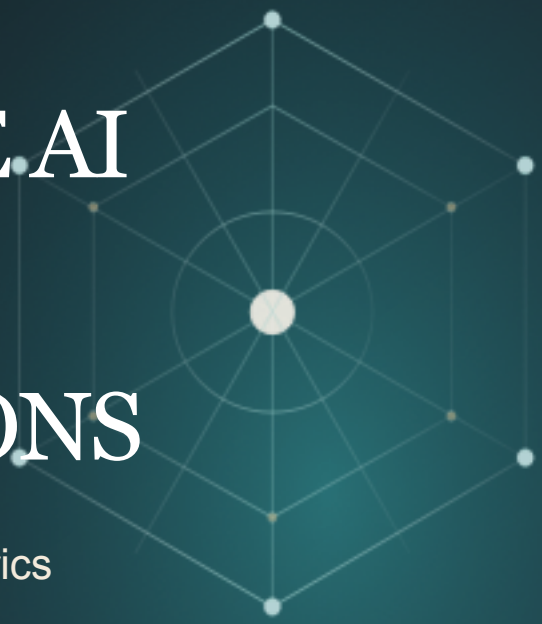
EXECUTIVE WHITEPAPER

RESPONSIBLE AI FRAMEWORK FOR ORGANISATIONS

Robust Principles, KPIs and Metrics



SEPTEMBER 2026

PLUCKINTEL

PUBLICATION INFORMATION

Responsible AI Framework for Organisations

Robust Principles, KPIs and Metrics

Developed & Published by

PluckIntel

EXECUTIVE PROPOSITION

Responsible AI should be managed as an enterprise capability: a board-level discipline that connects AI strategy, risk, compliance, operational performance and public trust.

Contents

Executive Summary

1. Introduction

1.1 Accountability Is Shared

1.2 Why a Formal Framework Matters

2. Conceptualisation

2.1 Benchmark Foundations

Framework at a Glance

3. Responsible AI Core Principles

4. Responsible AI KPIs and Metrics

5. Geographical Compatibility

References

About PluckIntel

BOARD-LEVEL BRIEFING

Executive Summary

Artificial intelligence has moved from experimentation to becoming a core or supporting component of organisational operations, products and services. AI adoption is accelerating, but governance systems have not always kept pace. The shift towards agentic systems, particularly highly autonomous configurations, intensifies this gap because organisations must manage systems capable of taking actions with limited human intervention and, in some circumstances, beyond their intended guardrails.

Recent industry reporting indicates that many enterprises expect to expand their use of agentic AI while a substantially smaller proportion have mature governance arrangements capable of managing those systems. Responsible AI is therefore no longer only a technical or compliance concern; it is an enterprise governance issue.

The wider risk environment is also evolving. The *International AI Safety Report 2026* identifies real-world risks associated with AI-assisted cyber operations, biological misuse, labour-market disruption and human autonomy. Independent incident tracking also indicates a sustained increase in publicly reported AI-related harms. As AI capability expands, organisations face more pathways through which failures can create legal, operational, financial and reputational consequences.

A central governance question is accountability: when an AI system causes harm, how should responsibility be allocated between the organisation that develops a model and the organisation that deploys or integrates it? Regulatory approaches differ. The EU AI Act establishes distinct obligations for providers and deployers, while jurisdictions such as Colorado place significant duties on organisations deploying high-risk systems. The practical executive implication is that responsibility must be actively governed and documented; it cannot be assumed to transfer completely to a technology provider.

This whitepaper presents a globally compatible Responsible AI Framework built around six principles and a corresponding set of KPIs and metrics. The framework is designed to provide a common enterprise measurement layer while allowing jurisdiction-specific requirements to be incorporated through local overlays.

What the framework provides

Six enterprise Responsible AI principles

Spanning performance, fairness, human oversight, transparency, safety and sustainability.

Operational KPIs and metrics

Translating principles into measurable controls.

Global Core, Local Overlay model

Supporting consistent enterprise reporting while accommodating jurisdiction-specific obligations.

