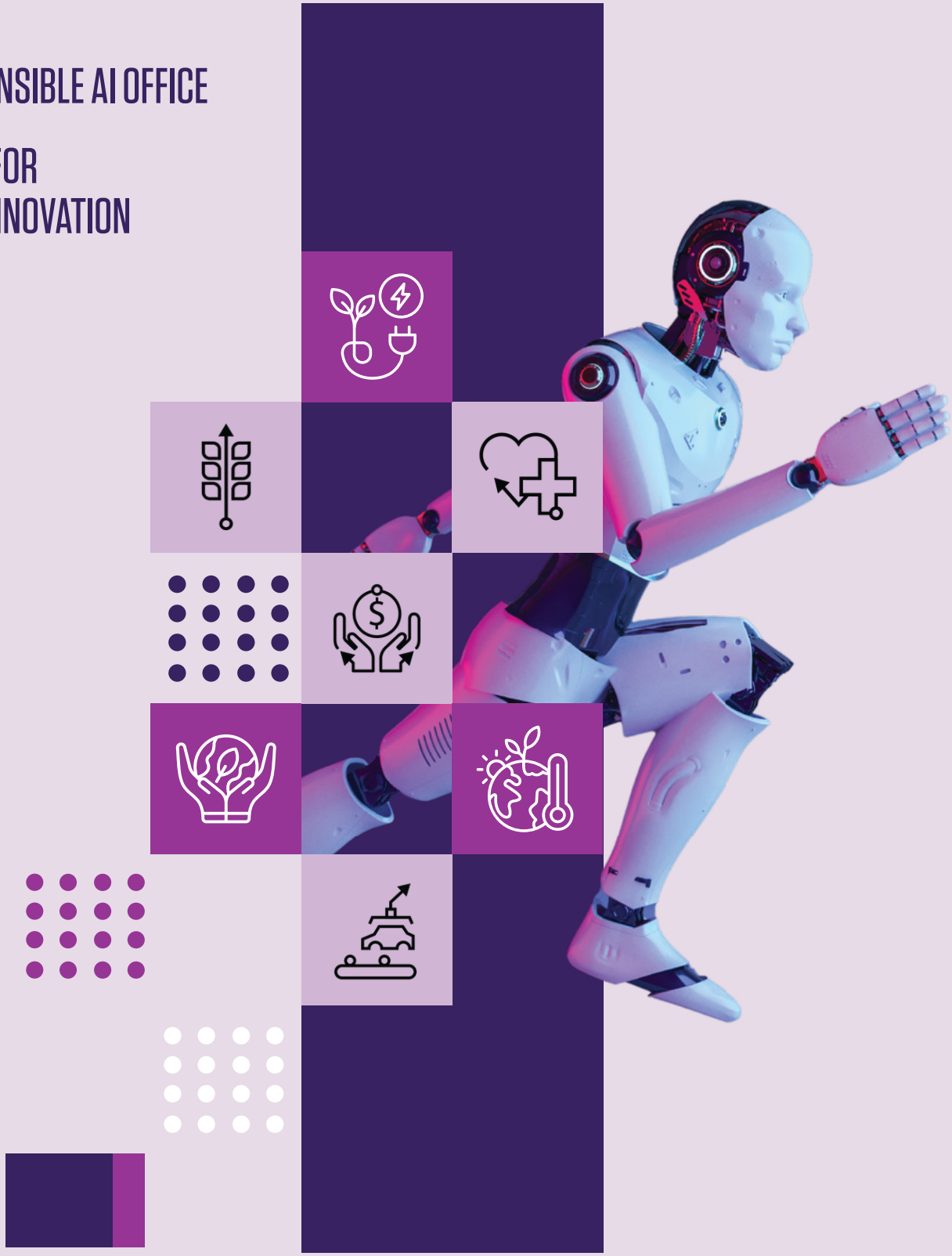
UK-INDIA AIXCELERATE

2025-26

INFOSYS RESPONSIBLE AI OFFICE

STARTER PACK FOR RESPONSIBLE INNOVATION



UK Government

The Dialogue®
INFORM ENGAGE IDEATE

Infosys®
Navigate your next

Foreword

Artificial Intelligence is transforming economies and societies at unprecedented speed, making robust and forward-looking AI governance a shared priority for India and the UK. As AI advances across healthcare, bioengineering, climate, fintech, and agriculture, ensuring that innovation is ethical, transparent, safe, and inclusive is critical. Infosys is proud to work closely with the British High Commission to strengthen India–UK cooperation in responsible AI, fostering collaboration between policymakers, researchers, startups, and enterprises. Together, we are contributing to a governance-driven innovation ecosystem that balances technological progress with accountability and public trust.

Central to this effort is the Infosys Responsible AI Starter Pack, designed to translate AI governance principles into practical action. Through structured frameworks, assessment tools, and implementation guidance, the Starter Pack enables AI solutions to embed fairness, transparency, safety, and compliance from the earliest stages of development. By operationalizing responsible AI at scale, it empowers organizations to innovate confidently while reinforcing trust, strengthening bilateral collaboration, and delivering sustainable value for both nations.

Syed Ahmed
Global Head
Infosys Responsible AI Office

Foreword

Artificial Intelligence (AI) is transforming how we deliver public services, grow our economies, and respond to shared global challenges. As its influence expands, so too does our collective responsibility to ensure that AI systems are developed and deployed in ways that are ethical, transparent, and worthy of public trust. The United Kingdom and India share deep democratic values and a strong tradition of innovation, placing us in a unique position to work together in shaping responsible and forwardlooking AI governance.

The British High Commission in India is pleased to support closer collaboration between our two nations in this important area. Our partnership with Infosys reflects a shared belief that responsible innovation flourishes when policymakers, researchers, startups, and industry leaders work side by side. By strengthening dialogue and cooperation, we are helping to build a trusted AI ecosystem that balances ambition with accountability and innovation with inclusion.

The 'UK-India AIxcelerate Programme' is a unique flagship initiative under the 'India–UK Joint Centre for Artificial Intelligence', established as part of the UK-India Technology Security Initiative (TSI). This bilateral platform with The Dialogue and Infosys aims to promote responsible, inclusive, and high-impact AI through collaboration in research, innovation, and policy. The programme aligns with the India-UK Vision 2035 and supports the digital and technology ambitions under the UK-India Free Trade Agreement (FTA).

The Responsible AI Starter Pack developed by Infosys as part of this programme is a practical and timely contribution to our shared journey to advance responsible and trustworthy AI across health, engineering biology, climate, energy, agritech, and fintech. By translating governance principles into clear, actionable guidance, it supports organizations in embedding fairness, transparency, safety, and accountability from the earliest stages of development. Initiatives like this demonstrate how India and the UK can move beyond dialogue to delivery advancing AI that drives progress, earning and sustaining public confidence.

Joshua Bamford
Head of Science, Tech and Innovation
British High Commission, New Delhi

Table Of Contents

Executive Summary 5

Introduction: AI Threats and Risks – Why Responsible AI Matters.....6

Responsible AI Principles.....7

Principles in Global and National Regulations.....7

Responsible AI Governance.....9

Embedding Responsible AI Into the AI Lifecycle.....12

Data Preparation.....13

Model Development & Training.....13

Model Validation.....13

Model Compliance & Approval.....14

Deployment.....14

Production & Monitoring.....15

Retirement / Phase-Out.....15

Metrics for AI Solution.....16

Responsible AI Principles – Assessment Questionnaire.....19

Responsible AI Principles – Adherence Assessment.....21

Regulatory Landscape by Sector – India & UK.....24

Responsible AI in Healthcare.....24

Responsible AI in Agriculture.....26

Responsible AI in FinTech.....28

Responsible AI in Climate & Energy.....30

Responsible AI in Engineering Bioengineering.....33

AI Development with Ethics: Recommended Toolkits.....35

References.....36

Contribution.....38

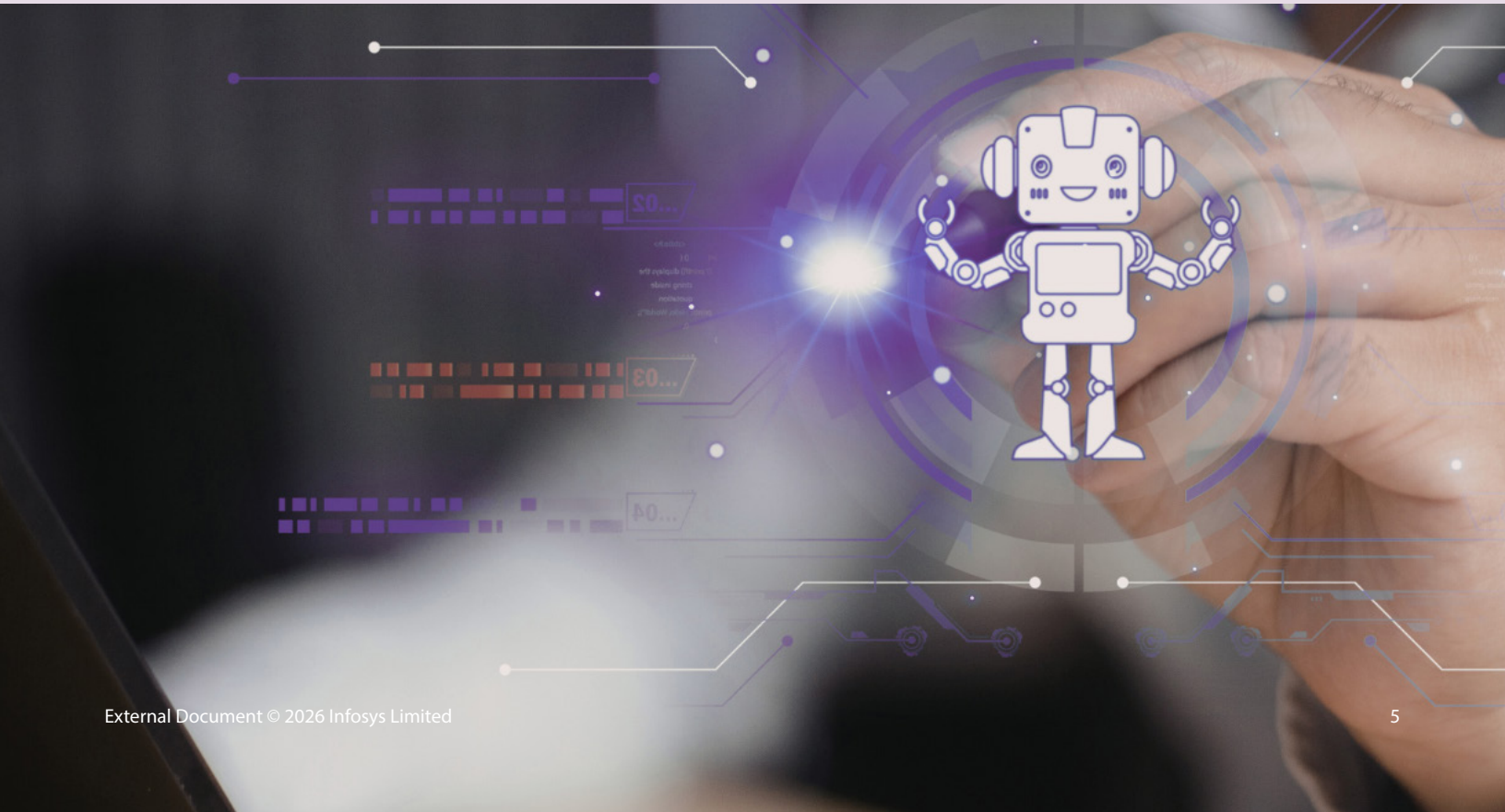
Executive Summary

Artificial Intelligence (AI) is reshaping industries, economies, and societies at an unprecedented pace, unlocking innovation and efficiency while also introducing profound ethical and governance challenges. As AI systems increasingly influence human lives and critical decisions, the need for Responsible AI, AI that is transparent, fair, accountable, safe, and aligned with human values, has become essential. Responsible AI ensures that technology development not only advances capability but also safeguards privacy, promotes inclusivity, and upholds trust across diverse communities.

