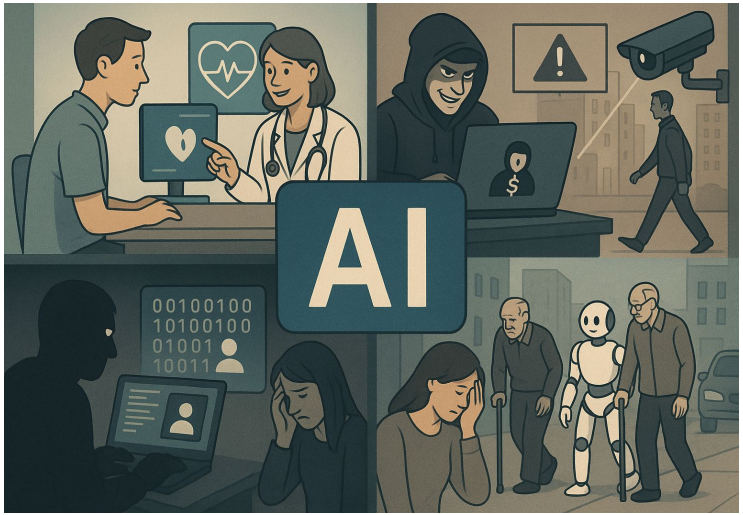
AI Activate | Mitigate

A strategic primer for national AI policy & planning to support nation building and economic transformation

Taking a national strategic approach to AI



Artificial Intelligence is being adopted in diverse, innovative and increasingly pervasive ways: across the globe & by all aspects of society creating significant national opportunities and challenges for governments everywhere (Human Development Report, 2025). Inaction isn't a viable choice, as it risks vulnerability to unprecedented threats & being left behind. Responsible action is critical and urgent.

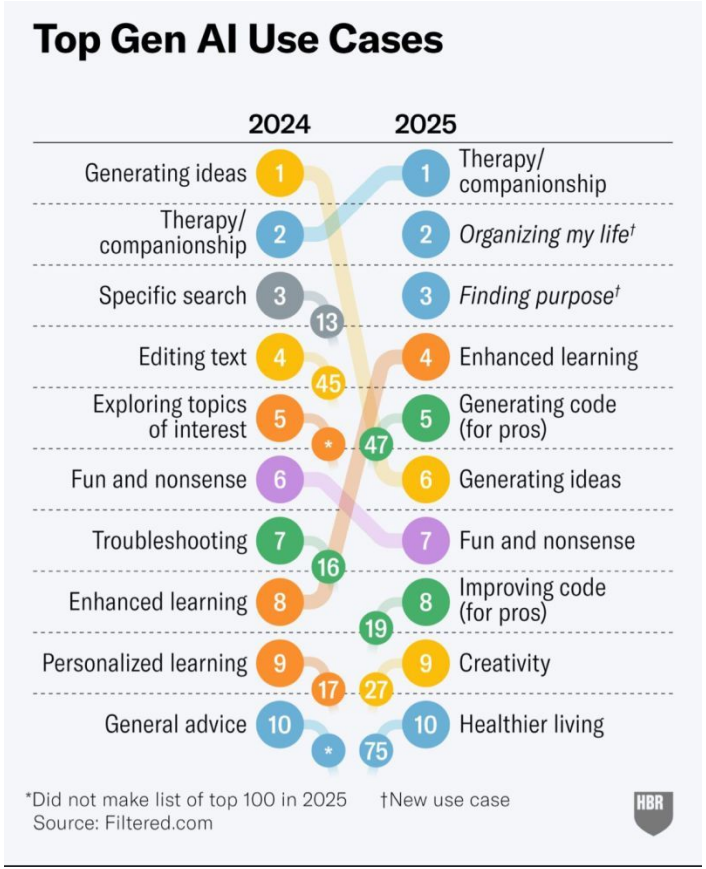
Governments that take a strategic approach can enable responsible & productive adoption of AI across all sectors (including gov), and can learn how to effectively respond to unprecedented threats from the weaponization of AI.