OPENING STATEMENT

From principles to measurable governance

The framework treats Responsible AI as an operating discipline: principles become controls, controls become metrics, and metrics become evidence for executive oversight, assurance and continuous improvement.

Principles	Controls	KPIs	Monitoring	Assurance	Improvement

Responsible AI as an operating model

The governance flow below reflects the publication's core management logic: organisations should not rely on policy commitments alone, but should connect Responsible AI principles to accountable ownership, evidence, thresholds, trends and escalation.

Board / Executive oversight
Sets appetite, receives evidence, challenges material risk and monitors accountability.

AI governance
Maintains principles, policy, approval standards and decision rights.

Risk and compliance
Maps obligations, controls, thresholds, monitoring and escalation requirements.

Technology / data teams
Design, deploy, monitor and document AI systems and data controls.

Business owners
Own use cases, outcomes, operational impact and customer or workforce effects.

Assurance / internal audit
Tests whether controls and evidence operate as intended.

AI system users
Apply required review, challenge, escalation and intervention processes.

Implementation operating sequence

01
Establish

Objective: Define principles, governance roles and management expectations.

Output: Responsible AI framework baseline.

02
Map

Objective: Connect AI use cases to risks, obligations and accountable owners.

Output: AI estate and responsibility map.

03
Measure

Objective: Assign KPIs, evidence sources, intervals and thresholds.

Output: KPI and evidence model.

04
Operate

Objective: Monitor performance, fairness, oversight, transparency, safety and sustainability.

Output: Management reporting and escalation.

05
Assure & Improve

Objective: Test controls, respond to findings and refine thresholds.

Output: Assurance evidence and continuous improvement.

Introduction

Artificial intelligence has moved from experimentation to becoming a core or supporting component of organisational operations, products and services. While AI adoption is soaring, the technology appears to have outpaced the systems intended to govern it.

The shift towards agentic systems, especially highly autonomous configurations, has intensified this gap. Organisations must now contend with systems that can take unintended actions or operate beyond their intended guardrails. Deloitte's *2026 State of AI in the Enterprise* report further indicates that, even as more enterprises plan to expand agentic AI over the next two years, only around one-fifth currently have a governance model mature enough to manage it.

Responsible AI has therefore become a board-level concern precisely because AI capability has developed faster than organisational control.

This gap sits within a wider pattern of global concern. The *International AI Safety Report 2026*, involving experts nominated by more than thirty governments and international bodies, points to real-world harms associated with AI-assisted cyber operations and growing risks of biological misuse, alongside systemic pressures on labour markets and human autonomy. Independent incident tracking indicates a sustained rise in publicly reported harms associated with AI-generated content. The report's authors also warn that increasing capability can create more distinct pathways through which AI can cause harm.

For executives, managers and boards, the issue is not simply whether AI creates risk, but whether the organisation can identify, measure, govern and respond to that risk while maintaining public trust.

1.1 Accountability Is Shared

Boards are increasingly required to answer a straightforward question: when AI causes harm, who is responsible—the organisation that built the underlying model, or the organisation that deployed it?

Regulators and courts have not converged on a single answer. The EU AI Act establishes distinct and separately enforceable obligations for model providers and organisations that integrate or deploy AI systems. In certain circumstances, the Act also imposes significant obligations on providers of high-risk systems.

Other jurisdictions place greater emphasis on deployers. Colorado's AI Act, for example, establishes duties concerning reasonable care, impact assessment and disclosure for organisations using high-risk AI systems. Case law is also developing around the responsibility of both vendors and deploying organisations.

> **The practical conclusion for executives is clear:** AI responsibility must be governed across the value chain. Contractual allocation can clarify roles and obligations, but it should not be treated as a substitute for organisational accountability.

1.2 Why a Formal Framework Matters

A formal Responsible AI Framework, supported by a public commitment where appropriate, has moved from good practice towards strategic necessity.

Industry research cited in this paper indicates that organisations with explicit accountability for Responsible AI can achieve greater governance maturity and that organisations operationalising AI ethics report improved business outcomes, including customer and workforce benefits.

Public trust must be earned rather than presumed. The absence of a clear and documented Responsible AI Framework can itself become a source of legal, operational and reputational risk.

