Ministry for Regulation
Te Manatū Waeture

AI Guidance

Responsible AI in Action

Practical steps for Regulators leading AI innovation with confidence

May 2026

Guidance Owner: Ministry for Regulation

Version: 1.0

Classification: Public

Published in May 2026 by the Ministry for Regulation, Wellington, New Zealand. ISBN 978-1-991372-13-0 (Online)

This document is available on the Ministry for Regulation website: **regulation.govt.nz** Crown copyright © 2026

This work is licensed under the Creative Commons Attribution 4.0 International licence. You are free to copy, distribute, and adapt the work, as long as you attribute the work to the Crown and comply with the licence terms. To view a copy of this licence, visit **https://creativecommons.org/licenses/by/4.0/**

Please note that no departmental or governmental emblem, logo or Coat of Arms may be used in any way which infringes any provision of the Flags, Emblems, and Names Protection Act 1981. Attribution to the Crown should be in written form and not by reproduction of any such emblem, logo or Coat of Arms.

Acknowledgements

We'd like to thank the following organisations for sharing their knowledge. Allen + Clarke, Datacom, Hutt City Council, Department of Conservation, MartinJenkins, Objective Corporation, University of Queensland, The Government Digital Delivery Agency, and Victoria University of Wellington.

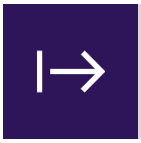
The Ministry for Regulation acknowledges the contribution of LensenMcGavin AI, whose independent peer review of this guidance helped strengthen its technical grounding, clarity, and practical applicability.

AI Roundtable

In November 2025 the Ministry for Regulation and Department of Internal Affairs held an AI roundtable with regulatory leaders on how AI can impact regulatory practice. We thank the attendees for their contribution.

What this guidance covers – at a glance

Understanding AI in a regulatory context
A simple, plain English explanation of what AI is – and isn't – and how it fits within regulatory systems.



Starting safe: low-risk ways to begin
How regulators can experiment with AI using public information, consumer tools, and reversable pilots before scaling.

Building capability & confidence
The skills, literacy, and organisational conditions needed for responsible AI use across regulatory functions.

Working with AI vendors safely & smartly
Practical guidance on engaging vendors, protecting independence, and avoiding lock-in.

Human-in-the-loop (HITL)
When and why humans must stay in the decision cycle – especially for high-stakes or rights affecting decisions.

Iterative unleashing: start small, learn fast
A practical approach to testing, refining, and scaling AI responsibly in regulatory settings.

Ethical use of AI in regulation
Core principles, obligations, and expectations drawn from NZ public sector frameworks – with a practical ethical checklist.

Case examples & tools
Real-world use-cases, lightweight prototypes and scenarios to help leaders and practitioners see what's possible.

Contents

Why this? Why now?

Who is this guidance for?

This guidance is for senior leaders and management teams in regulatory organisations who are responsible for regulatory systems, statutory decision-making, and regulatory performance.

It complements existing government-wide frameworks and guidance on AI, including the Public Service AI Framework, the Government Chief Digital Officer's Responsible AI Guidance for the Public Service, the Algorithm Charter for Aotearoa New Zealand, and the Privacy Act 2020. This guidance focuses specifically on the regulatory context, where decisions affect rights, obligations, and public confidence. Regulators exercise the state's power to compel and make decisions that have real consequences for people and organisations. That creates distinct considerations for how AI is used.

AI is advancing quickly, and some regulators are already experimenting with it. The challenge for regulatory leaders is knowing where AI can genuinely support regulatory practice, how to apply it responsibly, and how to maintain accountability, transparency, and human judgement. This guidance supports regulatory leaders to make confident, informed choices about AI use within their organisations.

Why AI matters for regulation

AI matters for regulation because it can help regulators act earlier, target effort where it matters most, and make better use of limited resources. Used well, AI can support regulatory practice by helping regulators to:

- detect patterns of non-compliance and emerging risk earlier
- prioritise monitoring, inspections, and interventions based on risk
- support decision-making with timely, evidence-based insights
- identify patterns and anomalies across large datasets that may not be visible through manual analysis
- reduce administrative burden for both regulators and regulated parties.

Understanding AI in a regulatory context

Not all AI is the same, and the differences matter for regulators.

The two types most relevant to regulatory work are **predictive AI** and **generative AI**.

Predictive AI learns patterns from existing data to classify, score, or forecast outcomes. A system that flags high-risk licence applications based on historical compliance data is an example. These systems can be powerful, but they inherit the biases of the data they were trained on. If past decisions were inconsistent or unfair, a predictive AI system will tend to repeat and amplify those patterns.

Generative AI produces new content, including text, summaries, draft documents, and responses to questions. Large Language Models (LLMs) like those behind tools such as Copilot and ChatGPT are generative AI. They are well suited to tasks like summarising information, drafting correspondence, or identifying themes across large volumes of text. However, they can also produce outputs that sound confident but are factually wrong. This is known as hallucination, where the system generates plausible-sounding content that has no basis in fact. Without human review, this risk can go undetected.

Both types of AI share an important characteristic: they learn from data, and their outputs reflect the quality, completeness, and representativeness of that data. Incomplete or biased data produces unreliable outputs.

Regulatory practice depends on something AI cannot provide: careful judgement, legal interpretation, discretion, and decisions that affect people's rights and public trust. AI can do powerful work, analysing at scale, drafting at speed, and surfacing patterns people would miss. The judgement, legal interpretation, and accountability that regulatory decisions demand stay with people.

Additional resources and supporting AI guidance:

- Responsible AI Guidance for the Public Service: GenAI
- Public Service AI Framework | NZ Digital government
- Generative AI guidance for lawyers
- Government algorithm transparency and accountability
- Strengthening the rule of law in Aotearoa New Zealand

Information on privacy principles from the Office of the Privacy Commissioner:

- Artificial Intelligence and the Information Privacy Principles

Resources on Māori data and Māori data sovereignty:

- Te Mana Raraunga | Māori Data Sovereignty Network
- Māori Data and AI – guidance for business
- Co-designing Māori data governance

Section one: set the foundations

If your regulatory foundations are strong, AI can help amplify good systems and practices by reducing administrative effort, improving targeting, and supporting faster, better-informed decisions. If those foundations are weak, AI will not fix them. It is more likely to expose and amplify existing problems.