Building and deploying AI responsibly requires strong governance frameworks and adherence to international and national standards. Globally, initiatives such as the OECD AI Principles, ISO/IEC 42001, NIST AI Risk Management Framework, and the EU AI Act establish clear expectations for risk classification, data transparency, and human oversight. Nationally, India's NITI Aayog's Responsible AI for All, MeitY's AI Ethics Guidelines, and the Digital Personal Data Protection (DPDP) Act 2023, set the foundation for ethical and secure AI adoption. Similarly, the UK's PRA SS1/23 and ISO-aligned governance practices, and emerging U.S. state-level regulations such as those in Texas, emphasize accountability and safety in sectoral AI use.

Responsible AI must be embedded throughout the AI lifecycle, from ethical data collection and model development to deployment and monitoring. This integrated approach transforms responsible practice from a compliance requirement into a continuous process of ethical assurance. The following sections of this document further examine global standards, market scan insights, and sectorspecific regulations across India and the UK, offering organizations practical pathways to align innovation with accountability and build AI systems that are trustworthy, explainable, and beneficial to society.

This starter pack serves as a practical guide for developers to implement AI solutions responsibly. It provides actionable insights, checklists, and governance frameworks that translate high-level principles into dayto-day practices across the AI lifecycle. By following the guidance outlined here, developers can ensure ethical data handling, robust model design, transparent decision-making, and continuous monitoring, helping them build AI systems that are compliant, fair, secure, and aligned with both Indian and UK regulatory expectations. Ultimately, it equips developers with the tools and knowledge to create AI solutions that are trustworthy, inclusive, and socially beneficial.



Introduction: AI Threats and Risks - Why Responsible AI Matters?

Artificial Intelligence (AI) has rapidly evolved from a promising technology to a pervasive force shaping industries, economies, and societies. While its potential to drive innovation and solve global challenges is immense, AI also introduces a complex web of threats and risks that cannot be ignored. The urgency for Responsible AI arises not only from ethical imperatives but also from the need to safeguard human rights, social stability, and long-term trust in technology.

One of the most immediate concerns is automation-driven job displacement, as intelligent systems replace human labor across sectors, potentially widening socioeconomic divides and exacerbating inequality. At the same time, algorithmic bias resulting from poor or unrepresentative data can lead to unfair treatment, discrimination, and loss of public trust. Privacy violations have become increasingly common, as AI systems rely on vast amounts of personal data, often without explicit consent or transparency.

The rise of deepfakes and AI-driven misinformation poses serious risks to democracy and social cohesion by blurring the line between truth and manipulation. Likewise, psychological harm, from overreliance on AI companions to exposure to manipulative content, raises new questions about human well-being in digital environments. AI's integration into weapons systems and military automation amplifies the potential for unintended or unethical use in conflict zones. Furthermore, market volatility driven by algorithmic trading and AI-enabled criminal activity, including cyberattacks and exploitation of children, highlights how technological advancement can outpace regulation and oversight.

These threats underscore why Responsible AI is not an optional consideration but a foundational necessity. It calls for a balanced approach, harnessing AI's transformative capabilities while embedding safeguards for fairness, accountability, transparency, and human dignity. Only through deliberate design, governance, and cross-sector collaboration can societies ensure that AI serves humanity, rather than endangering it.

Responsible AI Principles

Responsible AI principles define how AI should behave and be governed in real-world contexts. They form the foundation of regulatory compliance, ethical design, and public trust, setting clear expectations for how AI must be built, governed, and deployed. Acting as both a moral compass and operational guardrail, these principles ensure innovation does not outpace safety, transparency, or accountability. From financial models to agricultural advisory tools, climate forecasting systems, and medical diagnostics, they guide developers, policymakers, and organizations in creating AI systems that are trustworthy, inclusive, and resilient.

Principles in Global and National Regulations

Across the world, Responsible AI principles have evolved from ethical ideals into enforceable regulatory standards. Leading governments and global organizations now define how these values must translate into practice:

- India – India's Principles for Responsible AI (2021) and National AI Strategy (2018) identify seven foundational principles: transparency, accountability, fairness, privacy, inclusiveness, human-centricity, and sustainability.
- United Kingdom – (ICO) embeds transparency, fairness, and accountability directly into compliance obligations through its AI & Data Protection Guidance.
- Global – The (OECD) AI Principles (2019) and the AI Act (2024) establish structured governance and risk-based oversight for trustworthy AI.
- Standards – International standards such as ISO/IEC 23894 (AI Risk Management) and ISO/IEC 42001 (AI Management Systems) transform these guiding principles into auditable processes and certification pathways.

Together, these frameworks broadly point toward eight commonly cited Responsible AI principles-though expressed in varying ways, they appear to emphasize transparency, accountability, safety, fairness, privacy, human-centricity, inclusiveness, and sustainability.

Responsible AI Principles	Explanation
Transparency (incl. explainability)	AI systems must clearly communicate how they work - including their data sources, decision logic, and limitations - in a way understandable to users, regulators, and auditors.
Accountability & Governance	Humans remain legally and ethically responsible for AI outcomes. Clear roles, documentation, and oversight structures must be in place to ensure compliance and liability.
Safety & Robustness	AI must operate reliably under different conditions, resist adversarial attacks, and degrade safely if failures occur. Ongoing monitoring and risk mitigation are required.
Fairness & Non-Discrimination	AI should not reinforce or create bias. Systems must ensure equal treatment for similar cases and undergo testing to identify and correct disparities.
Privacy & Data Governance	Personal data must be collected and processed lawfully, securely, and with user consent. Rights such as rectification and deletion must be respected.
Human-Centricity & Values	AI should support - not replace - human decision-making in critical areas, and respect dignity, freedom, and democratic values.
Inclusiveness & Accessibility	AI must be designed to benefit diverse groups, considering linguistic, socioeconomic, regional, and ability-based differences to ensure equal access.
Sustainability & Social Impact	AI should minimize environmental harm and contribute positively to longterm social well-being and sustainable development goals.

Responsible AI Governance

Responsible AI Governance establishes the foundation for ethical, transparent, and accountable AI systems by defining how organizations design, deploy, and monitor AI responsibly. It sets out the structures, policies, and oversight mechanisms that ensure AI operations align with legal, ethical, and societal expectations. By embedding clear accountability, risk management, and transparency processes, Responsible AI Governance enables organizations to balance innovation with safety, foster public trust, and promote consistent, trustworthy AI practices across national and global contexts.

The significance of Responsible AI Governance lies in its ability to turn ethical intent into actionable practice. Without strong governance, even well-intentioned AI systems can unintentionally amplify bias, compromise privacy, or erode public trust. A robust governance framework establishes clear mechanisms for accountability, compliance, and risk management—ensuring that AI innovation remains grounded in human values and societal expectations. It enables organizations to go beyond mere regulatory adherence, fostering a culture of integrity, transparency, and responsible stewardship that builds public confidence and supports the long-term sustainability of the AI ecosystem.

Global / Multi-lateral AI Regulations, Frameworks & Standards

Instrument	Summary	Issuer	Year	Category	Key takeaways	Source
EU Artificial Intelligence Act	First horizontal, risk-based AI law with bans, high-risk obligations, and GPAI requirements.	EU	2024	Law	Classify systems; build documentation, logging, risk controls, and human oversight; prepare for 2025– 26 compliance milestones.	Link
UN GA Resolution 78/265 on AI	First UN-wide consensus resolution urging safe, secure, trustworthy AI for SDGs.	United Nations	2024	Reso-lution / Principles	Baseline for international cooperation, rights-based AI, and ethical governance.	Link
G7 Hiroshima AI Code of Conduct	Non-binding lifecycle risk management code for advanced AI developers.	G7	2023	Guidelines	Use as blueprint for red-teaming, safety evaluation, incident reporting, transparency.	Link
OECD AI Principles	Foundational intergovernmental framework for trustworthy AI.	OECD	2019 (updated in 2024)	Principles / Frame-work	Global ethics baseline - transparency, accountability, robustness, human rights.	Link
ISO/IEC 42001	First certifiable AI Management System standard.	ISO/IEC	2023	Standard	Set up AI governance processes: roles, controls, audits, risk registers.	Link
ISO/IEC 23894	Standard for AI risk management.	ISO/IEC	2023	Standard / Guidance	Use for structured risk identification, treatment, monitoring.	Link
NIST AI RMF 1.0	Voluntary risk management framework to govern, map, measure, manage AI.	NIST (USA)	2023	Frame-work	Practical lifecycle controls aligning with ISO and EU AI Act.	Link
NIST SP 800 -218A	Secure software dev profile for GenAI & dual-use foundation models.	NIST (USA)	2024	Guidance	Integrate security, provenance, evals, red-teaming, response.	Link
Singapore Model AI Governance Framework for Generative AI	Practical 9-dimension governance framework for GenAI.	Singapor e IMDA	2024	Frame-work / Guidelines	Concrete controls for provenance, content safe-guards, evaluations.	Link

India

Instrument	Summary	Issuer	Year	Category	Key takeaways	Source
Digital Personal Data Protection Act, 2023	India's baseline personal data protection law.	MeitY / India	2023	Law	Consent, purpose limitation, rights enablement, governance integration in AI lifecycle.	Link
Responsible AI for All Framework	A national framework promoting inclusive, ethical, and locally validated AI development, especially in sectors like agri-tech. It emphasizes fairness, transparency, and indigenous innovation.	NITI Aayog / India	2021	Principles / Policy	Focus on inclusion, fairness, and local validation. Use indigenous tools, self-assessment checklists, and governance frameworks. Align with Safe & Trusted AI Pillar under IndiaAI Mission	Link
MeitY AI Advisory 2024	Interim advisory on labelling, disclaimers, and accountability for AI.	MeitY / India	2024	Advisory	Add "under-trial" labels, ensure testing, election & risk safeguards.	Link
IndiaAI Governance Guidelines (Consultation)	National consultation toward responsible AI governance framework.	India	2025	Draft Framework	Align early with OECD/EU/ NIST practices; expect future compliance alignment.	Link