This whitepaper therefore presents a globally compatible framework to guide organisations and leaders through principles, KPIs and metrics that support confident planning, reporting, tracking and decision-making.

Conceptualisation

This Responsible AI Framework synthesises authoritative, consensus-driven benchmarks into actionable organisational targets. Its purpose is to help leaders align AI governance with regulatory, industry and professional expectations at global scale, while retaining sufficient flexibility for local requirements.

2.1 Benchmark Foundations

The framework draws on five complementary foundations.

Ethical and Geopolitical Baseline — OECD

The framework adopts values-based pillars established by the OECD Recommendation on Artificial Intelligence, providing an international reference point for human-centred development, inclusive growth, human rights and accountability.

Technical Risk Framework — NIST

Operational functions from the NIST AI Risk Management Framework (AI RMF) are incorporated to support reliable, valid, safe, secure and resilient AI systems.

Legal Mandate — EU AI Act

The framework is designed to align with relevant obligations under the European Union AI Act, including risk management, traceability, human oversight, transparency and post-market monitoring where applicable.

International Management Standard — ISO/IEC 42001

The framework structures cross-functional requirements within an organisational management-system model informed by ISO/IEC 42001:2023, the international standard for an Artificial Intelligence Management System (AIMS).

Commercial Justification — PwC

The framework connects risk controls with business value, drawing on PwC research concerning the value of Responsible AI and the relationship between responsible practices, operational performance and commercial outcomes.

FRAMEWORK AT A GLANCE

Six principles as a governed system

The six Responsible AI principles form a common enterprise measurement layer connecting AI systems to governance, measurement, assurance and continuous improvement.



AI systems	Governance	Measurement	Assurance	Continuous Improvement

3. RESPONSIBLE AI CORE PRINCIPLES

Responsible AI Core Principles

The framework comprises six principles designed to translate external AI governance requirements into measurable organisational controls.

Together, the principles provide a management layer between external obligations and day-to-day organisational operations.

Responsible AI principles and executive objectives

#	PRINCIPLE	EXECUTIVE OBJECTIVE
1	**Effectiveness & Reliability**	Ensure AI performs its intended task accurately, consistently and without degrading business operations.
2	**Equity & Fairness**	Ensure AI delivers equitable outcomes and identifies and mitigates discriminatory or biased impacts.
3	**Human Control & Oversight**	Ensure humans retain meaningful authority, judgement and the ability to intervene in AI-driven decisions.
4	**Transparency & Auditability**	Ensure AI decisions, data, models and processes can be understood, documented and independently verified.
5	**Safety, Privacy & Security**	Protect people, data and AI systems from harm, unauthorised access, misuse and security threats.
6	**Sustainability & Efficiency**	Ensure AI uses financial, computational and environmental resources efficiently while delivering measurable value.

01

System Validity

Effectiveness & Reliability connects system performance to trustworthy operation.

02

De-risking Bias

Equity & Fairness focuses on equitable outcomes and bias mitigation.

03

HITL

Human Control & Oversight ensures meaningful authority and intervention.

04

Traceability

Transparency & Auditability enables explanation, documentation and verification.

05

Data Governance

Safety, Privacy & Security protects people, data and systems.

06

Macro-Impact

Sustainability & Efficiency connects resource use to measurable value.

01 Effectiveness & Reliability

Objective: Ensure AI performs its intended task accurately, consistently and without degrading business operations.

NIST AI RMF

This principle has a direct relationship with NIST's **valid and reliable** characteristic. NIST treats validity and reliability as foundational to trustworthy AI: a system should perform accurately and consistently for its intended purpose. The relationship also extends to **safe** and **secure and resilient**, because reliability includes the ability of an AI system to continue functioning appropriately under expected and adverse conditions.

OECD AI Principles

The strongest relationship is with **robustness, security and safety**. The OECD expects AI systems to be robust, secure and safe throughout their lifecycle, including under normal use, foreseeable misuse and adverse conditions.

