workplace culture; the environment; **intellectual property** and creators; and physical safety. A range of thoughtful safeguards can help ensure AI systems work well and responsibly. By better understanding the implications of using and developing AI systems, businesses can choose mitigations that are appropriate for their context and feel more confident in taking advantage of the varied and potentially significant benefits of leveraging this

technology. Over time, the Guidance can be built on and supported through supplementary resources and materials, case studies, and toolkits.

This Guidance is not an exhaustive list of all considerations that may be relevant for a business. International standards, such as those developed by the International Organization for Standardization (ISO), can also provide useful direction and information, for instance. The Guidance also does not remove the need to follow all relevant and applicable New Zealand legislation and regulations (or those of international jurisdictions where a business operates). Examples of relevant laws are mentioned later in this guidance (see Compliance and legal obligations).

Who is this Guidance for?

This Guidance is for all New Zealand businesses, regardless of:

- where they sit in the AI supply chain –users and deployers, developers, or both. All **AI actors** can benefit from thinking about the considerations outlined in this Guidance when it comes to AI. However, the majority of New Zealand businesses will be current or potential AI users or deployers, and this Guidance is focused accordingly.
- what kind of work they do – rather than attempt a 'one-size-fits-all' approach, the Guidance sets out internationally accepted concepts and practices to consider across industries. Sector-specific entities may provide their own guidelines, which can be considered in tandem with this Guidance.
- scale or AI expertise – all sized businesses increasingly facing expectations around responsible business practices and due diligence with respect to AI. Bigger businesses may be better-equipped to assess and manage AI. However, small businesses also have much to gain from a responsible approach to adopting AI – including opening new markets, improving access to finance, and staff recruitment and retention. Size is not always indicative of opportunity when it comes to responsible AI adoption.

Why use this Guidance?

AI can involve changes in business processes, the types of products and services offered, and the way businesses interact with customers and stakeholders.

It is useful to consider whether existing management approaches and firm capabilities remain fit-for-purpose in an AI enabled environment, including whether an AI application or system may expose vulnerabilities or exacerbate existing risks. This Guidance outlines considerations for businesses to help put good processes in place, so you can spot and fix issues early.

This Guidance supports a responsible AI approach that can help businesses:

- build better, safer, and more fit-for-purpose AI products and services
- earn the confidence and trust of customers and other **stakeholders**
- be ready for challenges and avoid costly mistakes.

This Guidance draws on the Organisation for Economic Cooperation and Development (OECD) values-based AI Principles: inclusive growth, sustainable development and well-being; human rights and democratic values,

including fairness and privacy; transparency and explainability; robustness, security and safety; and accountability.

It also aligns with international equivalent guidance materials, while taking into account New Zealand values and interests in:

- creating a flexible and adaptable environment for technology experimentation and innovation
- preserving and taking account of our diversity, including by taking care of New Zealanders' data, exercising good practice engagement with impacted communities, and enabling equitable impact of AI developments where possible
- enabling all New Zealanders to benefit from AI, and that benefits from AI are fairly and appropriately distributed
- respecting te reo Māori (Māori language), Māori imagery, tikanga, and other mātauranga (knowledge) and Māori data
- supporting sustainability, including of industry, reflecting that it is in our interest as a small island nation to ensure we are able to innovate with and benefit from our resources for years to come.

How should it be used?

Some parts of this Guidance are always important when using AI, while others might not apply to a business's specific situation.

Businesses can decide the extent to which different parts of the Guidance matter in their context – based on where they sit in the AI life cycle, how the business works, who their customers are, and their business goals and limits.

Even if some parts don't seem to fit right now, reading the Guidance in full will help with understanding the big picture and what to watch out for when using, deploying, or developing AI systems.

This Guidance builds on things many businesses already do—like planning, keeping customer information safe, following the law, and managing risks.

Reflects that, in many cases, managing AI systems responsibly builds on things many businesses already do – like planning, engaging with customers and keeping their information safe, risk management, and otherwise following the law.

Any voluntary guidelines or standards, including this Guidance and supplementary resources, should be used in conjunction with relevant legal requirements as well as any context-specific guidance that may be available from industry bodies or peers, as well as AI suppliers and vendors.

Navigating the Guidance

This Guidance is structured around three layers as detailed below:

1. **'Understanding your why for AI'** – encouraging a clear understanding of your purpose, principles and objectives for the use and/or development of an AI system, as well as attention to continuous monitoring and improvement.
2. **Good business foundations for responsible AI** – including leveraging existing business expertise and functions to account for the use and/or development of an AI system (proportionate to the context and/or use case).
3. **AI system specific considerations** – in line with the AI life cycle including steps businesses can take to manage identified risks/issues.

If you are a business thinking about using AI, the Quickstart Guide for Business Users available at https://www.mbie.govt.nz/assets/businesses-using-ai-tools.pdf may be a useful place to begin.

Some content is specific to **GenAI**, and this will be made clear and highlighted in red throughout.

TIPS

Throughout the section on 'AI specific considerations,' **tips** for mitigations are highlighted in blue boxes

Hypothetical scenarios to demonstrate examples of AI risks and their management are included in yellow boxes.

Other Appendices provide additional information and resources to aid businesses with understanding and implementing the Guidance:

- Appendix One: Glossary
- Appendix Two: Checklist resources
- Appendix Three: Options to ethically source datasets including copyright works
- Appendix Four: Other guidance and resources

Understanding your ‘why’ for AI

There are different ways AI could add value to your organisation. You can use AI systems to, for example, classify data, make predictions or recommendations, produce new content, or, in the case of **agentic AI** take action on your behalf. Sector bodies and organisations may have more information or recommendations for industry-specific uses and offerings.

It is good to be clear on what your organisation is looking for AI to achieve. When determining the purpose of AI use – ensure that its intention is lawful, including that it does not impede on any human or **commercial rights, including privacy.**

When considering how to incorporate **AI solutions** into operations, or what an AI system purpose could be, your business could take steps to:

- understand where there are areas of inefficiencies or repetitive, rule-based tasks
- consider whether there is organisational **data** or assets that you can lawfully process with AI to produce new products or services, or a new business model (see fit-for-purpose data for more detail on appropriate data use)
- draw on **stakeholder** insights
- conduct a **cost-benefit analysis** (including any intangible or indirect costs the organisation may be interested in, for example potential environmental impact)
- consider similar case studies in the relevant industry to draw on
- explore opportunities to experiment safely via ‘AI sandboxes’ where AI systems can be tested in a way that is isolated from real-world operations
- consult industry peers or peak bodies on recommended solutions or use cases.

Cross-functional groups or teams established as part of your business governance practices (detailed in the next section) could have strategic conversations, for example to define:

- priorities, values, and approach to AI, and check these complement and enhance existing policies about information security, data permissions and management, legal compliance, and privacy
- guiding principles to help understand what is important as a business and steer decisions made around data and AI use
- common language and terminology to help with getting started on the right foot.

Good business foundations for responsible AI

This part of the Guidance has a focus on functions and processes that may already be in place in a business, namely governance and accountability, supporting capabilities, and stakeholder relations. They are the foundation of what will allow a business to effectively and safely use or develop trustworthy AI systems.

Take the time to understand your organisational context – including existing governance structures and processes across legal, IT, security, and software development. Then, you can determine what can be used or adjusted to account for AI, or what else needs to be put in place to ensure sound oversight and governance.

Governance and accountability

Oversight across the AI life cycle and operations helps ensure risks are adequately monitored and managed, compliance with any regulations and rules, and clear lines of accountability.

Assembling a team

A variety of roles and expertise is useful to support responsible AI. Depending on the size of the organisation, there may not be the human resources to have individuals dedicated to AI-specific roles. **AI governance** responsibilities may only be a portion of an individual's role. However, if a team can be assembled, it can be useful to bring together a range of diverse perspectives and expertise.

Leadership around responsible use and/or development of AI systems *could* include staff with responsibilities or expertise across, for example:

- any strategic leadership, ethics, or AI specific functions your organisation may have available – e.g. you could engage a senior lead to drive responsible AI efforts
- *Security specialist* – to provide cybersecurity and other security expertise such as vulnerability management, security assurance and governance, threat detection, data, and infrastructure protection and more (see IT and Cybersecurity)
- *Data and/or AI Governance* – to provide advice on appropriate measures to understand and assess data quality, security, algorithms, access permissions, and proper usage of AI systems procured or developed (see Data & Modelling)
- *Technology / Data Science* – to provide data, IT, and technology expertise and deliver on responsible AI processes through operations and/or system development
- *Legal and compliance* – to provide an updated understanding of relevant laws and regulations and how they may apply, to help secure any required legal permissions, as well as help navigate any contractual arrangements with third party data or system suppliers (see Compliance and legal obligations)
- *Privacy* – to give guidance on building trust and managing risks related to information privacy laws
- *HR/Training* – to take responsibility for supporting internal AI education and upskilling, as well as diversity of workforce (see Skills and knowledge building)

- *Communications* – to understand and help deliver on the best way to discuss the AI approach with customers and other identified stakeholders, providing comfort that the best possible service is being provided, and any concerns or risks are being controlled adequately (including through triaging AI related feedback, queries etc.) (see Stakeholder interactions)

These individuals should have the capability to be able to adequately steer responsible AI use and development in the context of the project/s being considered and on an ongoing basis as needed. This includes encouraging alignment across teams, for example towards:

- strategic information sharing
- shared decision-making
- overarching oversight mechanisms
- development of AI policies and supporting resources (helping to mitigate potential risk explored in AI system specific considerations).