This strategic AI primer has been drafted as a way to support policy makers, thought leaders, public servants and civil society to plan and adopt a holistic and national approach to AI that acts to **activates** national benefits and **mitigates** national threats. The dualistic approach proposed (activate:mitigate) isn't just desirable, but necessary to ensure nations are able to keep pace and thrive in an increasingly AI-enabled world, and realise a better and more inclusive future for all.

The uncomfortable truth about AI…

…is that AI is most horrifying when it reflects, perpetuates & accelerates the (often horrifying) status quo.

A responsible approach to AI often involves establishing more humane and human centric approaches to policy making, success definition, funding, prioritisation, professional culture & operations. It's also important to understand and plan for the environmental impacts of widespread AI usage, and to take an approach that acknowledges AI systems continuously change outside of traditional IT change management.



What do we mean by Activate | Mitigate?

Below is a brief overview of national AI opportunities and challenges.

Examples of national opportunities and potential benefits from responsible use of AI	**Examples of national threats and potential harms from malicious & weaponised use of AI**
• **Economic diversification:** opportunities to transform traditional industries and local economic diversification, improving domestic access to global markets. • **Productivity gains:** general access automation and synthesis capabilities provide a next generation "spell check", supporting content synthesis & generation. • **Better decision making:** blending information, data & analysis into easier evidence-based decision making. • **Public engagement:** policy participation at scale. • **Patterns analysis at scale:** can help understand the real impact of policies, identify/remediate bias & inequality, & to support better policy modelling & design. • **Issue identification:** advanced, scientifically sound predictive analytics and computer vision can help identify threats, illicit goods & healthcare diagnoses.	• **Well funded criminals & terrorists:** increased funding from increasingly convincing scams & social engineering, also significantly impacting communities. • **Radicalisation:** ease of identifying, targeting and radicalising vulnerable people and communities, either for ideology (eg, undermining democracy, terrorism) or profit (eg. 'manosphere' influencers, harmful sales). • **Social fragmentation:** lonely kids/adults turning to AI companions, furthering their isolation & risk of influence. • **Automated attacks:** an unprecedented ability to generate malicious software at scale, enabling effective attacks on sensitive information & critical infrastructure. • **Deep fakes:** videos & images of anyone can be used to generate fake content for nefarious reasons (fake porn, fake endorsements, deep fake evidence, etc).

Recommended Pillars for a National AI Strategy

Each pillar requires clear ownership and delegation of authority to act

Policy

Pillar Goal: Clear, frameworks to assure safe, inclusive AI.

- Safety Standards
- Dynamic risk mgmt
- Regulatory coherence
- AI Economy framework
- Digital Inclusion Index
- Government AI Strategy
- Reporting/monitoring
- Assurance/governance
- Technical guardrails
- Transparency

DPI 4 AI

Policy Goal: DPI investments for a safe & inclusive AI Nation.

- DPI framework extended to enable safe & responsible AI
- Agentic AI security
- Safe AI sandboxes
- Digital regulation & policy infrastructure
- Sovereign AI initiatives
- Data/Cloud strategies
- Multi-lingual LLM

Capability

Policy Goal: A skilled & innovative domestic workforce & society.

- Baseline AI literacy to the general public
- Sector based training available (inc for gov)
- Multi-disciplinary policy
- AI literacy for all school & university curricula
- Talent pipeline
- Innovation enablement, R&D and partnerships

Reform

Policy Goal: Effective economic transition of traditional sectors.

- AI for Human Development
- Sector-based projects, prototyping & reform
- Government service delivery & adaptive policy innovation
- Exploration of AI+IoT for emergency mgmt
- Startup support

Resilience

Policy Goal: National coordination & response to AI threats.

