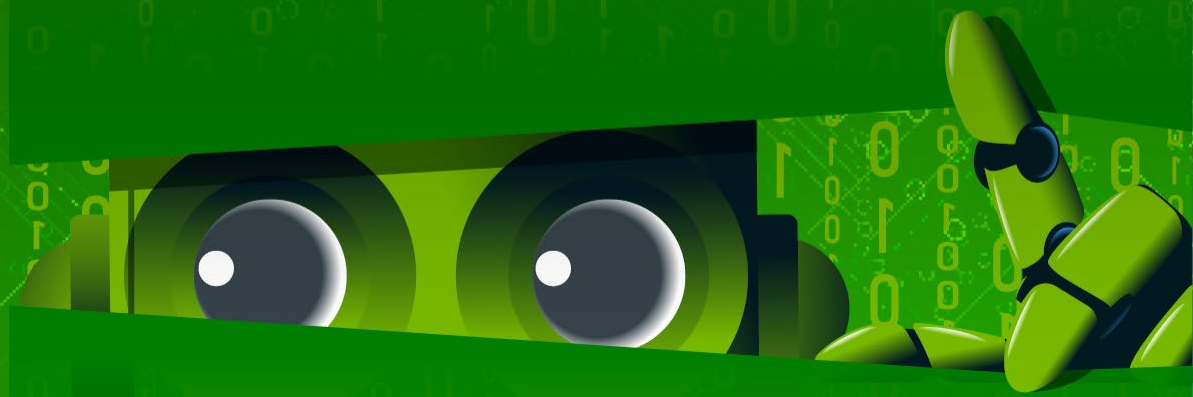
ALLEN + CLARKE

AI RISK MANAGEMENT

AVOIDING THINGS GOING WRONG



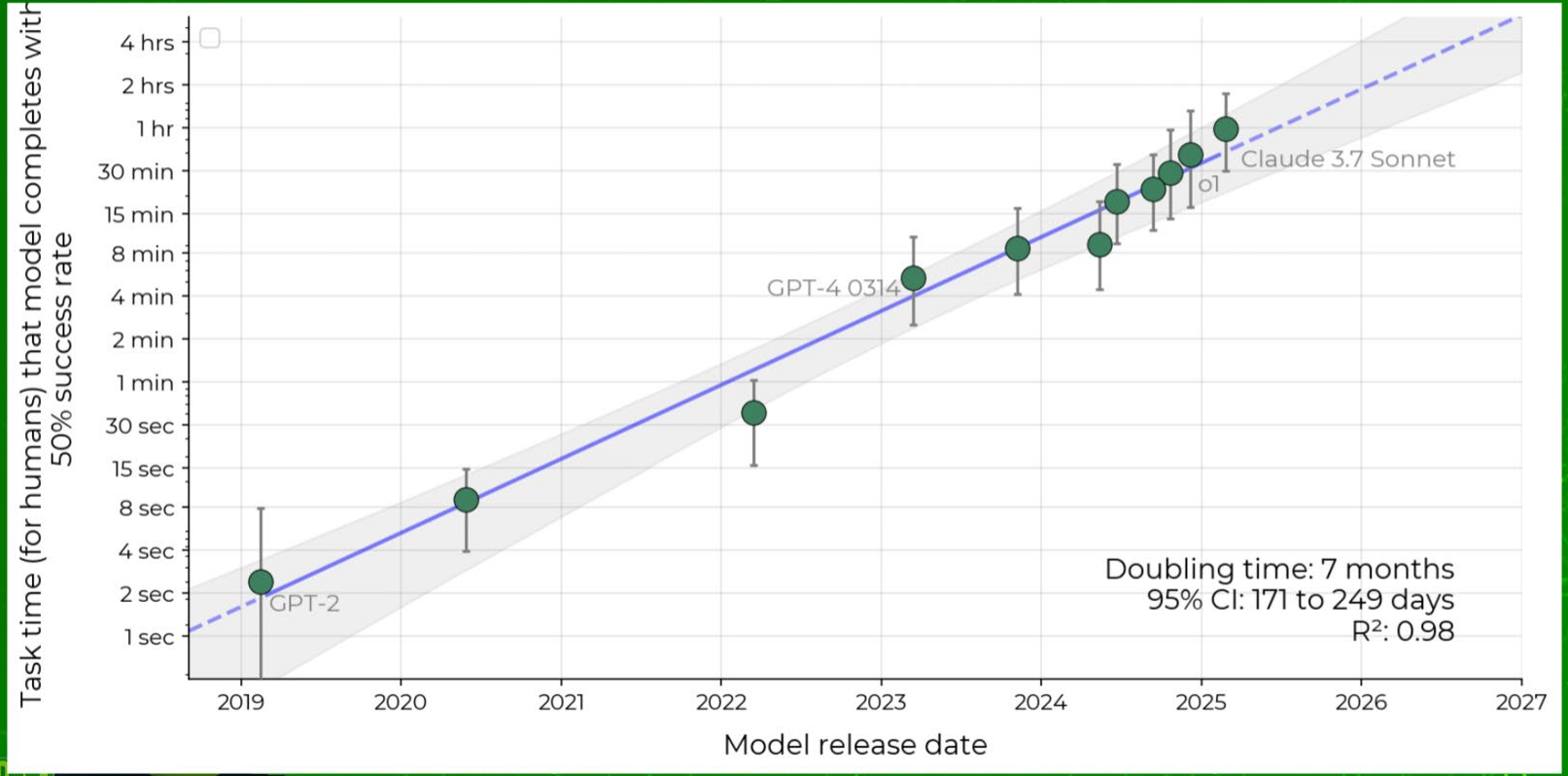
CONTENTS

Seven
main AI risks, from largest to smallest

Four
organisational gaps these relate to

16 steps
for closing these gaps and managing the risks

LENGTH OF TASKS AI AGENTS CAN COMPLETE AUTONOMOUSLY

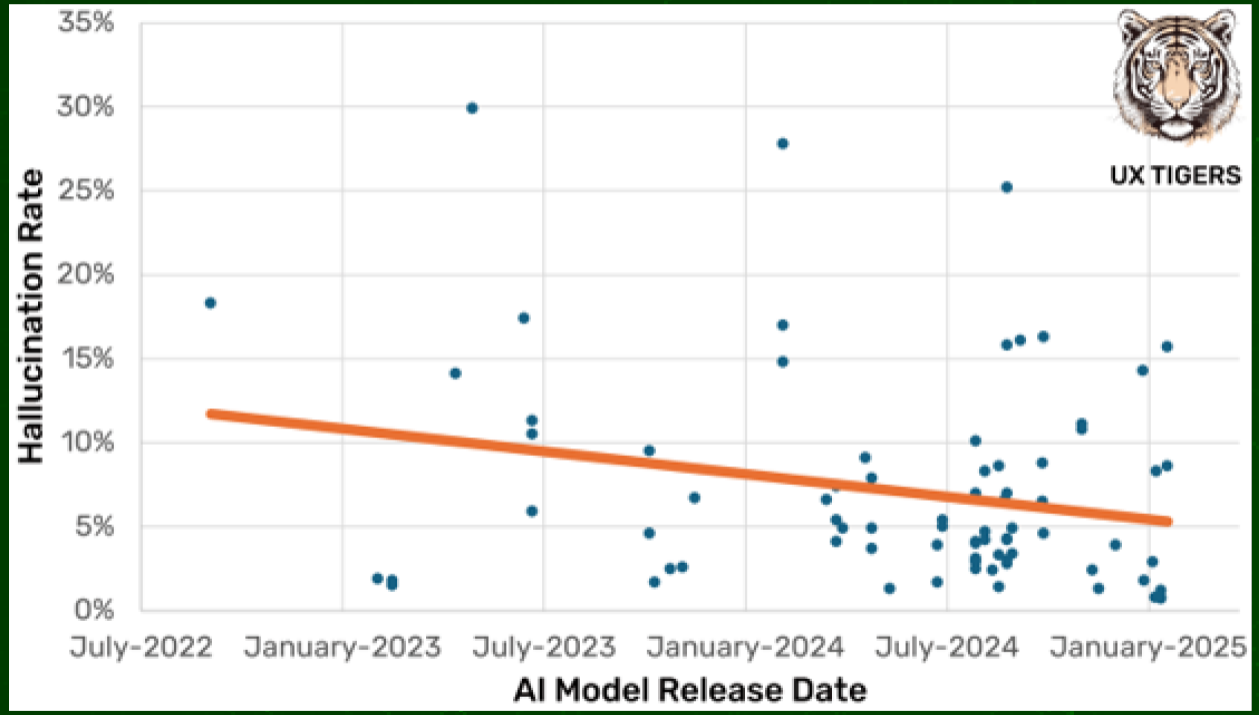


7 BIGGEST AI RISKS – next 3 years

Seven main AI risks, from largest to smallest → **Four** organisational gaps these relate to → **16 steps** for closing these gaps and managing the risks

RISK (Highest to Lowest Impact)	WHAT THE RISK IS	WHY IT'S RANKED HERE
1. Deskilling / Over-reliance	People increasingly rely on AI to think, structure, analyse, or draft for them. Judgement weakens and capability declines.	The biggest long-term threat to quality, judgement, and institutional capability. Deskilling compounds silently becoming an organisational dependency that is extremely hard to reverse.
2. Bias & Fairness	AI learns using the internet and user input. This means it reproduces and amplifies inequities, stereotypes, and historical biases found in this data.	Bias risk can only be managed through specific model training. So until we have country specific models, weighted to minimise data this is a persistent, subtle, and material risk.
3. Transparency / Explainability	Difficulty explaining how AI contributed to a recommendation, conclusion, or piece of content. Lack of provenance or traceability.	As AI becomes more embedded in workflows, the requirement to defend decisions becomes more important, whilst becoming harder to explain.
4. Accuracy / Hallucination	AI makes stuff up by confidently producing fabricated, or misleading outputs that look authoritative.	A major operational risk today that affects every user. While hallucination rates will continue to decline with stronger models, things like agents compound this risk by extrapolating on previous hallucinations.
5. Privacy & Confidentiality	Sensitive or personal information is entered into AI systems without proper controls or understanding of data handling boundaries.	A top-tier risk in non-enterprise settings — but drops significantly when using tools like Copilot + Purview.
6. Copyright & IP	AI-generated text or images may inadvertently reproduce copyrighted.	Important but not something you can control in relation to training data. Legal and reputational risk can be managed with the right internal controls.
7. ROI Failure	AI investments fail to deliver value due to poor implementation and governance.	Any moment you might need your own creative tune – a workshop or a webinar!