By "foundations", we mean the basics of good regulatory practice: clear delegations, sound record-keeping, reliable information management, consistent decision reasoning, and effective oversight. Where these are unclear or inconsistent, introducing AI increases risk. Decisions become harder to explain, errors easier to scale, and accountability harder to trace. AI is also highly dependent on the quality of underlying data. If inspection, licensing, or compliance datasets are incomplete, inconsistent, or poorly structured, AI systems will tend to reinforce those weaknesses rather than produce reliable insights.

AI also introduces new categories of risk. Bias can be hidden inside a model's training data and invisible in its outputs. Decisions can drift toward automation without anyone making a conscious choice to allow it. AI outputs can be difficult to explain to the people they affect, creating real challenges for transparency and appeal rights. None of this transfers responsibility away from your organisation. Accountability remains with the legally authorised decision-makers, not with the tool and not with the vendor. Getting the governance arrangements right from the start makes this easier to uphold in practice.

Regulatory leaders will want to be confident they understand what AI is doing in their systems, and why, before placing significant weight on its outputs.

This is why AI adoption should not be treated as an IT project

A new case management system changes how people do their work. AI can change what work people do, how decisions get made, and where accountability sits. Getting genuine organisational buy-in is what makes the difference. Without it, AI tools tend to be adopted at the margins, used by some and avoided by others, without delivering consistent value or being properly governed.

AI adoption needs to be aligned to regulatory priorities, supported by clear governance, and backed by real investment in people. Regulatory leaders will need to make considered decisions about where to direct limited resources, whether that is maintaining business-as-usual capability, improving existing systems, or investing in AI that could meaningfully

change how work gets done. Those decisions are easier to make, and easier to defend, when the whole organisation understands the direction and why it matters.

The practical guidance that follows covers strategy and governance, vendor engagement, workforce readiness, and ethical use. Getting these settings right early makes it easier to scale AI safely, and harder for risks to become embedded.

1.1 Start strong: strategy, governance, and guardrails

Regulatory leaders need to be clear on the problem AI is solving, how it supports system outcomes, and what guardrails must stay in place to protect trust and accountability. AI adoption works best when it has genuine organisational buy-in and is treated as a strategic regulatory capability, not a standalone IT project.

Start by setting three foundations:

1. Align AI to purpose and priorities

AI initiatives should link directly to your regulatory priorities and be reflected in strategic planning and performance frameworks. This keeps adoption grounded in outcomes, rather than driven by novelty or vendor promises. Before committing to an AI initiative, two questions are worth working through:

- Is this a proportionate investment? What is the likely benefit relative to the cost, effort, and maturity of your organisation's current capability?
- What is the risk profile? What happens if the AI system performs poorly or produces unreliable outputs, and how does that compare to the risks of your current approach?

For some use cases, a simpler technical solution or an improvement to existing processes will be the more sensible choice. AI should be adopted where it genuinely serves regulatory purpose.

2. Put governance and decision rights in place early

Every AI-assisted regulatory activity should have clear ownership, decision rights, and escalation pathways. In regulatory settings it is also important to design for meaningful human oversight in decision-making, and to ensure people can request review, including an appeal to a human where AI has influenced an outcome.

3. Design for learning and adaptability

AI systems will change, and so will the risks. Regulators should take an iterative approach, starting small, getting comfortable with AI, and then refining based on evidence and feedback. Build in flexibility so AI systems can be adjusted, paused, or withdrawn if risks emerge or benefits do not materialise. Regulatory leaders may also choose to pilot AI systems through regulatory sandboxing or other controlled pilots to learn quickly, test safeguards,

and refine approaches before scaling. See the **MfR guidance on regulatory sandboxes** for more detail.

AI use needs to be integrated into how regulatory work is designed and delivered, not developed in isolation. This means aligning with legal obligations and working across organisational functions including policy, operations, data, privacy, procurement, and digital. Involving these perspectives early helps ensure risks are understood, responsibilities are clear, and systems remain accountable in practice.

Many regulators benefit from establishing an AI governance group to:

- set safeguards so decisions remain traceable and explainable
- oversee risk assessments and assurance activity
- monitor staff capability and confidence
- review performance and impacts over time.

Build in regular review points to check whether AI systems remain fit for purpose, aligned with delegations and legal obligations, and continue to deliver value without undermining fairness or trust.

This includes reviewing the human-in-the-loop arrangements themselves. Having a human in the process is necessary, but it is not sufficient on its own. Over time, reviewers can become over-familiar with AI outputs and begin accepting them without adequate scrutiny, a pattern known as automation bias. Regular review should look not just at what the AI is producing, but at how human reviewers are engaging with it. Designing oversight processes that actively guard against automation bias is as important as the technical performance of the AI system itself.

1.2 Procure with purpose: AI vendor engagement

Implementing AI is not just a technology investment. It affects decision-making, trust, privacy, and accountability. Getting procurement right helps ensure AI systems are safe, explainable, and fit for a regulatory setting.

When engaging vendors, be clear about the regulatory problem you are solving, what decisions the tool will support, and what guardrails must remain in place. Vendors should be able to demonstrate how their tool manages data, reduces bias, supports transparency, and can be monitored over time.

Practical steps for engaging AI vendors

1. Start with the basics

Follow the NZ Government Procurement Rules and procurement principles of integrity, transparency, and accountability. Before procuring new tools, check whether suitable All-of-Government AI capabilities are already available through the Marketplace or other shared public service platforms.

2. Apply public service AI guidance

Apply relevant government guidance to test whether the tool is responsible and suitable for public sector use, including fairness, explainability, and privacy expectations.

3. Do vendor due diligence

With the support of your digital services and procurement teams, ask for evidence of real deployments in similar contexts. Vendors may not be able to share everything, as access to a model's internal workings is rarely granted, but the focus should be on what the system actually produces in practice and whether that is good enough for your context.

As a starting point, confirm:

- data storage, access arrangements, and data sovereignty, including where data is held and who can access it
- intellectual property and information management settings
- how the model is tested, updated, and monitored, including what testing and benchmarking data has been used
- the provenance of the data the model was trained on, where this is relevant to your use case
- what contractual commitments the vendor will make regarding bias levels and ongoing performance.