Robust oversight and governance that promotes transparency and explainability of AI use or development enables business users to demonstrate responsible use, and developers to demonstrate that AI systems have been developed in a trustworthy way.

They also assist your ability to collaborate with staff, stakeholders, or third parties to assess, explain or demonstrate, and/or externally review responsible use of AI.

Depending on project scale and resourcing, it could also be appropriate to include impacted or external **stakeholders** as part of broader engagement efforts, or to ensure diversity in decision-making (for example, through Māori representation). (See Engagement and consultation).

Policies, contracts, or process documentation may need to be updated to match how you are using or developing AI systems.

It is helpful to clearly document who is in charge of what, so everybody knows their role.

Compliance and legal obligations

Businesses should be aware of their legal obligations in relation to any of their operations, including sector-specific regulations.

It is valuable to identify, catalogue and understand all legal obligations that are relevant for use or development (as appropriate) of the AI system. This can include, but is not limited to elements of the following:

Legislation	Example of potential relevance for AI
Commerce Act 1986	Ensuring AI systems do not engage in practices that restrict competition (e.g. algorithmic pricing collusion)
Companies Act 1993	Upholding Directors duties including due care and diligence and legal and ethical obligations.
Consumer Guarantees Act 1993	Ensuring businesses can uphold obligations to retail customers of goods or services.

Legislation	Example of potential relevance for AI
Contract and Commercial Law Act 2017	Ensuring use of AI does not impede the governance of contracts and transactions.
Crimes Act 1961	Ensuring AI is not used to support fraudulent activity, theft, or other crimes.
Fair Trading Act 1986	Avoiding misleading or deceptive conduct related to outputs or use of AI tools.
Human Rights Act 1993 and ***Bill of Rights Act 1990***	Upholding requirements to avoid discrimination on sex, race, and other protected grounds in protected areas, and protecting civil and political rights.
The Privacy Act 2020*, and any applicable code of practice*	Ensuring responsibilities are upheld for handling **personal information** (which may be included as part of AI inputs or outputs). **Any business interacting with personal information is required to have a privacy officer.** Privacy officers are responsible for ensuring compliance with the Privacy Act, dealing with requests for access or correction to personal information, and working with the Privacy Commissioner during complaints investigation. More information is available here.
Intellectual Property Law including: ***Designs Act 1953, Copyright Act 1994.***	Understanding ownership, protection, and licensing of AI outputs, datasets, and underlying algorithms.
Media Law including: ***Harmful Digital Communications Act 2015, Films Videos and Publications Classification Act 1993, Broadcasting Act 1989***	Being aware of the potential for use of tools to generate or share abusive and/or unlawful electronic communications. And, if relevant, being prepared to handle takedown obligations.

Other laws may be relevant depending on the intended and potential uses of the AI system. This could include sector specific laws, regulations, and standards.

In New Zealand, other resources and guidelines are available to help understand and comply with various obligations, for example the Privacy Commissioner's guidance on Information Privacy Principles' applicability to AI.

If your organisation is operating multinationally or globally, it is important to keep up to date on local and international developments. These may be different to New Zealand requirements.

Various forms of AI-specific regulation and governance are emerging internationally (complementing more general applicable regulations, such as the EU's General Data Protection Regulation). For instance, the EU AI Act was the first AI-specific regulation to be enacted, with others emerging internationally. Countries like Australia, Singapore, the UK, and the US have developed resources to support lawful AI practices.

Additionally, institutions such as the OECD and World Economic Forum are also influencing global expectations and standards. Some businesses are also aligning themselves with international technical standards, such as ISO/IEC 42001:2003 Artificial Intelligence Management System.

Professional legal advice should always be sought if needed.

Scenario A: Appliance Alliance

Three mid-sized companies dominate New Zealand's smart thermostat market (the Appliance Alliance). Each of them separately decide to adopt a third-party AI-based pricing tool (AI Price Software), to stay competitive and optimise revenue.

AI Price Software was marketed as a dynamic pricing solution. It analysed non-public pricing data from each of the three companies and used a shared algorithm to recommend optimal prices for each smart thermostat model. The goal was to maximise margins by responding to market trends in real time.

Initially, the tool appeared to work well. Prices stabilised, and each company reported improved profitability. However, six months in, market analysts noticed something unusual: prices across all three brands were rising in unison and remained high, even during periods of low demand. Competition in the market seemed to vanish, which led to higher prices, poorer quality, and poorer service for consumers and suppliers across the entire market.

A closer look revealed that AI Price Software had effectively facilitated algorithmic collusion. By pooling sensitive pricing data and recommending uniform prices, the software reduced competitive pressure and enabled coordinated pricing – without any direct communication between the companies.

This raised serious legal concerns. The Commerce Commission launched an inquiry, and the companies were found to be at risk of breaching section 30 of the Commerce Act 1986, which prohibits cartel conduct. This type of breach is a criminal offence attracting substantial financial penalties and potential imprisonment.

In response, companies took decisive action:

- immediately ceasing use of AI Price Software
- being accountable to customers about the risk that presented itself and how it is being managed
- legal and technical staff developed procurement guidelines for their respective businesses that help
- training staff to recognise and report irregularities and/or non-compliance in AI systems and their use, including on Commerce Act obligations.

Risk management

AI use and integration can expose vulnerabilities and exacerbate existing risks. That's why it's important to have strong ways to spot and fix risks early. Good risk management helps businesses stay safe and make the most of what AI can offer.

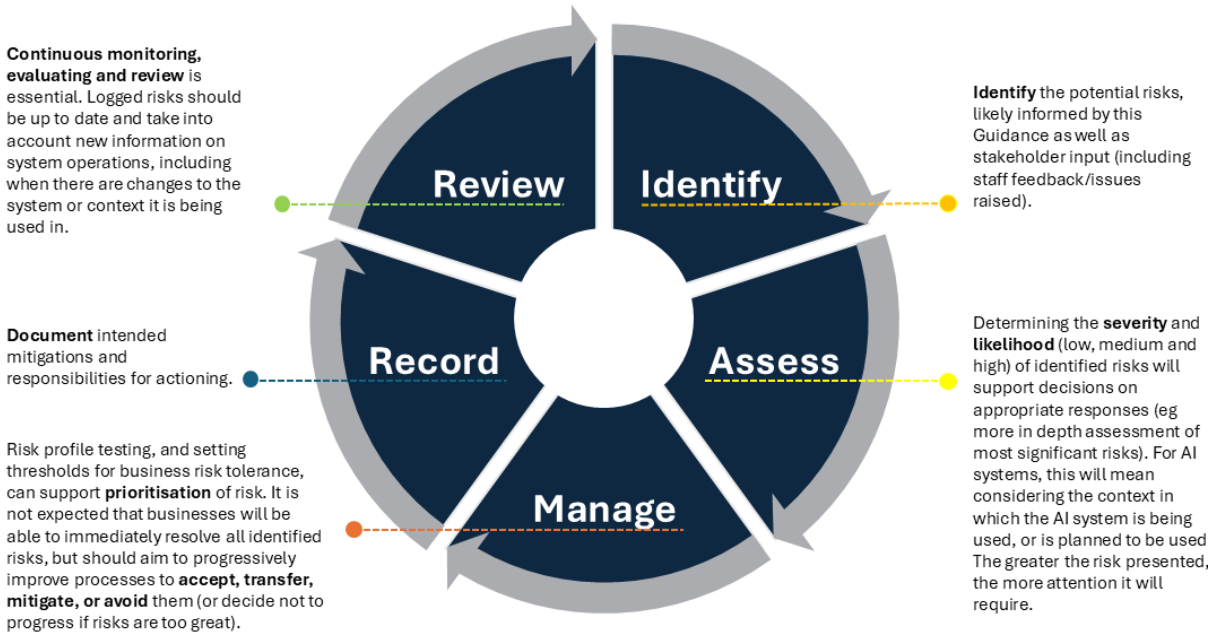
Whether using or developing AI systems, businesses should take a balanced approach to risk management. This means taking account of how AI is being used, what could go wrong, and what resources are available to manage it. Conducting a stakeholder impact assessment can help with also identifying *who* is impacted and possible mitigations.

This Guidance can support businesses to identify and understand what risks might be relevant for their context. These could span ethical, legal, operational, reputational, and technical risks. Some examples of common AI related risks businesses may face include:

- compromise of personal or otherwise sensitive or confidential information through security vulnerabilities and attacks
- unfair treatment of those impacted by AI-informed actions or decisions due to unwanted bias in AI systems and/or human overreliance on those AI systems
- lack of transparency with users or customers as to AI use
- misinformed action or decision-making based on GenAI 'hallucinations.'

The Massachusetts Institute of Technology AI Risk Repository is a comprehensive open resource that identifies and categorises AI risk for public use. It reflects that a number of operational, business, compliance and reputational risks can be exacerbated by AI.