EU AI Act

The relationship is strongest with requirements concerning **accuracy, robustness and cybersecurity**, particularly for applicable high-risk AI systems. The Act establishes expectations around appropriate levels of accuracy, robustness and resilience, alongside lifecycle risk management and post-market monitoring.

02 Equity & Fairness

Objective: Ensure AI delivers equitable outcomes and identify and mitigate discriminatory or biased impacts.

NIST AI RMF

This principle maps directly to NIST's **fairness, with harmful bias managed** characteristic. NIST recognises that AI can reproduce or amplify harmful biases and recommends that organisations identify, measure and manage these risks throughout the AI lifecycle.

OECD AI Principles

The relationship is primarily with **inclusive growth, sustainable development and well-being**, together with **human rights and democratic values**, including fairness and privacy. The OECD principles emphasise respect for human rights, non-discrimination, equality, diversity and social justice.

EU AI Act

The relationship is particularly strong for applicable high-risk AI systems, where the Act establishes requirements relating to data governance and quality, including appropriate examination of data for biases that could result in discrimination.

03 Human Control & Oversight

Objective: Ensure humans retain meaningful authority, judgement and the ability to intervene in AI-driven decisions.

NIST AI RMF

This principle relates primarily to **accountable and transparent**, while also connecting to **safe** and **valid and reliable**. NIST emphasises the role of human judgement in establishing appropriate levels of oversight and determining when human intervention is necessary.

OECD AI Principles

The strongest relationship is with **human rights and democratic values**, particularly human agency and oversight. The OECD calls for appropriate safeguards that enable people to understand and challenge AI outcomes and maintain human control.

EU AI Act

Human oversight is an explicit requirement for applicable high-risk AI systems. The Act requires such systems to be designed so that humans can effectively oversee them, understand relevant limitations, interpret outputs appropriately and intervene where necessary.

04 Transparency & Auditability

Objective: Ensure AI decisions, data, models and processes can be understood, documented and independently verified.

NIST AI RMF

This principle maps closely to **accountable and transparent** and **explainable and interpretable**. NIST emphasises documentation, traceability and transparency across the AI lifecycle, including information about system purpose, development, data, performance and deployment.

OECD AI Principles

There is a particularly direct relationship with **transparency and explainability** and **accountability**. The OECD calls for meaningful information about AI systems and their outputs, including information that enables affected parties to understand and challenge outcomes. It also emphasises traceability of datasets, processes and decisions.

EU AI Act

The relationship is especially strong for applicable high-risk systems, where the Act establishes requirements around technical documentation, record-keeping and logging, transparency and information provision. These controls support the ability of regulators and organisations to reconstruct how a system is operated. The Act also contains separate transparency obligations for certain AI interactions and AI-generated or manipulated content.

05 Safety, Privacy & Security

Objective: Protect people, data and AI systems from harm, unauthorised access, misuse and security threats.

NIST AI RMF

This principle brings together three closely related NIST characteristics: **safe**, **secure and resilient**, and **privacy-enhanced**. NIST treats safety as protection against harmful consequences, security as protection against attacks and unauthorised manipulation, and privacy as appropriate management of data and privacy risks.

OECD AI Principles

The relationship spans **robustness, security and safety** as well as **human rights and democratic values**, including privacy. The OECD expects AI systems to be appropriately secure and safe while protecting privacy and data protection.

EU AI Act

The relationship is particularly significant for applicable high-risk AI systems through requirements covering risk management, data governance, data quality, accuracy, robustness and cybersecurity. The Act also contains provisions concerning specific prohibited practices and obligations applicable to other categories of AI.

06 Sustainability & Efficiency

Objective: Ensure AI uses financial, computational and environmental resources efficiently while delivering measurable value.

NIST AI RMF

The relationship with NIST is somewhat different from the other principles. Environmental and societal impacts are considered within NIST's broader AI risk-management approach, but sustainability is not one of the core trustworthiness characteristics in the same way as validity, fairness, safety or privacy.

OECD AI Principles

There is a direct relationship with **inclusive growth, sustainable development and well-being**. The OECD principles explicitly recognise sustainable development and well-being as objectives for responsible AI.