United Kingdom

Instrument	Summary	Issuer	Year	Category	Key takeaways	Source
UK AI White Paper	Principles-led, regulator-driven AI governance model.	DSIT / UK	2023–24	Policy / Principles	Align to 5 principles: safety, transparency, fairness, accountability, contestability.	Link
ICO Guidance on AI and Data Protection	Practical guidance and toolkit for AI & data protection.	ICO / UK	2023–24	Guidance	Run AI DPIAs, manage explainability, lawful basis, fairness, rights handling.	Link

Embedding Responsible AI Into the AI Lifecycle

A Developer's Guide with UK & India Alignment.
Responsible AI is not merely a one-time compliance checklist but a continuous design principle that must be embedded throughout the AI lifecycle. From data collection and preparation to model development, deployment, and ongoing monitoring, developers should integrate Responsible AI practices to ensure trust, fairness, transparency, and safety at every stage. This approach promotes ethical value delivery, mitigates risks of bias or unintended harm, and ensures that AI systems remain accountable and aligned with both organizational values and regulatory requirements. By making Responsible AI a core, iterative part of development, organizations can build solutions that are reliable, compliant, and socially responsible, fostering stakeholder confidence and long-term sustainability.

The first step involves clearly defining the AI system's purpose, scope, and its potential short-term and long-term risks to ensure transparency and accountability from the outset. To support this, an ethical committee should be established comprising domain experts, social scientists, data scientists, and key stakeholders. This committee will be responsible for assessing the social, economic, and ethical implications of AI deployment, ensuring that potential adverse impacts are identified early. Additionally, the committee should recommend practical guidelines and measures for risk mitigation, fairness, and inclusion, fostering responsible and equitable AI adoption across all stages of the system's lifecycle.

To ensure accountability and trust, organizations should establish a robust grievance redressal mechanism for stakeholders who may be affected by AI system decisions. It is essential to identify and document explainability requirements tailored to different user groups, ensuring that AI decisions can be understood by both technical and non-technical audiences. For scenarios involving high-risk or high-impact outcomes, human-in-the-loop protocols must be implemented to maintain oversight and prevent unintended harm. Additionally, the AI system should be designed around fairness, equality, and inclusion goals, carefully considering the social cost of exclusion and promoting equitable access and outcomes for all segments of society.

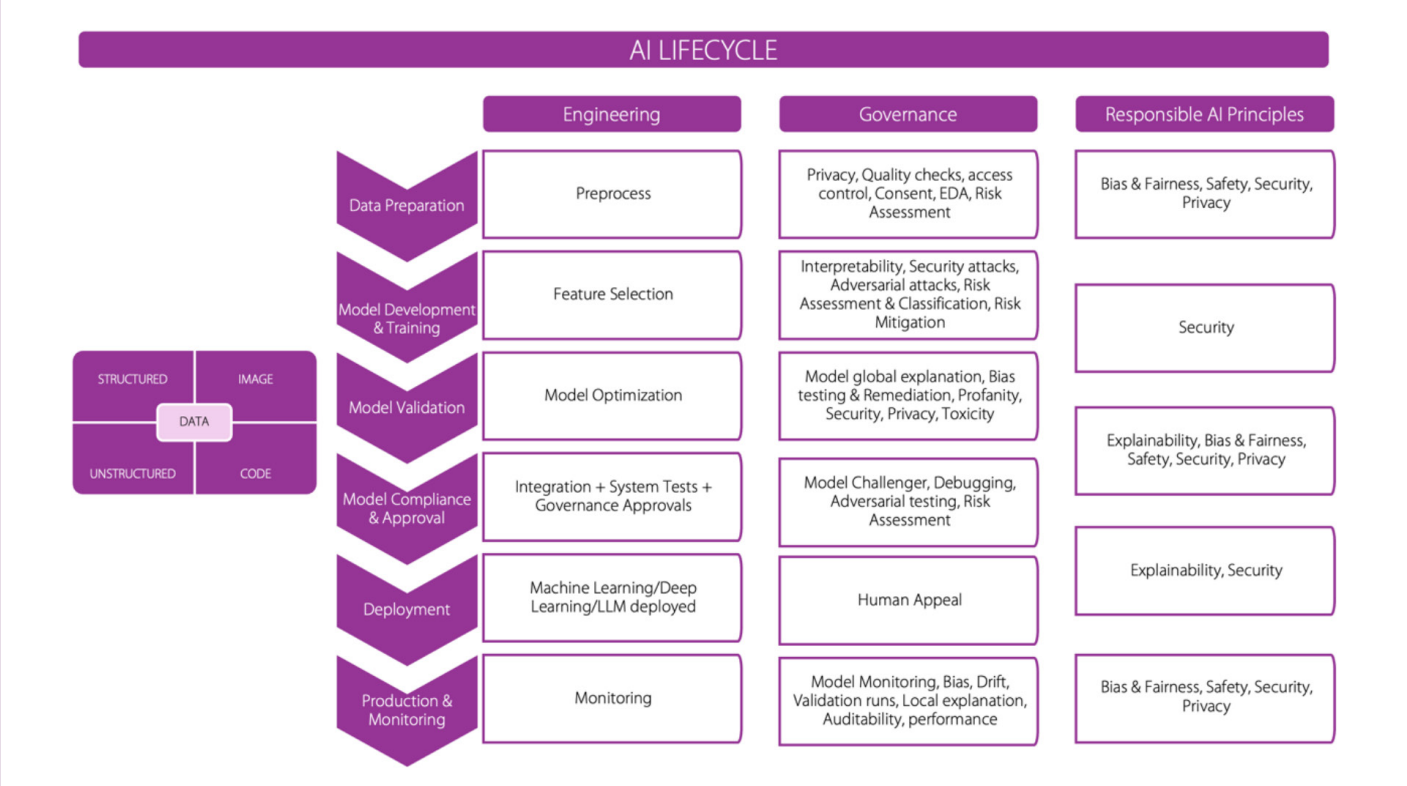


Figure: AI Lifecycle for AI use-case implementation

Follow UK AI Toolkit principles for transparency, fairness and accountability; NITI Aayog principles for inclusivity, risk management, and stakeholder engagement.

Data Preparation

- Collect data in accordance with local laws, ethical guidelines, and informed consent requirements.
- Ensure diversity and representativeness of datasets to reflect real-world demographics and use cases.
- Protect privacy through anonymisation, encryption, and controlled access mechanisms.
- Maintain detailed documentation of all data sources, attributes, and collection parameters.
- Standardise annotation and labelling processes, incorporating inter-annotator reliability checks.
- Mitigate labelling bias by engaging diverse annotator pools and conducting iterative review cycles.
- Apply preprocessing techniques that preserve fairness and eliminate proxy or indirect biases.
- Conduct statistical analysis to detect anomalies, correlations, and skewed data distributions.
- Perform fairness audits and record identified ethical risks and mitigation strategies.
- Identify representational gaps to reduce potential discrimination in model outputs.
- Incorporate Responsible AI checks for bias, explainability, fairness, and data

Adhere to UK GDPR and India's Personal Data Protection Bill for privacy and fairness.

Model Development & Training

- Objective: Build and test AI models, refine data strategy, perform feature selection/engineering, and ensure adherence to Responsible AI principles.
- Integrate ML-Ops pipelines embedded with applicable Responsible AI guardrails and validations.
- Verify AI capabilities using independent and representative test data before advancing to validation.
- Ensure adherence to non-functional requirements (NFRs) and objective functions and perform impact and risk assessments for safety and reliability.
- Document model versions, training parameters, and reproducibility measures for audit readiness.
- Conduct peer reviews and internal evaluations before proceeding to formal validation.

Adhere to UK AI governance principles for accountability and traceability; and to NITI Aayog's AI for All framework emphasizing inclusivity, reliability, and explainability.

Model Validation

- Objective: Independently test and validate model performance, robustness, and fairness.
- Perform quantitative validation across multiple metrics including accuracy, precision, recall, F1-score, and confidence intervals.
- Conduct qualitative validation for interpretability, explainability, and ethical alignment.
- Execute stress and adversarial testing to evaluate resilience against misuse or unexpected inputs.
- Undertake bias and fairness testing to ensure outputs do not discriminate across demographic or contextual groups.
- Validate compliance with data protection, cybersecurity, and regulatory requirements.

- Document validation outcomes, testing datasets, evaluation reports, and any identified gaps or corrective actions.

Align with UK AI Assurance Framework for validation rigor and transparency; and NITI Aayog guidelines on ethical assurance, testing, and risk management.

Model Compliance & Approval

- Objective: Ensure that AI systems meet internal, regulatory, and Responsible AI compliance standards before deployment.
- Conduct Responsible AI compliance reviews against frameworks such as UK AI Regulation, NITI Aayog, and NASSCOM AI guidelines.
- Obtain signoffs from Legal, Data Protection Officer (DPO), Responsible AI Office, and IP teams where applicable.
- Ensure that all risk assessments, bias audits, and model validation reports are completed and approved.
- Document ethical considerations, including explainability levels, fairness objectives, and inclusion metrics.
- Archive all approvals, compliance documentation, and stakeholder feedback for future traceability.

Follow UK AI Regulation principles for auditability and legal compliance; and NITI Aayog Responsible AI recommendations on governance and multi-stakeholder

Deployment

- Objective: Install, release, and configure validated AI systems for operation in the target environment.
- Define the target operating environment and distinguish between software component and model deployment.
- Review and enhance risk management processes; update risk registers as needed.
- Ensure Responsible AI principles are embedded, and telemetry is configured for real-time monitoring.

Deployment Guidelines – AIML:

1. Data Pipeline: Preprocess, clean, and secure data; use distributed frameworks.
2. Model Training & Evaluation: Employ scalable infrastructure, version control, and cross-validation.
3. Model Deployment: Implement via microservices, containerization, and API endpoints.
4. Monitoring & Performance: Track accuracy, latency, drift, and resource utilization.
5. CI/CD: Automate deployment and testing; roll out incremental updates safely.
6. Feedback Loop: Gather user feedback, retrain models, and apply active learning.
7. Scalability & Resource Management: Monitor workloads and enable auto-scaling.
8. Security & Compliance: Apply encryption, periodic audits, and ensure regulatory compliance.
9. Collaboration & Documentation: Maintain end-to-end documentation and host knowledge workshops.