The level of scrutiny should reflect the risk. For lower-risk applications, standard due diligence may be sufficient. Where AI will inform decisions that affect people's rights or entitlements, a higher standard of interpretability should be required. This means the vendor should be able to explain, in terms your decision-makers can understand, why the system produces the outputs it does. If a vendor cannot provide adequate assurance for a higher-risk application, that is important information. It may mean the product is not suitable for that use case, even if it works well in other settings.

4. Identify and manage key risks early

Common risks include bias in model outputs, security vulnerabilities, and vendor lock-in. However, the risks associated with any particular AI system will depend heavily on the regulatory context it is being used in. A generic risk checklist is unlikely to be sufficient.

Procurement and digital teams should work closely with regulatory staff to develop a full risk profile for each system under consideration. Regulatory staff understand the decisions the AI will be informing, the people those decisions affect, and what good and poor outcomes look like in practice. That operational knowledge is essential for identifying risks that may not be visible from a purely technical or commercial perspective. Once risks have been identified, require vendors to explain how they will be mitigated and reviewed over time. Risk management should not end at procurement.

5. Lock in accountability through the contract

Include clear clauses on:

- performance measures and audit rights
- security standards
- ongoing monitoring and reporting
- lifecycle management, including updates and decommissioning
- certainty on where data is stored
- data portability and access to outputs generated using agency data
- knowledge transfer if the vendor relationship ends
- exit and transition plans to avoid lock-in
- cost controls, including protections against significant price increases over the contract term.

That last point warrants specific attention. The pricing of AI models, particularly those at the frontier of capability, can change considerably over time. An AI system that is affordable at the point of procurement may become significantly more expensive as vendor costs change. Building price certainty into contracts, or retaining the ability to exit if costs become unsustainable, is prudent risk management.

These arrangements support the Digital Government Target State and whole-of-government digital capability initiatives led by the Government Chief Digital Officer.

Procurement References:

Government Procurement Rules	***AI Procurement Guide – AI Forum NZ***	***MBIE AI Procurement Checklist***

1.3 Getting people ready for AI

AI can change workflows, roles, and how decisions get made. Without genuine organisational buy-in, even well-designed AI systems can fail to deliver value, or create new problems rather than solving existing ones.

Getting people involved early matters. Consulting staff before significant adoption decisions are made helps identify practical concerns that leaders may not have anticipated. It can surface sticking points in existing workflows, highlight where proposed AI use sits uncomfortably with regulatory obligations, and occasionally reveal that the problem being solved does not actually require AI at all. Finding these things early is considerably better than finding them after implementation.

Engaging staff early also builds the confidence and shared understanding needed for AI to be used well. People who understand why a tool is being introduced, what it will and will not do, and how their role fits within the new process are better placed to exercise the judgement that keeps AI-supported decisions sound. The goal is to make a genuinely informed decision about whether, and how, AI can improve regulatory practice, with the people who will be closest to it.

Why organisational buy-in matters

1. Cultural shift

AI introduces new ways of working, including supervising AI-supported processes and actively questioning outputs rather than accepting them at face value. The foundation for this is AI literacy. Staff do not need to be technical experts, but they do need a working understanding of how AI systems produce their outputs, why those outputs can sometimes be wrong, and what the limitations of the tools they are using are. Without this, human review can become superficial. People tend to defer to outputs that look authoritative, particularly under time pressure, and that deference is precisely what good oversight is meant to prevent.

Research consistently shows that organisations underestimate the groundwork required to bring staff along, and that return on AI investment is very hard to achieve without it. Building AI literacy is an ongoing part of maintaining the critical thinking that keeps AI-supported decisions sound.

2. Role redefinition

AI may change the balance of some roles over time, potentially reducing time spent on certain manual tasks while increasing the importance of oversight, assessment, and interpretation. How significant that shift will be, and how quickly it happens, will depend on the tools adopted and the regulatory context they are used in. Clear communication about what is changing, and what is not, helps reduce uncertainty and keeps teams engaged. Role-specific support is more useful than general reassurance.

3. Capability building

Digital capability across the public sector is still developing. The 2025 Public Service Census shows that while most staff feel confident learning new digital skills (88%), AI use remains low in practice. Only 12% of public servants use AI daily or weekly, and 67% have never tried using AI for work at all.

AI literacy is not simply a new digital skill to add to an existing training programme. Understanding how AI systems work, why they produce the outputs they do, and how to critically evaluate those outputs requires a different kind of learning, one that is as much about judgement and critical thinking as it is about using a tool. Effective capability building is practical, targeted to specific roles and use cases, and built into day-to-day work.

4. Ethical and governance alignment

Principles such as human-in-the-loop oversight, transparency, and the right to appeal to a human decision-maker only work when people understand what they mean in practice. Embedding these expectations consistently across teams and decision points is an important part of getting people ready for AI. Human-in-the-loop is covered in more detail in section 2.5.

5. Stakeholder confidence

Regulators operate under public scrutiny. A structured approach to AI adoption supports confidence by demonstrating that decisions remain accountable, reviewable, and aligned with the public interest.

What regulatory leaders can do

Sponsor visibly. Staff take their cues from leadership. If senior leaders are not actively engaged with AI adoption, governance arrangements tend to weaken and accountability becomes unclear. Assign clear ownership, stay close to how AI is being used in practice, and keep reinforcing that safe and accountable use matters.

Explain the "why". Before introducing any AI tool, be clear about what problem it is solving and what will change for staff and regulated parties. If you cannot articulate a clear answer, that is a signal to pause. People engage more constructively when they understand the purpose, and are more likely to raise concerns early when they feel part of the process.

Build role-specific learning. Generic AI training has limited value. Focus learning on the specific tools staff will use, the tasks those tools will support, and how to critically evaluate outputs in that context. Include practical guidance on known limitations and what to do when something does not look right.

Pilot and iterate. Start with low-risk use cases where outputs are easy to check and errors are unlikely to cause harm. Be honest about what you are learning. Scale only when the value is demonstrated and the safeguards are working. Ending a pilot that is not delivering is the process working as intended.

Create feedback loops. Build in regular opportunities for staff and affected groups to raise concerns and share what they are observing. Act on what you hear. If risks or issues emerge, adjust quickly rather than waiting for a formal review cycle.

