Whitepaper

LIFE SCIENCES AI

VB INSIGHTS

Ethical & Trustworthy Artificial Intelligence

A Governance Framework for Life Sciences

From ethical principles to risk-based operating controls for safe, responsible, and innovation-enabling AI adoption across the drug development lifecycle.

May 2026

Author: Dr. V. "Bala" Balasubramanian, PhD, MBA
Strategic Advisor | VB Insights, LLC

Patient Safety | Human Oversight | Data Integrity | Regulatory Readiness | Public Trust

Executive Summary

Trustworthy AI needs an operating governance model

Artificial intelligence is moving from isolated discovery tools into an enterprise capability across research and development, clinical development, regulatory operations, pharmacovigilance, manufacturing, and healthcare delivery. That expansion creates enormous opportunity, but it also increases the need for governance that is clear, practical, and continuously maintained.

This whitepaper translates the ethical AI principles issued by the World Health Organization and emerging global health authority expectations into a life sciences AI governance framework. The intent is to help organizations innovate faster while preserving patient safety, data integrity, regulatory credibility, and public trust. A more detailed analysis of the principles and governance framework described here can be found in a companion paper by the author.

Core message

Regulatory compliance alone is not sufficient. Life sciences companies need governance that embeds ethical AI principles into the culture, evidence model, control environment, and daily decisions of the enterprise.

How to Use this Whitepaper

A guide for leaders, builders, and end users

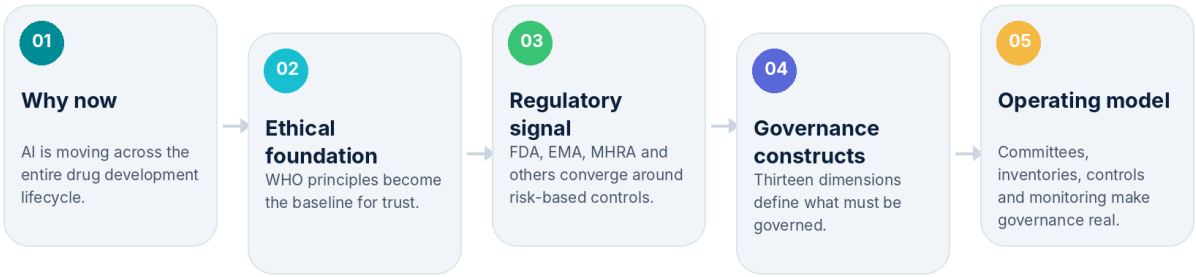
This whitepaper is written for business leaders, clinical and regulatory teams, end users, data scientists, technology teams, legal and compliance teams, quality professionals, and governance bodies. It is intentionally organized as a practical blueprint rather than a purely conceptual policy document. A more detailed analysis of the principles and governance framework described here can be found in a companion paper by the author.

Executive leaders Use the framework to set sponsorship, risk appetite, accountability, and investment priorities for responsible AI adoption.	**Clinical, regulatory, and quality teams** Use the constructs and risk catalog to evaluate whether AI outputs are safe, reliable, explainable, and auditable.
Data, technology, and product teams Use the model card, triage canvas and monitoring metrics to design AI solutions that are fit for their intended use.	**AI governance committee** Use the roadmap and operating dashboard to move from policy creation to repeatable review, evidence, and oversight.

This whitepaper follows five questions: why is governance urgent now; which ethical principles should anchor decisions; what do global regulators expect; what dimensions must be governed; and how can a life sciences organization operationalize governance without creating unnecessary friction?

A practical journey from principles to operational trust

The whitepaper helps executives, quality leaders, data teams and product owners move from policy intent to controlled adoption.



Reader map: from ethical principles to operational governance

1. Introduction

AI innovation is accelerating the drug development lifecycle

Artificial intelligence has played an increasingly prominent role in early discovery activities within life sciences for more than a decade. Beyond genomics, proteomics, systems biology, target identification and molecular screening, AI applications now extend across the full spectrum of R&D, including clinical development, pharmacovigilance, and manufacturing [1,2].

AI is also being integrated into healthcare systems to optimize diagnostics, treatment planning, and patient monitoring. More recently, generative AI has provided a basis for broader experimentation and adoption across the entire value chain.

As organizations increase experimentation, due attention must be paid to guardrails and governance that prevent questionable or harmful uses of AI inside and outside the enterprise. Exploration of use cases should be matched by equal emphasis on ethical AI principles, risk management, and operational governance.

Design principle

AI governance should not be positioned as an innovation blocker. It should be embedded into existing governance, quality, and decision processes so that trustworthy AI becomes easier to implement than unmanaged AI.

2. Ethical AI Principles as the Foundation

Principles become meaningful when they shape decisions

The WHO issued ethical AI guidance for health beginning in 2021 and continuing through 2024 [9,10,11]. Health authorities have also begun publishing position papers, draft guidance and principles addressing AI in medical products and drug development. Examples include good machine learning practice principles from FDA, Health Canada, and MHRA, as well as the FDA draft guidance on AI to support regulatory decision-making [3,4].