Deployment Guidelines – LLM/RAG:

1. Deployment Guidelines – LLM/RAG:
   RAG Deployment: Use a separate retrieval service integrated with the LLM; deploy on containerized, load-balanced infrastructure.

2. Fine-Tuning: Maintain version-controlled fine-tuning pipelines; test in sandbox environments before production rollout.

3. New LLM Deployment: Utilize distributed training, compatibility layers, A/B testing, auto scaling, and continuous improvement cycles.

Comply with UK AI deployment guidance on safety and accountability; and India's AI ethics guidelines on inclusivity, explainability, and risk management.

Production & Monitoring

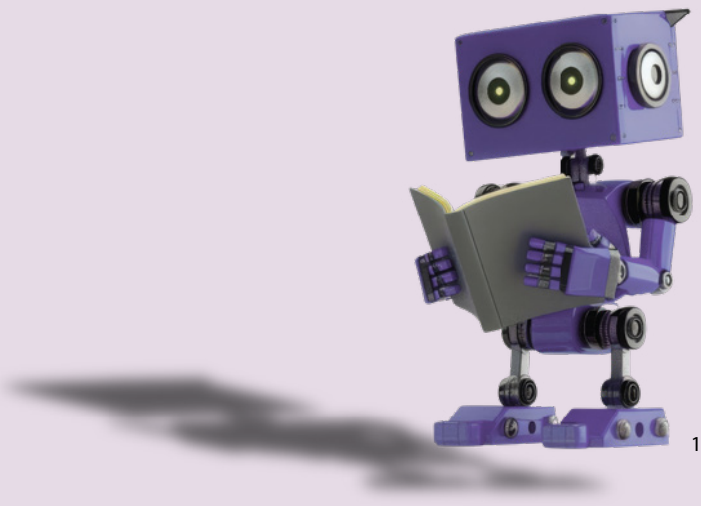
- Continuously track model performance metrics including accuracy, precision, latency, throughput, and drift detection.
- For LLM use cases, monitor text quality (faithfulness, hallucination), response time, diversity, and ethical compliance.
- Implement human oversight mechanisms to review outputs and address anomalies promptly.
- In continuous learning setups, perform incremental validation with updated data and models.
- Periodically re-evaluate system performance to determine whether retraining or decommissioning is required.
- Maintain Responsible AI guardrails to ensure sustained ethical and compliant behavior.

Follow UK AI Assurance and Monitoring principles for transparency and accountability; align with NITI Aayog and NASSCOM frameworks for lifecycle

Retirement / Phase-Out

- Objective: Retire AI systems that are outdated, underperforming, or non-compliant.
- Obtain approvals from product owners, business sponsors, and the AI Governance Council prior to decommissioning.
- Communicate the rationale for retirement and offer alternative solutions where necessary.
- Ensure security of sensitive data and intellectual property during retirement.
- Perform data backup and secure disposal of datasets, models, and configuration files.
- Document lessons learned and improvement opportunities for future AI lifecycle implementations.

Follow UK AI governance guidance on ethical system retirement; align with NITI Aayog principles on sustainability, transparency, and responsible phase-out.



Metrics for AI Solution

Metrics for AI solutions play a crucial role in ensuring that models are not only highperforming but also ethical, reliable, and aligned with Responsible AI principles. They provide measurable ways to evaluate aspects such as data quality, model accuracy, fairness, transparency, robustness, and compliance across the AI lifecycle, from data preparation to deployment and monitoring. By tracking these metrics, developers can detect issues like bias, drift, hallucinations, or security vulnerabilities early, allowing timely interventions and continuous improvement. Incorporating Responsible AI metrics alongside traditional performance indicators helps balance innovation with accountability, ensuring that AI solutions remain trustworthy, explainable, and beneficial to all stakeholders.

Table: Metrics for AI Solution

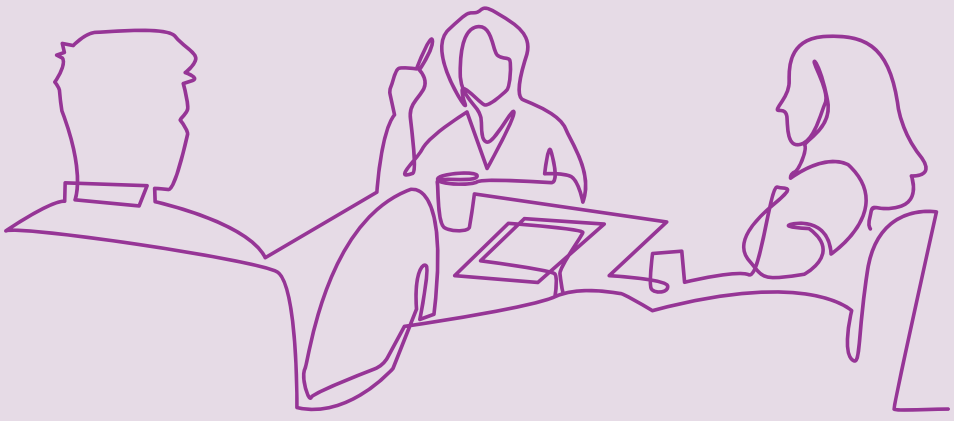
Metric	Stages of Measurement	Purpose
Error Rate	Data Preprocessing, Modelling & Retraining	This is the proportion of data that contains errors across dataset.
Missingness Rate	Data Preprocessing, Modelling & Retraining	This is the proportion of data that is missing across dataset.
Duplication Rate	Data Preprocessing, Modelling & Retraining	This is the proportion of data that is duplicated across dataset.
Aging Rate	Data Preprocessing, Modelling & Retraining	This is the age or freshness of the data.
Feature Importance score	Data Preprocessing, Modelling & Retraining	It helps us to understand the importance of each feature in data
Coverage Rate	Data Preprocessing, Modelling & Retraining	This signifies the coverage extent of the data for the use case.
Variability Rate	Data Preprocessing, Modelling & Retraining	This signifies the degree of change or variation in the data across dataset.
Fairness	Data Preprocessing, Modelling & Retraining	There should not be any bias in data.
Biasness	Data Preprocessing, Modelling & Retraining	There should not be any bias in data.
Perplexity score	Deployment	It is a measure of how confidently the model can predict the sequence of words. The higher the perplexity, the less confidently the model predicts the observed sequence.
Accuracy	Deployment	Accuracy is measured by how many are true amongst all the predictions

Metric	Stages of Measurement	Purpose
Precision	Deployment	Precision is measured by the proportion of true positives to the number of total positives that the model predicts
Recall	Deployment	Recall is measured by how many we have correctly predicted as true amongst all the data points to be predicted as true
Error rate	Deployment	Error rate is simply defined as 1 -Accuracy. If the accuracy is 90%, then error rate is 10%.

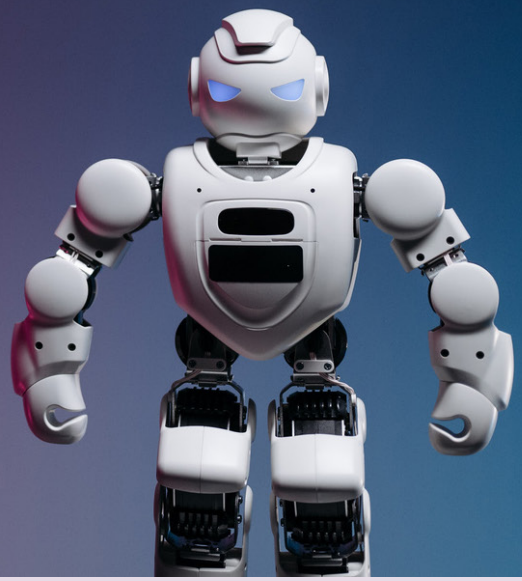
Table: Various Scenarios for Risk Category.

Risk Category	Risk Definition	Mitigation Measures
Job elimination	If the AI use case eliminates existing jobs partially or completely.	Assess impact on livelihood, provide reskilling opportunities and sufficient time, and offer appropriate severance as per legal requirements.
Mental harassment	If the use case causes illegal, disorderly, or socially harmful acts.	Validate prompt inputs against prohibited words to prevent harmful outputs.
Decision making bias	If AI output favors any group or section unfairly.	Apply prompt engineering to guide model context; prioritize controlling training data where possible.
Unauthorized data processing	If personal data is misused, denied, or not deletable.	Enable data deletion, modification, and end-user access; maintain transparency on data processing.
Poor data quality	If data is incomplete, biased, or unsuitable for AI.	Use data quality tools, SME validation, sample reviews, and schedule regular updates to prevent obsolescence.
Financial loss	If AI malfunction causes financial damage.	Implement human switchover and inbuilt safety measures to prevent catastrophic outcomes.
Toxic chemical release	If AI malfunction leads to toxic chemical release.	Ensure human switchover and safety mechanisms are in place.
Critical equipment failure	If AI malfunction causes critical asset failure.	Implement human switchover and safety measures to avoid major incidents.

Risk Category	Risk Definition	Mitigation Measures
Hallucination	If AI generates misleading or incorrect outputs.	Use RAG mechanisms, prompt engineering for context, and human verification; cannot fully eliminate hallucination.
Confidential information breach	If private information is disclosed without consent.	Enforce contracts for model training, anonymize PII, maintain safeguards, and implement automatic deletion when needed.
Data security breach	If unauthorized parties access sensitive info.	Contracts for open-source use, data safeguards, anonymization, transparency, and deletion protocols.
Identity theft	If personal info is exposed without consent.	Similar to data security breach: contracts, safeguards, anonymization, and deletion guidelines.
Model drift	If model loses predictive accuracy over time.	Regular automatic updates of training data, with manual SME validation where needed.
Model unwanted bias	If model output shows favoritism or discrimination.	Regularly review training datasets and integrate bias-reduction APIs.
Unexplainable model result	If decisions are not understandable to users.	Provide reference documentation and implement explainability APIs for visibility into decision-making.
Unreliable model results	If predictions or recommendations are consistently inaccurate.	Control model temperature for consistency and implement backup systems to maintain business logic.
Loss time injury	If AI malfunction causes injury or death.	Ensure human switchover and in-built safety measures to prevent catastrophic events.



Responsible AI Principles – Self-Assessment Questionnaire



The questionnaire serves as a practical tool to help developers and other stakeholders ensure that Responsible AI principles are consistently applied throughout the AI lifecycle. It guides teams in verifying that data handling, model design, testing, and deployment processes align with ethical, transparent, and compliant practices. By following this checklist, developers can identify potential risks early, maintain accountability, and ensure adherence to both Indian and UK regulatory standards. Ultimately, it promotes the creation of AI solutions that are reliable, fair, and aligned with organizational and societal expectations. Below is a checklist that should be answered by the developers and stakeholders.