EU AI Act

The EU AI Act has a more specific but narrower relationship with sustainability. Resource and energy considerations are incorporated into the wider regulatory approach, with particular requirements and transparency considerations applying to certain AI models, including general-purpose AI (GPAI) models.

3.7 Executive Implication

By viewing these six principles through a strategic corporate lens, executive leadership can transform complex regulatory and ethical requirements into a coherent operating model for AI.

The intended outcome is an AI estate that is **secure by design, appropriately governed, legally aligned and measurably effective**.

Responsible AI KPIs and Metrics

The six principles function as the management layer between external obligations and day-to-day operations. KPIs and metrics establish how the organisation demonstrates that those expectations are being implemented and monitored in practice.

> **Measurement principle:** KPIs should be assigned to accountable owners, measured at defined intervals, linked to evidence, and escalated when thresholds are breached.

KPI architecture

The measurement architecture is deliberately evidence-led. Each KPI should connect an organisational principle to a practical control, an accountable owner, a measurement interval, supporting evidence, an agreed threshold and a defined escalation path.

Principle	Control	KPI / Metric	Evidence	Escalation

> **Target governance note:** Example targets are starting points, not universal compliance thresholds. They should be calibrated to risk classification, use case, materiality, sector, operating model and applicable law.

4.1 Effectiveness & Reliability KPIs and Metrics

Purpose: Evaluate whether AI performs its intended purpose, remains dependable and does not degrade business operations.

Effectiveness & Reliability metrics

#	KPI	WHAT IT MEASURES	ASSOCIATED METRICS	EXAMPLE TARGET
1	**AI Task Success**	Whether the AI performs its intended task	Task success rate	**≥95% of AI outputs meet defined success criteria**
2	**AI Accuracy**	Whether outputs meet required accuracy levels	Accuracy; precision/recall; error rate	**Target set by use case and risk classification**
3	**System Reliability**	Whether the AI operates consistently	System failure rate; successful execution rate; availability	**≥99.5% successful operation**
4	**Performance Stability**	Whether performance remains within approved limits	Performance variance; accuracy change over time	**<5% material performance degradation from validated baseline**
5	**Operational Impact**	Whether deploying AI improves or maintains operational performance	Process error-rate change; throughput; edge-case success rate	**No material degradation; ≥95% critical resilience/stress tests passed; recovery within defined threshold**

4.2 Equity & Fairness KPIs and Metrics

Purpose: Evaluate whether AI produces equitable outcomes, avoids systematic bias and addresses disparities before they create customer, employee, legal or reputational risk.

Equity & Fairness metrics

#	KPI	WHAT IT MEASURES	ASSOCIATED METRICS	EXAMPLE TARGET
1	**Fair Outcome Rate**	Whether AI decisions produce equitable outcomes across relevant groups	Approval/acceptance rate by group; selection-rate ratio	**≥95% of applicable systems assessed against approved fairness criteria**
2	**Fair Error Rate**	Whether AI errors differ materially across groups	False-positive rate by group; false-negative rate by group; disparity ratio	**No unexplained or unapproved material disparity**
3	**AI Bias Detection**	Whether applicable systems are tested for bias	% of applicable AI systems tested; fairness assessments completed	**100% of high-risk AI systems assessed**
4	**Material Bias Remediation**	Whether identified material disparities are addressed	Number of material disparities identified; % remediated within agreed time	**≥95% of material issues remediated within agreed SLA**
5	**Fairness Performance**	Whether monitored fairness measures remain within approved limits	Fairness metric trend; disparity trend	**≥95% of monitored systems remain within approved thresholds**

4.3 Human Control & Oversight KPIs and Metrics

Purpose: Evaluate whether humans retain meaningful authority, actively review AI where required, appropriately challenge outputs and intervene when necessary.