The European Union AI Act and related EMA expectations add a risk-based legal and regulatory context for high-impact uses, particularly those influencing clinical decisions, trial outcomes, patient selection, or diagnostic pathways [14,15]. Other regions, including ASEAN, are also developing AI governance and ethics guidance [17].

This whitepaper synthesizes those ethical and regulatory signals into governance constructs, a risk management lens, and an implementation framework for life sciences enterprises.

Ethical AI principles issued by the WHO should form the foundation for trustworthy and responsible AI use in healthcare and life sciences. The principles address autonomy, safety, transparency, accountability, inclusiveness, equity, sustainability, and responsiveness [9,10,11].

In practice, responsibility and accountability are distributed across a broad ecosystem: foundational technology developers, application developers, deploying organizations, end users, health authorities, governments, world bodies, and the public. Each stakeholder carries a role in preventing misuse and building trust.

Figure 1. WHO ethical AI principles translated into operational guardrails

Ethical AI training should reach beyond specialist AI teams. Developers, business owners, clinical and regulatory users, legal and compliance teams, quality functions, and executives should all understand the principles well enough to identify red flags and escalate concerns.

Going from passive principles to active controls

A frequent failure mode in AI governance is the creation of high-level principles that sound appropriate but do not change behavior. This paper highlights the need for organizational awareness and operationalization so that ethical principles do not remain passive statements [12].

Autonomy -> human control

Document who can accept, reject, or override AI outputs. Confirm consent and data access controls where patient or participant data is involved.

Safety -> harm minimization

Define validation thresholds, safety escalation criteria, and review points for false-positive and false-negative recommendations.

Transparency -> intelligible evidence

Create model cards, plain-language AI labels, and decision rationale that non-technical stakeholders can understand.

Accountability -> named owners

Assign business, technical and quality owners; define escalation paths and preserve audit trails for review and redress.

Equity -> bias testing

Test outputs across relevant demographic, clinical and operational subgroups; document limitations and remediation plans.

Sustainability -> responsible use

Assess environmental and societal impacts, and prohibit use cases that could create unacceptable public harm.

Governance interpretation

Every principle should map to four things: a policy expectation, a control, a piece of evidence and a named owner.

3. Health Authority Regulatory Frameworks

Regulators are converging around risk-based evidence

The early regulatory foundation for AI in health products includes the Good Machine Learning Practice principles jointly issued by FDA, Health Canada, and MHRA for medical device development [3]. Those principles emphasize robust software engineering, data quality assurance, cybersecurity, independent training and testing data, human oversight, clear user information, and continuous monitoring of deployed models.

In January 2025, FDA issued draft guidance on the use of AI to support regulatory decision-making for drug and biological products [4]. The framework begins with the question of interest and context of use, then moves to model risk, credibility planning, execution, documentation, and evaluation.

Figure 2. FDA draft guidance flow for AI credibility assessment in regulatory decision-making

Four aspects are especially important: the question being addressed, how AI will be used, the risk and credibility of the AI model, and the adequacy of results for the defined context of use. The draft guidance repeatedly emphasizes early and ongoing consultation with the regulatory agency [4].

Common expectations with different legal and strategic emphasis

FDA, EMA, and MHRA are all moving toward risk-based frameworks for AI in the medicinal product lifecycle, though their approaches reflect different legal environments and strategic priorities. FDA and EMA increasingly emphasize human-centered design and context of use; EMA operates within the broader EU AI Act environment; and MHRA emphasizes a pro-innovation, principles-based approach [13,14,15,16].