Transparency & Explainability

1. Can the system's purpose, scope, and limitations be clearly communicated to end users and regulators?
2. Are the data sources, features, and decision logic documented and accessible for audit?
3. Does the system provide interpretable outputs or explanations for key decisions?
4. Are users informed when they are interacting with an AI system?
5. Is there a mechanism for stakeholders to request clarification or explanations about model behavior?
6. (Adherence: Yes / Maybe / No → High / Medium / Low)

Accountability & Governance

1. Are clear roles and responsibilities assigned for managing, reviewing, and approving AI outcomes?
2. Does the project maintain documentation of key design and decision-making steps throughout development?
3. Is there a governance process or review board overseeing ethical and regulatory compliance?
4. Are escalation or appeal mechanisms defined for addressing AI-related harms or complaints?
5. Are human reviewers empowered to override or correct AI system outputs where necessary?
6. (Adherence: Yes / Maybe / No → High / Medium / Low)

Safety & Robustness

1. Has the system been tested under varied operational and environmental conditions?
2. Are there monitoring and fallback mechanisms in place for detecting and mitigating failures?
3. Does the model handle adversarial inputs or unexpected data gracefully without harmful behavior?
4. Is there a plan for continuous monitoring and updates postdeployment?
5. Has the system undergone stress testing or red teaming to identify vulnerabilities?
6. (Adherence: Yes / Maybe / No → High / Medium / Low)

Fairness & Non-Discrimination

1. Has the dataset been reviewed for demographic, linguistic, or regional representativeness?
2. Are fairness metrics or bias audits conducted before deployment?
3. Is there a process to correct or mitigate identified biases?
4. Are decisions explainable and equitable across different population groups?
5. Has stakeholder feedback from potentially impacted groups been incorporated into design or testing?
6. (Adherence: Yes / Maybe / No →High / Medium / Low)

Privacy & Data Governance

1. Is all personal or sensitive data collected and processed lawfully with user consent?
2. Are privacy-preserving methods (e.g., anonymization, encryption) implemented across the data lifecycle?
3. Can users access, correct, or delete their personal data if required?
4. Are data storage and sharing practices compliant with relevant legal frameworks (e.g., DPDP Act, GDPR)?
5. Is there a process to review third-party data sources or APIs for compliance risks?
6. (Adherence: Yes / Maybe / No→High / Medium / Low)

Human-Centricity & Values

1. Does the system enhance rather than replace human decision-making in critical contexts?
2. Are ethical considerations and societal values explicitly integrated into design choices?
3. Can users provide meaningful feedback or contest automated outcomes?
4. Are potential risks to human dignity, rights, or autonomy identified and mitigated?
5. Has the AI been designed with user well-being and inclusivity as key priorities?
6. (Adherence: Yes / Maybe / No→High / Medium / Low)

Inclusiveness & Accessibility

1. Has the system been tested for usability across diverse demographic and linguistic groups?
2. Does the interface or output accommodate users with disabilities or varying digital literacy levels?
3. Is there representation from diverse user groups during design and validation stages?
4. Are local contexts, regional languages, or cultural nuances considered in deployment?
5. Is the AI designed to ensure equitable access regardless of geography or socio-economic background?
6. (Adherence: Yes / Maybe / No→High / Medium / Low)

Sustainability & Social Impact

1. Has the system's environmental impact (energy use, compute intensity) been assessed?
2. Are sustainability considerations integrated into model training and deployment choices?
3. Does the AI system contribute positively to long-term social or economic wellbeing?
4. Are there mechanisms to identify and mitigate unintended societal harms?
5. Does the project align with broader sustainability or Responsible Innovation goals?
6. (Adherence: Yes / Maybe / No→High / Medium / Low)

Responsible AI Principles – Adherence Assessment

This section helps developers and other stakeholders in the project teams evaluate how well their AI systems align with core Responsible AI (RAI) principles. Each principle reflects ethical, technical, and societal expectations essential for trustworthy AI. The adherence assessment uses a simple Yes / Maybe / No scale to indicate the extent to which the system meets each principle, representing High, Medium, or Low adherence respectively. By completing this assessment, teams can identify strengths, gaps, and areas requiring further mitigation or oversight. The weighted scoring provides a transparent view of overall RAI compliance and guides decisions on whether a system is ready for deployment, review, or redesign.

Principle	Explanation	Yes – Fully Aligned	Maybe – Partially Aligned	No – Not Aligned
Transparency & Explainability	AI systems must clearly communicate how they work - including data sources, decision logic, and limitations - in a way understandable to users and regulators.	System logic, data, and limitations are well documented and explainable to non-technical users.	Some documentation exists but not accessible or easily understood.	Model operation unclear; no documentation or explainability mechanism.
Accountability & Governance	Humans remain responsible for AI outcomes. Clear roles and oversight ensure compliance and liability.	Roles, review process, and accountability are formally documented and enforced.	Roles partially defined; oversight inconsistent.	No ownership or review accountability established.
Safety & Robustness	AI must operate reliably, resist attacks, and fail safely under unexpected conditions.	Model validated for robustness; monitoring and fallback plans are in place.	Limited testing; partial fallback or resilience measures.	No robustness testing or monitoring framework.
Fairness & Non-Discrimination	AI should not reinforce bias; must ensure fair outcomes across groups.	Fairness metrics tested and bias mitigation methods implemented.	Fairness testing planned or partially conducted.	No fairness or bias evaluation done.
Privacy & Data Governance	Data must be collected, stored, and processed lawfully, securely, and with consent.	Data complies with DPDP/ GDPR; consent and deletion rights ensured.	Partial privacy compliance or unclear retention practices.	No privacy safeguards; data provenance uncertain.
Human-Centricity & Values	AI should assist, not replace, human judgment and respect human rights and values.	Human-in-the-loop defined; ethical guardrails embedded.	Human oversight present but inconsistently applied.	No human control or ethical consideration integrated.
Inclusiveness & Accessibility	AI must serve diverse groups and enable equal access across regions and abilities.	Inclusive design principles applied; accessible interfaces for all users.	Some inclusive design features present; limited localization.	Excludes or disadvantages certain groups or languages.
Sustainability & Social Impact	AI should minimize environmental harm and enhance longterm societal benefit.	Resource-efficient; supports SDG or positive community outcomes.	Some efficiency measures or limited social benefit.	High environmental cost; no social responsibility measures.

Scoring Logic

Response	Score Value	Meaning
Fully Aligned	3	Fully aligned / High adherence
Partially Aligned	2	Partially aligned / Needs improvement
Not Aligned	1	Not aligned / High risk or unclear

Overall Assessment

Total Score (sum of 8 principles)	Adherence Level	Interpretation & Next Steps
21 – 24	● High Adherence	Strong alignment. Ready for deployment with routine monitoring and governance.
15 – 20	● Partial / Moderate Adherence	Some gaps present. Require focused remediation: docs, testing, governance, privacy checks.
8 – 14	● Low Adherence	Major gaps across principles. Immediate remediation required before deployment.

Note: This assessment framework is intended solely as a guidance tool. It is derived from practitioner experience and has not undergone formal validation with industry or standards bodies. Users are advised to contextualize the results and not rely on them as a definitive measure of compliance or performance.

Stakeholder	Core Responsibilities to Watch For
Domain Experts	• Ensure business rules and sector boundaries are respected. • Assess if AI outputs align with domain logic and intended use. • Identify risks or harm to vulnerable groups. • Review labels, context, and edge cases for correctness and fairness. • Flag potential misuse or misinterpretation scenarios early.
Data Scientists	• Use representative, diverse, and unbiased datasets. • Evaluate models for fairness, robustness, and explainability-not just accuracy. • Document assumptions, limitations, and retraining triggers. • Perform stress and drift testing before deployment. • Recommend retraining when patterns or behaviors change.
Policy & Legal Experts	• Classify use cases into appropriate risk categories (aligned with regulations). • Ensure compliance checks (DPIA / AI-DPIA) and accountability roles. • Confirm transparency, explainability, and human oversight requirements. • Maintain audit-ready documentation and policy sign-offs. • Communicate regulatory boundaries and restricted uses clearly.
Developers	• Implement privacy, security, and safety guardrails in the build. • Enable explainability and traceability (model cards, logs, versioning). • Monitor performance and fairness continuously post-deployment. • Log deviations and outcomes for audits. · Ensure fallback mechanisms and human review for high-impact decisions.

Stakeholder	Core Responsibilities to Watch For
Language & Cultural Experts	• Validate linguistic nuances and cultural sensitivities in outputs. • Detect and mitigate bias or stereotyping in language models. • Ensure inclusivity and accessibility across languages. • Advise on tone and cultural appropriateness for global deployments.
UX Designers / Human Factors Specialists	• Design interfaces that make AI decisions understandable to users. • Ensure human-in-the-loop pathways for critical decisions. • Prevent dark patterns or misleading UI that could cause misuse. • Promote transparency and user control over AI-driven actions.
Ethics & Risk Officers	• Define ethical principles and risk thresholds for AI use. • Conduct ethical impact assessments before deployment. • Monitor for unintended consequences and escalate issues. • Ensure alignment with organizational values and Responsible AI frameworks.
Security & Privacy Specialists	• Implement data protection measures (encryption, anonymization). • Ensure compliance with privacy laws (GDPR, DPDP Act, etc.). • Monitor for adversarial attacks or model exploitation. • Validate secure handling of sensitive or personal data.

Regulatory Landscape by Sector – India & UK

Responsible AI in Healthcare

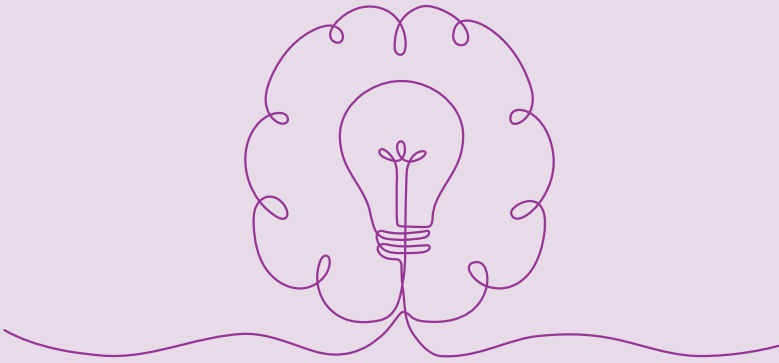
Artificial Intelligence is redefining healthcare, from clinical diagnostics and medical imaging to drug discovery, hospital operations, and patient engagement. But unlike many other industries, the stakes are uniquely high: lives, safety, and trust.

Healthcare AI must not only be accurate but accountable, explainable, and auditable. Trust in AI-enabled healthcare solutions depends on strong governance and regulatory alignment. Global and national authorities are now embedding Responsible AI principles directly into healthcare-specific legal frameworks and standards.

While general Responsible AI principles apply across all industries, healthcare demands a higher standard of assurance.

AI systems here directly impact patient lives and clinical decisions — leaving no room for error. This makes safety, human oversight, patient rights, and clinical validation not just important, but nonnegotiable pillars of responsible AI use in this sector.

