



WHITE PAPER

# Building Responsible AI at Scale through Platform Engineering

[VULTR.COM](https://vultr.com)

# Setting the stage

---

As enterprises increasingly integrate artificial intelligence (AI) into their core operations, the expansion of AI to include both predictive and generative AI is redefining digital innovation. This evolution offers unprecedented potential for growth and efficiency but also introduces significant risks, such as questionable training data provenance, model drift and bias, and heightened data security concerns.

This transformation is playing out as jurisdictions worldwide debate and enact legislation aimed at mitigating harms from problematic AI practices. Thus, the urgency to maintain integrity and trust in AI systems has never been greater. Enterprises urgently need a scalable, robust approach to responsible AI across all operations.

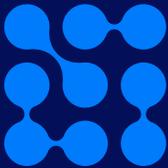
Responsible AI demands comprehensive observability and rigorous data governance throughout the machine learning (ML) model lifecycle. This cannot be achieved with makeshift adaptations of existing IT infrastructures or generic cloud solutions. Instead, it requires platform engineering specifically designed for the complexities and demands of AI operations.

Such platform engineering must manage the entire ML model lifecycle – from development and training to deployment and monitoring – ensuring that model observability and data governance are integral, foundational elements of the infrastructure. This approach enables businesses to meet current demands and scale AI operations responsibly and efficiently.

# Proactively Addressing the Coming AI Legislation

Legislative bodies worldwide are moving swiftly to establish frameworks that govern the development and deployment of AI technologies. One significant piece of legislation in this realm is the [EU AI Act](#), which sets benchmarks for responsible AI practices.

The EU AI Act has several key provisions related to [transparency](#) and [observability](#) requirements:



Providers must ensure transparency around the data sets used to develop high-risk AI systems, including their sources and any limitations or biases.

There must be thorough documentation and traceability of the data lineage - the journey of data from origin to endpoint. This creates an audit trail to demonstrate compliance.

Providers must implement robust governance frameworks to manage data throughout its lifecycle, including establishing processes for data management, documentation, and accountability.

## Global implications

The EU AI Act is just the beginning. Enterprises worldwide must anticipate and adapt to similar legislation in other regions. Proactively integrating observability and governance into AI workflows is not only about compliance but also about securing a competitive advantage by establishing trust and reliability in AI systems.

## Looking ahead

Other jurisdictions, including the United States, are also considering AI-specific regulation. President Joseph Biden's executive order on [artificial intelligence](#), which seeks to balance the needs of cutting-edge technology companies with national security and consumer rights, is one such example. Others include proposals for transparency in AI decision-making processes and mandates for audit trails in AI systems, reflecting a growing global consensus on the need for regulatory oversight.

## Proactive steps for enterprises

To stay ahead of these regulatory changes, enterprises should:

- Build observability into their AI operations from the outset.
- Ensure that data governance frameworks can adapt to new regulations as they arise.
- Engage with legal and compliance teams to anticipate changes and prepare accordingly.

By taking proactive steps today, enterprises can ensure that they not only comply with upcoming regulations but also lead the way in responsible AI development.

# Capabilities of advanced observability solutions

The growing complexity of AI operations and their integration into business workflows has heightened the demand for advanced observability solutions and platform engineering designed to scale responsible AI. These tools are crucial for maintaining control, performance, and compliance as they provide a holistic view of AI models throughout their lifecycle.

Key capabilities of advanced observability solutions include:



## Comprehensive monitoring

The integration of metrics, events, traces and logs in observability solutions provide a real time, full view of an AI model's performance. This integration is vital for detecting anomalies, tracking changes, and assessing the impact of modifications on system behavior.



## Anomaly detection and pattern recognition

Purpose-built generative AI models can learn normal behavior and quickly identify deviations, making observability solutions more proactive in flagging potential issues.



## Explainability and root cause analysis

These tools identify deviations in model performance and pinpoint the reasons behind them. This capability is essential for addressing issues related to model drift, data quality, or operational anomalies.



## Centralized management

Observability platforms act as a centralized hub for monitoring all AI models within an organization, facilitating collaboration among data scientists, ML engineers, and DevOps teams and ensuring that insights are effectively shared.



## Scalable infrastructure

Designed to scale with the needs of the enterprise, these solutions can handle an increasing number of models and complex data streams without sacrificing performance.



## Instrumentation and automation

Automating the instrumentation of observability reduces the manual workload, allowing teams to concentrate on strategic tasks. Automated alerts and notifications enable quick responses to critical issues, ensuring system integrity.