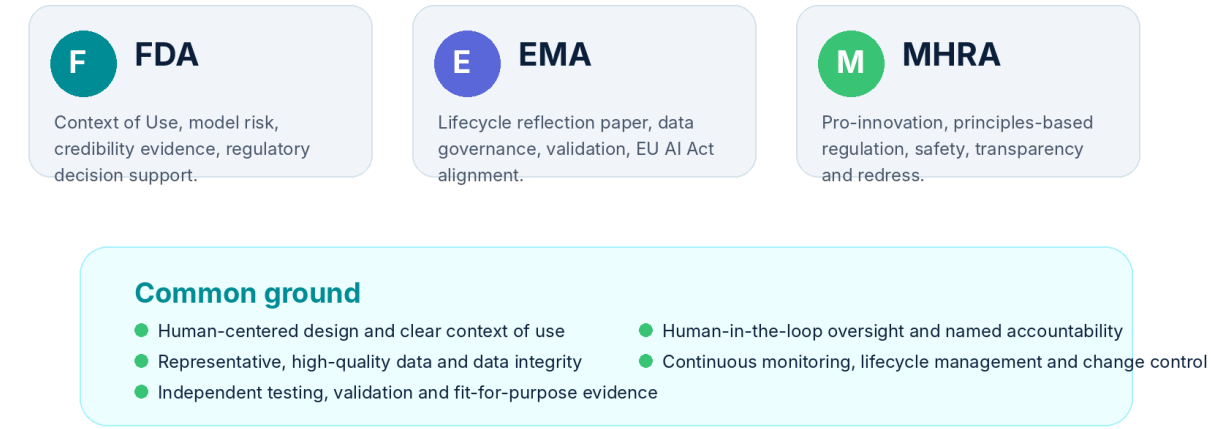


Figure 3. Regulatory convergence around risk-based, human-centered AI governance

Dimension	FDA	EMA	MHRA
Primary focus	AI throughout drug and biological product development; regulatory decision support.	AI throughout the medicinal product lifecycle; EU legal alignment.	Pro-innovation AI regulation across sectors; AI as a medical device experience.
Risk approach	Model influence and consequences drive risk assessment and credibility needs.	High patient risk or high regulatory impact requires stronger scrutiny.	Context-based, proportionate approach weighing benefits and risks.
Data expectations	Data quality, reliability, relevance, representativeness, privacy, and provenance.	Balanced and representative data; documented acquisition and processing.	Representative and independent data; QMS traceability and bias controls.
Oversight	Human accountability through planning, development, use, modification, and discontinuation.	SOPs for high-risk/high-impact cases and human-centric approach.	Effective human oversight and clear lines of accountability.
Innovation mechanisms	Pilot programs and engagement pathways.	Regulatory sandboxes and controlled use environments.	Multi-regulator AI sandbox and international collaboration.

Good AI Practice can become a governance checklist

The FDA and EMA have jointly released guiding principles of good AI practice in drug development [13]. Although concise, the principles provide a practical anchor for future guidance and for internal governance design. They can also help organizations harmonize use-case review across clinical, regulatory, quality, data, and technology teams.

Good AI Practice: ten principles to translate into controls

The FDA/EMA guiding principles create a compact checklist for drug development use cases.

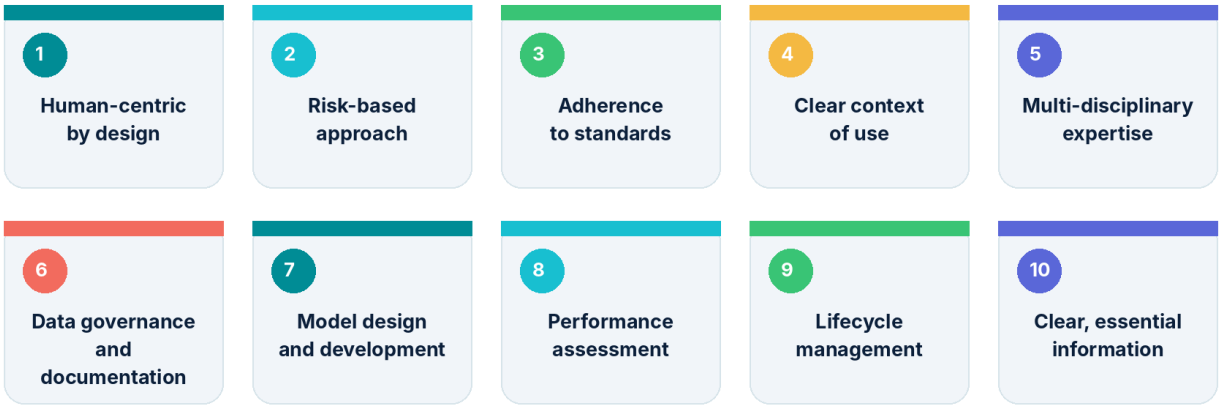


Figure 4. Ten principles for Good AI Practice in drug development

Governance translation

For each AI use case, ask: which Good AI Practice principles apply, what evidence proves they have been addressed, who owns that evidence, and how will it be refreshed after deployment?

4. Key Ethical AI Governance Constructs

Thirteen constructs define the AI governance perimeter

The ethical principles and regulatory frameworks point to a common set of governance constructs. These constructs answer the question: what dimensions must be governed for AI use to be trustworthy in life sciences?

The constructs apply across the ecosystem. Technology platform developers, application developers, end users, life sciences companies, healthcare companies, health authorities, governments, world bodies, and the public all participate in the trust chain.

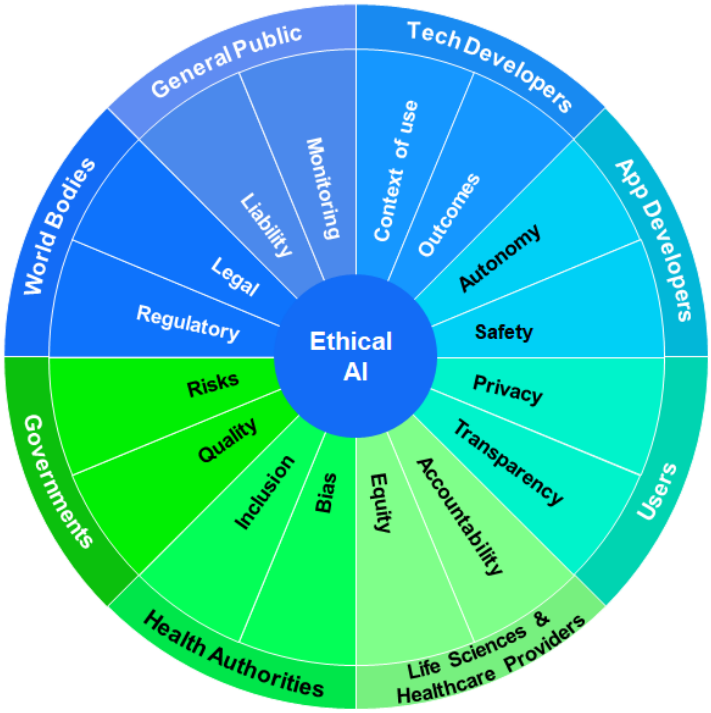


Figure 5. Key ethical AI governance constructs and stakeholder responsibility network