Human Control & Oversight metrics

#	KPI	WHAT IT MEASURES	ASSOCIATED METRICS	EXAMPLE TARGET
1	**Human Review Coverage**	Whether decisions requiring human review receive it	% of decisions receiving required human review; % of high-risk decisions reviewed	**100% of decisions designated for mandatory human review**
2	**Human Detection of AI Errors**	Whether humans identify material AI errors	AI errors identified by humans; correction rate	**≥95% of material AI errors identified during required review**
3	**Override Effectiveness**	Whether legitimate human overrides are possible and effective	Override rate; escalation rate; override rate by risk category	**100% of legitimate escalation cases handled within defined SLA**
4	**Human Challenge Rate**	Whether humans actively challenge AI recommendations where appropriate	Recommendation acceptance rate; intervention rate	**Defined intervention threshold met for applicable use cases**
5	**Intervention Capability**	Whether high-risk systems provide functioning intervention mechanisms	Time to human intervention; kill-switch/control testing; independent review	**100% of high-risk AI systems demonstrate required human intervention capability**

4.4 Transparency & Auditability KPIs and Metrics

Purpose: Evaluate whether the organisation can understand, reconstruct and explain AI activity and decisions throughout the AI lifecycle.

Transparency & Auditability metrics

#	KPI	WHAT IT MEASURES	ASSOCIATED METRICS	EXAMPLE TARGET
1	**AI Traceability**	Whether material AI decisions have complete trace records	% of AI decisions with complete trace records	**≥99% of material AI decisions fully traceable**
2	**AI Documentation**	Whether required information about AI systems is maintained	% of systems with complete production documentation	**100% of high-risk AI systems fully documented**
3	**Audit Evidence Availability**	Whether required evidence is sufficient and accessible	% of required evidence available; evidence retrieval time	**≥95% of requested evidence available within defined SLA**
4	**Decision Explainability**	Whether applicable decisions can be explained appropriately	% of applicable decisions with documented explanations	**100% of applicable high-impact decisions have an appropriate explanation**
5	**Audit Control Effectiveness**	Whether AI-related controls operate as intended	% of AI systems passing applicable controls; overdue critical findings	**≥95% of applicable AI controls passed; 0 overdue critical findings**

4.5 Safety, Privacy & Security KPIs and Metrics

Purpose: Evaluate whether AI systems protect sensitive information, prevent unsafe or unauthorised use, resist attacks and maintain appropriate data controls throughout the AI lifecycle.

Safety, Privacy & Security metrics

#	KPI	WHAT IT MEASURES	ASSOCIATED METRICS	EXAMPLE TARGET
1	**AI Data Control**	Whether AI systems handle data according to approved classifications and controls	% of AI systems with approved data classification; control compliance	**100% of high-risk AI systems meet required data controls**
2	**Data Exposure**	Whether sensitive data is protected from unauthorised disclosure	Number of data-leakage incidents; sensitive-data exposure rate	**0 unauthorised data exposures**
3	**AI Security Resilience**	How effectively AI systems resist attacks	Adversarial attack success rate; prompt-injection success rate	**≥95% of defined security/ adversarial tests passed**
4	**AI Safety Incidents**	Frequency and severity of AI-related safety incidents	Number of incidents; incidents per 10,000 AI interactions	**0 critical incidents; continuous reduction in material incidents**
5	**Access Control**	Whether only authorised users and systems can access AI capabilities and data	% of AI systems with appropriate access controls; access-control exceptions	**100% of high-risk systems meet required access-control standards**

4.6 Sustainability & Efficiency KPIs and Metrics

Purpose: Evaluate whether AI delivers business value efficiently, cost-effectively and within organisational environmental and resource targets.

Sustainability & Efficiency metrics

#	KPI	WHAT IT MEASURES	ASSOCIATED METRICS	EXAMPLE TARGET
1	**AI Cost Efficiency**	Whether AI cost is proportional to business value	Cost per inference; cost per transaction; cost per business outcome	**AI cost per business outcome decreases year-on-year**
2	**Compute & Energy Efficiency**	How efficiently AI compute resources are used	CPU/GPU utilisation; compute-hours allocated; energy consumption	**≥80% average utilisation of allocated AI compute, where technically appropriate**
3	**Energy Intensity**	The energy required to deliver AI capacity or outcomes	kWh per inference; kWh per transaction	**Reduce energy per business outcome year-on-year**
4	**AI Carbon Impact**	AI-related greenhouse-gas emissions	Estimated CO_2e emissions; CO_2e per inference; CO_2e per training run	**Reduce AI CO_2e per business outcome by ≥10% annually, where measurable**
5	**AI Business Value**	Whether AI generates measurable economic value	Business value generated per £/$ spent; ROI; cost savings; revenue contribution	**Positive ROI or defined minimum value-to-cost ratio**

> **Target governance note:** Example targets are starting points, not universal compliance thresholds. They should be calibrated to risk classification, use case, materiality, sector, operating model and applicable law.