HALLUCINATION RATES OVER TIME



Source UX Tigers using data from the Hugging Face Hallucination Leaderboard.

A & C APPROACH TO BUILDING AN OPERATING MODEL

Seven main AI risks, from largest to smallest

Four organisational gaps these relate to

16 steps for closing these gaps and managing the risks

Operating model

Governance Lens

How those involved organise and operate.

- Transparency / Explainability
- Privacy & Confidentiality
- Copyright & IP

Cultural Lens

The cultural values and principles that guide and underpin the behaviours essential to enable the desired outcomes.

- Deskilling / Over-reliance
- Bias & Fairness
- Privacy & Confidentiality (residual)

Capability Lens

The processes expertise, resources, information, tools, technology & infrastructure required to underpin objectives & activities

- Accuracy / Hallucination
- Bias & Fairness (shared)
- Transparency / Explainability (shared)

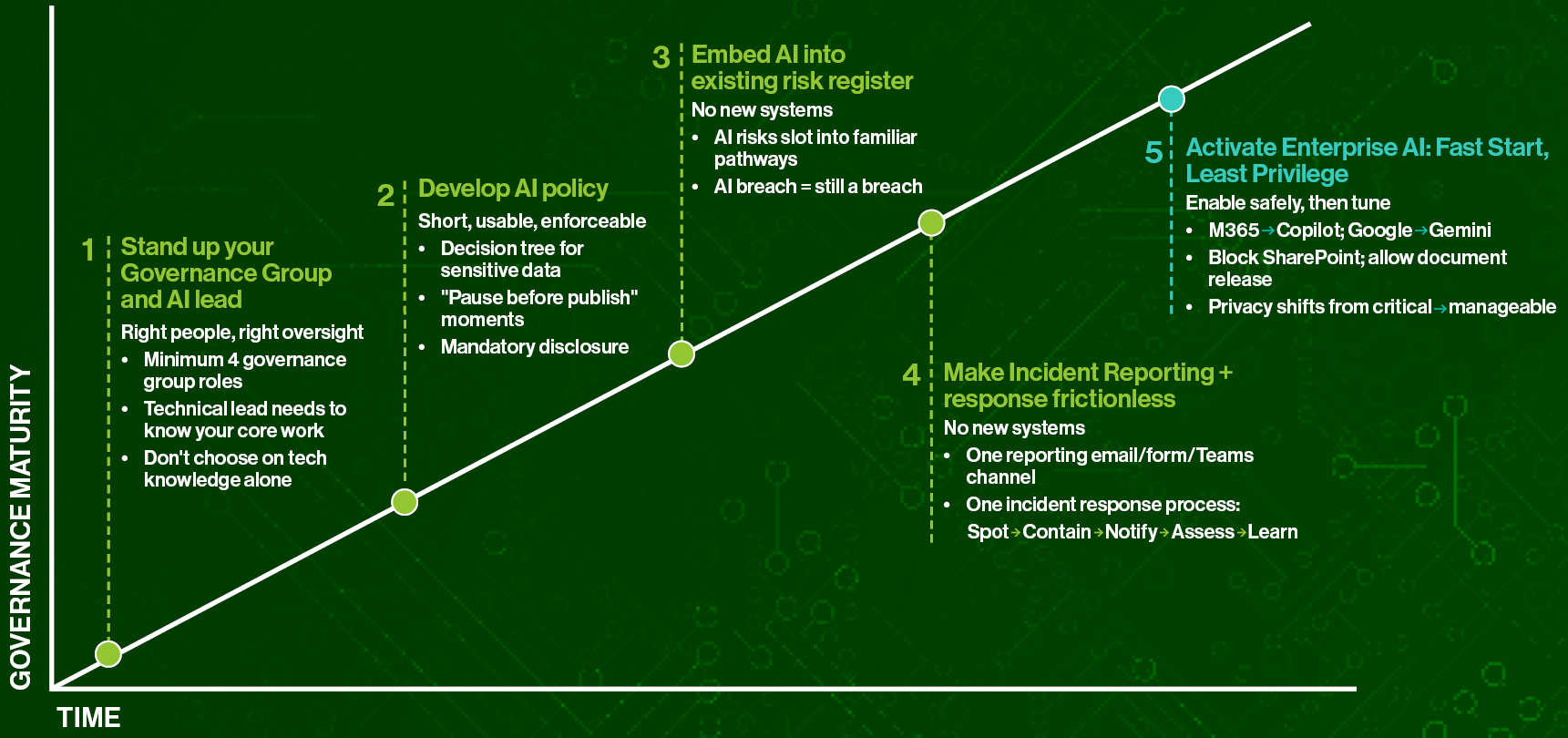
Value Lens

The value contributed and the activities that create that.