Section One: Summary

- AI amplifies what is already there. Strong foundations scale well; weak ones are exposed.
- Good foundations mean clear delegations, good information management, consistent decisions, and effective oversight.
- Accountability never shifts to AI or vendors. It stays with authorised decision-makers.
- AI must be treated as a regulatory capability, not an IT project.
- Success depends on early alignment to purpose, strong governance, careful procurement, and workforce readiness.

Section two: use AI well (and keep humans in control)

People who do regulatory work need the confidence, curiosity, and practical know-how to test, oversee, and apply AI systems safely within their authorising environment.

2.1 Emerging use, limited impact

Generative AI tools like ChatGPT and Copilot are available across much of the public sector, but actual use remains limited. The 2025 Public Service Census found that only 12% of public servants use AI daily or weekly, and 67% have never tried it for work at all. Where AI is being used, it tends to support individual productivity rather than changing how regulatory work is delivered.

This is a predictable starting point. Most organisations are still in an early stage of understanding what AI can and cannot do in their specific context. The gap between having access to AI tools and using them to meaningfully improve regulatory outcomes is real, and closing it requires more than making tools available.

Research on AI adoption consistently shows that realising value from generative AI requires significant investment in people, culture, and workflow redesign. MIT's Project NANDA report, The GenAI Divide: State of AI in Business 2025, suggests that despite significant global investment, 95% of organisations studied saw no measurable return from AI initiatives. The core issue was not the technology itself, but the absence of deliberate integration, workflow redesign, and organisational capability to support it. Without deliberate leadership, targeted capability building, and a clear sense of where AI genuinely supports regulatory purpose, tools tend to be used occasionally and at the margins.

The practical steps below are intended to help regulatory leaders move beyond early experimentation toward more deliberate and purposeful use of AI.

1. Treat early adoption as the starting point

Widespread use of tools like Copilot builds familiarity, and that is a reasonable place to begin. But familiarity with a tool is not the same as understanding it well enough to use it safely in a regulatory context. The next step is developing a working understanding of how tools actually produce their outputs, where they are reliable, and where they are not. This technical

grounding does not need to be deep, but it needs to be sufficient for staff to exercise genuine judgement when reviewing AI outputs rather than simply accepting them. From that foundation, regulatory leaders can start to identify where AI fits within specific regulatory functions and where small, safe trials could add value, without moving faster than the organisation's capability to oversee what it is doing.

2. Create an authorising environment for safe experimentation

AI is unlikely to move beyond personal productivity without visible leadership support. Staff need to know that experimentation is legitimate, and that trying something that does not work out will not reflect badly on them. But freedom to experiment needs to sit within a clear framework. Leaders need to be confident about what kinds of AI use are appropriate in their regulatory context, and what safeguards are in place. Encouraging experimentation without that foundation risks AI being used in ways that create problems for accountability and public trust. The goal is an environment where staff feel confident to try AI in low-risk settings and share what they learn, within boundaries that leaders have thought through.

3. Look for regulatory use cases

Ask practical questions like:

- Where are our backlogs, delays, or bottlenecks?
- Where do decision-makers need better information sooner?
- Which high-volume tasks slow down core work?

These questions help identify where AI might add value. But before settling on an AI solution, it is worth asking whether AI is actually the right tool. Some problems are better solved through process improvement or simpler automation, which carry less risk and are easier to govern. AI should be considered where it offers a genuine advantage over these alternatives.

4. Start small and let progress compound

Given that most of the public sector is still in the early stages of AI use, starting small is the right approach. Small, low-risk pilots build the organisational confidence, capability, and practical understanding needed to use AI well at any scale. This kind of progress requires investment in people alongside investment in tools. Training, communication, and genuine leadership engagement are what turn early experimentation into embedded capability. Each successful small-scale trial creates a foundation for the next one.

5. Build confidence before considering bigger changes

The early goal is building trust that:

- humans remain the decision-makers
- the risks are understood and manageable
- the process remains lawful, fair, and reviewable.

Scaling AI use before this confidence is established creates governance risks that are difficult to unwind.

6. Learn from AI adoption across the regulatory system

Early AI experiments often stay local to teams or agencies. Progress accelerates when regulators share what they are testing, what worked, and what did not. Peer networks and communities of practice help build collective capability, reduce duplication, and develop a shared understanding of good practice. This learning should feed back into formal guidance and frameworks. Lessons are most valuable when they are captured, shared systematically, and used to update the guidance that shapes how AI is adopted more broadly. The Ministry for Regulation will support this by updating guidance as practice develops.

2.2 Examples of AI in regulatory practice

The following examples show how some regulators are starting to apply AI in practical ways. They demonstrate how small experiments can build capability, generate learning, and point to where AI may improve regulatory practice.

Case Study: AI in permissions & concessions

Organisation: Department of Conservation (DoC)

Function: Licensing, approvals, and permissions (statutory decisions on activities on public conservation land)

AI type: AI decision-support assistant (statutory document search and navigation)

Context:

DoC manages a large and complex set of statutory and policy documents that underpin permissions, concessions, and operational decision-making. Many of these documents are hundreds of pages long and often exist as scanned PDFs, making them difficult to search or interpret quickly. For permissions advisors, manually reviewing and applying these requirements to concession applications was slow and resource-intensive, contributing to delays and backlogs. This administrative load also pulled skilled staff away from higher-value regulatory judgement and DoC's core mission of protecting Aotearoa New Zealand's natural environment.

Action taken:

DoC implemented an AI assistant (built by MakerTech) to help permissions advisors find and interpret relevant statutory and policy requirements more efficiently. The tool allows staff to ask questions in plain language and receive grounded answers with citations back to official documents. The tool provides decision support only; permissions advisors and delegated decision-makers remain accountable for the final statutory decision.

Results:

Early feedback from staff suggests the tool is helping reduce the time spent searching through lengthy statutory and policy documents. Permissions advisors report that it can support quicker navigation of complex requirements and help surface relevant information more efficiently. Staff have also noted that the tool can help newer permissions advisors close experience and knowledge gaps, supporting confidence when working through complex applications. The intention is that tools like this allow staff to spend less time navigating large volumes of documentation and more time focusing on the more complex and nuanced aspects of concession applications. Initial uptake has been steady. Of a potential user group of around 80–100 staff, approximately 30 are repeat users. The tool went live in June 2025 and has since been operationalised. It also supports a broader shift in approach from managing application backlogs by adding people to changing how work is done, enabled by strong executive backing and a more deliberate risk appetite for innovation.