Risk management processes generally will move through the following steps throughout the AI life cycle to support integration or development of AI, and are most effective if started early and performed regularly:



Guidelines are available to support implementation of risk management processes more generally in your organisation, which can be built on for AI contexts, such as:

- the risk management plan worksheet from Business.govt.nz, as part of guidelines for laying the groundwork for good governance
- the Stats NZ Algorithm Impact Assessment toolkit (designed for government agencies but also helpful for business)
- international standards and frameworks, including ISO/IEC 42001: 2023: Information technology – Artificial Intelligence – Management system, ISO/IEC 23894:2023 IT-AI Guidance on risk management, ISO 31000:2018 Risk management guidelines, and other country frameworks listed at Appendix Four.
- other sector-specific risk management strategies and frameworks that may be available (some are listed in Other Guidance and resources).

Additional tips for AI risk management:

- Create and regularly update a **'risk inventory'** to itemise potential risks that can be then prioritised for management (see the Business.govt.nz risk management plan worksheet). This Guidance can be used to support understanding of where risks might lie.
- Predict and prepare. Some problems can be fixed after they happen, while others need planning to stop them before they start. Consider what proactive steps can be taken to prevent foreseeable risks and demonstrate responsible AI use or development.
- Share your plan. Risks and associated mitigations should be communicated to relevant teams and/or third-party providers, and some information could be published, if appropriate and within reason, for transparency.
- Especially for GenAI, usage policies and standards are a good way to communicate the organisation's stance on AI and rules for its use. Through user education and regulation, usage policies can be used to support mitigation of potential risks (including around cybersecurity, privacy, and misinformation). More on this, and potential content to include in usage policies, is available at Use & Outputs.
- Watch out for new risks. Sometimes fixing one problem can cause another. For example, changing how an AI system works to reduce bias might make it less accurate. Think carefully before making changes.
- Contingency, business continuity, and disaster recovery plans can enable fast reactions in case of AI system failure, and an exit strategy allows for safely phasing out AI systems if needed. Consider if data, code, or training can be transferred to a different tool or system if needed.
- Remember to consider dependencies on the tools your organisation invests in and ensure it is able to continue to meet obligations made to customers, as well as the risks of all relevant contracts (including risks within End Use Licence Agreements).

Recordkeeping

As with most business development, harnessing AI successfully requires strategic thinking, planning and, importantly, documenting of business actions and processes.

Recordkeeping and documentation supports accountability, helping you to understand where and how decisions have been made. AI-specific documentation like **AI model cards** are useful to record (for builders and developers for maintaining and communicating, and to users and deployers for understanding) detailed model purpose, data sources, training methodologies, performance metrics, and potential biases.

For larger businesses, several technologies and digital solutions may already be logged in a central place. Businesses may need to be able to refer to where AI is being used and how, to help with answering customer queries, for example, or complying with any required audits.

It's valuable to follow good documentation practices throughout the AI lifecycle. See the Recordkeeping Checklist for more.

Supporting capabilities

Business functions and controls like procurement, cybersecurity, and staff training are needed to cover a host of different areas, including AI.

Procurement

If your organisation is looking to procure an AI system, you will want to plan appropriately for the size and complexity of the project.

As with any procurement – it is important to be clear on business needs and conduct thorough market research on suppliers and their offerings. Consider (and document as relevant):

- customer needs and benefit to your customers
- business needs and the purpose the AI system needs to fulfil
- the costs, risks, and benefits of using an AI system to fulfil that purpose
- evaluation criteria for successful procurement.

When assessing AI solution/s to use, consider requesting a trial (isolated from other technical systems) to figure out if the system is right for your organisation. Seek to understand and clarify the items outlined in AI Procurement Checklist.

Remember and consider that other jurisdictional laws may be different to those in New Zealand – for example requiring providers and vendors to on-share your data.

The AI Forum NZ also has AI Procurement Guides available to support vendor and product assessment. AI model or service cards, and/or third-party responsible AI assessment services can provide information to help determine whether a product is the right fit.

IT and Cybersecurity

The integrity and protection of all systems and datasets related to AI system operations is essential for their effective and uninterrupted operations. This includes assessing and navigating security vulnerabilities, jurisdictional risk (where data is stored or processed outside of New Zealand, and subject to another country's laws), and privacy protection where personal information is involved.

Early discussions with IT security teams help business decision-makers establish level of preparedness and conduct a security risk assessment to determine whether the AI system is appropriate.

A **secure-by-design approach** can support physical and digital resilience, protecting against attack and enabling reliability and consistency in system operations. Digital datasets and systems, including AI systems, can be exploited by **malicious actors.** Attackers could look to harm AI system integrity (to deceive users) or availability (disrupting its use), or to compromise the confidentiality of the training data, intellectual property, or input data.

Security risk of any system depends on several factors, including:

- the information the system has access to
- permitted users
- whether it was developed in-house or procured from a third party

- external data sharing
- attacker motivation for disruption or interference.

Any technology can contain vulnerabilities, which could be compromised and exploited by malicious actors for various reasons. Without careful consideration and management, vulnerabilities can lead to information leaks or unauthorised disclosure, 'poisoning' (of the training dataset), injection attacks, and other attacks.

In line with the National Cyber Security Centre's Cyber Security Framework, businesses should consider taking steps to ensure any security or privacy breaches can be noticed, contained, assessed, and responded to quickly, to mitigate any potential harm to individuals or company intellectual property and comply with Privacy Act obligations where necessary (including notification). This includes forming an incident response plan, as well as monitoring system behaviours and inputs for security risks, and ensuring clear and effective feedback loops for reporting of any system vulnerabilities and risks.

New Zealand has resources available for cybersecurity practitioners to understand good practice in AI cybersecurity particularly, including guidance that New Zealand's National Cyber Security Centre has developed jointly with international partners on:

- Engaging with Artificial Intelligence
- Guidelines for Secure AI System Development
- Deploying AI Systems Securely.

Additionally, an informational video series has been developed to support general online security for businesses – 'Unmask Cyber Crime'.

The Government Chief Digital Office also provides guidance to support government agencies' jurisdictional risk assessment, which can also be useful for private sector consideration.

Given its expanded **attack surface** and often opaque processes, GenAI can also bring some additional cybersecurity risks (see Skills and knowledge building and Use & Outputs for tips to mitigate these at point of use):

- GenAI systems can leak secure information if supplied with certain prompts or other data. For example, Large Language Models have been known to suggest real passwords and API keys found in training data.
- GenAI outputs can be a security risk, for example if used for code generation. Verify that any generated code is sufficiently trustworthy and free of errors with quality control processes.
- GenAI **prompt injection** manipulates AI model behaviour with specifically crafted instructions.
- GenAI models can be more susceptible to data poisoning, given the quantity of data they are trained on, and that this is often public. Extra care is needed to ensure retrieval repositories for **Retrieval Augmented Generation** are protected.
- GenAI models require more **compute** and so can be particularly susceptible to **denial-of-service attacks.**
- There are also related sociotechnical risks to information security (further elaborated on in Use & Outputs.

Privacy

As well as protecting systems and data (including personal information) from cyber threats, businesses also need to consider responsible and lawful management of data more generally.

Legally, businesses need a privacy officer appointed if dealing with any personal information (including collecting, use, or storage). Details on that role are available from the Office of the Privacy Commissioner. See Legal and Compliance for more.

Artificial Intelligence can exacerbate privacy risks, and a privacy-by-design approach can be valuable to help build in privacy protection to information systems, business processes, products, and services from the start. Privacy is often also considered as part of processes around risk management.

As a starting point, it is important that data is classified appropriately so you know what is personal information and should be treated appropriately.

More information on how to support privacy protection as part of building and using AI models is included in AI specific system considerations.

Skills and knowledge building

Staff need the capability to competently perform their roles and uphold responsible AI practices as required. Growing **AI literacy** within a business is important.

To support this, businesses can:

- document what competencies different roles or groups of staff require (in relation to their involvement with AI system development or operations and risk management practices and policies). Consider –
  - foundational responsible AI training/education for all staff, including understanding the fundamentals of AI, using GenAI responsibly, what are allowed business uses, and its limitations
  - tailored training for those developing or governing AI as part of their roles (see Assembling a Team), including on how to mitigate AI system bias and evaluate AI outputs
  - training for end users and other operators.
- put in place regular training for staff. This could cover technical education as required but also upskilling around other areas of risk that have been identified (e.g. ethical obligations, privacy, cybersecurity, intellectual property). It may also cover incentives for complying (or consequences for non-compliance) with defined policies (including AI usage policies and/or standards) or procedures.
- adapt and improve training programmes based on participant feedback where possible. A clear mechanism for continuous feedback on training, experiences deploying training, or gaps that need to be filled will help with understanding issues as they present themselves and identify improvements.
- partner with (AI) specialists to help bridge any skills gaps.
- join or leverage collaborative initiatives to share experiences, tools, and practices, including for ensuring responsible AI use and/or development. The OECD has catalogued Tools and Metrics for Trustworthy AI (such as technical tools to remove bias, audit AI systems, or measure fairness), which can be browsed, filtered or searched as required.

To embrace the opportunity GenAI offers in a safe and responsible way, businesses can build skills, capability, and diversity across teams, including through education (see Appendix Three: Other Guidance and resources) on:

- strengths and weaknesses of GenAI

- how GenAI tools work, and how to use them most effectively (especially with **sensitive information**) including through prompt engineering (see Use & Outputs)
- when GenAI can augment aspects of human work and/or decision-making and where AI may be able to perform tasks with minimal human involvement (see Human-in-the-loop decision-making).