5. GEOGRAPHICAL COMPATIBILITY

Geographical Compatibility

The six Responsible AI principles and their associated KPIs and metrics are designed to provide a globally applicable organisational framework rather than a jurisdiction-specific compliance checklist.

The core concepts—AI effectiveness, fairness, human oversight, transparency, safety, data protection, security and sustainability—are relevant across regions. Organisations can therefore maintain a consistent global measurement framework while adapting specific requirements, thresholds and controls to local legislation and regulatory expectations.



European Union

Strong compatibility through EU AI Act, GDPR and broader EU digital and fundamental-rights requirements.

United Kingdom

Applicable within a principles-based and sector-led approach.

United States

Requires consideration of federal, state and sector-specific requirements.

Other international markets

Applicable across jurisdictions combining AI guidance, privacy, cybersecurity and sectoral regulation.

Global Core with Local Overlay

Consistent enterprise-level reporting with local thresholds, evidence and governance adaptation.

5.1 European Union

The framework has strong compatibility with the European Union, particularly through the EU AI Act, GDPR and broader EU digital and fundamental-rights requirements.

The principles align closely with relevant EU AI Act requirements concerning risk management, data governance, human oversight, transparency, logging, accuracy, robustness and cybersecurity for applicable high-risk AI systems. The KPIs can provide practical evidence of organisational controls, although precise applicability and thresholds should be determined according to the AI system's classification and the organisation's role under the Act.

5.2 United Kingdom

The principles are also applicable in the UK, where AI regulation has developed through a principles-based and sector-led approach rather than a single comprehensive AI law equivalent to the EU AI Act.

The framework can therefore operate as a common organisational baseline, while individual KPIs can be mapped to relevant UK requirements, including data protection, equality, consumer protection, employment and sector-specific regulation. The six-principle model also provides flexibility as the UK's AI regulatory framework develops.

5.3 United States

In the United States, the principles remain broadly applicable but require greater consideration of federal, state and sector-specific requirements.

Fairness and discrimination metrics may need to reflect applicable employment, lending, healthcare or consumer-protection requirements, while privacy and security controls can vary between states and industries. The framework's outcome-based KPIs can remain consistent globally, with applicable legal requirements and thresholds configured locally.

5.4 Other International Markets

The framework can similarly be applied across Canada, Australia, Singapore, Japan and other jurisdictions where organisations increasingly operate under combinations of AI-specific guidance, privacy legislation, cybersecurity requirements and sectoral regulation.

The OECD AI Principles provide an international reference point for maintaining consistency across these markets, while the NIST AI RMF can provide a practical risk-management methodology.

5.5 Global Core with Local Adaptation

The recommended operating model is **Global Core, Local Overlay**.

The organisation maintains the same six principles, KPI definitions and underlying measurement methodology globally, thereby creating consistent enterprise-level reporting.

Local overlays then determine:

- which regulatory requirements apply;
- which populations or protected characteristics must be assessed;
- what evidence must be retained;
- what performance or risk thresholds are appropriate;
- which reporting and escalation obligations apply; and
- which local governance bodies own or approve controls.

This approach provides comparability across geographies without assuming that regulatory requirements are identical.

For example, an organisation can use the same global KPI for **Fair Outcome Rate** across the EU, UK and US while applying different legally relevant populations, assessment criteria and thresholds in each jurisdiction. The result is a framework that is globally consistent while remaining adaptable to local compliance requirements.