Case Study: AI Volution - decision support

Organisation: Hutt City Council (HCC)

Function: Compliance triage, licensing/enforcement support, and operational safety monitoring.

AI type: AI decision support assistants (summarisation/triage/document navigation) and targeted computer vision pilots for real time safety signals (HITL).

Context:

HCC's regulatory teams handle high volumes of documents, emails, and policies across compliance, licensing, and enforcement workflows. Manually triaging, summarising, and extracting requirements from disparate sources slowed turnaround times and created variation in practice, stretching capability; particularly for newer staff. In parallel, site safety oversight on large infrastructure projects relied on human observation and after the fact reviews, limiting real time prevention and response.

Action taken:

HCC launched the "AI Volution" programme (late 2024) to embed AI as decision support, not decision making, under explicit HITL oversight and structured governance. Early waves focused on low risk, high volume tasks - summarising submissions and correspondence, document organisation, and policy lookup - to speed triage while keeping statutory decision-making with authorised officers. Leadership sponsorship from the Chief Executive Jo Miller and an AI governance group enabled a safe but ambitious scale up; a short, structured trial phase (with vendor Price Waterhouse Cooper) provided quick wins and built confidence. In parallel, HCC tested computer vision safety pilots with partners (e.g., RUSH/Te Ara Tupua Alliance) for real time detection of hazards (falls, PPE non-use, people in no-go zones), again with human oversight.

Results:

Faster triage and document handling: GenAI assistants reduce manual effort for summarisation and first-pass analysis, helping teams route cases and focus attention sooner on higher risk matters.

Capability uplift and confidence: the programme deliberately closes experience gaps by giving staff (including newer advisors) a guided way to find policy and precedent, while training leaders and teams to use AI safely.

Measured productivity benefits: HCC set (and partially realised) a working assumption of ~30 minutes per person per day in time savings, tracked via usage and surveys; gains were also achieved by not requiring back-fill for roles.

Human-led safeguards embedded: Human-in-the-Loop (HITL) guardrails and an AI governance group ensure AI is an input only; authorised officers remain accountable for statutory decisions and can override or roll back at any time.

Real time safety signals: Computer vision pilots provide onsite alerts (e.g., falls detected, PPE compliance, people in restricted zones), improving situational awareness for humans in the loop.

Sector leadership and transparency: HCC openly shares frameworks and lessons with Other councils and agencies, positioning itself as a practical exemplar for public sector AI

2.3 AI Literacy: safe and purposeful use

Regulators who develop genuine AI literacy are better placed to identify appropriate use cases, ask the right questions of vendors, recognise when outputs should not be trusted, and maintain accountability for AI-supported decisions.

AI literacy requires some level of technical understanding. Not deep expertise, but enough to understand how AI models work in practical terms, why they produce the outputs they do, and where their limitations lie. The 2025 Public Service Census found that 67% of public servants have never tried AI for work, and only 12% use it regularly. For most regulatory organisations, developing this foundation is where the work begins.

In a regulatory context, AI literacy means understanding:

- how AI models are trained and why that affects the reliability and fairness of their outputs
- what AI can and cannot do, and where human judgement must lead
- how to question outputs for bias, accuracy, and relevance
- how to recognise hallucination and other common failure modes
- how human oversight and the right to appeal to a human decision-maker are preserved
- how AI fits within risk-based and responsive regulation.

Where this understanding is absent, practical risks follow. Outputs get over-trusted or dismissed without good reason. Vendors fill the gap left by weak internal judgement. Progress stalls.

AI literacy is an ongoing organisational capability. As tools develop and regulatory use cases expand, the understanding required to use them well needs to keep pace.

2.4 AI as a support for regulatory thinking

There is a risk in using AI that is easy to overlook. The more people rely on AI to do their thinking, the less they exercise their own judgement. Research in cognitive science, including the 2025 MIT study Your Brain on ChatGPT, suggests that offloading reasoning to intelligent systems over time can erode the analytical capability, recall, and sound judgement that good decision-making depends on. The tools get better while the people using them become less confident in their own assessments.

This matters acutely in a regulatory context. Regulation is not mechanical. It requires weighing evidence, applying legal judgement, exercising discretion, and being accountable for decisions that affect people's rights and obligations. These are capabilities that need to be actively maintained.

The consequences of getting this wrong are not always immediately visible. When AI outputs are accepted without adequate challenge, institutional knowledge quietly erodes, domain expertise gets substituted by outputs that sound confident but may be incomplete or wrong, and human reasoning becomes reactive rather than deliberate. This is sometimes called automation bias, the tendency to defer to machine outputs without applying independent judgement. It is a well-documented risk in any setting where people work alongside intelligent systems. For regulators, where public trust depends on confidence that decisions are made by people who understand what they are deciding and why, automation bias carries real consequences for organisational credibility and social licence to operate.

AI should be making regulatory thinking sharper, not substituting for it. Where regulators cannot explain how an AI output informed a decision, or why it was accepted or rejected, that is a signal that the balance has shifted in the wrong direction.