Stakeholder interactions

Clear and effective communications with impacted **stakeholders** - including but not limited to staff, customers and clients, shareholders, and other impacted parties – is important for building and maintaining trust and confidence in how AI is being used and developed.

A stakeholder impact assessment can help identify and understand your organisation's AI stakeholders. This includes steps to:

- brainstorm individuals, groups, communities, or other organisations that have an interest in, can affect, or can be affected by what your business does (and what the AI system does). This could include internal employees, management, customers and clients, shareholders, connected providers, Māori customers and users, and other interest groups. Stakeholders may include, for example, individuals whose personal information and/or IP are used in training or input data; **data annotation** workers; and anyone being monitored by AI systems.
- determine the level of impact the project will have on identified stakeholders
- determine the relative impact of each stakeholder on the success of project implementation
- define known rights, needs, expectations, concerns, risks, and areas of interest they may have, and if or how organisational-level AI policies and procedures will or can address these.
- understand what engagement approaches will best suit different stakeholders – including frequency, channel, and relationship 'owner.'

AI systems can affect different people in different ways. Special consideration needs to be given to the impact that AI systems may have (either directly, or indirectly) on specific and/or underrepresented communities such as non-English speakers, disabled people, older people, LGBTIQ communities, Māori, or Pacific peoples.

For example, some speech recognition systems might not understand different accents or languages. This can stop impacted people from engaging with those systems, limiting customer/user base, and exacerbating inequities. Facial or image recognition poses similar challenges.

Engagement and consultation

Considering the needs of impacted stakeholders, and involving them in decision-making as appropriate, helps you build and offer products and services that people actually want and need. Meaningful stakeholder engagement can inform all steps of the AI life cycle.

Stakeholders (identified as part of any stakeholder impact assessment) can be involved through internal discussions deciding what the AI system should do, designing how it works, selecting the right training data, as and/or testing and improving the system. Where you can't speak directly with stakeholders, you can try to consult independent experts, peak representative industry groups, unions, and civil society groups could be consulted.

It is especially important to be mindful of specific and/or underrepresented communities who may have unique considerations to take into account. For example, where Māori community/ies are impacted by a project, businesses could engage with Māori expertise through channels such as your customer base or

governance board. This helps to identify and respond to any problems early that impact service or product offerings, and/or disproportionately affect certain groups.

A **stakeholder engagement plan** can help businesses stay organised - outlining who to talk to, when to talk to them, and what to share.

Effective engagement is equitable, safe, two-way, value-adding, and conducted in good faith. Careful and strategic planning about the most effective method and timing of engagement is recommended, particularly for larger companies who may be deploying several different systems in different areas simultaneously with overlapping stakeholder groups.

Transparency

People want to know when and how your business uses AI. Being clear builds trust and helps avoid problems.

You can share information on what your AI systems do, how they work, and what risks they might have. Check out the AI Transparency Checklist to help think this through.

Different groups may need different levels of detail. Consider and document the level of transparency required, method and type of communication, for different audiences. Consultation with representatives from relevant stakeholder groups can help to understand what type and level of engagement might be best.

In general, it is good practice to let people know when/where a GenAI system is being used or included in functionality (e.g. a chatbot on a website). Labelling AI generated content, or providing disclaimers about AI use, supports transparency and helps identify GenAI use to customers. **AI watermarking** is an emerging technology to identify AI generated images in a more robust way that can support authentication, though not immune to manipulation or watermark removal.

Consider digital **accessibility** and readability needs. GenAI tools can inherently support improved communication, for example with those who have speech or hearing impairments.

Reflect any transparency requirements of third-party suppliers and other partners in relevant documentation and contracts.

Feedback and complaints

Businesses may already have ways for people to ask questions or make complaints about AI use or development. If not, it could be as simple as a contact email address and/or phone number, and a clear plan for how information received will be responded to.

Some Guidance is available on business.govt.nz about the steps in a complaints process and training staff to handle complaints.

Effective channels for stakeholder queries/feedback/complaints:

- are easy to find and use - think about groups with specific accessibility needs and how they can be supported to provide meaningful feedback
- explain how decisions or outcomes of the AI system can be reviewed
- let people talk to a real person if needed (a human-in-the-loop)
- support timely and effective responses
- tell people when they can expect to receive a response.

Recordkeeping is crucial, as proportionate to the risk, to support effective handling of stakeholder issues.

AI system specific considerations

Data & Modelling

Data is the 'fuel' that makes AI work. It is used to train the AI model so it can, for example, spot patterns, make predictions and choices, create content, or perform another function for which it was built. The better and more fit-for-purpose the data, the better the AI will work.

Sometimes, even AI developers don't fully understand how AI systems arrive at their outputs, especially with GenAI. That's why it's important to prioritise data quality from the start, so outputs can be as reliable as possible.

Data and modelling processes are primarily relevant for those building or developing AI models. However, those using or deploying existing, already-trained AI solutions, should be aware of the considerations relating to data and modelling – to better understand whether the system is right for its intended use and fits with business values. This is increasingly relevant as more GenAI applications are released.

IT and Cybersecurity discusses the importance of ensuring good cybersecurity practice to protect AI training data and model integrity.

Fit-for-purpose data

Training data and datasets should be good quality – accurate and 'clean,' complete, lawfully obtained or accessed, structured appropriately (including to support transparency and explainability), relevant and representative of the environment that the AI model may function in.

Poor quality training data doesn't just cost money – it can lead to mistakes, upset customers, waste money, and damage business reputation.

Improper **data collection and processing** can mean potential privacy and confidentiality breaches, intellectual property rights violations, or potential data sovereignty and human rights impacts. Consequences may be reputational, financial, punitive, or obstructive (such as via compliance, cease and desist, or asset freezing orders).

AI datasets and systems are likely to have varying degrees of **bias** (for example, reflecting patterns in historical decision-making that may be based on inappropriate or harmful biases, representation, or how methodologies are implemented).

Special attention needs to be paid to potential bias when training data contains information about people – and especially when model outputs and associated decision-making can directly impact individuals, families and whānau and wider communities.

AI systems can amplify inaccurate or unreliable outputs, **unfairness,** discrimination, and/or harm to individuals (including those potentially *not* receiving services due to system discrimination), and/or damage to business reputation.

TIPS

- Understand the data, the context it was collected in, and what uses are permitted.
- Document (or require documentation on) how the data was sourced or collected, any modifications made and known biases or accuracy levels. Evaluate its reliability and limitations.
- Consider what open data sets are available and relevant for developing or fine-tuning an AI model. This includes data that doesn't have personal information but has valuable geospatial, health and other data sets made available free of charge. The Open Government Data Programme takes a collaborative approach to making government data available for reuse.
- Consider the environment in which the AI model will be deployed, and whether the data is suitable and relevant for that context. For example – a facial recognition model to be used in New Zealand would likely be more accurate and effective if trained on images representative of the New Zealand population.
- Document (or require documentation on) the legal basis for training data to be collected, used, and/or stored (including consent mechanisms for customer data collection).
- Remove inaccuracies, duplicates, and errors from datasets (including correcting typos, checking for correct formats, valid ranges, and logical consistency).
- Where relevant, take care that data from various sources is combined in a cohesive manner, and you are able to attribute those sources.
- Keep data updated to reflect new information.
- Label, tag, and annotate data for better accuracy and performance, and to support any attribution, traceability, and audit needs.
- Document (or require documentation on) any **preprocessing steps** for the data (e.g. **cleaning, splitting, transformation, or augmentation**).
- **Data fields** containing sensitive or **protected attributes** should be defined and handled appropriately to minimise harm.
- Diversity of thinking (including a range of backgrounds, perspectives, skills, abilities, experiences, and identity characteristics) in teams involved in governance, deployment, use, and review of AI system/s can help identify and support resolution of bias, inequity, and discrimination.

Scenario B: BigBuild

BigBuild is a New Zealand construction firm looking to speed up their recruitment processes. They decide to trial a CV screening AI tool to scan applications and rank them based on desired skills, experience, and other key words. Those that fall below a certain threshold are filtered out and do not progress further.

Following a risk assessment, BigBuild puts in place some risk mitigations, including being transparent with applicants about AI involvement in the shortlisting process, and implementing a human reviewer for all shortlisted applications.

After deployment, 50 candidates apply for a position at BigBuild. They are advised that AI is being used. The human reviewer notices that, despite there being an equal amount of women who applied, only 3 of the 10 shortlisted applications were women. This is not in line with BigBuild's goals for gender diversity in hiring. They also receive correspondence from rejected female candidates questioning the AI system results, and upon evaluation their applications were at least as deserving as those that were shortlisted.

On investigation, BigBuild finds that the AI model was trained on historical 'successful hire' data – which underrepresented women, likely due to a mixture of historical human bias and societal factors. Therefore, despite the AI system being fed 'gender blind' CVs, it viewed factors and language more so included on men's CVs as positive (such as 'executed'). Meanwhile, it assessed similar factors that appear more so on women's CVs (for example, language like 'supported' and 'co-ordinated') negatively. Career gaps (that may have been taken due to maternity leave), and certain hobbies, roles, or institutions (such as 'volleyball,' 'girl's school') also contributed to the system being more likely to rank female candidates lower.

To address the issue, BigBuild apologises to impacted candidates and paused use of the tool while it works with the provider to retrain the model using more representative, bias-audited data. They also introduce a diverse set of 'success profiles' representing a range of career pathways and experiences and lower the shortlisting threshold, so more applications move to human review.

