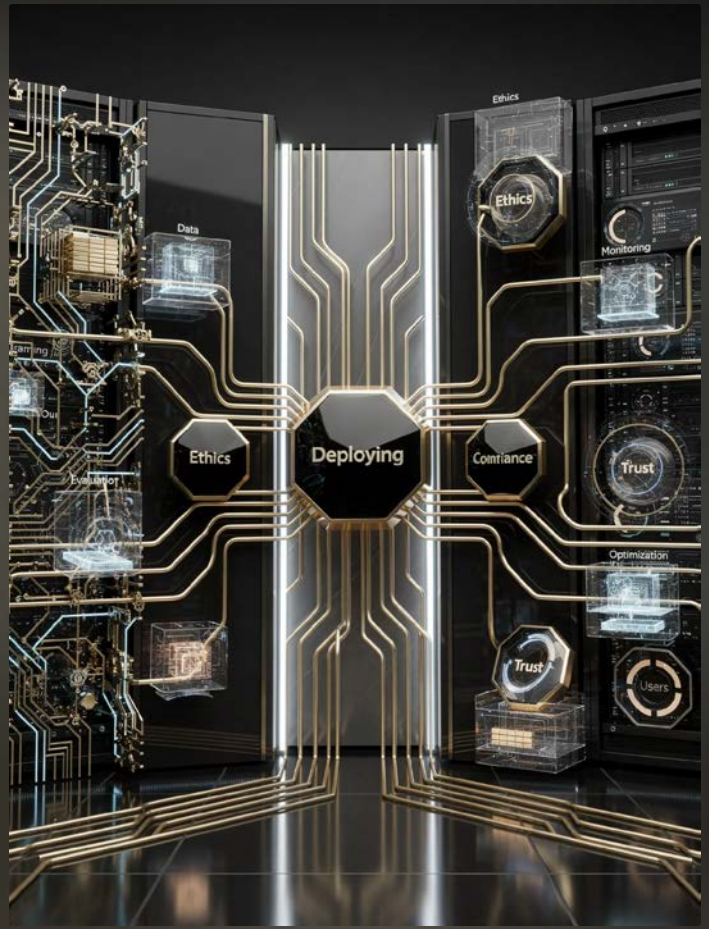


Responsible AI Product Management for Production LLM Lifecycles

A practical framework for building, deploying, and operating LLM-based products integrating ethics, compliance, trust, and operational controls into everyday AI delivery workflows.

By: **Vijayalakshmi Narasimhan**

All views and opinions discussed in this presentation belong solely to me and do not reflect the views/opinions of eBay.

Why Traditional PDLC Breaks Down for LLMs

The Core Problem

Traditional PDLC was built for **deterministic software**: specify once, validate once, and maintain predictably. LLM systems are probabilistic, adaptive, and far less controllable.

Probabilistic Outputs

Same prompt. Different answer.

Model Drift

Performance shifts after release.

Limited Explainability

Hard to trace why a model decided.

Responsible AI as a Product Discipline

Responsible AI is not just a technical or regulatory concern — it is fundamentally a **product management discipline**, sitting at the intersection of strategy, risk, governance, and ethics.

Trust-Aware Requirements

Fairness, explainability, auditability as first-class objectives

Cross-Functional Alignment

Engineering, legal, compliance, UX, and leadership unified

Lifecycle Accountability

Governance ownership beyond deployment

LLM System Characteristics

Probabilistic Outputs

Outputs vary by context, temperature, and retrieval no single "correct" answer

Behavioral Drift

Retraining, provider updates, and user behavior shift model performance over time