Leadership implications

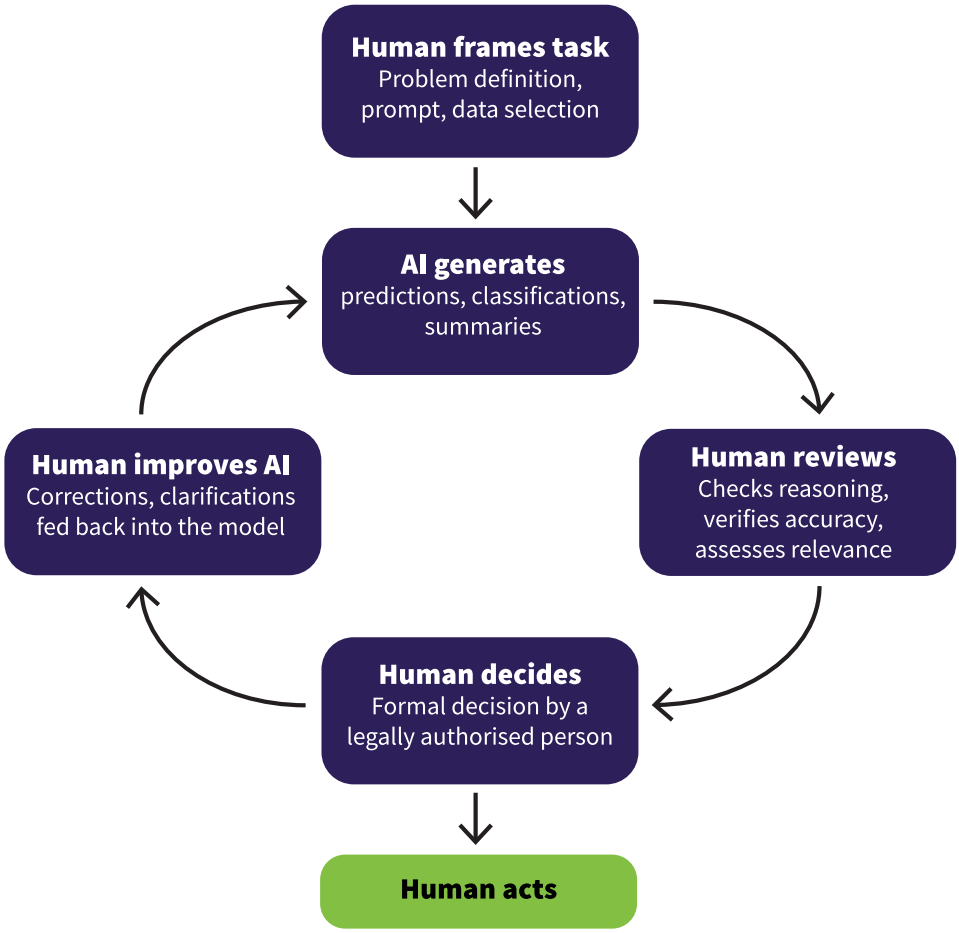
Regulatory leaders set the tone for how AI is used across their organisation. Being clear that understanding the law, applying discretion, and making decisions remain human responsibilities is important, but the more powerful signal comes from behaviour. Leaders who visibly interrogate AI outputs, treat AI suggestions as inputs rather than answers, and ask how decisions were reached rather than just what was decided, create the conditions for AI to be used well throughout the organisation.

Where AI has influenced a decision affecting a member of the public, that person has a right to seek information about how that decision was reached. This is an existing obligation under the Official Information Act 1982, and it has direct implications for how AI-supported processes are designed and documented.

2.5 Human-in-the-Loop: keeping decision-making human

Human-in-the-Loop (HITL) means ensuring people remain actively involved in decisions that are supported by AI. In a regulatory context this is critical. Decisions that affect rights, obligations, or public trust must be made by authorised decision-makers, not automated systems.

Example of decision-making in a human-in-the-loop process



AI can help by scanning large volumes of information, identifying patterns, or highlighting risks. It does this by recognising statistical patterns in the data it was trained on, not by understanding context, applying discretion, or weighing legal and ethical considerations. That distinction matters. An AI system can produce an output that looks authoritative and well-reasoned while being based on patterns that do not apply to the situation at hand. Practitioners who do not understand how AI produces its outputs are poorly placed to recognise when this is happening.

It is also worth noting that humans shape AI long before it produces an output. People choose the tools, define the problem, design governance settings, select data, and interpret results. Human judgement influences AI at every stage, not just at the point of review.

HITL is therefore not simply about having a human present in the process. It requires practitioners who understand what the AI system is doing well enough to genuinely question its outputs, identify where those outputs should not be trusted, and make decisions they can explain and defend independently of what the AI suggested.

HITL is not a complete safeguard on its own. As noted in section 2.4, automation bias can undermine human review even when it is formally in place. Regulators should design HITL processes that actively require reviewers to interrogate outputs rather than simply confirm them. Oversight mechanisms should periodically review how human reviewers are engaging with AI, not just how the AI system itself is performing.

HITL matters most at points of consequence. This includes licensing decisions, enforcement action, compliance assessments, and any decision that may be reviewed, appealed, or challenged. AI is most appropriate in lower-risk regulatory functions where statutory decision-making or judgement about a person's or entity's suitability is not required. Examples include:

- triaging applications based on risk indicators
- checking applications for completeness and missing information
- validating data against known rules or thresholds
- flagging inconsistencies for human review.