BigBuild regularly monitor and evaluate the tool's performance throughout their trial and continue to provide feedback to the developer, as necessary.

Legal and ethical data

Collecting and processing data ethically, in ways that respect people's rights and privacy, helps protect from legal risk and keep reputations strong.
Some types of data need extra care.

Sensitive, proprietary, or personal information

If any data includes **proprietary, confidential** or personal information, special consideration needs to be given to risks of it being accidentally shared - which could result in trust or privacy breaches, legal liabilities, commercial harm, and reputational damage. Such data could include confidential business information (such as access keys, source code, or billing details), trade secrets, customer data, or personal information. AI systems learning from this data could retain patterns and relationships within it that may surface information that you don't want it to or to people that are not supposed to be able to access it.

Even use of personal information that is already 'publicly available' may be considered unethical, illegal and/or damage reputation in some contexts. For example, '**web scraping'** practices can be seen as intrusive and could erode customer trust and threaten customer privacy. These considerations are important to both an AI system's underlying training data, and to any data supplied by a user to an AI system (see Use & Outputs).

TIPS

- Conduct a **Privacy Impact Assessment** at the design stage to support a privacy-by-design approach. Privacy-by-design helps ensure privacy protection is built into information systems, business processes, products, and services from the start. See Privacy for more.
- As part of your Privacy Impact Assessment, check that **Information Privacy Principles** (IPPs) are applied (see OPC Guidance on Privacy Impact Assessments, and AI and the IPPs). This includes being able to answer questions like:
  - Why is the information required for the AI-enabled (business) purpose?
  - Is only necessary data being collected for that intended business purpose?
  - Has the information been collected in a way that is 'fair' (with the individual having an appropriate level of understanding and choice)?
  - Is it reliable enough and accurate?
  - If/how are individuals able to access and correct relevant data? If the model has already been trained on it, can it be corrected?
  - Do those handling the information have proper training, including to exercise required confidentiality?
  - Is the storage and privacy of personal information being actively monitored?
  - Will data inputs be transferred to companies located overseas (out of New Zealand)?
- Practices such as **data anonymisation**, **encryption** and secure storage can be employed to help protect personal, confidential, and other sensitive information.
- Robust model evaluation processes can help mitigate against sensitive, proprietary, or personal information used for training being reverse-engineered or otherwise extracted from the system.
- The use of privacy-enhancing technologies can help maintain privacy without impacting the AI system's functionality. See information provided by the OECD on Privacy enhancing technologies for more details.

Ownership and intellectual property rights

Training data and models can be sourced from proprietary datasets, open data platforms, or public content (including via web scraping practices). Each source may come with specific licensing agreements that describe who may access them, how they can be used, and/or how they must be labelled or attributed. It is good practice to disclose these details about the source(s) of training datasets, including any **intellectual property rights** licences entered into if applicable.

The 'black box' nature of some AI systems puts pressure on expectations and understanding around intellectual property protections when it comes to training data. Open-source datasets may be protected by intellectual property rights and have certain conditions that need to be met in order to copy and/or use those datasets (e.g. Creative Commons licences), including **attribution requirements**, or restrictions on derivations or commercial use. Similarly, terms of use restrictions in publicly available or proprietary datasets might prohibit web scraping or use for AI training.

GenAI, in particular, has been largely reliant on copyright works for its development. Fairly attributing and compensating creators and authors of copyright works can support continued creation, sharing, and availability of new works to support ongoing training and refinement of AI models and systems.

Some datasets will include proprietary databases - where authorisation is needed to access and copy information from those databases, which may only be granted for specific purposes. For example, publishers of medical journals may have created a database of medical articles and provide licenced access to academics for

research purposes, but not for other purposes (such as for software developers training AI tools). Besides raising questions around potential infringement of IP rights, it may also involve a breach of contract (terms and conditions related to access to the database).

While this section focuses on data and modelling, ownership, and intellectual property considerations around GenAI *outputs* is included in Use & Outputs.

The World Intellectual Property Organisation also has GenAI specific Guidance available here.

TIPS

- Conduct a licensing assessment to ensure appropriate copyright permissions are in place. These considerations are also relevant to issues outlined in GenAI user inputs.
- A **licensed-by-design** approach can ensure fair, lawful, and ethical compliance is built into business practices and systems from the start. New and emerging options are outlined in Appendix Four.
- Businesses will want to obtain appropriate permissions to copy and/or use data (or verify that proper authorisation has been acquired) for training or other means, alongside maintaining **data provenance**, to avoid infringement and creating commercial or reputational harm.
- Some customers may prefer to use or procure AI systems, particularly GenAI tools, trained on ethically obtained or accessed data (with appropriate consent/s). There are options available to ensure training data is ethically and lawfully obtained or accessed, including with the explicit permission of the owners of the datasets used – for example, through 'collective licensing schemes.' See Appendix Three for more options.

Māori and other indigenous data

Māori data refers broadly to digital or digitisable data, information or knowledge that is about, from or connected to Māori. It includes data about people, language, population, place, culture, and environment.

Producing, using, or handling Māori data in your organisation may warrant special considerations. AI systems can enable misrepresentation, misappropriation, or misuse of data and mātauranga Māori and other Indigenous knowledge. This can mean inappropriate commodification of that data, disregard for indigenous protocols around that data, or reinforcement of stereotypes which perpetuate inequality and harm.

Businesses who may be using Māori data can avoid its misuse or exploitation, with appropriate safeguards, consultation, and cultural considerations.

Guidance from the Centre of Data, Ethics, and Innovation provides further detail on Māori data and AI.

TIPS

- Know the difference between data that is non-sensitive or **noa** data and **tapū** Māori data (which has sacred meaning to Māori, and where Māori have a strong interest in being involved in deciding how it can be protected with appropriate tikanga, if collected or used at all). More information on tapū and noa data is available on data.govt.nz.
- Having Māori data as part of datasets reinforces the importance of ensuring appropriate guardrails are put in place and processes are in line with good practice as outlined in this Guidance including around: data accuracy; ethical and legal data collection and sourcing; supporting creators to economically benefit from their IP; protecting personal information; ensuring cybersecurity; minimising bias; reflecting diversity; committing to data provenance and recordkeeping; and being transparent.

- Involve Māori in AI development and decision-making that could impact them (for example, as part of any governance board). This can help to understand what data could be considered **tapū** and/or require more extensive measures to appropriately manage that data, including potentially not using that data. Build these relationships not just to manage immediate risks, but also to create room for innovation and leadership into the future.
- Your organisation may choose to store data supporting its AI system on New Zealand servers and within New Zealand's legal jurisdiction, which may better support data sovereignty than offshore storage.[1]
- The Waitangi Tribunal case 'WAI262' also delves into issues around intellectual property rights in the context of protection of taonga Māori. While primarily relevant for the Crown, this can provide some direction around what is considered to be respectful collection, definition, storage, and use of Māori data.
- The Principles of Māori Data Sovereignty – Te Mana Raraunga, the Māori Data Sovereignty Network are a helpful guide. While developed for government agencies, businesses may also find the discussion relevant for their activities.

Model efficacy

Those building and developing AI models will want to be mindful of exactly what they want it to do. Developers can work to minimise the risk of misleading, biased, or inaccurate outputs or decision-making, or of cybersecurity weaknesses, at the outset, while there is opportunity to improve or correct it before potential harm is caused.

The type of system architecture that is the best fit for an AI system will primarily depend on what the system needs to learn from the data in order to solve the business problem.

To make sure AI systems work properly, they should be tested often and actively monitored. The type/s of testing will depend on the objectives of the system, and can include a focus on accuracy, privacy, or **explainability** (understanding how the AI system makes decisions) for example.

AI systems can be adjusted to improve how they work. Refinement methods such as fine-tuning, constrained sampling, and post-processing filters (for example, for grammar correction, offensive content, or filtering out personal information) can be used to improve outcomes.

Before model release, developers or deployers should be satisfied that it performs adequately, is reliable, and safe. A number of vendors have guidance and tools to support continuous evaluation of performance and responsibility metrics before systems go live.

TIPS

- Establish success metrics and thresholds to evaluate the model against and monitor performance over time.
- Avoid unnecessary model features to support better model performance and explainability and reduced computational costs.
- Developers can consider feature flags to be able to turn some functionality off without deploying new code.

[1] Information on New Zealand based data centres can be found at https://www.datacentermap.com/new-zealand

- Run ‘model scenarios’ with a pool of human evaluators representing your customer or end user perspective. By adapting communications or the model itself in response to its performance, your organisation can support efforts to build stakeholders’ (including workforce and any users) understanding of how to interpret AI system outputs (and related decisions) as it relates to their use.
- Models can be evaluated on safety characteristics such as prompt stereotyping (encoded biases for gender, socioeconomic status, etc), factual knowledge, and toxicity.
- Model safety can be assessed through penetration testing, ‘**red teaming,’** threat modelling, and/or audits.
- AI models can change over time, so it is important to regularly test and audit the system to support continued assurance of its accuracy and usefulness and understand if there are any emerging or new risks to consider.
- Document the model version and dataset used for each model.
- Developers can produce an **AI model (or service) card** detailing model purpose, data sources and provenance, training methodologies, performance metrics (see the OECD’s Catalogue of Tools and Metrics for Trustworthy AI), and potential biases. Deployers and users can request or require model cards to understand and assess these things.
- **Data drift** can impact a model as it takes hold. Active monitoring (and plans to retest and retrain the model as needed) can help detect deviations.