Constructs 1-7: define purpose, preserve control, and protect people

1. Context of use and outcomes

Define the specific role and scope of the AI model, the question of interest, intended outcomes and foreseeable unintended outcomes. Use model cards and plain-language labels to communicate limits.

2. Autonomy

Ensure users retain control over AI applications, model selection, and outputs. Provide accept, reject, and override mechanisms; respect consent and data access rights.

3. Patient safety and harm minimization

Test AI applications to prevent errors that could harm patients or product quality. Establish safeguards for high-impact scenarios such as patient recruitment or remote monitoring.

4. Privacy and data protection

Protect patient data across its lifecycle. Ensure training data integrity, anonymization where appropriate, cybersecurity and compliance with applicable privacy regulations.

5. Transparency and explainability

Document data sources, logic, decision methods, and rationale. Make AI-supported decisions understandable to clinicians, regulators, patients, and reviewers.

6. Accountability and responsibility

Identify the human-in-the-loop and the person accountable for the output, including override authority, review requirements, and escalation duties.

7. Equity, fairness, and non-discrimination

Ensure AI does not perpetuate or amplify bias. Test training data and outputs across relevant populations and operational contexts.

Constructs 8-13: make AI reliable, compliant, and continuously accountable

8. Quality and reliability

Use high-quality data and ensure outputs are consistent, repeatable, auditable, and traceable. AI quality depends directly on data quality.

9. Regulatory and legal compliance

Track and comply with local, regional, and global AI requirements, including draft guidance, final regulations, and jurisdiction-specific legal obligations.

10. Risk mitigation

Catalog legal, ethical, regulatory, safety, data, model, and operational risks. Create mitigation plans that reflect AI-specific risks such as opacity, drift, and hallucination.

11. Liability

Identify accountable individuals with sufficient authority and training to address legal, ethical, and financial consequences of AI recommendations or decisions.

12. Monitoring

Continuously monitor model performance, training data quality, output quality, hallucinations, drift, vendor changes, and undesirable outcomes. Conduct periodic AI audits.

13. Sustainability and societal impact

Assess environmental and societal impact, align AI recommendations with public interest, and prohibit harmful uses such as chemical or biological weapon development.

Training implication

Every stakeholder should understand these constructs well enough to recognize misuse, challenge questionable outputs and escalate potential harm.

5. Risk Management

AI risk management should be a core regulatory pillar

Risk management is a significant part of governance. In drug development, AI risk management should move beyond a technology concern and become a core regulatory and quality pillar. Formal approaches such as the NIST AI Risk Management Framework can help organizations structure risk work [21].

AI-specific risk identification must look beyond traditional clinical or software risks. It should include data bias, model drift, opacity, hallucination, privacy leakage, inadequate traceability, overreliance on generated content, vendor change and regulatory uncertainty.