- Protection for critical infrastructure & against social manipulation
- Consumer protections
- Social cohesion
- Participatory gov
- Validation mechanisms for critical information
- Disrupt criminal funding from AI-enabled scams

Malaysia: an exemplary & adaptive strategic approach

From the 1st AI strategy in 2021, Malaysia has effectively driven & evolved their approach

National Vision	Make Malaysia a nation where Artificial Intelligence augments JOBS, drives NATIONAL COMPETITIVENESS, encourages INNOVATION & ENTREPRENEURSHIP to bring economic prosperity, social good and improves PEOPLE'S WELL BEING (Artificial Intelligence Roadmap 2021-25)					
Mission	To create a THRIVING NATIONAL AI ECOSYSTEM that allows EVERYONE (government, business, & people) to capitalise on the BENEFITS OF AI in a SECURED AND SAFE manner for economic prosperity & social well-being"					
2021-2025 strategic approaches	**Establish AI Governance** including national coordination	**Advance AI R&D** through robust AI R&D ecosystem	**Digital Infrastructure to Enable AI** to target investment	**Fostering AI Talents** to grow local skills & workforce	**Acculturating AI** to increase awareness & adoption	**Kick-Start AI Innovation Ecosystem** with a national innovation hub
2026-2030 strategic approaches	**Extend National AI Governance** National Regulatory Framework, Ethics	**National AI Research** report on national AI trends, publish AI data, AI impact study for gov	**National AI Technology Action Plan** to target investment & national planning	**Fostering AI Talents** continue to mature national & specialist skills, communities of practice	**Accelerate AI** sector based focus on health, transport, agriculture & public sector adoption	**International engagement** critical analysis, collaboration, sharing and showcasing best practices
Leadership	**Activation leadership** The Malaysia National AI Office (NAIO) provide strategic and policy leadership, national coordination for implementation the national AI Strategy and economic outcomes.			**Mitigation leadership** The Malaysia national security Agency provides proactive leadership on identifying, mitigating and responding to emerging threats and criminal misuse of AI.		

Defining your north star, a national destination

Engage with the political, civil service and public aspirations for AI

As most policy makers know, you can't copy and paste a strategy from another country. Every nation has different context, constraints, capabilities and domestic market conditions, so whilst it is helpful to draw inspiration from others, it is critical to forge your own destiny.

It is helpful to refer to (or take the time to develop) a broad aspirational future for the country, including political and public context. Every nation has a different political and cultural context, with different values & ideas for what "good" means for their society & economy. A clear "north star" helps to ensure a strategic and values-aligned approach to navigating new technologies, threats and opportunities, particularly through growing change & uncertainty.

Ideally, a national vision would be developed with public participation, representing the diversity of voices, needs and aspirations from across the entire population. UNDP provides support and toolkits for many nations to enable effective public participation.

A clear national vision provides both a broadly agreed destination to aspire to, which acts to unify & converge efforts across sectors, while also helping policy makers to keep their eye on the horizon while tackling new and ongoing challenges & opportunities.

Activation & mitigation need different programmes of work

(some examples are linked throughout the text)