Use and Outputs

AI tools can be helpful, but they don’t always get things right. That’s why it’s important to think carefully before using AI results to make big decisions – especially ones that affect people’s lives. AI should support the work people do, not replace their judgement. It’s best when humans and AI work together, with people staying responsible for the final choices.

GenAI user inputs

GenAI tools work by responding to prompts – the questions or instructions you give them. The information included in any prompt (e.g. the images you provide, or text you type into an LLM) can affect what the tool gives back, and sometimes even how it learns and responds over time.

Some GenAI tools, particularly those that are publicly available or free of charge, may share prompt data (including attachments) with external parties (including developers or other users). This means information provided through prompts could be exposed in future through a security breach or the model itself.

Even if your prompts or attachments don’t include names or are otherwise **anonymised** or **depersonalised**, personal, sensitive, or classified information could be joined up with other information over time to re-identify someone.

TIPS

- Educate users on prompting limitations, requirements, or guidance.
- Prompt engineering courses are available from a variety of providers, many of which are available online and/or free of charge. These can be particularly effective if paired with organisational guidance that takes into account the unique business context.
- Business usage policies can help communicate to staff what tools are endorsed, and how they can be used.
- Providers of GenAI tools can support responsible use by proactively providing the parameters of their tools, and information on whether prompts will be reused as training data or otherwise stored.
- Consider the source of any prompts and data being inputted and whether doing so breaches any rights (privacy, intellectual property, confidentiality). Inputting your own copyright works could risk your control or exclusivity of the input or works generated from AI systems trained on that input data.
- Tools are available to flag inappropriate prompts.

Advice for prompting

- Be specific and prescriptive in queries, including in regard to relevant context and the desired tone or style.
- If using an LLM, consider asking for step-by-step instructions to help with understanding the response and pinpoint if and where errors are being made.
- Provide examples of what you are looking for to guide the type and structure of response.
- Include instructions to be clear if/when the model is unsure of a response.
- Ask for information sources and references so you can verify output accuracy.

Use of GenAI outputs

GenAI can be used in a variety of ways including to improve customer experience, for marketing and content creation, and for productivity gains. However, there are some potential risks when it comes to GenAI outputs and their use:

- LLMs are designed to generate "statistically probable language patterns," which means different (though often similar) answers are likely to be given to the same prompt. Businesses should therefore not rely on any specific answer being produced in response to a specific prompt.
- LLMs may be susceptible to errors, omissions, training bias, and factual inaccuracy. While responses can seem well developed and credible, they may confidently present opinion as fact, skip important details depending on the prompt given, and in some instances fabricate information as truth. This tendency is referred to as 'hallucination.'
- Depending on the training data and prompt used, outputs may also reflect prevalent, often biased, out-of-date, or unethical viewpoints. These tendencies could impact Māori and other Indigenous, minority, or otherwise disproportionately disadvantaged communities (including women, older people, young people and children, disabled people) disproportionately.
- Some GenAI tools that generate realistic images or videos of a person or their voice, can be used maliciously. These are called 'deepfakes' – realistic but fabricated audio, video, or images that convincingly mimic real individuals. Malicious actors can use them to spread misinformation, defame

individuals, or enact scams and exploitation. The technology has also been misused to create explicit content.

- GenAI can be, and is often used, to generate new and novel content and material. However, outputs may lack commercial protection or ownership (including given issues assessing level of 'human authorship' and establishing originality). There is also always a risk that model outputs may (intentionally or inadvertently) be substantially similar to existing copyright works (whether or not those works were used to train the model) and therefore be potentially subject to plagiarism and copyright infringement claims.
- GenAI has been used to replicate components of Māori culture and traditional knowledge without appropriate permission and attribution. New Zealand's intellectual property laws include provisions for the protection of mātauranga Māori, in relation to patents and trademarks. These provisions help prevent the registration of trade marks or granting of patents that would be considered offensive by Māori or contrary to Māori values. Te Puni Kōkiri (the Ministry of Māori Development) is working to address regulatory barriers to commercialisation of mātauranga Māori (traditional knowledge).

TIPS

- Offer information and training on:
  - the components of a good prompt, to support usefulness of GenAI outputs. For example, asking an LLM to act as its own editor or proofreader can help to reduce errors. See GenAI user inputs for more.
  - fact-checking and cross-referencing LLM outputs as appropriate, especially where they are used for decision-making or other work where there is significant negative impact associated with inaccuracy.
  - necessary steps to treat sensitive, personal, or otherwise protected information that may be produced by GenAI with appropriate care to prevent unauthorised sharing, disclosure, or use.
- Check terms and conditions and specifications of GenAI models before use, and/or establish written agreements with AI providers to create certainty of output ownership. See Procurement and Legal and ethical data for more.
- Be transparent about who 'owns' any outputs (responses or content) and any conditions on what those outputs may be used for.
- Consider **AI watermarking** as a way to support the public to identify what is real and what is AI-generated (though noting it can be susceptible to manipulation and/or watermark removal).
- Consider how the system will be used operationally and aspects like user interface – for example, it is less likely a user will review a generated email before sending if it is entered directly into the email window and they only have to click 'send.'
- If using GenAI, potentially to assist in IP production, you may wish to keep your own records of how it is used (prompts, edits, decisions etc.). Internationally, there are AI systems with inbuilt tools to help track and document human authorship behind AI-assisted creations (e.g. Invoke AI's Provenance Records).
- Uses where IP rights are not essential could be for internal business use, idea generation, production planning, or scheduling, brainstorming, or mind-mapping, and streamlining.
- Work closely with Māori to manage and understand potential impacts, especially where Māori data is used, on services or products that affect Māori communities.

Scenario C: ChoiceConsulting

ChoiceConsulting is a small independent policy consultancy in New Zealand that has started using generative AI to increase efficiency in producing background briefs, literature scans, and contextual overviews for clients in the public and private sectors. The team finds that the AI tool significantly speeds up their workflow, allowing them to generate first drafts of content in minutes.

While preparing a report for a government client, one of the consultants uses GenAI to produce a summary of recent international research. The AI-generated draft includes several citations and a compelling statistic attributed to an internationally renowned organisation. The consultant includes the information in the client deliverable with only a light review, trusting the AI's output.

However, the client attempts to follow up on the statistic and finds that the cited report does not exist. Upon further investigation, it turns out the statistic and reference were entirely fabricated by the AI, a known phenomenon called a hallucination. The client raises concerns about the credibility of the work and requests a full review of all sources and claims in the report. This unexpected rework costs ChoiceConsulting additional time and damages their professional reputation.

In response, ChoiceConsulting revises their internal processes. All AI-generated outputs are now reviewed by a human with subject-matter knowledge, and every citation must be verified against original, reputable sources before inclusion. They introduce a step in their workflow to replace or remove any unverifiable content and incorporate a disclaimer in all draft materials that clearly explains how GenAI was used and what human verification was undertaken.

Human-in-the-loop decision-making

AI systems, and particularly those using advanced machine learning methods such as deep learning, are often treated as 'black boxes' due to their complex inner workings.

Depending on the way an AI system is being used, some level of real-time human intervention may be needed. In most situations, the level of human review will depend on the proven level of reliability of the system to deliver desired outputs, combined with a business's assessment of their specific tolerance for the impact of inaccuracy.

If an AI system is doing something low-risk – like suggesting a movie or a product – it might not need a person to check its work. This is often referred to as 'human-out-of-the-loop.'

If an AI system is helping to make big decisions – like those about money, health, or the law – a person should always be involved. This is called 'human-in-the-loop.' The level of human involvement will likely depend on the risk associated with those decisions, or the cost of harm of a 'wrong' decision. This can range from having a human consistently make the final decision (informed by an AI output) or having human oversight over particularly high-risk or low-confidence AI results.

For example, an AI system may be used to analyse transaction patterns to detect fraud but given the potentially high cost of incorrect decisions (which could lead to financial losses, regulatory penalties, or reputational damage), high-risk transactions are flagged for human review. This can be considered 'augmented intelligence' – the AI assists with scanning transactions for suspicious patterns, but the fraud analyst will make the final decision.

Humans can over-rely on automated systems or algorithms and accept outputs as correct without critical evaluation. This is 'automation bias,' and can undermine the purpose of having human oversight as individuals are not effectively acting as an independent check.

Overall, regardless of the supporting technology used, businesses are responsible for their decisions, so robust checking and due diligence, including adequate recordkeeping and documentation, is crucial.

TIPS

- Humans-in-the-loop should understand the risk of automation bias, identify other biases they may bring, and be well equipped to serve the purpose of their involvement (for example, have the subject matter expertise to recognise inaccuracies).
- Confidence, certainty, or accuracy percentages can be provided alongside outputs to help gauge whether it should be used or relied upon.
- Where possible, systems and processes should be designed to support **explainability**. This creates understanding of how an AI system has come to a conclusion, helps build confidence in its outputs, and can support responses to any customer queries. Saliency maps, 'what if' scenarios, and other explainability features can support this.
- Where there is very low risk appetite for inaccuracies, tasks can be divided into intermediary outputs that can be easily checked or verified by humans – so it is clearer how an end output was arrived at and what checks and balances have taken place. Mandatory verification steps can also be put in place.