- ROI Failure

SIX STEPS TO CLOSING THE GOVERNANCE GAPS

Seven main AI risks, from largest to smallest → Four organisational gaps these relate to → 16 steps for closing these gaps and managing the risks



1 Stand up your Governance Group and AI lead

Right people, right oversight

- Minimum 4 governance group roles
- Technical lead needs to know your core work
- Don't choose on tech knowledge alone

2 Develop AI policy

Short, usable, enforceable

- Decision tree for sensitive data
- "Pause before publish" moments
- Mandatory disclosure

3 Embed AI into existing risk register

No new systems

- AI risks slot into familiar pathways
- AI breach = still a breach

4 Make Incident Reporting + response frictionless

No new systems

- One reporting email/form/Teams channel
- One incident response process: Spot → Contain → Notify → Assess → Learn

5 Activate Enterprise AI: Fast Start, Least Privilege

Enable safely, then tune

- M365 → Copilot; Google → Gemini
- Block SharePoint; allow document release
- Privacy shifts from critical → manageable

Closing governance gaps

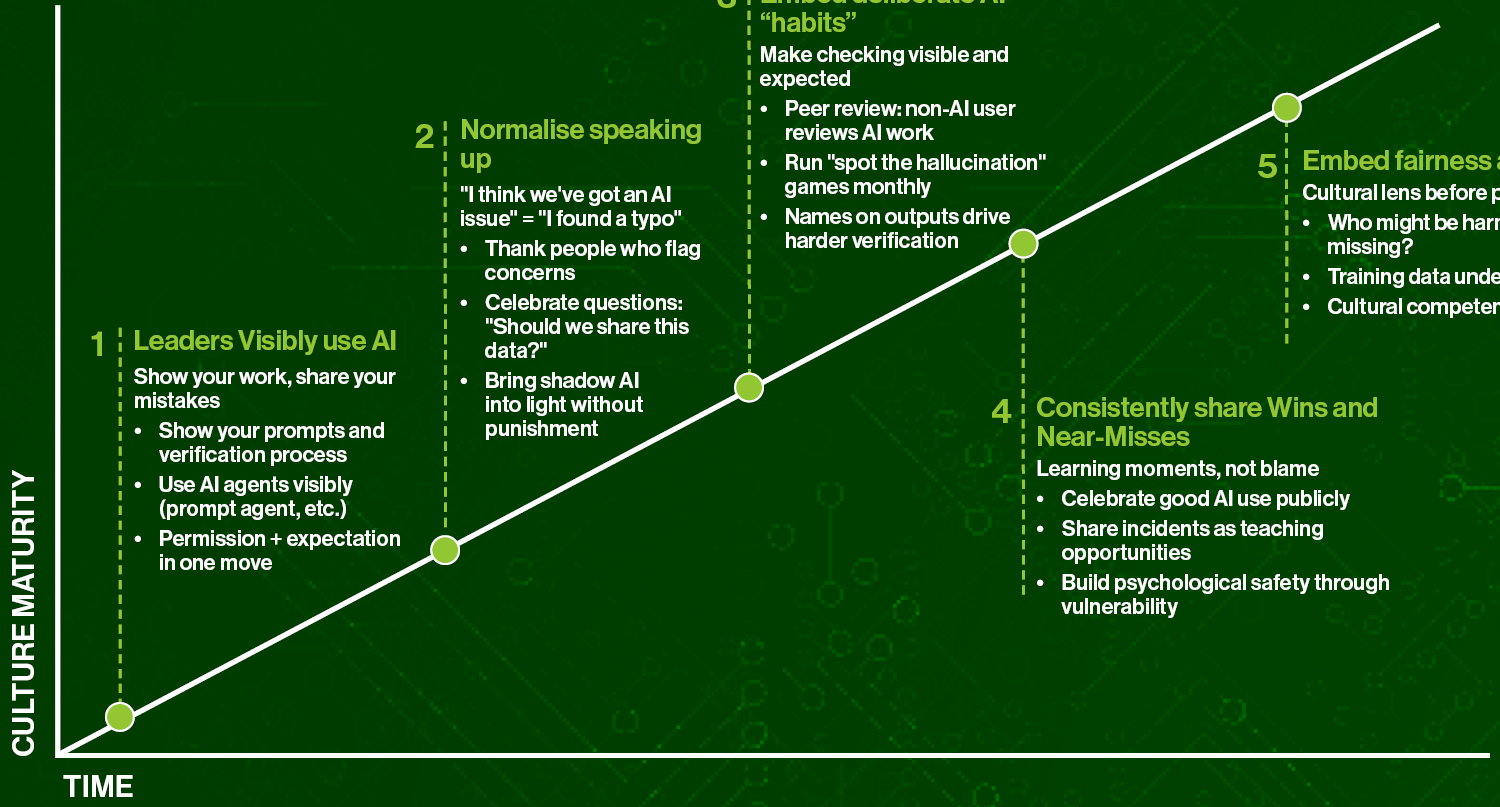
Stand up governance group | **Develop AI policy** | Embed AI into risk register | Develop AI incident response plan | Implement enterprise AI

