Ministry for Regulation
Te Manatū Waeture

Responsible AI in Action

Summary of the guidance for senior leaders and management teams in regulatory organisations

Overview

Used well, AI can strengthen regulatory practice. It can analyse large amounts of information quickly, draft text at speed, and highlight patterns people may miss.

But AI cannot interpret legislation, weigh fairness, or take responsibility for decisions. Those responsibilities must always remain with people.

This summary of the ***Responsible AI in Action*** guidance is for regulatory leaders who are accountable for decisions, systems, and public trust. It explains how to make confident, informed choices about when and how AI is used in regulatory work.

At a leadership level, responsible AI use comes down to three things:

1. Setting strong foundations.
2. Keeping humans in control.
3. Navigating AI ethically.

1. Set the foundations

Why this matters

AI does not fix weak systems. It amplifies what is already there.

- If regulatory foundations are strong, AI can scale good practice.
- If foundations are weak, AI makes problems bigger, faster, and harder to explain.

Leadership message
Treat AI as a regulatory capability, not an IT project.

What we mean by ‘foundations’

Good regulatory foundations include:

- Clear delegations and decision rights.
- Sound record-keeping and audit trails.
- Reliable, well-managed information and data.
- Effective oversight and consistent reasoning in decisions.

When these are unclear, AI increases risk. Errors scale quickly, accountability becomes blurred, and decisions become harder to defend.

AI is also only as good as the data behind it. Incomplete, biased, or poorly structured data produces unreliable outputs.

Two types of AI most relevant to regulators

Predictive AI
Used to classify, score, or prioritise cases (for example, flagging higher-risk licence applications).

- Strength: Helps focus effort where risk is higher
- Risk: Can repeat or amplify past bias if historical data is flawed

Generative AI
Used to produce text or summaries (for example, drafting guidance or summarising submissions).

- Strength: Saves time on high-volume information tasks
- Risk: Can generate confident-sounding content that is incomplete or wrong

Leadership message

Different AI uses carry different risks. Governance and oversight must scale accordingly.

Three foundation areas leaders need to get right

1. Strategy, governance, and guardrails

Leaders should ensure that:

- AI use is linked to regulatory priorities, not novelty or vendor claims.
- Ownership, decision rights, and escalation paths are clear.
- AI systems can be paused, adjusted, or withdrawn if risks emerge.
- New uses are tested using sandboxes or controlled pilots before scaling.

2. Procurement with purpose

When buying AI tools:

- Follow Government Procurement Rules and check All-of-Government options.
- Identify and manage key risks early with procurement, digital and regulatory staff.
- Require clarity on data sources, testing, and data sovereignty.
- Require greater explainability where AI supports higher-impact decisions.
- Lock accountability into contracts (audit rights, monitoring, exit terms).

3. Getting people ready

AI adoption fails without leadership:

- Leaders must visibly sponsor responsible use.
- If you cannot clearly explain *why* AI is being used, that is a signal to pause.
- Role-specific learning matters more than generic AI training.
- Stopping a pilot that is not delivering value is success, not failure.

Where AI can genuinely help regulators

- Detect non-compliance and emerging risks earlier.
- Prioritise monitoring and inspections based on risk.
- Surface evidence to support decisions.
- Identify patterns across large datasets manual analysis would miss.
- Reduce administrative burden for regulators and regulated parties alike.

New categories of risk leaders should watch

AI introduces risks that traditional regulation may not fully cover, including:

- Hidden biases in training data.
- Drift toward automation without explicit approval.
- Gaps in explainability that affect appeal rights.
- Vendor lock-in that constrains future regulatory choices.

2. Keep humans in control

Why this matters

Access to AI tools does not equal capability.

Without deliberate oversight, people can over-trust AI outputs, especially when they look authoritative or save time.

Leadership message
AI should support human judgement, not replace it.

AI literacy is the core organisational capability

Staff do not need to be technical experts, but they must understand:

- What AI can and cannot do, and where human judgement must lead.
- How to recognise hallucination and other common failure modes.
- How AI produces outputs.
- Why those outputs can be wrong.
- Where the limits sit.
- How bias can appear.
- How appeal rights and review are preserved.

Without this understanding, human review becomes superficial and responsibility quietly shifts to the tool.

What human-in-the-loop means

Human-in-the-loop is not just having a person involved.

It means:

- People understand the AI well enough to challenge outputs.
- They know when outputs should not be trusted.
- They can make and defend decisions independently of AI suggestions.

This is essential for legality, fairness, and public trust.

Where AI fits best (lower-risk uses)

AI is generally safest when used to:

- **Triage or prioritise cases.** Helping staff decide which cases need attention first, while humans still decide what action to take.
- **Check applications for completeness.** Identifying missing information or obvious gaps before a human assesses the application.
- **Validate data against clear rules.** Checking information against defined thresholds or criteria, where the rules are stable and transparent.
- **Flag inconsistencies for human review.** Highlighting anomalies or conflicting information so a person can investigate and decide what it means.

Decisions that affect rights, obligations, or livelihoods must always remain with authorised humans.

Watch for automation bias

Automation bias is the tendency to defer to machine outputs without independent judgement.

Leaders should ensure:

- Reviewers are required to question AI outputs.
- Oversight checks *how* people engage with AI, not just how the system performs.

Iterative unleashing: scaling safely

Rather than large, irreversible deployments, leaders should authorise:

- Safe experimentation within clear boundaries.
- Small pilots focused on real regulatory problems.
- Reversible steps that build confidence and learning.
- Sharing lessons across the regulatory system.

3. Navigate AI ethically

Why ethics matter more in regulation

Regulatory decisions affect rights, safety, and livelihoods. Those impacts do not change when AI is used to support decisions.

Because of this, regulators must use AI in ways that protect fairness, legality, transparency, and trust. Ethical considerations are not an add-on to AI use; they are central to maintaining public confidence and the legitimacy of regulatory decisions.

Four ethical principles leaders should expect to see applied

1. Transparency and explainability

- Be open about when AI is used and the role it plays.
- Decision-makers must understand how outputs are generated and their limits.
- Affected parties have rights to understand how AI informed decisions.

2. Fairness and bias

- Bias can enter through data, design, or interpretation.
- Data must be checked for completeness and historical bias.
- Monitor for disproportionate impacts and be ready to pause or roll back.

3. Privacy and data governance

- Apply strong, proportionate data governance.
- Prefer low-risk or public data for early use to build confidence.
- Handle personal and sensitive data lawfully.

4. Te Tiriti o Waitangi

- AI does not change existing Te Tiriti obligations.
- Māori data is taonga.
- Engage Māori early and proportionately.

Leadership message

The bar for ethical AI use in regulation is high. Regulatory decisions affect rights, safety and livelihoods – and that does not change when AI is involved.

What next

Regulatory leaders are encouraged to:

- Read relevant sections of the complete **Responsible AI in Action** guidance.
- Support safe experimentation. See our **Sandbox guidance.**
- Share learning across the system so capability grows collectively.