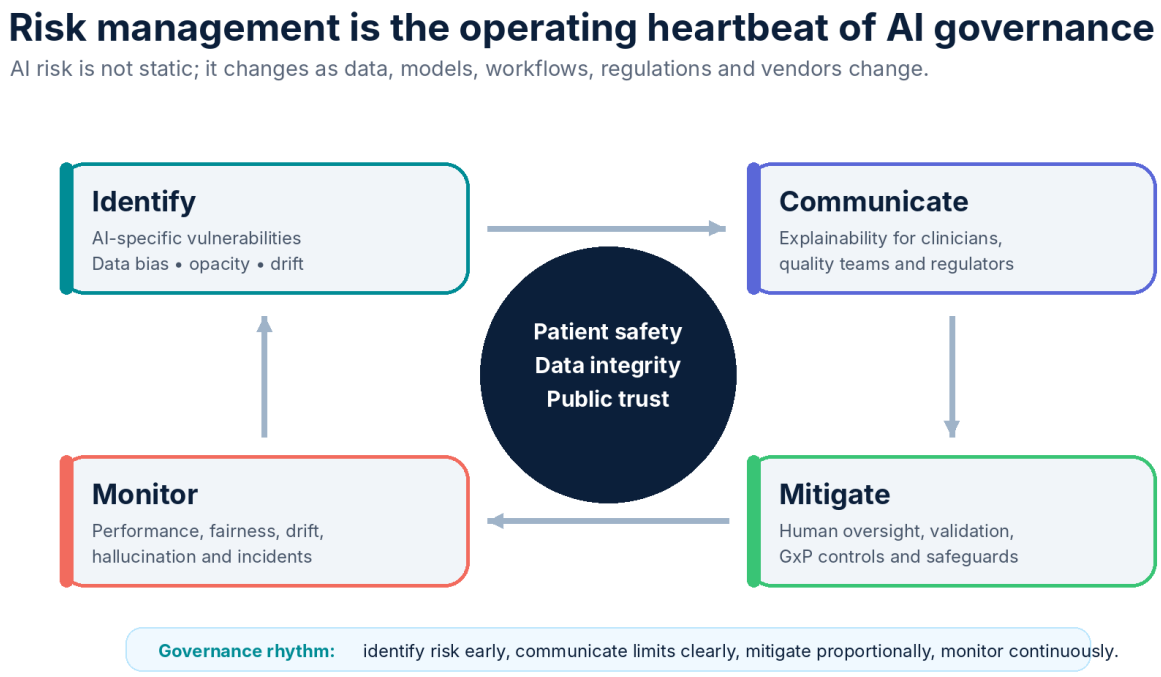


Figure 6. Continuous AI risk management loop

Risk communication is the bridge to regulatory trust. Explainability, model documentation, validation evidence and plain-language limitations help clinicians, quality teams and regulators understand when AI is fit for use and when it is not.

A practical risk control catalog for life sciences AI

Risk scenario	Why it matters	Example controls
Data bias	Training or real-world data under-represents clinically relevant populations.	Representativeness assessment, subgroup testing, bias remediation plan, and data stewardship review.
Model drift	Performance degrades as patient populations, processes or source systems change.	Drift metrics, alert thresholds, scheduled revalidation, and change-control triggers.
Hallucination	Generative AI produces plausible but incorrect content.	Human review, source grounding, output verification, restricted use for high-impact decisions.
Privacy leakage	Sensitive patient or proprietary data is exposed to unauthorized parties or public models.	Access controls, anonymization, approved model environments, vendor due diligence, and logging.
Loss of traceability	Outputs cannot be reconstructed, audited, or explained.	Prompt/output capture, model/version logging, data lineage, and evidence repository.
Overreliance	Users defer to AI despite uncertainty or limitations.	Training, confidence labels, required review, override rights and accountability statements.
Vendor/model change	External model or platform changes alter performance or risk.	Vendor change notifications, revalidation triggers, controlled model versioning, and rollback plans.
Regulatory misalignment	AI use lacks documentation for regulatory expectations or jurisdictional requirements.	Regulatory scanning, early engagement, submission evidence package, and global/local policy mapping.

High-impact rule

When an AI system can influence patient selection, clinical interpretation, safety assessment, regulatory evidence or product quality, assume a higher evidence burden until the governance committee determines otherwise.

6. Operationalizing AI Governance for Life Sciences

Governance cannot be standalone

A life sciences enterprise should create operational AI governance that aligns with WHO ethical principles, government policy, health authority expectations, and emerging global dialogue on AI governance. The United Nations resolution on global AI governance emphasizes international cooperation, safe and trustworthy AI, human rights, transparency, and human oversight [22].

AI governance should be incorporated into existing data governance, enterprise governance, IT governance, and GxP quality management systems. When governance is embedded, it becomes a practical operating discipline rather than an extra layer of bureaucracy.