SIX THINGS YOUR AI POLICY SHOULD HAVE

1. **Your approved tools list.** What can people use, what's banned. Be specific - "ChatGPT free tier is banned, ChatGPT Plus with training turned off is approved for non-sensitive work."
2. **A decision tree.** A simple flowchart - is this data sensitive? Is this output going external? What review is required? Make the decision path visible and simple for everyone to follow.
3. **Caution points.** I like to build in caution triggers. These are simple statements that you repeat over and over. They become the catch phrases that lead to behavioural implementation of your policies core principles. EG: Loading stuff in, have a hmm or pause before you publish. These micro-moments prevent macro-problems.
4. **Human review requirements.** For many of you this will mean mandating no AI output goes external without human verification. Be explicit about when sign-off is required.
5. **Mandatory disclosure.** Any AI-assisted work that gets published must include disclosure with the drafting team's names and positions. Not "AI was used." Actual names. "This report was drafted by [Name, Position] with AI assistance for X+Y" This creates clear ownership and responsibility. People verify harder when their name's on it.
6. **Who's accountable.** Name the person responsible for AI risk management. Not a committee. A person.

5 STEPS TO CLOSING THE CULTURE GAPS

Seven main AI risks, from largest to smallest → Four organisational gaps these relate to → 16 steps for closing these gaps and managing the risks



1 Leaders Visibly use AI

Show your work, share your mistakes

- Show your prompts and verification process
- Use AI agents visibly (prompt agent, etc.)
- Permission + expectation in one move

2 Normalise speaking up

"I think we've got an AI issue" = "I found a typo"

- Thank people who flag concerns
- Celebrate questions: "Should we share this data?"
- Bring shadow AI into light without punishment

3 Embed deliberate AI "habits"

Make checking visible and expected

- Peer review: non-AI user reviews AI work
- Run "spot the hallucination" games monthly
- Names on outputs drive harder verification

4 Consistently share Wins and Near-Misses

Learning moments, not blame

- Celebrate good AI use publicly
- Share incidents as teaching opportunities
- Build psychological safety through vulnerability

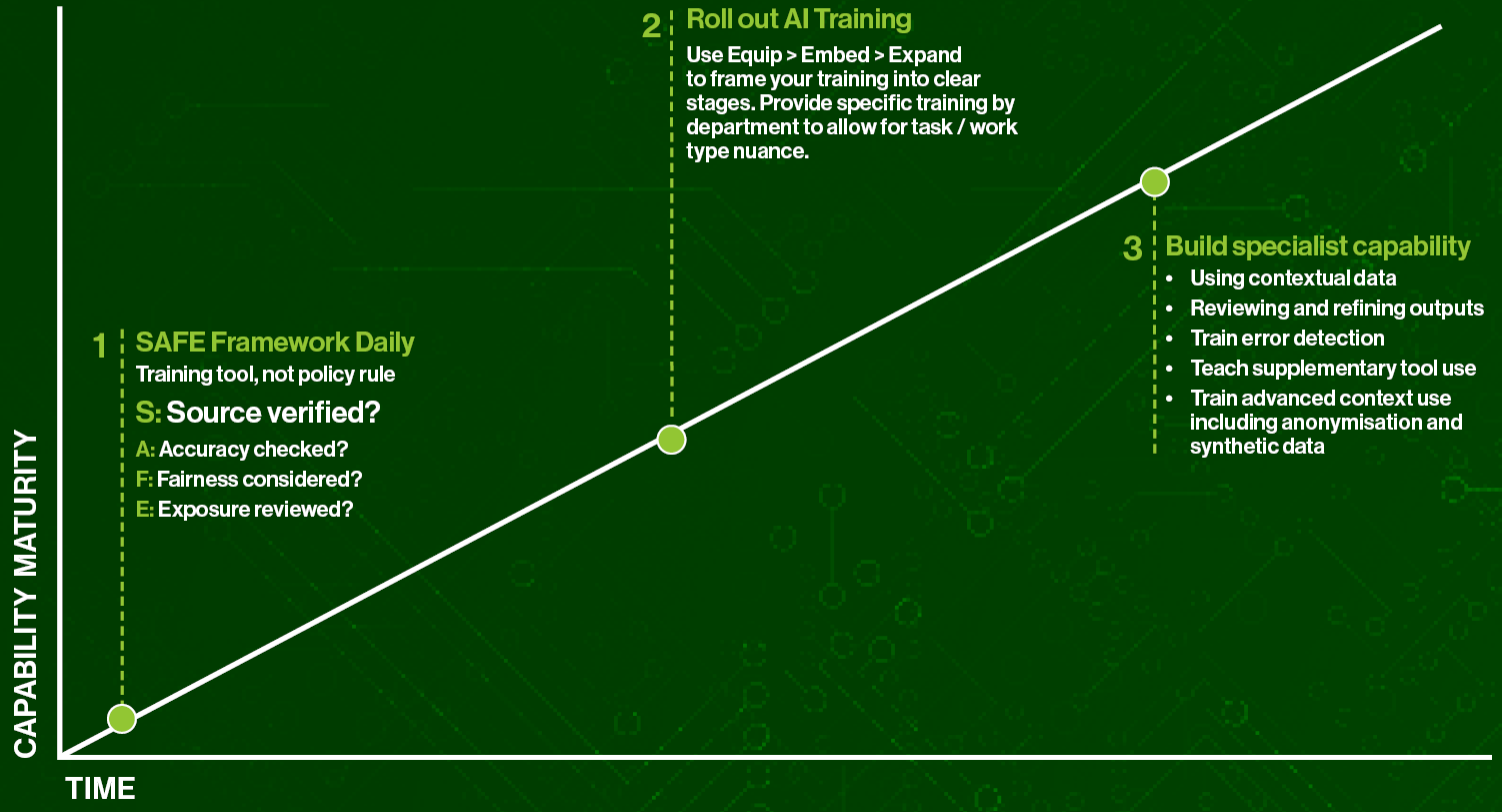
5 Embed fairness and Equity Reviews

Cultural lens before publication

- Who might be harmed? Whose voices missing?
- Training data underrepresents communities
- Cultural competency review non-negotiable