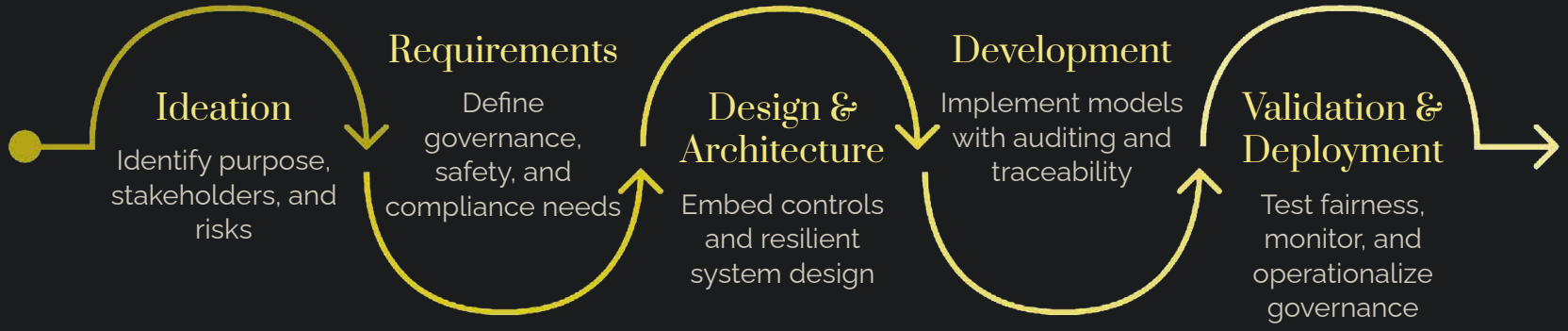
Ethical Risk Exposure

Harmful, biased, or misleading outputs carry legal and reputational consequences

Human-AI Complexity

Users dynamically influence outcomes requiring oversight and expectation design

Responsible AI Across the PDLC



Governance must be embedded at every phase not bolted on at the end. Architectural decisions made during discovery often determine long-term governance outcomes.

Trust-Aware Requirements

AI requirements must expand beyond functionality. Product managers must define measurable objectives across nine trust dimensions:

Fairness & Safety

Identify impacted populations; define unacceptable output categories

Explainability & Auditability

Transparency for users, regulators, and auditors; full logging standards

Reliability & Privacy

Accuracy thresholds, fallback strategies, and data protection controls

Governance Artifacts: Model Cards & Datasheets

1

Model Cards

Intended use, known limitations, performance characteristics, ethical considerations, and risk exposure

2

Datasheets for Datasets

Dataset origins, labeling practices, known biases, data quality limitations, and privacy considerations

Why They Matter

These artifacts maintain institutional knowledge, support audit readiness, and improve transparency across engineering, legal, and governance teams.

Embed documentation requirements into development workflows not as afterthoughts, but as delivery standards.

Validation Beyond Traditional QA

1

Bias Testing

Demographic, linguistic, and cultural scenario evaluation

2

Robustness

Adversarial prompts, injection attacks, edge cases

3

Hallucination Detection

Factual consistency, source grounding, citation accuracy

4

Human Evaluation

Helpfulness, tone, ethical alignment, trust perception

Operational Governance in Production

Deployment Is Not the Endpoint

Operational risk often **increases after release**. Production AI requires continuous governance mechanisms.

- Monitor response quality, hallucination trends, bias indicators, and prompt abuse
- Detect drift via benchmark comparison, shadow testing, and longitudinal evaluation
- Define escalation thresholds and retraining policies

AI Incident Response Demands Immediate Escalation

AI incidents are not like conventional software outages: they can affect users, data, and compliance at the same time, so they require rapid, cross-functional escalation.

Harmful Outputs

Unsafe, biased, or misleading responses that reach users and require immediate review, containment, and rollback.

Data & Security

Prompt injection, unauthorized data exposure, or compromised model behavior that demands security triage and access control checks.

Compliance Failures

Policy violations, regulated-output errors, or missing audit evidence that require legal, risk, and governance escalation.

Applying OECD & NIST Frameworks

OECD AI Principles

Emphasize inclusive growth, human-centered values, transparency, robustness, and accountability. Operationalize through requirement templates, vendor evaluations, and monitoring standards.

NIST AI Risk Management Framework

A four-stage structured approach product managers can embed directly into delivery workflows:

Compliance-by-Design

Regulatory expectations continue to evolve globally. Integrate compliance from the **start** not after deployment.

01

Early Legal Collaboration

Involve legal and compliance during discovery to identify high-risk use cases, data obligations, and consent requirements

02

Documentation Readiness

Maintain risk assessments, governance approvals, evaluation methodologies, and incident records

03

User Transparency

AI disclosures, confidence indicators, source citations, and human escalation pathways

Human-Centered AI Design

Human Oversight

Review mechanisms, override capabilities, and verification checkpoints for high-impact decisions

Expectation Management

Clearly communicate limitations, confidence uncertainty, and appropriate use cases

Accessibility & Inclusion

Multilingual support, cognitive accessibility, cultural sensitivity, and assistive compatibility

Building a Responsible AI Operating Model

Governance Structure

Executive sponsorship, cross-functional review boards, risk classification frameworks, and escalation pathways

Standardized Lifecycle Processes

Repeatable workflows for ideation, validation, deployment approvals, monitoring, and retirement planning

AI Observability Platforms

Prompt tracking, drift analytics, safety violation detection, and model performance dashboards

The Future of Responsible AI Product Management

"The future of AI product management will not be defined solely by technical capability it will be defined by the ability to build systems that users, regulators, enterprises, and society can **trust**."

Near-Term

Real-time governance automation, adaptive policy enforcement, agentic system oversight

Strategic Advantage

Organizations with mature Responsible AI models achieve stronger resilience, customer trust, and regulatory readiness

Thank You