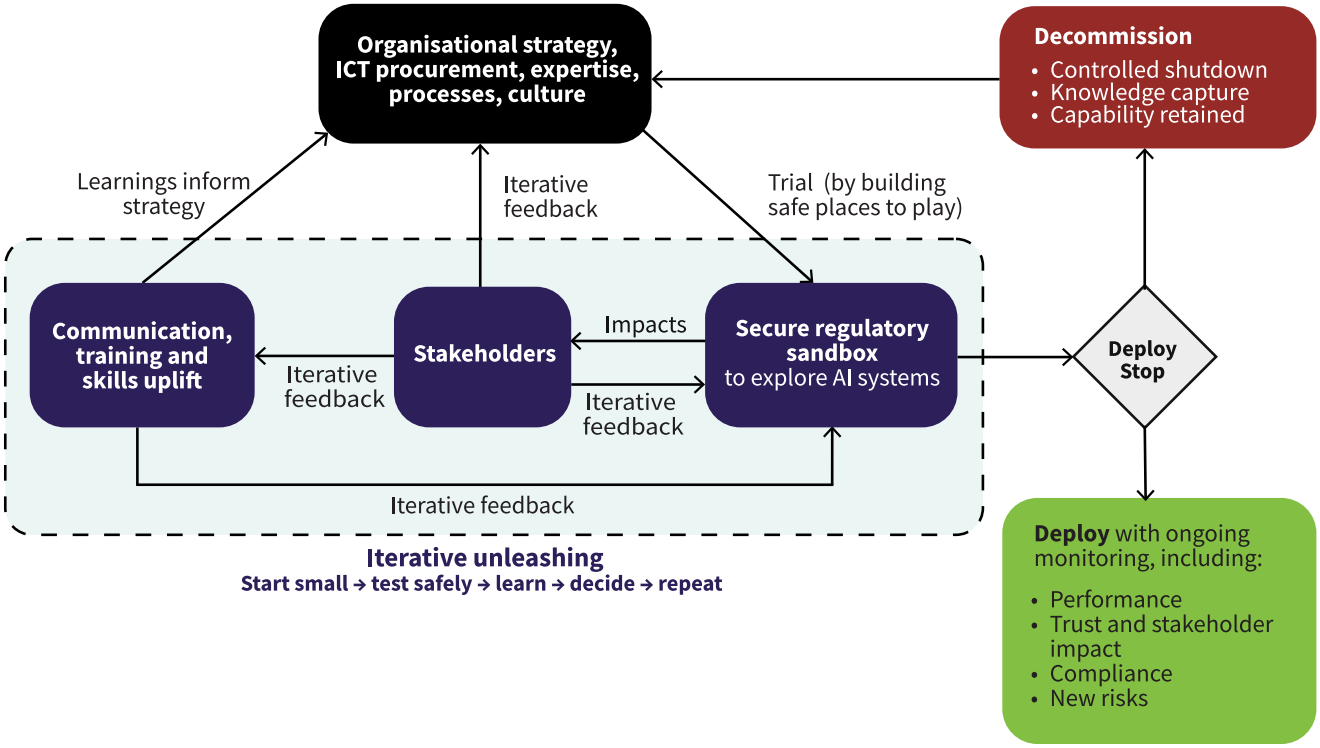
Where AI has influenced a decision affecting a member of the public, that person has a right under section 23 of the Official Information Act 1982 to seek information about how that decision was reached. HITL processes should be designed with this obligation in mind. The role AI played in any decision must be clearly documented and capable of being explained. Accountability remains with the organisation and the legally authorised decision-maker.

2.6 Iterative unleashing: learning safely by doing

Across conversations with regulators, academics, and system stewards, one idea comes up consistently: capability does not grow from designing the perfect plan. It grows from doing small things well, learning from them, and building from there.

Adjunct Professor Dr S. Kate Conroy (QUT Centre for Robotics, Queensland University of Technology) describes this approach as iterative unleashing. It means starting with contained, low-risk experiments, watching closely what happens, and keeping the ability to pause, adjust, or stop if something does not work as expected. This is a practical method for identifying where AI genuinely adds value, what safeguards are needed, and how to build the organisational confidence to extend use responsibly over time.

Feedback loops involved in iterative unleashing



This diagram was originally presented by Adjunct Professor Dr S. Kate Conroy to the AI roundtable with regulatory leaders.

Regulators routinely pilot, test, evaluate, and adapt. Bringing AI into regulatory practice should follow the same pattern. Starting in lower-risk areas, where outputs are easy to check and errors are unlikely to cause harm, allows organisations to build real experience with AI before extending its use to more consequential functions. This approach is consistent with the AI system lifecycle described in the Public Service AI Framework, which emphasises ongoing testing, monitoring, and improvement rather than one-off deployment.

Why this matters for regulators

Working iteratively helps organisations separate genuine usefulness from hype. It gives staff direct experience of AI's strengths, such as speed, pattern recognition, and triage, alongside its weaknesses, including hallucination, bias, and lack of contextual understanding. Early pilots build the staff confidence and organisational knowledge that are prerequisites for using AI well in higher-stakes settings.

How iterative unleashing works in practice

A regulator starts with a narrow, low-risk problem where AI might help. Examples include organising large volumes of public information or identifying themes in submissions or

complaints. These are tasks where outputs can be checked easily, errors are unlikely to cause harm, and the work would otherwise consume significant staff time. It is worth considering the social licence implications of using AI for tasks where the public may expect human review, particularly in higher-risk regulatory settings. Pilots work best when they are designed to generate learning, not just demonstrate success. That means running them in controlled conditions, watching how the tool behaves in practice, and being honest about what the results show. A pilot that reveals problems with underlying data quality, or that staff are not yet confident enough with the tool to use it well, has done exactly what it should.

This is closely related to the principles behind regulatory sandboxing. Just as a sandbox creates a bounded space to test whether a regulatory change works before committing to it, an AI pilot creates a bounded space to test whether a tool works in a specific regulatory context before extending its use. The same discipline applies; clear limits, active monitoring, and a genuine commitment to acting on what is learned, including stopping if the evidence points that way.

Each pilot should be designed so it can be paused or stopped without disrupting core regulatory functions. Ending a pilot that is not delivering is the process working as intended, and it produces information that a large-scale deployment never could. Each cycle finishes with reflection. What worked? What did not? What risks emerged? What governance or oversight needs strengthening? These lessons feed into the next pilot and gradually lift organisational maturity.

Linking back to HITL

Iterative unleashing reinforces HITL in practice. Early use cases focus on functions where AI supports people rather than replacing judgement, such as triage, completeness checks, or information synthesis. Statutory decisions about suitability, approvals, or enforcement remain with authorised decision-makers. Starting small and low-risk allows regulators to build confidence in how human oversight works before AI is used in decisions that affect rights, obligations, or public trust.

The capability impact

Iterative unleashing is as much about people as technology. It builds the habits of questioning AI outputs, surfacing risks early, and learning from what does not work. When regulators share what they learn across the system, including what worked, what failed, and what surprised them, that learning benefits the whole regulatory community.

Section Two: Summary

- Early and limited AI use is a predictable starting point, but deliberate leadership is needed to move beyond it.
- Leadership sets the tone: without visible sponsorship and clear governance, AI tends to stay at the margins.
- Small, low-risk, and reversible use cases are the safest way to build confidence and capability.
- Regulators progress faster when they share learning, risks, and examples across the system.
- AI literacy is essential so leaders and practitioners can judge quality, bias, and risk appropriately.
- Human judgement and accountability must remain central, with AI supporting decisions rather than replacing them.

Section three: navigating AI ethically

3.1 Setting ethical boundaries

Regulators make decisions that affect people's rights, safety, and livelihoods. The bar for ethical AI use in a regulatory context is therefore higher than in many other settings. Getting the technical and governance foundations right, as earlier sections describe, is necessary but not sufficient. Regulators also need to be clear about the principles that should guide how AI is used, and where it should not be used at all.