Among these, Safety & Robustness is paramount, healthcare AI must demonstrate clinical-grade reliability, withstand real-world variability, and ensure patient well-being at every stage of its lifecycle. Leading AI Regulators in Healthcare.



Regulatory Body	Summary of AI Regulatory Work	Country / Region	Year (Latest)	Type	Source
World Health Organization (WHO)	Issued global ethical & governance guidance for AI in health, including large medical models and patient safety frameworks.	Global	2023 – 2025	Guidance / Principles	Link
US Food and Drug Administration (FDA)	Draft lifecycle & marketing submission guidance for AI/ML in medical devices (SaMD), GMLP principles, registry of authorized AI devices.	USA	2025	Regulatory Guidance	Link
Medicines and Healthcare products Regulatory Agency (MHRA)	Developed the "AI Airlock" sandbox and updated regulatory strategy for AI-enabled medical devices (AIaMD).	UK	2024 – 2025	Sandbox / Strategy	Link
International Medical Device Regulators Forum (IMDRF)	Finalized international Good Machine Learning Practice (GMLP) principles for AI/ML-enabled devices - adopted by multiple regulators.	Multilateral	2025	Global harmonization guidance	Link
Indian Council of Medical Research (ICMR)	Issued national Ethical Guidelines for AI in Biomedical Research & Healthcare - governance, consent, transparency, and oversight.	India	2023	Guidelines / Ethical Framework	Link

Sector impacts

AI is transforming healthcare and biotech through faster genomic analysis, drug discovery, diagnostics, and personalized treatments. But risks like biased data, model drift, privacy concerns, and lack of explainability demand responsible governance to ensure safety, trust, and fairness.

Mitigation

- Run bias audits across ethnicity, age, and gender.
- Publish explainability packs for clinical tools.
- Deploy drift detectors and retrain with new biomedical data.
- Enforce human-in-the-loop for diagnosis and treatment.
- Apply privacy-by-design for genomic and health data.
- Combine simulation + expert review for drug discovery models
- Maintain audit logs and feedback loops for clinical AI systems.

Use Case / Incident Examples

Use case	Incident & impact	Mitigation (developers' practices)	Red Flags	Green Flags	Source
Genomic Risk Prediction Bias	A model trained on Western genomic datasets misclassified disease risk for South Asian populations.	Diversify genomic datasets, run fairness audits, validate across ethnic groups.	Homogeneous data, no demographic slicing.	Diverse datasets, fairness slicing.	Link
Drug Discovery Model Drift	AI model failed to adapt to new pathogen variants, leading to ineffective drug candidates.	Deploy drift detectors, retrain with updated biological data, include expert review.	Static models, no retraining.	Seasonal updates, expert-in-loop.	Link
Diagnostic AI Opaqueness	Clinicians rejected AI diagnostic tools due to lack of interpretability in cancer detection.	Add explainability layers, provide clinical rationale, include visual outputs.	Black-box predictions, no rationale.	Explainable AI, clinician-friendly outputs.	Link
Clinical Trial Misrepresenta tion	AI excluded older adults from trial recommendations due to biased training data.	Audit inclusion criteria, balance age representation, validate with diverse cohorts.	Age bias, exclusionary filters.	Inclusive sampling, transparent criteria.	Link
Biosecurity Risk in Protein Design	AI-generated proteins had dual-use potential, raising biosecurity concerns.	Proposed pre-deployment risk evaluations. Advocated for targeted AI safety protocols and biosecurity screening.	No misuse safeguards, open access.	Risk screening, ethical oversight.	Link

Responsible AI in Agriculture

Artificial Intelligence is transforming the agriculture sector, from precision farming and crop yield forecasting to pest detection, smart irrigation, logistics, and market intelligence. Unlike many other industries, the stakes here involve food security, livelihoods, and environmental stability. AI-driven agriculture solutions must be accurate, explainable, inclusive, and context-aware, especially for smallholderfarmers who form the backbone of the sector in countries like India. Trust in these solutions relies on strong governance, data ethics, and local validation to avoid amplifying inequities or environmental risks. While general Responsible AI principles apply across sectors, agriculture demands special attention to transparency, inclusiveness, and sustainability. Models must account for local soil, climate, crop, and socioeconomic diversity, not just global datasets. Among these principles, fairness, explainability,and sustainability are particularly critical to ensure equitable benefits and resilience against climate shocks.

Leading AI Regulators in Agriculture

Regulatory Body	Summary of AI Regulatory Work	Country / Region	Year (Latest)	Type	Source
Food and Agriculture Organization (FAO)	Issued guidance on AI & digital transformation in agrifood systems — focusing on equity, data governance, and environmental sustainability.	Global	2023–2025	Guidance / Principles	FAO AI in Agriculture
Organisation for Economic Cooperation and Development (OECD)	Released AI governance guidelines specific to agriculture, focusing on responsible scaling, market integrity, and climate adaptation.	OECD Member States	2025	Guidelines / Governance Note	Link
Ministry of Agriculture & Farmers Welfare, India	Launched national AI initiatives and guidelines for precision agriculture, crop health monitoring, and digital farming innovation.	India	2024	Policy	Link

Sector Impacts

AI in agriculture is changing the entire value chain — improving yield prediction, irrigation scheduling, pest detection, and market intelligence. But when poorly governed, the same systems can deepen rural inequality, cause crop mismanagement, or harm ecosystems.

- Data & Representation Gaps: Models trained on commercial farm data may misfire for smallholders.
- Model Drift: Seasonal variability and climate shocks can lead to inaccurate forecasts or irrigation failures.
- Lack of Explainability: Farmers may not trust opaque recommendation systems.
- Privacy Risks: Geospatial farm data linked to personal identities.
- Market Manipulation Risks: Unmonitored trading or forecasting models can distort pricing.

Mitigation

- Run bias & slice audits across crop types, region, and farm size.
- Provide clear explainability in advisories and subsidy decisions.
- Deploy drift detectors and retrain models seasonally.
- Keep human-in-the-loop for irrigation, subsidy, and policy-impacting actions.
- Adopt privacy-by-design for farm-level geodata.
- Maintain immutable audit logs for trading and forecasting systems.

Use Case / Incident Examples

Use Case	Incident & Impact	Mitigation (Developer Practices)	☒ Red Flags	☑ Green Flags	Source
Crop Yield Forecasting Drift	Seasonal anomalies led to over-irrigation and wasted resources for small farmers.	Seasonal retraining, drift detection, contingency overrides.	Static thresholds, outdated models.	Dynamic thresholds, drift alarms, farmer feedback loop.	Link
Fertilizer Advisory Bias	Fertilizer recommendations favored high-yield regions, excluding marginal farmers.	Bias slicing by geography, explainability dashboards, inclusion testing.	Single-region training, opaque rules.	Transparent dashboards, multi-region calibration.	Link
Market Forecast Manipulation	Trading algorithms amplified price spikes, hurting producers.	Anomaly detection, independent monitoring, audit logs.	No audit trail, opaque triggers.	Immutable logs, regulator alerts.	Link
Farmer Data Exposure	Geolocation & subsidy data unintentionally leaked.	Privacy-by-design, anonymization, consent frameworks.	No consent, exposed identifiers.	Data minimization, encryption.	Link

Key Takeaway

Responsible AI in agriculture means building trustworthy, explainable, and inclusive systems that respect both the environment and rural communities. Developers must align with FAO–OECD guidance, NITI Aayog's RAI principles, and ISO standards to ensure responsible scaling of AI solutions in agriculture.

Responsible AI in FinTech

Artificial Intelligence is reshaping financial services — from fraud detection, credit scoring, and lending to KYC automation, payments, wealth management, and algorithmic trading. Unlike most industries, financial AI operates in one of the most tightlyregulated spaces. Models influence access to credit, determine fraud risks, and shape regulatory compliance — which makes trust, explainability, and accountability critical.

While general Responsible AI principles apply across domains, FinTech demands heightened fairness, explainability, security, auditability, and governance to ensure financial stability and consumer protection. This is why regulators in India, the UK, and globally have moved fast to define AI governance frameworks for financial services.

Leading AI Regulators in FinTech

Regulatory Body	Summary of AI Regulatory Work	Country / Region	Year (Latest)	Type	Source
Reserve Bank of India (RBI)	Introduced RBI Digital Lending, IT Outsourcing, and FREE-AI Guidelines to ensure algorithmic transparency, vendor accountability, and AI risk governance.	India	2022– 2025	Regulation / Framework	Link1 Link-2
Bank of England Prudential Regulation Authority	PRA SS1/23 establishes strict expectations for AI/ML model governance, validation, and monitoring in financial institutions.	UK	2024	Supervisory Statement	Model Risk Management Principles for Banks
Financial Conduct Authority	FCA's Consumer Duty regulation embeds fairness, transparency, and accountability for AI-enabled financial decisions.	UK	2023	Regulation	AI Update

Why it matters: These measures are not optional recommendations — they are increasingly tied to supervisory audits, compliance reviews, and risk reporting cycles for financial institutions.

Sector Impacts

AI systems in FinTech are reshaping credit access, fraud prevention, and customer experience — but can fail dangerously if governance is weak.

- Bias can exclude vulnerable borrowers.
- Drift can weaken fraud defense.
- Opaqueness can block regulatory approvals.
- Vendor failures can lead to systemic exposure.