Activation involved creating the conditions for responsible use: guidance, capability, investment & regulation of willing entities.	**Mitigation involves active threat remediation: coordinated intelligence, cyber security, scam reporting, critical infrastructure, education.**
• **Leadership:** national AI activation leadership is critical, informed by a multi-sector, whole of society approach. • **Programs & Policy:** standards, risk based assurance, guidance & transparency requirements for public sector, safety standards, AI lifecycle guidance, & domestic/public cloud and data policies. • **Legal:** regulation for high risk AI systems, laws to prohibit harmful uses of AI generated content, to maintain privacy protection with AI, & upholding legal obligations of government. • **Skills:** training strategy for economy, education curriculum, labs, university and/or skilled migration targets for critical AI skills (eg, data scientists, data and cloud engineers, AI governance, etc). • **AI Infrastructure:** AI data/model monitoring & other technical guardrails, capability assessments, investment strategies, with clear requirements for AI that have high impact risk to people. • **Innovation:** establish a national AI function, responsible for exploring, testing & encouraging safe adoption and providing exemplary demonstrations & whole of AI lifecycle management. • **Market maturation:** invest in AI startups & industry adoption. • **Safe spaces:** Sandbox & testing areas to safely build maturity.	• **Leadership:** national AI mitigation leadership and intelligence coordination is also critical, ideally separate from activation leadership as national security is necessarily sensitive, and malicious actors rarely responding to policy or regulatory levers. • **Programs & Policy:** anti-scam initiative with public information campaign to interrupt the flow of funds to criminal networks & terrorists, social cohesion program to help identify and remediate the use of AI to target and radicalise vulnerable people, and a critical infrastructure / cyber security program. • **Legal:** legal authorities are required to support national AI threat management/remediation with appropriate oversight. • **Skills:** emergent threat insights can be fed into national capability uplift & skills needed to actively combat threats can be included in national skills framework for AI, including intelligence, analysis, cyber security, AI generated content/software. • **AI Infrastructure protection:** appropriate tools and capabilities to monitor, intercept, escalate and mitigate AI national threats. • **Partnerships:** intelligence sharing & collaboration is critical. • **Preparation:** continuous threat monitoring, planning & foresight.

Where to start with a national AI approach

For governments in the early phases of a national approach to AI, there are some tried and proven first steps that will help:

- Establish a **cross sector advisory group** to help define & design the national AI approach. This group should include AI experts, cross sector representation (including public sector), and community representation. Deliberations from this group would ideally be published publicly so that the general public can follow along, building confidence and feedback.
- Establish a **clear national AI lead**, ideally someone with appropriate expertise/credentials who is considered politically independent and of good reputation. This person should be supported by a cross sector team, authorised to create a program of work to address national capability uplift (including AI literacy), AI safety standards/assurance/guidance, procurement guidance (including privacy and copyright), any regulatory response required, & to create a centre of excellence to experiment with AI, supporting early adoption projects.
- Establish a **national AI threats lead (**or task the relevant national security agency) with responsibility for assessing, monitoring & responding to emergent national threats.
- AI is directly reliant upon data, so ensure there is a **clear national data lead** who is supported to assess & (if required) establish a program of work around data management & quality maturation, particularly in government.

Government use of AI as a policy lever

In most countries, government is the largest employer and controls the largest national ICT expenditure, so government adoption of AI can directly drive national AI maturity and capability uplift. As such, careful, responsible & innovative government adoption should be considered a key policy lever for national strategic AI planning, including:

- The development & public availability of safety & assurance standards
- Establish core principles for responsible AI adoption, such as workforce augmentation rather than displacement, taking risk-based approaches to regulation/assurance based on public impact, and balancing benefit with cost and environmental impact.
- Capability uplift in the government sector means a large workforce being AI literate, leading the way for others, ideally with training made publicly available
- AI adoption & development of solutions for government accelerates domestic sector capability and commercial investment in AI
- If AI usage in government is shown to be responsible & of public benefit, this can improve public confidence and usage of AI, & improve public trust in government
- Responsible use of AI in government can also improve the efficiency, efficacy and public participation in government programs, policies, services and public safety, providing real time insights, patterns analysis, automation & personalisation.

AI for digital transformation of government

"Digital transformation" is a term that can be used in many different ways, referring to anything from basic digitization of paper processes, to a completely reimagined socio-economic future.

The lack of a clear and consistently applied definition makes digital transformation hard to plan for and implement in a meaningful way.

The Digital Transformation Archetypes have been designed to support and clarify discussions on government ambitions for digital transformation with four distinct digital transformation types, each with strategic considerations & guidance for how to achieve them.