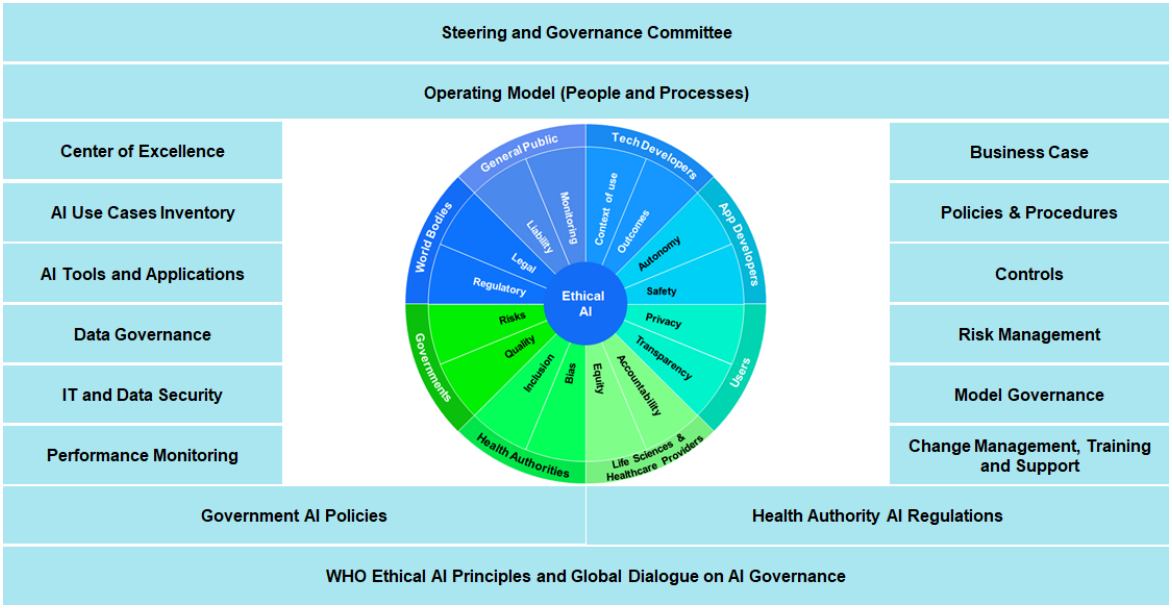


Figure 7. Ethical AI governance blueprint for life sciences enterprises

What each governance layer should do

Steering and governance committee Set risk appetite, approve high-impact uses, resolve cross-functional decisions and review enterprise AI performance.	**AI Center of Excellence** Provide reusable methods, templates, review support, training and guidance for use-case owners and builders.
Use-case inventory Track in-use, approved, planned, and rejected AI use cases; capture rationale to preserve institutional memory.	**Tool and model catalog** Maintain approved tools, model versions, limits, risks, vendor obligations and change notification requirements.

Policies and controls

Define allowed/prohibited uses, issue reporting, output verification, data quality checks, validation, and review requirements.

Model and data governance

Ensure data is fit-for-use, training and test sets are separate, model drift is monitored and changes are controlled.

AI auditors

Review evidence, controls, model cards, incident handling and compliance with ethical AI principles and governance processes.

Change management and training

Prepare business users, IT, developers, policy makers, legal, quality, and executives to verify outputs and raise concerns.

Implementation Playbook

The governance framework identifies fourteen implementation elements. For operational use, those elements can be grouped into four workstreams that leadership can assign, track and mature over time.

Workstream	Implementation elements	Governance outcome
Leadership and accountability	1. Establish AI steering and governance committee 2. Define operating model 3. Establish AI Center of Excellence	Set sponsorship, escalation, decision rights, cross-functional expertise, and program management.
Portfolio and policy	4. Define business case 5. Create AI use-case inventory 6. Catalog AI tools and applications 7. Create policies and procedures	Define allowed and prohibited uses, maintain institutional memory, document tool limits, and publish governance expectations.
Control environment	8. Ensure IT and data security 9. Implement controls 10. Ensure data governance 11. Ensure model governance	Protect data, validate outputs, enforce quality controls, govern model changes, and maintain evidence.
Assurance and adoption	12. Manage risk 13. Monitor performance 14. Implement change management, training, and support	Track risk, measure performance, audit AI applications, train stakeholders and create feedback mechanisms.

Operating model insight

The AI Center of Excellence should not own every AI decision. It should create common methods, reusable artifacts, review pathways, and expert support so business teams can adopt AI responsibly.

Roadmap: Make trustworthy AI measurable in the first 90 days

A governance framework becomes real when it is staffed, measured, and used. The first 90 days should focus on mobilizing leadership, standardizing intake and evidence expectations, and operationalizing dashboards and audit routines.

A pragmatic 90-day activation roadmap

Move from governance concept to repeatable AI review and assurance without slowing innovation.

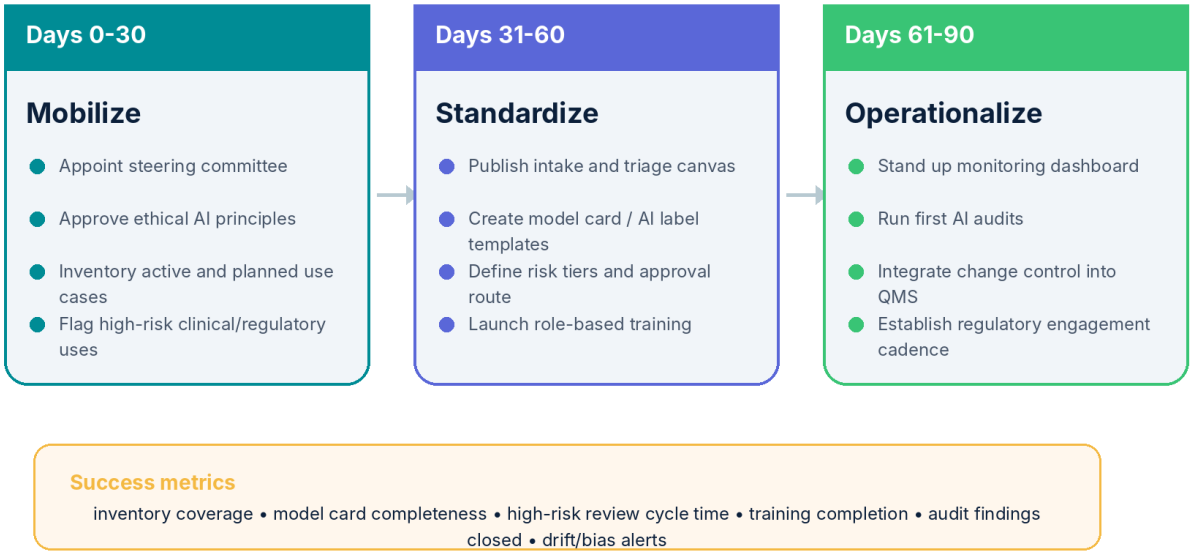


Figure 10. 90-day activation roadmap for life sciences AI governance

Operating dashboard: make trust measurable

The governance committee should see leading and lagging indicators, not anecdotes.