Scenario D: DermaDupe

DermaDupe is a rapidly growing skincare e-commerce brand. Within its first year, it had attracted a small but loyal customer base and built a sleek online presence using a third-party web storefront platform (ShoppingCommerce). To scale quickly, stand out in a competitive market, and boost conversions from website views to purchases, DermaDupe integrated AI tools from the ShoppingCommerce App Store into their website.

These included a chatbot for customer service, a conversion optimiser, a customer review generator, and a product description generator.

Not long after implementing these AI tools, the company received a number of phone calls from customers complaining of a number of issues including that:

- the website had informed them that there were only one or two units of their favourite product in stock, but had allowed them to purchase more (when they attempted to, worrying they would run out)
- some product descriptions were inaccurate – for example, claiming all natural ingredients when that was not the case –in contrast with the positive reviews on the same page
- when attempting to make their complaints via the online chatbot, they received incorrect information about their option to receive a refund and there was no option to speak to a human.

DermaDupe discovered that the AI tools they had implemented:

- were creating false scarcity claims, unrelated to actual stock levels, in order to trigger customers' psychological urgency to purchase the product
- made false and misleading claims about products in attempt to generate product descriptions that were appealing to the DermaDupe's customer base
- had automatically posted reviews under fake customer profiles
- did not take into account New Zealand regulation such as the Consumer Guarantees Act.

On learning this, DermaDupe disabled the ShoppingCommerce auto-publishing feature for product descriptions and reviews. They replaced the false reviews with verified customer feedback and ensured human review for all AI-generated descriptions prior to publication. They retrained the customer chatbot's

understanding of customers' options for return and ensured it provided options to speak to a human in the event it could not answer a customer query.

Luckily, DermaDupe were able to identify and rectify these issues early thanks to their customer feedback mechanisms and responsiveness. Otherwise, they risked breaching the Fair-Trading Act 1986 for misleading consumers – which could have attracted substantial financial penalties.

Continuing the conversation

AI is developing rapidly and approaches to safe and responsible AI adoption will likely evolve over time. Businesses should connect with one another where possible to share responsible AI approaches, patterns, and solutions. This will help grow our collective knowledge on how to best use these systems, and how to steward trustworthy AI that responds to social, legal, and ethical challenges.

We hope to improve and build on this Guidance in partnership with NZ businesses, and to reflect emerging international knowledge and practice.

This Guidance is part of a broader suite of Responsible AI Guidance including the *Responsible AI Guidance for the Public Service: GenAI* (published by the Government Chief Digital Office in the Department of Internal Affairs). Supporting material for Responsible AI Guidance for the Public Service will be housed on digital.govt.nz and grow over time.

Industry sectors are encouraged to consider what other resources would support implementation of the Responsible AI Guidance for Businesses.

The Ministry for Business, Innovation and Employment (MBIE) is also considering what support would be most useful for our small businesses to adopt AI successfully and in a responsible way, in line with this Guidance.

Acknowledgements:

This guidance has been developed by the Ministry of Business, Innovation and Employment (MBIE), alongside the Government Chief Digital Office and other government agency partners, in consultation with AI Forum Executive Council and Governance Working Group members. Many thanks to other organisations that have provided feedback and insights including BusinessNZ, Copyright Licensing NZ, NZTech, and Tech Users Alliance NZ.

Appendix One: Glossary

AI system
As defined by the OECD, 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.

(AI) Actors
Those who play an active role in the AI system lifecycle, including organisations and individuals that deploy or operate AI. These include, but are not limited to, those who:

- *Develop* AI systems, including through ideation, data gathering, model selection and testing
- *Sell* AI products to end-users
- *Deploy* AI systems in their business
- *Use* AI systems that have been deployed by others.

AI Lifecycle
According to the OECD, the AI system lifecycle involves the following phases which take place iteratively and are not necessarily one after the other: *i)* 'design, data and models' (encompassing planning and design, data collection and processing, as well as model building); *ii)* 'verification and validation;' *iii)* 'deployment;' and *iv)* 'operation and monitoring.' More information is available at oecd.ai/en/accountability.

Agentic AI
A system or program that can autonomously perform tasks on behalf of a user or another system by designing its workflow and using available tools. The system has "agency" (hence the name) to make decisions, take actions, solve complex problems, and interact with external environments beyond the data upon which the system's machine learning (ML) models were trained.

Algorithm
The procedure, set of rules or instructions that is used to solve a problem or otherwise come to an output.

Attack surface
The set of possible points where a malicious actor is able to access a system and extract data.

Audit
A comprehensive assessment of conformance to standards, policies or legal requirements for data gathering, storage and/or usage.

(AI) Bias
Bias allows AI systems to determine how to treat different situations accordingly and is therefore fundamental to its adaptive capacity when minimised and justified (so as to avoid unfairness).

Compute
The large-scale computer resources required for development and use of AI systems.

Confidential and proprietary information
Business data that a business wants to protect from public disclosure. This could include processes, formulas, designs, or other secret or tightly held information.

Cost-benefit analysis
Comparison of the pros and cons (including costs) of a decision or action, to help determine whether it is a valuable and worthwhile activity.

Data annotation
The process of labelling data to support machine learning algorithms to understand and classify it.

Data anonymisation
Modifying personal information in a way that it can no longer be linked to a specific individual, so as to protect privacy while still enabling data analysis and research.

Data augmentation
The process of artificially expanding a dataset, creating new data points through modifying or utilizing the existing data.

Data drift
When the statistical properties of distribution of data changes over time, impacting the performance of the model.

Data encryption
Scrambling data to mask sensitive information from unauthorised users.

Data poisoning
A type of cyberattack in which a malicious actor intentionally compromises a training dataset used by an AI model to influence or manipulate the operation of that model.

Data provenance
Recorded history of data, including its origin, transformations, and movements through various systems.

Deep learning
A subset of machine 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.

Denial of service attacks
A denial of service (DoS) attack aims to overload a website or network, with the aim of degrading its performance or even making it completely inaccessible. More information on DoS attacks and preparing for them is available from https://www.ncsc.govt.nz/protect-your-organisation/guides/preparing-for-and-mitigating-denial-of-service-dos-incidents/.

Explainability
As defined by the OECD, explainability encompasses efforts to enable people affected by AI system outputs and outcomes to understand how they were 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.

Generative AI
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.

(AI) Governance
Governance involves steering the development, deployment, and use of AI technologies throughout their lifecycle, in an organisation or jurisdiction by creating and implementing a range of tools such as voluntary guidelines, policies, rules, and regulations, amongst others.

Intellectual property
New or original innovations and creations of the mind, including but not limited to creative works, inventions, industrial designs, trademarks, and commercial names.

(AI) Literacy
The skills, knowledge and understanding to make informed decisions about AI use and development (as required by any individual).

Machine learning
A type of AI that allows machines to learn from data without being explicitly programmed. It does this by optimising model parameters (i.e. 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 machine learning model to learn and make predictions or decisions.

(AI) Model card
Model cards are short documents accompanying trained machine learning models that provide transparency about various aspects of the model: Claude 3 model card; AWS AI Service Cards; and Llama 3.1 model card.

Large Language Model (LLM)
A very large deep learning GenAI model that is pre-trained on vast amounts of data, allowing it to generate language responses to user inputs.

Personal information
Personal information (as defined by the Office of the Privacy Commissioner) is any information which tells us something about a specific individual. The information does not need to name the individual, as long as they are identifiable in other ways (e.g. through their home address).

(GenAI) Prompt
A specific user input into a GenAI tool to help convey what you want the output to do or be.

Prompt injection
Prompt injection manipulates an LLM's behaviour (e.g. to alter response style, retrieve hidden or restricted data, or disrupt interactions) by embedding specific instructions within a prompt. This approach exploits the model's tendency to follow instructions within the prompt sequence, even if instructions are unintended or malicious.

Red teaming
An organised process of generating malicious model inputs to test the system's reaction and/or ability to produce harmful behaviour as a result.

Retrieval augmented generation
A technique allowing LLMs to access and reference knowledge sources to inform responses as they are being generated, enabling more up-to-date outputs.

Stakeholder
Stakeholders (as defined by the OECD) are 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.

Unfairness
In an AI context, the unjustified differential treatment (which can occur as a result of data and model bias) that preferentially benefits (or disadvantages) certain groups more than others.

(AI) watermarking
Embedding unique identifiers into a GenAI model output to help identify it as being AI-generated, and often to track its origin. These are often only detectable by algorithms rather than visible to human users.

Web scraping

Automatic extraction and organisation of information from websites or online files into a structured format for use.