Guiding AI with Digital Transformation Archetypes

Governments have been investing in four broad types of digital transformation for many years, which can provide a helpful framework to guide AI adoption in government. Consider the digital transformation objectives of your government partners when discussing relevant AI use cases, as per below.

Service Delivery Reform is where the primary intention is to improve and/or make more efficient the government services to the public. AI use cases include: • AI-enabled multi-lingual chatbots and online services • Automation and simplification of processes/forms • Computer-vision assistance such as health or x-rays • Patterns of sentiment, complaints, helpdesk, FOI • Proactive management of fraud and cyber threats	**Adaptive Policy Reform** is where the primary intention is to better design and deliver policy outcomes in a complex & increasingly volatile world. AI use cases include: • Modeling of policy options, risk, projections, impact • Measure & prioritize investments for efficacy & impact • Scenario provocations, modeling, testing, planning • Continuous monitoring for & escalation of patterns • Automated continuous testing of regulatory settings
Institutional Uplift is where the primary intention is to make public institutions fit for purpose and effective in the modern age and into the future. AI use cases include: • Office productivity, spellcheck, summary, format, tone • Search and (shallow) synthesis of data and documents • Automation of simple, repetitive & testable tasks • Support for software development, deployment, testing • Improved coordination across functions & portfolios • Ability to enable more sensory programs & institutions	**Participatory Governance** is where the primary intention is to engage the general public in co-designing and governing policies and services. AI use cases include: • Engage the public in regular consultation (eg, SMS) using AI to synthesize ideas & sentiment • AI agents can engage widely in multiple languages • Quickly generate mockups/prototypes of public ideas • Supports more people to contribute more easily • Public visibility of programs, waiting times, trends, etc

Responsible AI requires

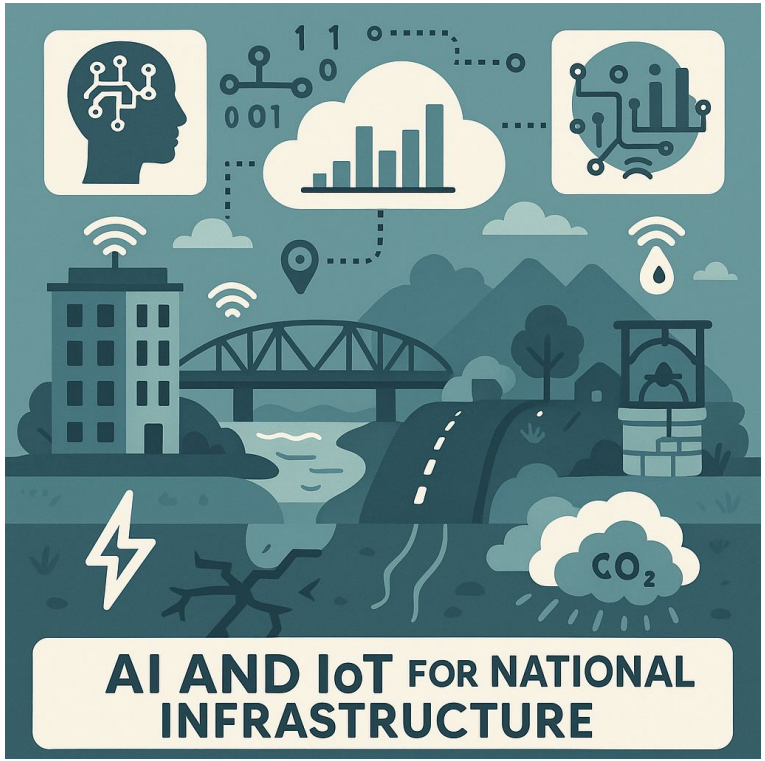
1. Policy cohesion, clarity and adaptive policy management.
2. Manage, monitor and remediate risk dynamically, including false positives.
3. Identify when explainability is a requirement.
4. Establish Feedback Loops & Impact Monitoring.
5. Transparency & participatory governance with *real guardrails*.
6. Design for People + AI and include humane measures of success.
7. Continuous change requires adaptive organisation.
8. Understand what you see (maintain high expertise, use the right tool for the right reason).