3 STEPS TO CLOSING THE CAPABILITY GAPS

Seven main AI risks, from largest to smallest → Four organisational gaps these relate to → 16 steps for closing these gaps and managing the risks



1 **SAFE Framework Daily**

Training tool, not policy rule

S: Source verified?

A: Accuracy checked?

F: Fairness considered?

E: Exposure reviewed?

2 **Roll out AI Training**

Use Equip > Embed > Expand to frame your training into clear stages. Provide specific training by department to allow for task / work type nuance.

3 **Build specialist capability**

- Using contextual data
- Reviewing and refining outputs
- Train error detection
- Teach supplementary tool use
- Train advanced context use including anonymisation and synthetic data

Closing capability gaps

Embed SAFE Framework | Roll-out AI training | Build specialist capability

SAFE FRAMEWORK

S: SAFE Where did this information come from? Can you verify it independently?	This is still your critical check even with enterprise AI. The AI doesn't cite sources accurately just because it's running on your tenant. Verify every citation against the original source.
A: Accuracy Is this factually correct?	Cross-check figures, quotes, claims. Enterprise AI reduces privacy risk but it doesn't fix hallucinations. Every AI output still needs human verification. For policy work, check against official documents. For data analysis, verify against source data.
F: Fairness Who could this harm or exclude?	Your enterprise setup doesn't solve bias. You still need to ask equity questions. Whose perspective is missing? Could this entrench discrimination? For community-facing work, this check is non-negotiable.
E: Exposure What data went into AI?	With enterprise AI and proper permissions, this risk is lower but not zero. Still review what information was shared. If it required special permission to release to the AI, document that decision.

Closing capability gaps

Embed SAFE Framework | Roll-out AI training | Build specialist capability

AI TRAINING - Capability Building Framework

EQUIP

Give people the foundations they need to use AI safely and effectively

- Equip + train all leaders, especially executive level to be AI champions
- Train staff on:
  - access and basic operation
  - verifying AI outputs
- Introduce SAFE as the core quality check
- Create simple “how-to” guides for common tasks

EMBED

Build AI-safe workflows so capability is not optional

- Implement org wide task automation and workflows, starting with leadership
- Make human review mandatory in workflows
- Use Pre-Flight Checks before significant AI tasks
- Integrate AI steps into team QA processes

EXPAND

Grow advanced capability across the org as confidence and maturity increase

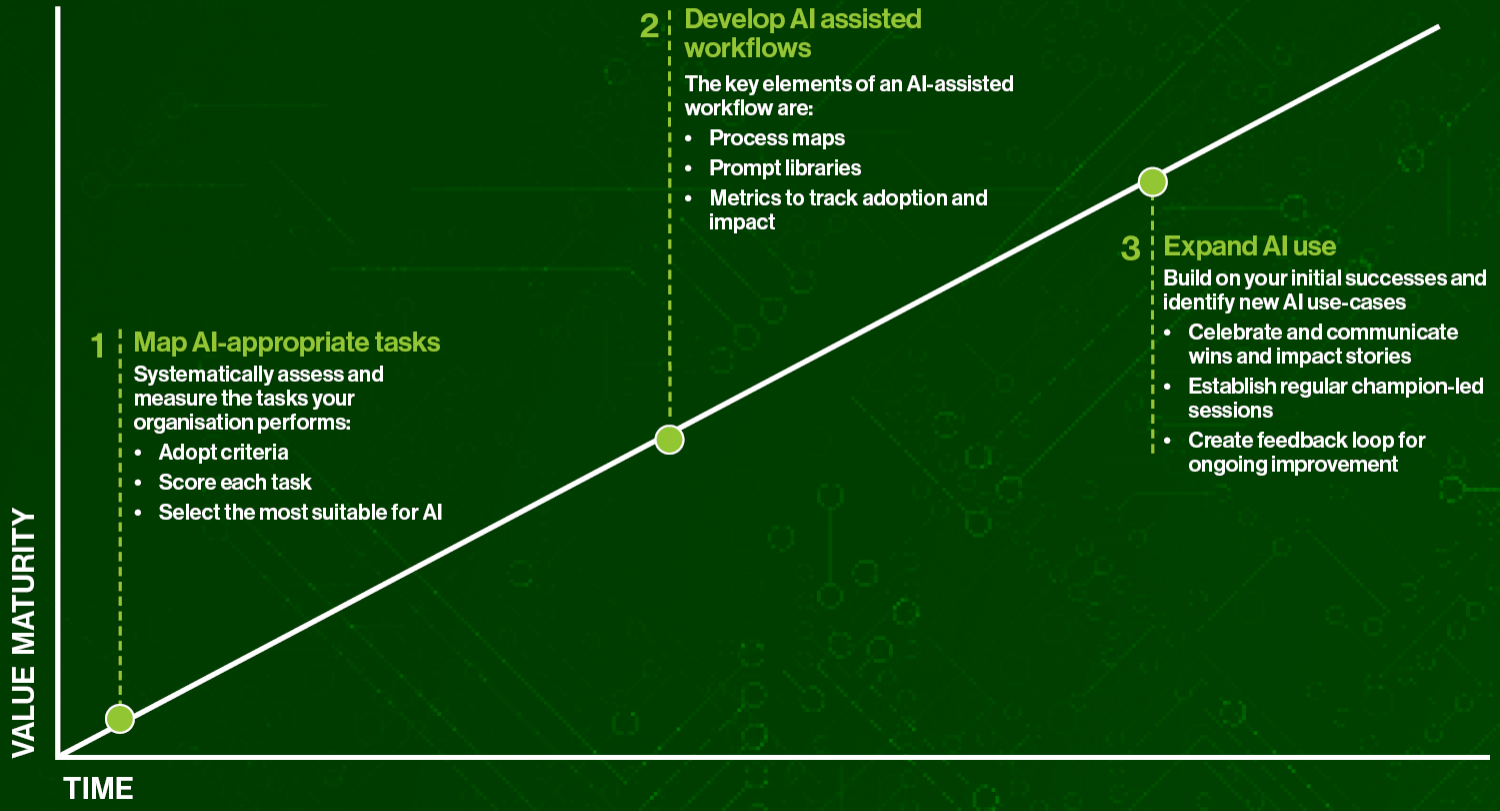
- Leadership to promote and demonstrate deeper skills in unique AI applications and oversight
- Develop specialists who support teams with complex tasks
- Expand use cases as capability matures
- Continuously refine guidelines based on real incidents and lessons learned