Appendix Two: Checklist resources

AI Procurement

When deciding on a supplier to use, and developing related agreements, contracts or other documentation with those suppliers, ensure you understand:

- ☐ supplier reputation (including any published infringement claims), capability, legal compliance, pricing, and contract value
- ☐ alignment with your business needs and criteria
- ☐ alignment with the purposes and principles set as part of your AI governance approach (see Understanding your 'why' for AI)
- ☐ supply chain and security risks, including the potential to fail or be misused/attacked and any incident response plans
- ☐ where operational and/or input data is stored, and how it is protected, retained, used, and destroyed
- ☐ model performance results (including around its accuracy and bias) and how it will be monitored and maintained
- ☐ ownership of model outputs and input data
- ☐ supplier terms of service/use, including for example around data governance and intellectual property. Consult industry peers, experts, or prior evaluations of the system to better understand system performance and any risks to consider.
- ☐ cost analysis – AI tools are presenting new commercial models and have different cost drivers to other known software models. Volume may be less relevant than computational chargers, which may not be fully transparent. Users need to be aware of and alert to the total cost of the tool and the impact of high usage, as well as the cost of ongoing support
- ☐ the risks of 'vendor lock-in' and how they will be managed, ensuring you understand if you can change providers, and whether its data is able to be easily migrated with any change.
- ☐ capacity to deploy the system and integrate with existing business systems and tools
- ☐ whether training data meets your business's, and other required, standards, including whether it's relevant to NZ as well as the issues set out in Data & Modelling
- ☐ what support arrangements with the provider would be in place should something go wrong

Deployers or developers of AI systems should in turn offer to share the following information with their suppliers of AI systems, models, or components:

- ☐ expected use of the AI system, model, or component
- ☐ where data privacy is a consideration, as much information as possible to highlight the issue and replicate the outcome without compromising data privacy or security such as data profiles or sample synthetic data
- ☐ issues, faults, and incidents that occur with the system
- ☐ a schedule of regular reviews and updates of this information.

AI transparency for deployers

You should aim to be transparent about:

- ☐ why/how AI is being used.
- ☐ when users or their data are, or will be, interacting with an AI system or AI-generated content
- ☐ when user/AI interaction may influence decisions about an individual or a group (see Human-in-the-loop decision-making.

It is also useful to set out:

- ☐ how an AI system may have been selected or procured, where applicable
- ☐ how training data was obtained or accessed, if possible
- ☐ general information about the AI system, including model performance metrics where possible (for example, around accuracy)
- ☐ how the AI system is being monitored
- ☐ identified risks and how they are being managed
- ☐ company AI policies and broader data use policies and guardrails
- ☐ how to provide feedback, report an issue, make a complaint, or request a correction (see Complaints and Feedback).

Additionally, consider being open and transparent about:

- ☐ what data is being used
- ☐ what measures are being taken to ensure data quality, sovereignty (where appropriate), security, and accountability
- ☐ where data is being stored (i.e. jurisdiction)
- ☐ if/how data can be corrected or erased
- ☐ governance arrangements in place, including if and how Māori have been consulted, and views represented
- ☐ see Data & Modelling for more details.

Recordkeeping

Consistent documentation is key when developing or using AI systems and will be essential for answering user or customer enquiries and/or any future audits or assessments (e.g. that may be required for international operations).

Each AI system should have an organisation-wide inventory which could include:

- ☐ purpose and business goals (Understanding your 'why' for AI)
- ☐ intended use case/s (Understanding your 'why' for AI)
- ☐ accountabilities and roles (Assembling a team)
- ☐ relevant legal obligations and requirements (Compliance and legal obligations, Privacy)
- ☐ identified risks, potential impacts, and relevant mitigations (Risk management)

- ☐ identified key stakeholders (Stakeholder interactions)
- ☐ capabilities and limitations of the AI system, including testing results (Procurement, Model efficacy)
- ☐ mechanisms for human control and oversight (Human-in-the-loop decision-making)
- ☐ technical specifications, requirements and components including training data and architecture (Procurement, Data & Modelling)
- ☐ any system audit requirements and outcomes, including dates of review (Model efficacy)

Appendix Three: Options to ethically source datasets including copyright works

Developers benefit from certainty around training data origins and that it is both legally and ethically sourced. Creators benefit from acknowledgement and remuneration for using their works – which in turn incentivises continued creation and sharing.

Increasingly various and dynamic options are becoming available to businesses to obtain permissions to use third-party proprietary datasets, including copyright works in training, refining, or prompting AI models. The table below outlines some of these.

Note these are examples only and not necessarily endorsed by MBIE or Government more widely.

Options	
Directly license copyright works	AI developers are increasingly striking partnerships with traditional publishers and media entities to license their extensive content libraries, in order to foster innovation and grow together. Consider creating opportunities to partner directly with media libraries, publishers, iwi, and other content creators, rightsholders and aggregators.
Access a collective license	AI developers can access traditional ways to licence use of copyright works through collective licensing schemes offered by various copyright management organisations. Collective licences can also be used to obtain permission to use overseas works, vastly increasing the available volume and variety of copyright works available to AI developers. There are examples of overseas collective licensing options in: • the UK – the Copyright Licensing Agency's GenAI licence permissions, and Text and Data Mining Licensing; and Authors' Licensing and Collecting Society's AI Licences • Australia – the Copyright Agency's extended Annual Business Licence for AI tools • the United States – Copyright Clearance Centre's Collective Licensing Solution for Content Usage in Internal AI Systems, and AI Systems Training License (United States) For New Zealand copyright works and business licence solutions, Copyright Licensing New Zealand intends to release a collective licensing scheme later in 2025 to partner AI developers with New Zealand rightsholders.
Use Fair Marketplaces	New marketplaces are emerging for creators and rightsholders to directly license their creative works for AI training, ensuring permission and remuneration. US examples include *Created by Humans* and *RHEI.*
Choose a fairly trained and commercially safe AI model	Examples of AI models that exemplify trustworthy AI and exclusively use licensed datasets include, but are not limited to: - ***Te Hiku*** – which used ethically sourced archival footage and audio to design an automatic speech recognition model that can transcribe te reo Māori with 92% accuracy. This has been used to run Kaituhi, an automatic bilingual transcription service. - ***Pro Rata AI*** – which enables attribution of contributing content and share revenues on a per-use basis. The process looks at the output of the GenAI content, analyses where the outputted content came from, and then shares half of the revenue with the rightsholders (similar to how content distributors like Spotify or YouTube compensate rightsholders on their platforms). - ***Adobe Firefly Video Model*** - which is trained exclusively on licensed content, and only public domain content where copyright has expired. You can also check the internet for reporting on AI model infringement claims and infringement checking tools are also emerging.

Appendix Four: Other Guidance and resources

Links to external resources are collated here for reader convenience, MBIE does not take responsibility for the content of external links.

NZ Context

- Cabinet agreed in July 2024 to a strategic approach to work on Artificial Intelligence.
- This included endorsement of the Organisation for Economic Cooperation and Development (OECD) AI Principles. These guide AI actors in efforts to develop trustworthy AI and provide policymakers with recommendations for effective AI policies. Countries use the OECD AI Principles and related tools to shape policies and create AI risk frameworks, building a foundation for global interoperability between jurisdictions.
- The New Zealand AI Strategy was released in July 2025.

Understanding how you can use AI tools

- Artificial Intelligence (AI) - YouTube – Business.govt.nz

Māori data

- Māori data and AI – Stats NZ

Governance and risk management

- Laying the groundwork for good governance – Business.govt.nz
- AI Forum NZ - AI Governance
- Algorithm Impact Assessment toolkit – Stats NZ (designed for government agencies but also helpful for business)
- Catalogue of Tools and Metrics for Trustworthy AI - OECD.AI

Privacy

- Information Privacy Principles' applicability to AI – OPC
- Privacy Impact Assessments – OPC

Cybersecurity

- Engaging with Artificial Intelligence – NCSC with international partners
- Guidelines for Secure AI System Development – NCSC with international partners
- Deploying AI Systems Securely – NCSC with international partners
- Unmask Cyber Crime: Business online security series – NCSC
- AI Data Security – NCSC with international partners

Public service

- Responsible AI Guidance for the Public Service: GenAI – GCDO, DIA
- Data toolkit – Stats NZ
- Algorithm Charter for Aotearoa New Zealand – Stats NZ
- Cloud jurisdictional risk guidance – GCDO, DIA

International

- ISO/IEC 42001:2003 Artificial Intelligence Management System
- ISO/IEC 23894:2023 IT-AI Guidance on risk management
- ISO 31000:2018 Risk management guidelines, and other country frameworks listed at Appendix Four
- EU - EU Artificial Intelligence Act
- United States - AI Risk Management Framework | NIST
- Singapore - AI Verify Foundation Resources
- Australia - The 10 guardrails | Voluntary AI Safety Standard | Department of Industry Science and Resources

Intellectual Property

- World Intellectual Property Office - **'***Generative AI: Navigating intellectual property*'
- Copyright Licensing Agency – *'The GAI Revolution: How professionals are using generative AI in the workplace and the copyright implications this creates'*

Industry specific guidelines and information:

- New Zealand Writers Guild – '*AI Advice: The NZWG Handbook for Screenwriters*'
- New Zealand Film Commission – '*Artificial Intelligence (AI) Guiding Principles*'
- Law Society – *'Generative AI guidance for lawyers'*
- Courts of NZ – *'Generative AI in Courts and Tribunals'*
- Institute of Directors – *'Governing AI'*
- Financial Markets Authority –'*Understanding Artificial Intelligence in Financial Services'*
- Royal Society – '*Guidelines for use of GenAI in research in Aotearoa New Zealand*'
- AI Forum - *Architecture, Engineering and Construction (AEC) Working Group*
- Engineering *NZ - Engineering and AI*

If your industry has issued guidance regarding use or development of AI please contact us at digitalpolicyteam@mbie.govt.nz.



New Zealand Government
Te Kāwanatanga o Aotearoa

AOG 11770