Exploring nation building with AI

Many people, vendors and frameworks consider and support the benefits of AI for individual projects, organisations or sectors, but a national strategy provides the opportunity to explore ways to positively **transform society, economy & way of life**.

To follow are four ways national policy makers might use AI as a trigger for **national socio-economic reform**.

Future #1: Sensory cities, resilient nations



AI can identify emergent or historical patterns in enormous data holdings, and as the price of IoT (Internet of Things) devices continues to drop, governments, development agencies and construction organisations could incorporate sensors into public infrastructure to create more evidence to inform investment. From buildings to bridges, roads and wells, wide spread sensors would provide enormous data to inform environmental planning, identify new threats, analyse the impact of investment and of course, to inform natural emergency planning. Imagine a living “national digital twin” where a sudden drop in water, or rise in toxic chemicals triggers a real time national response. Where damage to critical infrastructure could be understood for accurate remediation within minutes or hours of a natural disaster. AI with widely distributed sensors can support a truly resilient nation.

Future #2: Enabling participatory governance

What if AI was used to support more participatory governance, unlocking new ways for governments to easily, respectfully, inclusively and continuously engage citizens and residents across the community on their hopes, fears and ideas for nation building. Anyone could participate in assessing and modeling new policy options, and could participate in real time and participatory drafting of new legislation and regulations. Citizens could propose and gain support for community-led development initiatives and in setting national goals and priorities.

Future #3: AI helpers at home and at work

As people get more comfortable with using AI in their personal lives, they will increasingly explore the use of AI bots at work, not just to automate repetitive or mundane work, but to scale their efforts and expertise for greater impact. For example, coordination of major, time critical or complex programs/initiatives like border management or an emergency response could be supported by delegating certain tasks or outcomes to bots, directed and overseen by humans. Additionally, people are turning to AI helpers instead of official websites/apps, creating a need for service providers (including gov) to optimise for AI.

Future #4: Freeing up social capacity to live better

What if AI provided the opportunity to change our idea of the work week so that individuals had more free time to relax, create and contribute to their communities and civil society. Extra social capacity to participate in policy, education and voluntary efforts. Increased automation, productivity and AI-augmented staff could work more productively in less hours, enabling parents to spend with children and young people, taking the pressure of educators to fill vacuum currently often created by overworked parents, creating the opportunity to improve social & community cohesion and reduce the opportunities for youth radicalisation.

AI Resources and Food for Thought

- As AI giants duel, the Global South builds its own brainpower
- UN Inter-Agency Working Group on AI Principles for Ethical Use in the UN (2022)
- AI Hub for Sustainable Development
- UNDP policy framework for data centres and AI infrastructure
- AI Landscape Analysis (AILA) (example strategy in Sri Lanka)
- Human Rights Impact Assessment of AI systems
- UNDP RID blog on why AI needs to be shaped, not just deployed
- Various UNDP thought pieces on AI
- AI for Sustainable Development (AI4SD)
- Local Language Partnerships Accelerator for linguistic representation in AI
- The Strategic Workforce Planning initiative (OHR/BMS/CB)
- The Artificial Intelligence for Development Analytics (AIDA) system
- Public engagement in shaping AI: What can AI learn from the street?