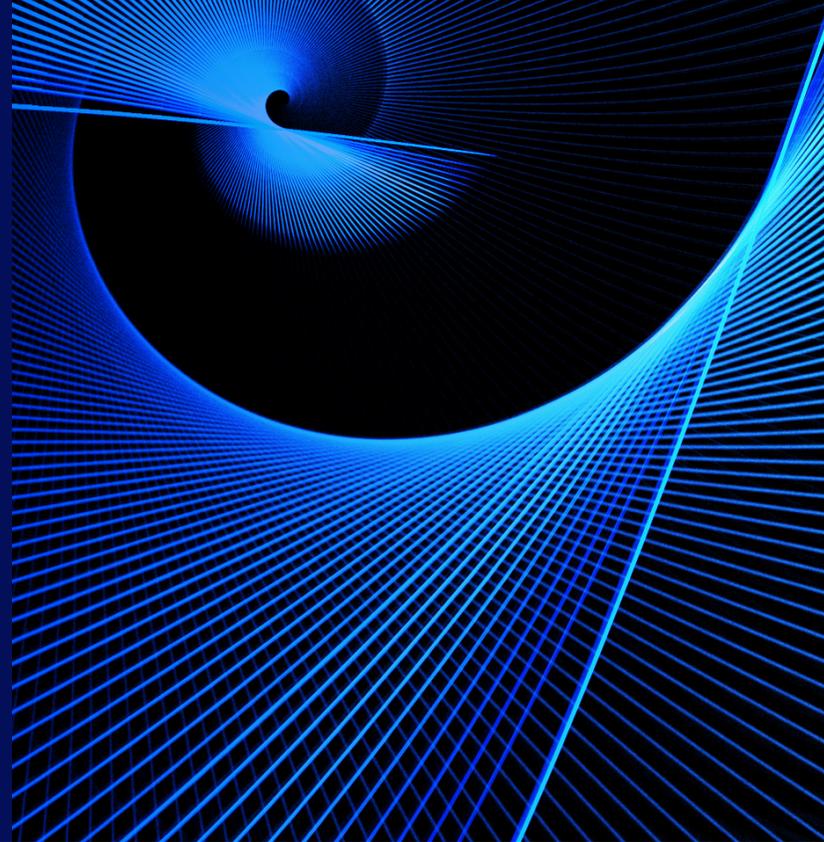
## Collaboration and feedback loops

These solutions promote a culture of continuous improvement with robust collaboration tools and feedback mechanisms, enabling ongoing adjustments to models based on operational feedback and business needs.

Responsible AI requires observability solutions that can provide deep visibility into AI model behavior, enable rapid issue detection and resolution, facilitate collaboration, and support the overall lifecycle of AI systems. These capabilities allow organizations to effectively monitor and manage their AI applications and uphold the accountability and transparency required in modern regulatory and business environments.

# Observability in the AI/ML lifecycle

Observability plays a critical role at every stage of the AI/ML lifecycle, ensuring that models are developed, deployed, and maintained effectively.



There are several key phases where observability needs to be built into the ML model lifecycle:

## ●● Model training, development, and fine-tuning

During these initial stages, observability tools monitor data quality, algorithm performance, and the impact of hyperparameter adjustments. This monitoring is crucial for ensuring the model's initial integrity and effectiveness.

## ●● Model deployment

As models transition from development to production, observability tools track their performance in real-world environments. This stage is vital for detecting early signs of model drift or performance degradation and for validating that the model operates within ethical and legal boundaries.

## ●● Continuous model monitoring

Post-deployment, continuous monitoring is necessary to ensure models perform consistently over time and adapt to new data or conditions without loss of effectiveness or accuracy.

## ●● Ongoing alignment with business goals

Observability supports ongoing improvements and refinements through structured feedback loops among all stakeholders, including data scientists, ML engineers, and business users. This collaboration ensures that models remain aligned with evolving business goals and regulatory requirements.

To ensure comprehensive coverage and functionality, integrating observability into AI stacks involves several critical strategies:

---

**Utilizing a centralized observability framework**

Implementing a centralized framework provides a unified view of all AI models across the organization, enhancing visibility and control.

---

**Leveraging cloud-native services**

Utilizing cloud-native services for observability ensures scalability and flexibility, allowing enterprises to dynamically adjust resources based on their current needs.

---

**Deploying observability instrumentation**

Instrumenting AI models and their operating environments to automatically collect data on performance, usage, and interactions is key to proactive management and issue resolution.

---

**Monitoring and alerting**

Systems set up to continuously monitor AI operations and alert teams to anomalies or deviations from expected performance enable rapid response to potential issues.

---

**Explainability and root cause analysis**

Tools that provide insights into model decisions and behaviors facilitate root cause analysis when anomalies occur. This capability is crucial for maintaining transparency and accountability in AI operations.

---



Through thoughtful integration of observability into the AI lifecycle via platform engineering solutions, enterprises can achieve not only operational excellence but also build trust and compliance into their AI initiatives.