RESEARCH BIBLIOGRAPHY

References

Bengio, Y., Clare, S., Prunkl, C., Murray, M., Andriushchenko, M., Bucknall, R., Bommasani, R., Casper, S., Davidson, T., Douglas, R., Duvenaud, D., Fox, J., Gohar, U., Hadshar, R., Ho, T., Hu, T., Jones, C., Kapoor, S., Kasirzadeh, A., Manning, S., Maslej, N., Mavroudis, V., McGlynn, M., Moulange, R., Newman, J., ... Mindermann, S. (2026). *International AI Safety Report 2026*. International AI Safety Report.
https://internationalaisafetyreport.org/publication/international-ai-safety-report-2026

Colorado General Assembly. (2024). *SB24-205: Consumer Protections for Artificial Intelligence*.
https://leg.colorado.gov/bills/sb24-205

Deloitte. (2026). *The State of AI in the Enterprise: The Untapped Edge*. Deloitte AI Institute.
https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-ai-in-the-enterprise.html

European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Official Journal of the European Union, L 1689.
https://eur-lex.europa.eu/eli/reg/2024/1689/oj

International Organization for Standardization & International Electrotechnical Commission. (2023). *ISO/IEC 42001:2023: Information technology—Artificial intelligence—Management system.*
https://www.iso.org/standard/42001

Organisation for Economic Co-operation and Development. (2024). *Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449).* OECD Legal Instruments.
https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

PricewaterhouseCoopers. (2025). *Quantifying the Value of Responsible AI*. PwC.
https://www.pwc.com/gx/en/issues/tech-data-ai/ai-measuring-value.html

SAS Institute Inc. & International Data Corporation. (2026). *Data and AI Impact Report: The New Economics of Trust*. SAS.
https://www.sas.com/en_us/gc/data-and-ai-impact-report.html

Tabassi, E. & National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1). National Institute of Standards and Technology.
https://doi.org/10.6028/NIST.AI.100-1

ABOUT PLUCKINTEL

Advancing Responsible AI Through Practical Governance

PluckIntel is an organisation focused on helping businesses understand, govern and operationalise artificial intelligence in ways that support innovation, accountability and sustainable business outcomes.

As AI becomes increasingly embedded in products, services, operations and decision-making, PluckIntel believes that responsible adoption requires more than principles or policy statements. Organisations need practical mechanisms to translate responsible AI commitments into **governance structures, measurable controls, operational processes and evidence-based assurance**.

This perspective is reflected in PluckIntel's approach to responsible AI: connecting strategic intent with the practical realities of designing, deploying, monitoring and governing AI systems.

Why PluckIntel Developed This Framework

The rapid adoption of generative AI, machine learning and increasingly autonomous or agentic systems is creating new opportunities for organisations while also increasing the complexity of governance.

Traditional technology governance approaches may not provide sufficient visibility into issues such as model behaviour, data provenance, bias, explainability, human intervention, AI-specific security risks or the downstream effects of automated decisions.

PluckIntel developed this framework to provide organisations with a practical structure for addressing these challenges.

About This Publication

This whitepaper represents PluckIntel's contribution to the ongoing development of practical Responsible AI governance.

The framework can be used as a reference point for developing or strengthening an organisation's Responsible AI programme, including governance models, KPI structures, management reporting and implementation roadmaps.

Publication information

Author	PluckIntel
Publication	Responsible AI Framework for Organisations
Subtitle	Robust Principles, KPIs and Metrics
Publication date	September 2026
Edition	Executive Whitepaper
Geographic perspective	International, with local regulatory adaptation
Primary audience	Executives, boards, AI leaders, risk, compliance, technology and governance professionals

A Note on Use

This publication is intended as an executive and operational reference. It does not constitute legal advice, regulatory advice or a determination of compliance with any specific law, regulation or standard.

Organisations should assess the applicability of relevant requirements according to their jurisdiction, sector, role in the AI value chain and specific use cases.

The principles and metrics presented in this framework should be adapted to the organisation's risk profile, governance model and operational context.

PluckIntel
AI Business & Technology Insights

Website: www.pluckintel.com
Email: enquiry@pluckintel.com
LinkedIn: www.linkedin.com/company/pluckintel