OUR FAVOURITE MODELS AND WHAT THEY'RE BEST AT

TOOL	WHY WE LIKE IT	BEST USE CASES
ChatGPT	The most versatile and user-friendly option. Has a wide range of more advanced features for more complex tasks when you need them.	Ideation + brainstorming, writing, image generation, able to export outputs into Word or Excel, internet based research.
Perplexity	Better than Google at internet searches, including formatting search results and follow up questions.	Detailed internet research, formatting of search results, verifying facts, exploring new topics.
Claude	Writes like a human with the ability to set custom writing styles. Great at understanding nuance and long documents.	Drafting, editing, reviewing all types of content.
Copilot	Helps resolve privacy concerns by working within your existing Microsoft environment. Accesses SharePoint/OneDrive for context-aware tasks.	Project planning, task tracking, document prep.
Gemini	Great if you use Google suite for personal or work as it connects to G Suite. Has amazing screen share mode – your personal tech support.	Drafting and editing Google Docs, slides, etc. real-time support, summarising spreadsheets
Notebook LM	Accurately summarises and reviews documents with citations. Has a cool podcast mode so you can discuss the documents you've added.	Reviewing reports, research or contracts, extracting key points, building FAQs
Udio	A fun way to create your own songs with your own lyrics	Any moment you might need your own creative tune – a workshop or a webinar!
Originality	Best tool for detecting AI-generated content.	Verifying written content authenticity down to individual sentences
Whisper	Transcribes speech clearly, even from tricky recordings. Nails Te Reo and NZ accents.	Transcribing interviews, meetings, webinars, or voicemails

3 STEPS TO CLOSING THE VALUE GAPS

Seven main AI risks, from largest to smallest → Four organisational gaps these relate to → 16 steps for closing these gaps and managing the risks



1 Map AI-appropriate tasks

Systematically assess and measure the tasks your organisation performs:

- Adopt criteria
- Score each task
- Select the most suitable for AI

2 Develop AI assisted workflows

The key elements of an AI-assisted workflow are:

- Process maps
- Prompt libraries
- Metrics to track adoption and impact

3 Expand AI use

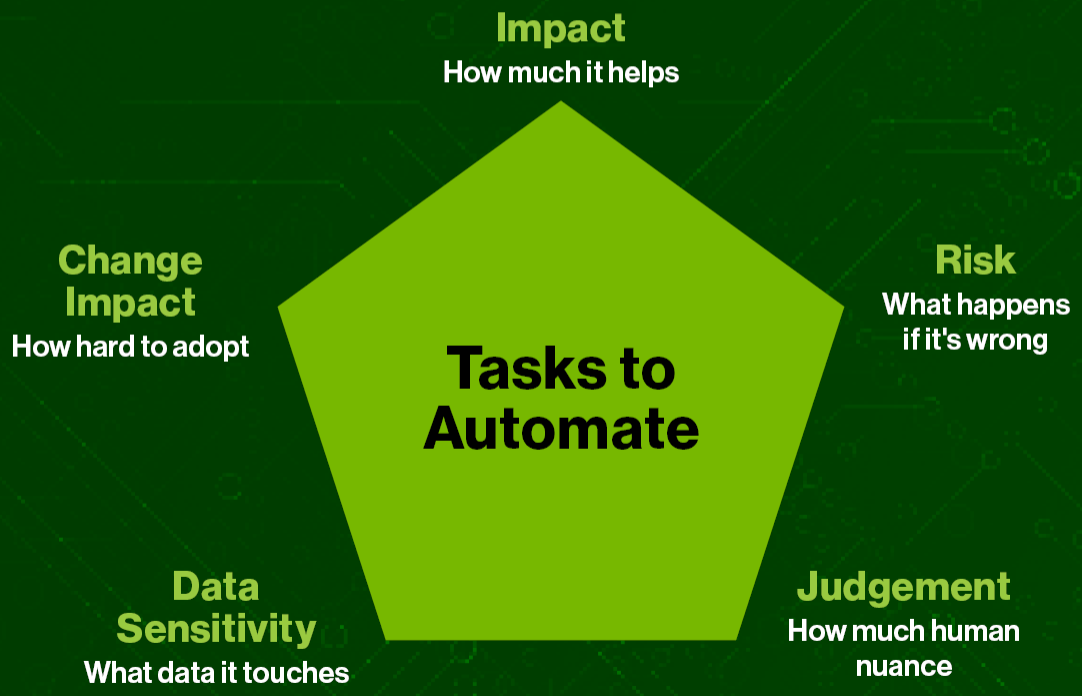
Build on your initial successes and identify new AI use-cases

- Celebrate and communicate wins and impact stories
- Establish regular champion-led sessions
- Create feedback loop for ongoing improvement

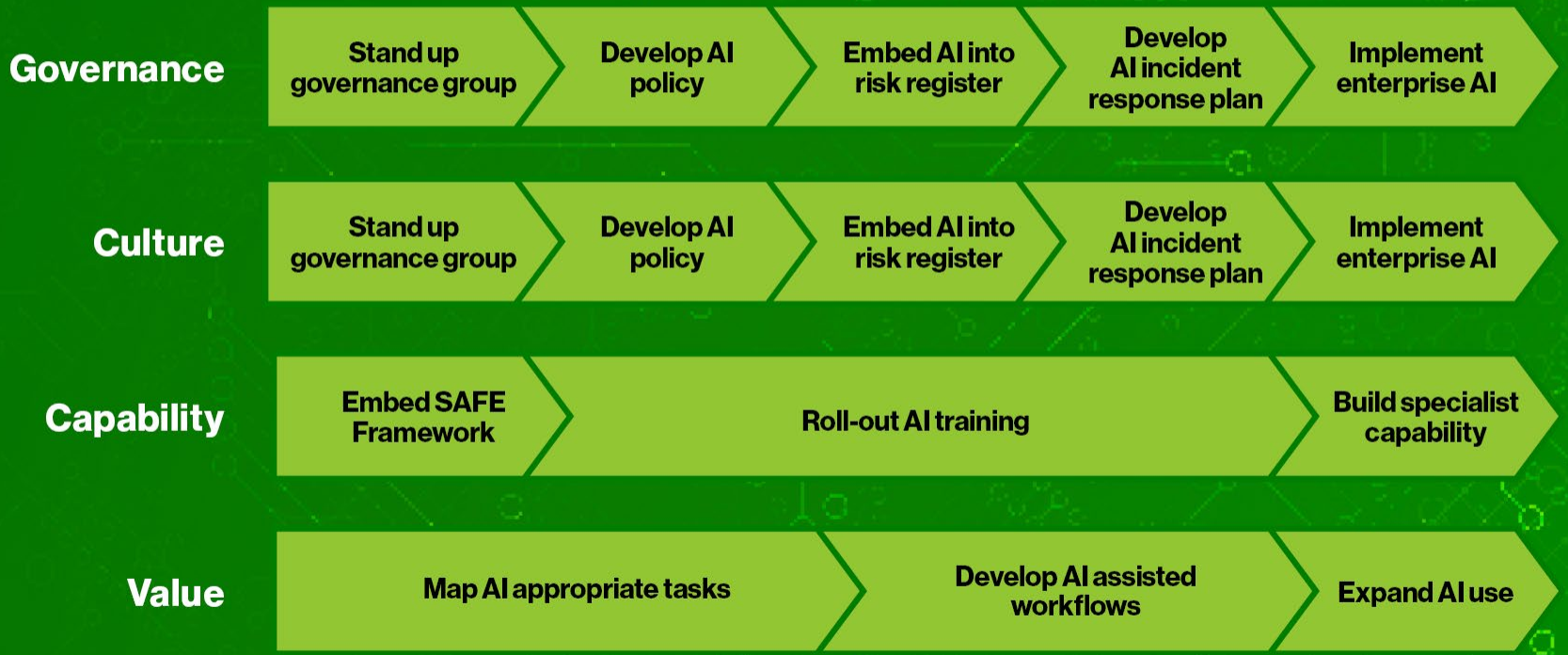
Closing value gaps

Map AI appropriate tasks → Develop AI assisted workflows → Expand AI use

MAPPING AI - appropriate workflows



RECAP AND RISK MITIGATION ROADMAP



ALLEN + CLARKE

Q+A

ALLEN + CLARKE

CONTACT ADAM EMIRALI

ADAM EMIRALI
aemirali@allenandclarke.com