# Responsible AI and platform engineering

Platform engineering is pivotal in accelerating enterprises' AI journeys. It empowers ML engineers with the necessary tools and infrastructure and integrates robust governance and observability across AI/ML operations. Here are specific steps platform engineering teams can take to embed responsible AI components effectively:



## Enable self-service access with integrated observability

Develop self-service portals that provide easy access to AI/ML infrastructure, like GPUs and vector databases, with embedded observability tools. These tools should allow ML engineers and data scientists to monitor model performance, data quality, and operational metrics in real time, fostering a culture of transparency and accountability.



## Bridge development and operations with observability tools

Act as an intermediary that not only provides rapid experimentation capabilities but also ensures these experiments are governed by enterprise-grade security and reliability standards. Implementing centralized observability frameworks within these tools helps maintain a clear view of all AI activities, ensuring compliance and facilitating cross-team collaborations.



## Offer curated templates with built-in governance and observability

Provide carefully vetted templates for common AI/ML workflows, such as model training and fine-tuning, that are pre-equipped with observability features. These templates ensure adherence to the organization's data privacy, ethics, and compliance standards, enabling developers to track and analyze model behavior throughout the lifecycle.



## Automate workflows with observability checks

Leverage AI and intelligent automation to streamline the AI development lifecycle, from testing to deployment, while integrating checks for model drift, bias detection, and ethical AI usage. This automation helps maintain a consistent application of responsible AI principles, reducing manual oversight and enhancing efficiency.



## Establish a responsible AI "red team"

For a team dedicated to testing and tuning AI models to eliminate errors, weed out biases, and validate models before they get moved to the enterprises' central model registry, build in test frameworks and automated access to models in development.



## Scale AI adoption cost-effectively with observability infrastructure

Guide the adoption of AI/ML infrastructure with an approach that balances scalability with cost-effectiveness, similar to successful cloud adoption strategies. Ensuring that this infrastructure supports expansive data monitoring and management capabilities, allowing enterprises to scale responsibly without compromising on observability or governance standards.

By embedding these observability-focused strategies within platform engineering, enterprises can ensure that their AI operations are not only efficient and innovative but also aligned with rigorous standards for responsible AI. This approach guarantees that AI systems are transparent, compliant, and ready to meet the challenges of a rapidly evolving technological landscape.

# What to look for in a scalable AI stack

In today's AI-driven market, choosing the right cloud-based infrastructure is critical for sustaining and scaling AI operations efficiently. Enterprises must consider a range of essential factors to ensure their AI stack is not only powerful and capable today but also prepared for future demands.

## Essential attributes of a scalable AI stack:

### Cloud-based architecture designed for the AI era

Look for an infrastructure explicitly built to handle the demands of AI applications, including high data throughput and intensive computational tasks.

### Public and private container registries

Access to both public and private registries allows for greater flexibility and security in providing enterprise-wide access to all ML models across various AI projects.

### Commitment to affordable resources

It is essential to partner with a vendor that is dedicated to keeping GPU and CPU resources affordable, ensuring that scaling AI operations does not become prohibitively expensive.

### Global footprint of data center locations

A vendor with data centers across the globe ensures that AI operations comply with data residency requirements and minimize latency, enhancing the performance of AI applications.

### Composable architecture

Composability enables businesses to adapt their tech stack swiftly and efficiently as their needs evolve, ensuring long-term scalability and flexibility.

### Predictable and affordable cloud costs

Beyond just affordability, predictable pricing models are crucial for long-term planning and budgeting, helping enterprises manage costs effectively as they scale their AI operations.

### Tightly integrated CPU and GPU stacks

Integration of processing units is crucial for maximizing performance and efficiency, especially in AI models that require intense computational power.

### Broad ecosystem of best-in-class providers

Vendors that offer a comprehensive range of tools, services, and applications can significantly enhance the capabilities of an AI stack, allowing enterprises to leverage the latest advancements in AI technology.

Selecting a cloud vendor that embodies these characteristics ensures that enterprises can leverage the full potential of AI technology, driving innovation while maintaining cost-efficiency and compliance with global standards.



# A final word

As regulatory scrutiny of AI applications heightens, it is crucial for CDOs, CAIOs, CISOs, and CTOs to embrace a platform engineering approach that supports responsible AI at scale. This approach should integrate robust model observability and data governance to ensure transparency and compliance. Partnering with a global cloud provider that upholds these standards is essential for maintaining operational excellence and future-proofing AI strategies against evolving global regulations.

---

To learn more about Vultr, visit [vultr.com](https://vultr.com) or [contact sales](#).

---

[VULTR.COM](https://vultr.com)

[CONTACT](#)