These principles draw on the New Zealand Public Service AI Framework and wider public sector guidance, but are framed specifically for regulators whose decisions have real and lasting consequences for people, organisations, and markets.

When using AI, regulators must set clear ethical boundaries that protect public trust and ensure decisions remain lawful, fair, explainable, and accountable. In practice this means paying particular attention to transparency and explainability, and to fairness, bias, and privacy.

Transparency and explainability

Regulators already have a duty to explain how and why decisions are made. That duty does not change when AI supports or informs those decisions. Where AI is used:

- regulators should be open about when AI is involved, what role it plays, and how particular models work
- decision-makers must understand how AI outputs were generated and their limitations
- affected parties should be able to understand how AI-informed decisions were reached.

Transparency supports fairness, protects public trust, and reduces legal and reputational risk. The Public Service AI Framework, consistent with OECD guidance, sets a clear expectation that people interacting with government AI systems should be aware of and understand how AI is being used. For regulators, this reinforces the importance of human judgement, clear reasoning, and explainable decision-making, regardless of whether AI is involved.

Fairness, bias, and privacy

When AI is used in regulatory contexts, fairness and privacy are not separate concerns. Both go to the heart of protecting people from unintended harm, especially where decisions affect rights, access to services, or regulatory outcomes.

Bias can enter AI systems in many ways: through training data, model design, feedback loops, or how outputs are interpreted by people. Regulators need to be alert to these risks and treat fairness as an ongoing responsibility. This includes:

- understanding, at a high level, how AI systems work and where their limitations lie
- using data that is relevant and appropriate to the regulatory context, including checking agency data for completeness and historical bias before it is used as an input to any AI system
- checking for disproportionate impacts, particularly on vulnerable or marginalised groups
- being able to pause, correct, or roll back AI use when issues emerge.

Privacy risks also increase when AI draws on large or sensitive datasets. Regulators must apply strong data governance and ensure personal information is handled lawfully, securely, and proportionately.

Where Māori data is involved, it should be treated as taonga. Māori data refers to data about Māori people, their communities, land, culture, and resources. Māori data sovereignty holds that Māori should have authority over how that data is collected and used, consistent with Te Tiriti o Waitangi. Regulators should refer to Te Mana Raraunga for guidance on Māori data governance expectations.

As a practical safeguard, early AI use should favour public or low-risk data sources and tasks where errors can be detected and corrected easily. Where AI-supported processes affect the public, flexibility is important so people are not disadvantaged by differences in digital access, literacy, or circumstance.

Ethical use of AI requires ongoing review and adaptation as systems, risks, and contexts change.

3.2 Te Tiriti o Waitangi in AI-enabled regulatory practice

Regulators have existing responsibilities under Te Tiriti o Waitangi to recognise Māori rights and interests, support equitable outcomes, and maintain trust and confidence.

The use of AI in regulatory practice does not change these responsibilities. It can, however, amplify impacts. Decisions supported by AI may affect how Māori experience regulation, including access to services, compliance outcomes, and regulatory oversight. For regulatory leaders, this means considering Te Tiriti implications early when exploring and adopting AI, in a way that is proportionate to the impact of the use case.

What this means in practice

When considering AI use in regulatory settings, regulators should identify and engage with the Māori individuals, groups, or organisations who are affected by, or have mana in relation to, the regulatory activity. For example, any Māori advisory groups, or committees used within your regulatory system, whether topic specific, or location based. The appropriate stakeholders will vary depending on the nature of the regulatory function and the type of impact.

Section Three: Summary

- The bar for ethical AI use in a regulatory context is high. Decisions affect rights, safety, and livelihoods, and that does not change when AI is involved.
- Regulators have a duty to explain how and why decisions are made. That duty extends to decisions that AI has supported or informed.
- Fairness, bias, and privacy are ongoing responsibilities, not one-off checks at the point of procurement.
- Māori data should be treated as taonga. Te Tiriti o Waitangi obligations apply to AI-enabled regulatory practice and should be considered early.
- Ethical AI use requires ongoing review and adaptation as systems, risks, and contexts change.

From guidance to practice: AI checklist

The checklist below is a practical guide to help leaders consider key issues when exploring the use of AI in regulatory work. Before proceeding with any AI initiative, leaders should be able to answer yes to the questions below. Where answers are unclear or absent, that is a signal to pause and address the gap before moving forward. The checklist is indicative rather than exhaustive, and some AI applications may require additional domain-specific considerations.

☑	Get clear on the 'what', 'when', and 'why'
☐	Is the purpose clear, justified, and linked to statutory or regulatory outcomes?
☐	Is there a measurable problem this AI system is intended to solve?
☐	Are the benefits realistic, relative to effort, cost, and maturity of a regulator's practice?
☐	Has a risk assessment been completed?
☐	Is the underlying data reliable, consistent, and suitable for the intended AI use?
☐	Is AI being used in a proportionate way for the decision type (e.g. support vs. determination)?
☐	Can the AI output or process be explained in plain language to a Minister, court, or regulated party?
☐	Are documentation and model assumptions available and understandable?
☐	Does the oversight level match the risk level of the AI-supported activity?
☐	Would the public reasonably trust this use of AI if it were publicly disclosed?

☑	Put robust processes in place for the ‘how’
☐	Has the AI tool been checked for unintended bias or disproportionate impacts?
☐	Are there processes in place to test bias regularly over time?
☐	Have impacts on Māori, data sovereignty, or cultural values been considered and addressed?
☐	Are privacy risks assessed and data checks documented?
☐	Is a **legally authorised** human decision-maker retaining responsibility for final decisions? Are points for human review, intervention, appeal and override clearly defined?
☐	Can the system be paused, reversed or rolled back if issues arise?
☐	Is there a plan to monitor AI performance, drift, errors or unintended consequences?
☐	Are update cycles, review triggers and responsibilities clearly assigned?
☐	Is there a mechanism for learning from incidents or feedback?
☐	Can a regulated party seek clarification or appeal to a human about an AI-supported decision?
☐	Are audit trails kept so that the decision path can be reconstructed?
☐	Does staff capability exist to implement, monitor, and manage this tool safely?
☐	Can this be trialled as a reversible pilot before scaling?
☐	Is vendor support appropriate and do you understand their data practices?