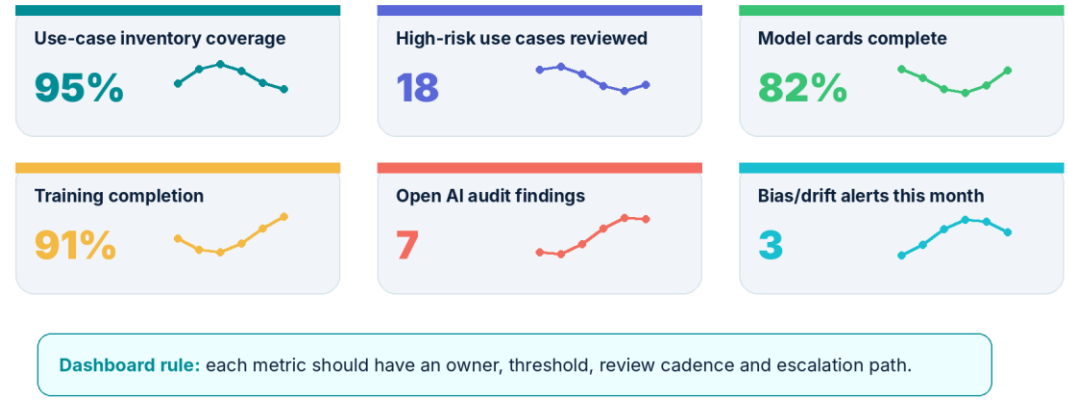


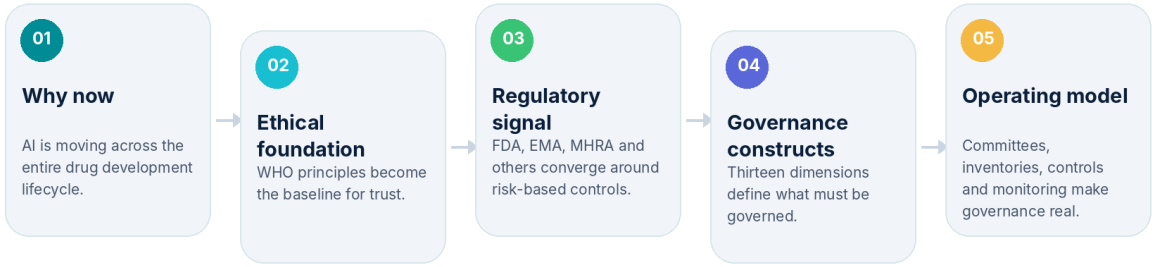
Figure 11. Example AI governance operating dashboard

From governance framework to trusted adoption

Trustworthy AI in life sciences is not a static policy. It should be a living system of principles, controls, evidence, accountability, and monitoring that evolves with technology and regulation.

A practical journey from principles to operational trust

The whitepaper helps executives, quality leaders, data teams and product owners move from policy intent to controlled adoption.



Governance should be repeatable as new use cases, models, data sources, and regulations emerge

Final operating mantra

Define the use. Classify the risk. Validate the evidence. Keep a human accountable. Monitor continuously. Improve relentlessly.

7. Conclusion

Responsible AI is a trust-building discipline

The life sciences industry can use AI to accelerate discovery, improve clinical development, strengthen regulatory operations, enhance pharmacovigilance, improve manufacturing quality, and support better healthcare delivery. That potential must be balanced with a proactive and adaptive governance approach.

This whitepaper synthesized WHO ethical AI principles, global regulatory frameworks, and ethical AI governance constructs into an implementation-oriented framework. The central message is straightforward: responsible AI cannot depend on good intentions alone. It requires clear context of use, high-quality data, robust validation, human accountability, continuous monitoring, change control and a culture that encourages escalation when AI use becomes questionable.

Organizations that prioritize ethical principles, data integrity, validation, regulatory compliance, and continuous oversight will be best positioned to build trust. This approach is not merely a compliance exercise; it is a critical step toward delivering safe and effective medicines and treatments while maintaining public confidence. For a more detailed analysis of the principles and governance framework described here, please contact the author.

Closing thought

The most effective AI governance programs make the right behavior the easiest behavior: clear intake, clear owners, unambiguous evidence, clear monitoring, and clear escalation.

Acknowledgements

The author appreciates feedback and comments on the original draft from industry colleagues Cary Smithson, Sridevi Nagarajan, James Averback and Tim Menard, and from reviewers and editors of TIRS Journal.