Use Case / Incident Examples

Use Case	Incident & Impact	Mitigation (Developer Practices)	☒ Red Flags	☑ Green Flags	Source
Credit Scoring Bias	Algorithm denied loans disproportionately to women and minorities.	Bias audits, fairness calibration, explainability packs.	No fairness checks, opaque scoring.	Transparent criteria, bias dashboards.	US CFPB Fair Lending Actions
Fraud Detection Drift	Model degraded after new transaction patterns, causing losses.	Drift detection, retraining, alerts.	Static thresholds.	Real-time drift triggers.	Link

Use Case	Incident & Impact	Mitigation (Developer Practices)	❌ Red Flags	✅ Green Flags	Source
Lending Model Opaqueness	Regulator rejected automated credit model due to lack of interpretability.	Explainability modules (reason codes), model cards.	Black box.	Explainable outputs.	Link
Prompt Injection Attack	Chatbot leaked customer data through malicious prompts.	Output scanning, guardrails, human-in-loop.	Unchecked outputs.	Guardrail filters, egress block-lists.	Link
Vendor Oversight Failure	Third-party fraud detection vendor breach exposed PII.	Contractual AI clauses, SOC2 checks, monitoring.	No oversight.	Strong AI vendor risk controls.	Link

Key Takeaway

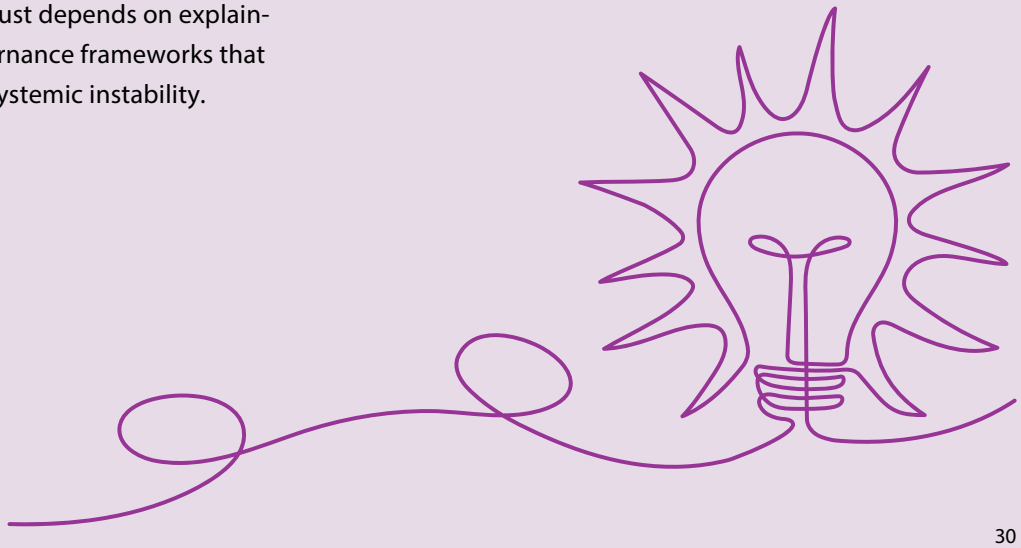
FinTech leads Responsible AI regulation, guided by RBI, EU AI Act, FCA, and ISO standards. AI systems must be fair, explainable, secure, and resilient to meet regulatory scrutiny.

Responsible AI in Climate & Energy

Artificial Intelligence is rapidly reshaping the climate and energy landscape — from optimizing power grids and tracking carbon emissions to forecasting climate risks and advancing renewable energy efficiency. Yet, in a sector that directly affects planetary health and human survival, the margin for error is razor thin.

AI in climate and energy must balance innovation with responsibility — ensuring models are transparent, scientifically validated, environmentally sustainable, and equitable across regions and populations. Trust depends on explainability, resilience, and robust governance frameworks that prevent greenwashing, bias, and systemic instability.

While foundational Responsible AI principles apply universally, climate and energy AI require deeper integration of sustainability, safety, and accountability. Models must not only perform efficiently but also ensure carbon-aware decisionmaking, energy efficiency, and environmental transparency. Among all principles, sustainability, fairness, and robustness are crucial to ensure that AI systems serve long-term climate goals without unintended harm.



Leading AI Regulators in Climate & Energy

Regulatory Body	Summary of AI Regulatory Work	Country / Region	Year	Type	Source
International Energy Agency (IEA)	Issued global recommendations on AI for energy transition — covering grid optimization, energy efficiency, and responsible carbon modeling.	Global	2024–2025	Guidance / Policy Brief	Energy & AI
United Nations Environment Programme (UNEP)	Developed ethical AI principles for environmental data use, emissions monitoring, and ecosystem modeling.	Global	2024	Ethical Principles / Framework	UNEP AI for Environment 2024
Ministry of Power & MNRE (India)	Launched AI roadmaps for renewable energy forecasting, smart grids, and climate risk mitigation aligned with national net-zero goals.	India	2024	Policy / Roadmap	MNRE AI in Renewable Energy 2024

Sector Impacts

AI in climate and energy is driving transformation across the value chain — from forecasting and optimization to carbon accounting and climate modeling. Yet, without responsible design, these systems can amplify existing risks: inequitable energy access, overreliance on opaque forecasts, or unsustainable resource consumption.

Key Risks:

- Data Bias & Representation Gaps: Climate models built on limited regional data mispredict conditions for developing regions.
- Model Drift: Changing weather and emission patterns degrade model accuracy over time.
- Opacity in Carbon Accounting: Unclear carbon footprint models enable greenwashing.
- Energy Intensity: Large AI models for climate simulation can consume excessive energy, undermining sustainability goals.
- Privacy & Data Integrity: IoT-based energy and emissions tracking expose household-level consumption data.

Mitigation

- Conduct bias and drift audits across geographies, seasons, and emission categories.
- Prioritize energy-efficient AI architectures and carbon-aware computing practices.
- Publish explainability dashboards for carbon accounting and grid optimization tools.
- Enforce human oversight for critical climate forecasts and energy grid automation.
- Adopt privacy-by-design for smart-meter and IoT-based datasets.
- Implement transparent reporting and immutable logs for emission and carbon-credit models.

Use Case / Incident Examples

Use Case	Incident & Impact	Mitigation (Developer Practices)	☒ Red Flags	☑ Green Flags	Source
Carbon Forecasting Drift	Climate prediction AI trained on outdated emission data failed to anticipate 2024 monsoon anomalies.	Continuous retraining with new climate data and validation by local meteorological agencies.	Outdated datasets, no revalidation.	Seasonal updates, human review.	Link
Grid Optimization Failure	Smart grid AI overloaded renewable circuits due to misaligned demand forecasting.	Incorporate real-time feedback loops and simulation testing before deployment.	No fail-safe, static load assumptions.	Simulation-tested, human override.	Link
Carbon Accounting Greenwashing	Corporate AI inflated emission offsets via opaque carbon-trading algorithms.	Mandate third-party audits and transparent calculation methodologies.	Opaque offset data, no audits.	Verified methods, public audit trail.	Link
Climate Model Bias	Global models underrepresented tropical regions, skewing disaster preparedness policies.	Include regional datasets and local calibration parameters.	Eurocentric datasets generalized forecasts.	Multi-region calibration, inclusive validation.	Link
IoT Data Leakage	Smart meter AI exposed household energy usage data to third-party vendors.	Enforce anonymization, encryption, and strict consent frameworks.	Plain-text data, broad data sharing.	Consent-based access, encrypted storage.	Link

Key Takeaway

Responsible AI in climate and energy is about designing transparent, efficient, and accountable systems that not only optimize operations but also uphold sustainability goals. Developers must align with FAO–OECD, UNEP, ISO, and NITI Aayog frameworks to ensure that AI becomes a driver of climate resilience, not risk. Building low-carbon, explainable, and fair AI systems is not just innovation - it's climate responsibility.

Responsible AI in Engineering Bioengineering

Artificial Intelligence is accelerating breakthroughs in synthetic biology, biomedical engineering, and genomics — enabling faster drug design, protein folding predictions, personalized medicine, and biosecurity monitoring. Unlike many domains, AI in bioengineering deals with living systems, high-stakes medical outcomes, and potential dual-use risks. A single error can trigger cascading biological, ethical, or security consequences. Responsible AI here requires high scientific integrity, security oversight, and strict governance. AI models must be transparent, safe, explainable, and traceable throughout their lifecycle — ensuring that innovation in gene editing, drug synthesis, or molecular design does not compromise public health or biosecurity. Among Responsible AI principles, safety, accountability, explainability, and security is paramount in this sector.

Leading AI Regulators and Standard Bodies in Engineering Biology

Regulatory Body	Summary of AI Regulatory Work	Country / Region	Year (Latest)	Type	Source
World Health Organization (WHO)	Issued global guidance on AI in health & life sciences, emphasizing safety, transparency, and patient protection.	Global	2023–2025	Principles / Guidance	WHO AI in Health
International Biosecurity Working Group (IBWG)	Developed AI biosecurity protocols covering dual-use research, lab safety, and model access controls.	Multilateral	2025	Biosecurity Guidance	IBWG Biosecurity Framework 2025
US Food and Drug Administration (FDA)	Introduced adaptive regulatory pathways for AI-enabled drug discovery and genomics tools.	USA	2025	Regulatory Guidance	FDA AI/ML Drug Discovery Guidance
UK Medicines and Healthcare products Regulatory Agency (MHRA)	Advanced regulatory sandbox for AI in synthetic biology and genomics, focusing on patient safety and traceability.	UK	2024–2025	Sandbox / Strategy	MHRA AI Strategy
Indian Council of Medical Research (ICMR)	Issued Ethical Guidelines for AI in Biomedical Research covering consent, genetic data protection, and oversight.	India	2023	Ethical Guidelines	ICMR AI Guidelines

Sector Impacts

AI in bioengineering enables powerful capabilities — but it comes with unique risks tied to human health, genetic privacy, and dual-use biological threats.

Key Risks:

- Data Bias: Genomic datasets often overrepresent specific ethnic groups, risking inaccurate diagnostics or treatments for underrepresented populations.
- Model Drift: Pathogen evolution or mutational changes can degrade the accuracy of diagnostic and drug discovery models.
- Opacity: Lack of explainability in molecular design can create regulatory blind spots and scientific risks.
- Biosecurity Threats: Generative protein design can be weaponized without proper governance.
- Genetic Privacy Risks: Sensitive genomic data is vulnerable to reidentification and misuse.

Mitigation

- Diverse genomic datasets and fairness audits to avoid biased models.
- Real-time drift detection to adapt to pathogen evolution and mutation patterns.
- Explainability layers for molecular design outputs to support regulator and clinician trust.
- Pre-deployment risk assessment to identify dual-use or bio security threats.
- Robust data governance with encryption, anonymization, and strict access controls.
- Human-in-the-loop oversight for all critical drug or gene-editing decisions.

Use Case / Incident Examples

Use Case	Incident & Impact	Mitigation (Developer Practices)	☒ Red Flags	☑ Green Flags	Source
Protein Design Biosecurity Risk	Open-access generative model was used to design proteins with dual-use potential.	Pre-deployment screening, access controls, ethical oversight.	No access control, public release.	Restricted access, screening pipeline.	Link
Genomic Bias in Diagnosis	AI trained on Western genomes failed to detect genetic markers in South Asian patients.	Diversify datasets, fairness audits, local validation.	Homogeneous data, no slicing.	Inclusive genomic data, bias testing.	Link
Model Drift in Pathogen AI	Diagnostic models underperformed on new pathogen variants.	Deploy drift detectors, retrain periodically.	Static training data, no updates.	Scheduled retraining, monitoring.	Link
Opaque Gene Editing Model	Clinicians couldn't interpret CRISPR model outputs, delaying approvals.	SHAP explainability, clear model documentation.	Black-box outputs.	Explainable rationale.	Link

Use Case	Incident & Impact	Mitigation (Developer Practices)	❌ Red Flags	✅ Green Flags	Source
Genetic Data Leak	Unsecured pipeline exposed genomic identifiers.	Encryption, privacy-by-design, consent frameworks.	Plain text storage.	Anonymized and secured data.	Link

Key Takeaway

Responsible AI in bioengineering is not just about accuracy, it's about biosecurity, genetic privacy, fairness, and regulatory trust. Developers must align their tools with WHO, IBWG, ICMR, MHRA, and ISO standards to ensure that innovation in life sciences remains safe, explainable, and ethically grounded. In this field, security and safety guardrails are as critical as innovation itself.