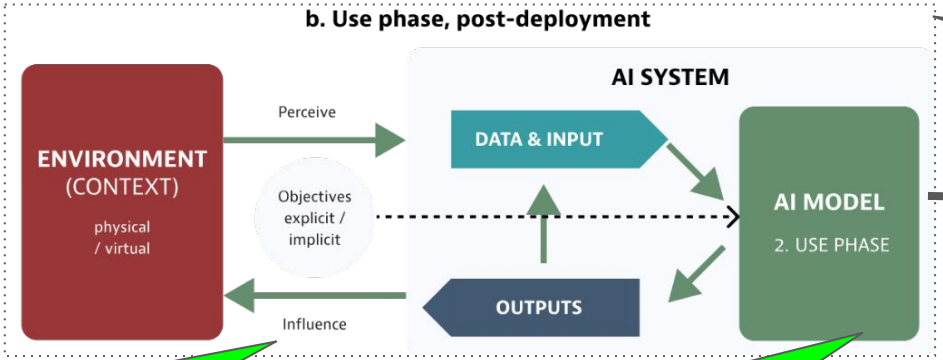
Appendix

1. Managing AI systems that have any risk of high impact
2. Balancing AI “Signage” vs Guardrails for AI governance

Managing high risk AI systems end-to-end

Many vendors will assure you they provide end-to-end management of AI systems you want to use, but high risk AI systems (most usefully defined as AI systems with **any risk of high impact** on the rights or livelihood of people) warrant extra protection with fully accountable, transparent and controlled end-to-end management to ensure AI systems don't go rogue or be manipulated to create negative impacts. The OECD AI system can be complemented by additional requirements for higher risk systems, which also happens to usefully limit some market options, where there is less transparency, accountability or guardrails.

High risk AI systems need full data supply chain transparency & monitoring to avoid data poisoning (note: generally not possible for many public AI tools, creating requirement for internal capabilities for high risk systems)



Operational systems (eg transactional, applications, etc) that use inputs from an AI model.

High risk AI systems need end to end internal AI expertise including data science, engineering & monitoring.

High risk AI systems should monitor for model drift, inputs and outputs for unexpected patterns, and be continuously tested against agreed thresholds.

High risk systems should include a "kill switch" to easily and quickly pause a high risk system if concerns are raised or something goes wrong.

The systems that use AI insights/inputs should have public and staff feedback loops, and the public impacts (intended and unintended) should be continuously monitored for adverse patterns.

More and less trustworthy uses of AI

More trustworthy

Identifying bias in social protection
(analysis ML)

Rules-based tax calculation

Internal knowledge LLM

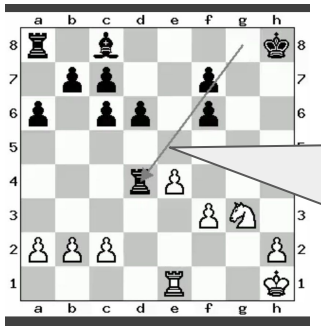
Internal and specialist AI/ML systems can be expertly trained and usefully constrained to provide meaningful insights or moves.

Less trustworthy

Biometric risk profiling
(analysis ML)

ML based tax calculation

Gen AI Chess



Gen AI tools don't explicitly follow rules, so anything rules based results in strange behaviours from a general public LLM.

AI Signage vs AI Guardrails

Signage: warnings & strategies to influence behaviours, but won't stop you falling off the cliff

Guardrails: actually stop or at least slow you falling off a cliff

Many AI programmes have a lot of signage with no guardrails, which is particularly dangerous for **high risk** use cases.

AI Signage

Policies & Ethics:

clear AI definition, usage patterns, co-design, policies, ethics, regulation

Assurance:

tests, compliance & requirements to go-live, security/privacy impact, gov standards

AI capabilities:

AI literacy for all staff, training, impact analysis, data/AI scientists & engineers

(People) Governance:

traditional oversight, escalation, risk mgmt, clear authorities/accountabilities

AI Guardrails

Test/Monitoring:

monitoring of models, outputs, impact, bias, pre-launch & continuous testing

Feedback Loops:

staff, public, advocates, courts, policy scenario testing & assurance

Full Lifecycle Management:

full data/code provenance, oversight, controls, supply chain assurance

(Realtime) Governance:

realtime reporting, notification/escalation, pause/kill switch, funded operational agility