8. References

1. Artificial intelligence in drug development: present status and future prospects, Kit-Kay Mak, Mallikarjuna Rao Pichika, 2018, https://pubmed.ncbi.nlm.nih.gov/30472429/
2. Using Artificial Intelligence & Machine Learning in the Development of Drug & Biological Products - Discussion Paper and Request for Feedback, FDA; 2025; https://www.fda.gov/media/167973/download
3. Good Machine Learning Practice for Medical Device Development: Guiding Principles, FDA, Health Canada, MHRA; 2021; https://www.fda.gov/media/153486/download
4. Draft Guidance on Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products: Maryland: FDA; 2025; https://www.fda.gov/media/184830/download
5. Harnessing AI and Automation in Regulatory: Insights from the Recent DIA RIM Intelligent Automation Survey, Cary Smithson, DIA Global Annual Meeting, June 2025.
6. Short Course on Introduction to AI Ethics, Regulations, and Governance for Life Sciences, V. Balasubramanian, Cary Smithson, DIA Global Annual Meeting 2025, June 2025.
7. A Global Perspective on AI for Life Sciences: Part 1: AI Applications, Ethics, Regulations, Guidances, and Business Impact; DIA Global Forum, August 2025, https://globalforum.diaglobal.org/issue/august-2025/a-global-perspective-on-ai-for-life-sciences-part-1-ai-applications-ethics-regulations-guidances-and-business-impact/
8. A Global Perspective on AI for Life Sciences: Part 2: AI Governance, Frameworks, and Implementation; DIA Global Forum, August 2025, https://globalforum.diaglobal.org/issue/august-2025/a-global-perspective-on-ai-for-life-sciences-part-2-ai-governance-frameworks-and-implementation/
9. Ethics and Governance of Artificial Intelligence for Health: WHO guidance, Geneva: World Health Organization; 2021; https://www.who.int/publications/i/item/9789240029200
10. Regulatory considerations on artificial intelligence for health. Geneva: World Health Organization; 2023; https://iris.who.int/bitstream/handle/10665/373421/9789240078871-eng.pdf?sequence=1
11. Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models: Geneva: World Health Organization; 2024; https://iris.who.int/bitstream/handle/10665/375579/9789240084759-eng.pdf?sequence=1
12. Developing a Set of AI Ethics Principles to Shape Ethical Behavior in Drug Development, Menard, Bramstedt, TIRS Journal, 2025; https://link.springer.com/article/10.1007/s43441-025-00766-2
13. Guiding Principles of Good AI Practice in Drug Development: FDA and EMA; 2026; https://www.fda.gov/media/189581/download
14. Reflection paper on the use of Artificial Intelligence (AI) in the medicinal product lifecycle, EMA; 2024; https://www.ema.europa.eu/en/documents/scientific-guideline/reflection-paper-use-artificial-intelligence-ai-medicinal-product-lifecycle_en.pdf
15. EU Artificial Intelligence Act: The European Approach to AI, EU; 2021; https://futurium.ec.europa.eu/sites/default/files/2021-10/Kop_EU%20Artificial%20Intelligence%20Act%20-%20The%20European%20Approach%20to%20AI_21092021_0.pdf
16. A Pro-Innovation to AI Regulation, MHRA Position Paper; August 2023; https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper
17. ASEAN Guide on AI Governance and Ethics, ASEAN; 2024; https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf
18. ISO/IEC DIS 22989(en), Information technology - Artificial intelligence - Artificial intelligence concepts and terminology, ISO, https://www.iso.org/obp/ui/#iso:std:iso-iec:22989:dis:ed-1:v1:en
19. FDA Digital Health and Artificial Intelligence Glossary - Educational Resource, FDA, 2024, https://www.fda.gov/science-research/artificial-intelligence-and-medical-products/fda-digital-health-and-artificial-intelligence-glossary-educational-resource
20. The transparency dilemma: How AI disclosure erodes trust, Oliver Schilke, Martin Reimann, Organizational Behavior and Human Decision Processes, 2025, https://doi.org/10.1016/j.obhdp.2025.104405
21. Artificial Intelligence Risk Management Framework (AI RMF V1.0), NIST, 2023, https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
22. Terms of reference and modalities for the establishment and functioning of the Independent International Scientific Panel on Artificial Intelligence and the Global Dialogue on Artificial Intelligence Governance; Resolution 79/325 adopted by the General Assembly, 26 August 2025; https://docs.un.org/en/A/RES/79/325

About the Author

Dr. V. “Bala” Balasubramanian, PhD, MBA is a Strategic Advisor for the Healthcare and Life Sciences (HLS) industry. Bala was chosen as one of NJ’s INNOVATE100 Community Leaders for 2025. He leads Drug Information Association’s (DIA) AI Consortium and RIM Reference Model Working Groups. Till 2025, he was also the Chair for TiE NJ’s HLS Special Interest Group. Till 2024, Bala was Senior Vice President and Global Head for HLS at Orion Innovation. He was the Co-Founder and CEO of Cabeus, a niche products and services firm specializing in regulatory for Life Sciences. Over his 37 years of experience, Bala has implemented innovative information management solutions for companies such as Bristol-Myers Squibb, Aventis, Roche, Merrill Lynch, AT&T, Bell Atlantic, and IBM. Bala got his Ph.D., and MBA from Rutgers University and MS in Computer Sciences from NJIT. Bala is an Industrial Board Advisor for Rutgers University. Bala can be reached at vbala@vbinsightsllc.com.