AI Development with Ethics: Recommended Toolkits

Ethical AI development and the use of recommended toolkits are essential to ensure technological progress aligns with societal values and safeguards human well-being. As AI systems increasingly influence critical decisions in healthcare, finance, education, and governance, ethical design helps prevent bias, privacy violations, misinformation, and loss of accountability. Leveraging trusted toolkits like Infosys Responsible AI Toolkit, IBM AI Fairness 360, TensorFlow Privacy, and Microsoft Presidio enables developers to embed fairness, transparency, privacy protection, and explainability into workflows. These frameworks not only promote compliance and stakeholder trust but also help identify and mitigate risks throughout the AI lifecycle— fostering systems that are both innovative and responsible.

Infosys Responsible AI Toolkit

A comprehensive toolkit to embed ethical principles across the AI lifecycle. It offers modules for bias detection, fairness assessment, explainability, and compliance monitoring, enabling developers to operationalize Responsible AI governance and maintain accountability.

Link: https://github.com/Infosys/Infosys-Responsible-AI-Toolkit

IBM AI Fairness 360

Focuses on fairness, transparency, and trustworthiness with tools for bias mitigation, explainability, and performance evaluation. Includes governance workflows to meet ethical and regulatory standards and supports cross-functional collaboration. Link: https://aif360.res.ibm.com/

TensorFlow Privacy

An open-source library integrating differential privacy into machine learning models. Helps safeguard sensitive data during training with mechanisms to control data leakage and maintain compliance.

Link: https://www.tensorflow.org/responsible_ai/privacy/guide

Microsoft Presidio

Detects, protects, and anonymizes sensitive data in AI workflows. Identifies personally identifiable information (PII) in text, audio, and other unstructured data, applying privacy-preserving transformations to reduce risks. Link: https://microsoft.github.io/presidio/

References

1. NITI Aayog. Responsible AI for All (2021). https://www.niti.gov.in/sites/default/fil es/2021-02/Responsible-AI-22022021.pdf

2. NITI Aayog. National Strategy for Artificial Intelligence (2018). https://www.niti.gov.in/sites/default/fil es/2023-03/National-Strategy-for-Artificial-Intelligence.pdf

3. FAO. AI & the Digital Revolution to Transform Agrifood Systems (2023). https://www.fao.org/newsroom/detail/f ao-highlights-the-potential-of-ai-and-the-digital-revolution-to-transform-the-world-and-its-agrifood-systems/en

4. OECD. How to Govern AI in Agriculture Responsibly (2025). https://oecd.ai/en/wonk/how-to-govern-ai-in-agriculture-responsibly-risks-tools-and-solutions

5. OECD. Sustainable Agricultural Productivity (2024). https://www.oecd.org/content/dam/oe cd/en/networks/network-on-agricultural-total-factor-productivity-and-the-environment/sustainable-agricultural-productivity-to-address-food-systems-challenges.pdf

6. ISO/IEC 23894:2023. Information technology — AI — Risk management. https://www.iso.org/standard/77304.ht ml

7. ISO/IEC 42001:2023. Artificial Intelligence Management System — Requirements. https://www.iso.org/standard/81230.ht ml

8. Xu, J. et al. HarvestNet: A Dataset for Detecting Smallholder Farming Activity (AAAI 2024). https://ojs.aaai.org/index.php/AAAI/arti cle/view/30251

9. MIT GEAR Lab. Low-cost Smart Irrigation Controller (2023). https://news.mit.edu/2023/gear-lab-creates-affordable-user-driven-smart-irrigation-controller-1025

10. MIT News (2025). Responding to the climate impact of generative AI. Available at: Responding to the climate impact of generative AI | MIT News | Massachusetts Institute of Technology

11. MIT News (2025). Explained: Generative AI's environmental impact. Available at: Explained: Generative AI’s environmental impact | MIT News | Massachusetts Institute of Technology

12. OECD (2019). OECD AI Principles. OECD Legal Instruments

13. European Union (2024). Implementation Timeline | EU Artificial Intelligence Act

14. ISO/IEC (2023). ISO/IEC 23894: Artificial intelligence — ISO - Environmental sustainability

15. ISO/IEC (2023). ISO/IEC 42001: Artificial intelligence — ISO - Energy

16. NITI Aayog (2018). Responsible AI for All. Available at: NITI AAYOG, India | Energy Aayog (2018). Responsible AI for All. Available at: NITI AAYOG, India | Climate Change & Environment

17. UK BEIS (2022-23). Net Zero Strategy & AI Governance Principles. Available at: Net Zero Strategy: Build Back Greener -GOV.UK

18. Stanford University / Harvard University (2023-25). AI in Climate Change Modeling and Environmental Sustainability: Leveraging Intelligent Systems for a Greener Future

19. NITI Aayog (2018). Responsible AI for All. Available at: NITI AAYOG, India | Economics & Finance-I (E&F-I)

20. NITI AAYOG, India | Economics & Finance-II (E&F-II)

21. RBI FREE-AI (2025): FREE-AI Committee Report – RBI [www.rbi.org.in]

22. RBI Digital Lending Guidelines (2022– 23): Notifications - Reserve Bank of India

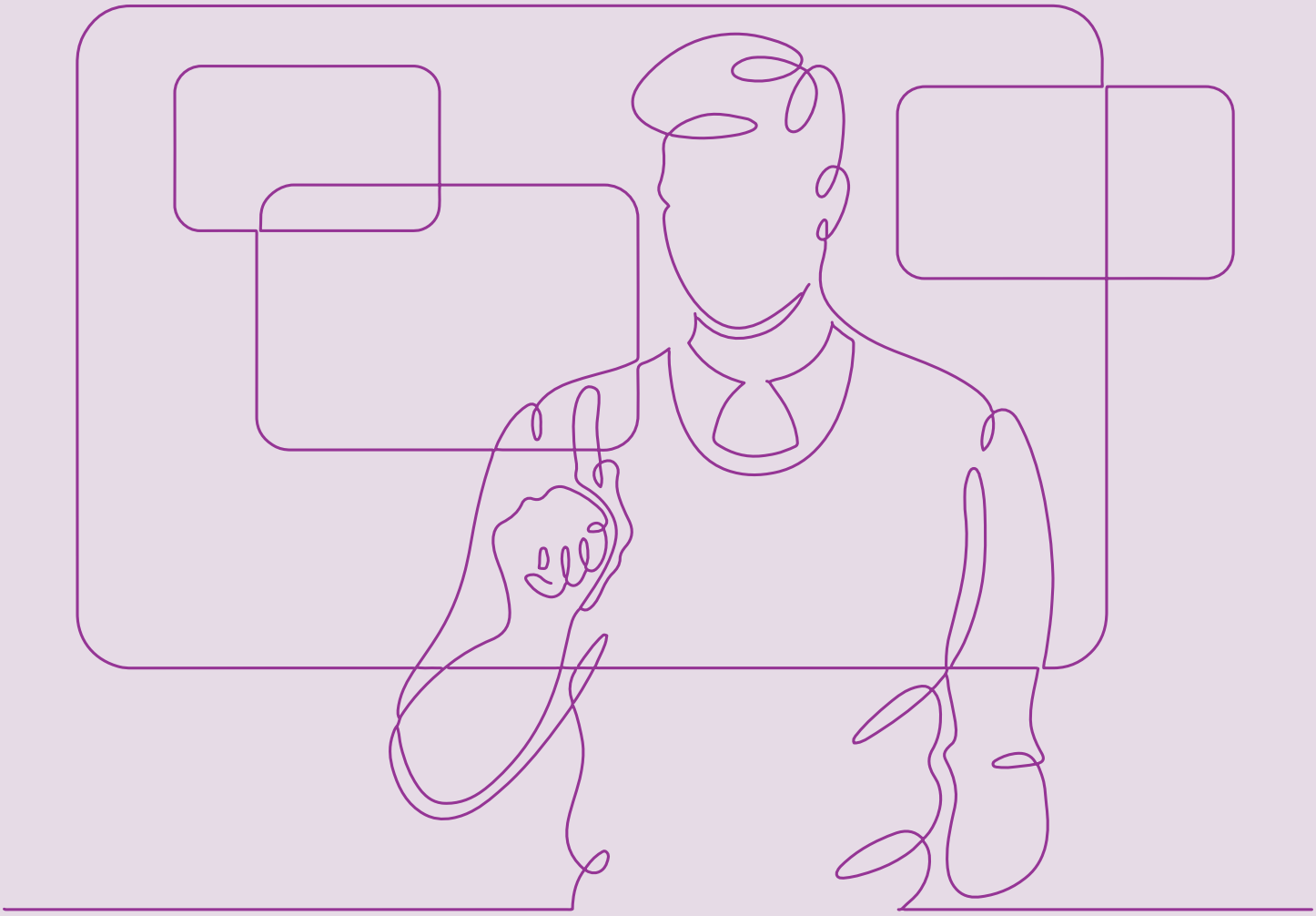
23. ICO AI & Data Protection Guidance: Guidance on AI and data protection | ICO

24. The Future Of AI In Financial Services: The Future Of AI In Financial Services

25. NITI Aayog (2018). Responsible AI for All. Available at: NITI AAYOG, India | Health and Family Welfare

26. ICMR Ethical Guidelines for AI in Biomedical Research and Healthcare (2023) ICMR Official Site [www.icmr.gov.in]

27. WHO AI Ethics & Governance for Health (2024) WHO Guidance [www.who.int]
28. OECD AI in Health & Biotech (2025) OECD AI in Health [www.oecd.org]
29. ISO/IEC 42001:2023 – AI Management System ISO Standard [ceur-ws.org]
30. ISO/IEC 23894:2023 – AI Risk Management ISO Standard [ceur-ws.org]
31. Frontiers Review on Responsible AI in Biotech (2025) Frontiers Article [www.frontiersin.org]



Contributors

Authors

Kiranmayee Janardhan (Responsible AI Consultant, Infosys Responsible AI Office)
Madhusmita Panda (Responsible AI Consultant, Infosys Responsible AI Office)

Reviewer

Ashish Tewari
Head – Infosys Responsible AI Office (India)

Build intelligent systems that protect tomorrow

Infosys Topaz is an AI-first set of services, solutions and platforms using generative AI technologies. It amplifies the potential of humans, enterprises, and communities to create value. With 12,000+ AI assets, 150+ pre-trained AI models, 10+ AI platforms steered by AI-first specialists and data strategists, and a 'responsible by design' approach, Infosys Topaz helps enterprises accelerate growth, unlock efficiencies at scale and build connected ecosystems. Connect with us at infosystopaz@infosys.com

For more information, contact askus@infosys.com

Infosys®
Navigate your next