




WHITE PAPER

Global Implications of the EU AI Act

A Proactive Approach for Distributed Enterprises

VULTR.COM



The fluid nature of AI regulation

Enterprises scaling AI operations globally already feel the impact of the [EU AI Act](#), the most significant legislation to date regulating AI practices. Other jurisdictions, from the UK to the U.S. and beyond, have [enacted similar legislation or are working on the same](#). Additionally, a [consortium of the EU plus an additional ten countries](#) participating in the recent AI Seoul Summit have agreed to collaborate “to make sure AI advances human well-being and help address the world’s greatest challenges in a trustworthy and responsible way.”

The clear message is that more AI regulation is coming, creating complexities for distributed enterprises working to scale their AI operations across geographies. Reacting to this evolving regulatory landscape is impractical and unsustainable. Instead, enterprises must proactively institute responsible AI practices, including model observability and data governance.

Doing so requires a systematic approach to embedding responsible AI into AI operations and deliberate choices of infrastructure, tooling, and processes to ensure compliance with current and future regulations while maintaining operational efficiency and innovation.

Infrastructure, tooling, and processes for responsible AI at scale

As enterprises scale their AI operations, they must ensure these operations remain compliant with evolving regulations. Doing so requires robust and flexible infrastructure, comprehensive tooling, and well-defined processes that support end-to-end observability and robust data governance throughout the AI model lifecycle. By strategically integrating these elements, enterprises can build a foundation for responsible AI practices at scale.

Global and composable AI infrastructure

A robust, flexible, and scalable AI infrastructure is essential for enterprises to manage distributed AI operations efficiently. There are several considerations global enterprises must take into account when choosing the optimal infrastructure to scale their AI operations:

Global distribution of data center locations

Enterprises need access to AI infrastructure wherever they do business. In almost all cases, this means aligning with a cloud partner that maintains a global footprint of data center locations configured for distributed AI's unique demands. This minimizes latency and ensures compliance with data residency requirements, which are crucial for effective and compliant AI inference at the edge.

Availability of state-of-the-art GPUs

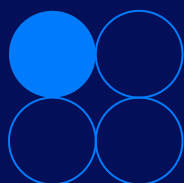
Access to state-of-the-art GPUs, such as the NVIDIA GH200 Grace Hopper™ Superchip and the NVIDIA H100, is essential for handling the intensive computational tasks required by modern AI models. A cloud provider should offer a suite of GPU resources that can match the computational power to the requirements of different AI models.

Tight integration of GPUs & CPUs

Efficient AI operations require the seamless integration of GPUs and CPUs. This integration optimizes workloads by intelligently assigning tasks to the appropriate compute resources, ensuring cost-effective processing and high performance.

Composability to assemble & reassemble the right stack

The rapid evolution of AI and machine learning technologies demands a flexible and adaptable tech stack. A composable architecture allows enterprises to select best-of-breed components at each layer of the AI stack, ensuring interoperability and future-proofing the organization against changing requirements or emerging technological capabilities. Composability empowers enterprises to customize their AI stacks for specific business requirements and switch out components as needs evolve without being limited by vendor lock-in constraints.



By leveraging a global and composable AI infrastructure, enterprises can ensure their AI operations are resilient, scalable, and compliant with local, regional, or national regulations. This foundation is critical for embedding responsible AI practices across all AI operations.

Platform engineering purpose-built for responsible AI at scale

To scale responsible AI, enterprises need platform engineering solutions purpose-built to incorporate end-to-end observability and strong data governance throughout the AI model lifecycle. This approach ensures that AI models operate ethically, securely, and in compliance with regulatory standards from development through deployment and beyond.



Self-service access with integrated observability

The baseline functionality of AI platform engineering is providing machine learning engineers and data scientists with self-service access to AI/ML infrastructure, including GPUs, CPUs, and vector databases. For responsible AI, platform engineering solutions must also integrate observability tools to enable real-time monitoring of model performance, data quality, and operational metrics, fostering transparency and accountability.



Curated templates with built-in governance and observability

Offering vetted templates for common AI/ML workflows, pre-equipped with observability and governance features, ensures adherence to data privacy, ethics, and compliance standards. These templates help streamline the development and deployment processes, making it easier for teams to implement responsible AI practices.



Automated workflows with observability checks

It is essential to 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. Automated workflows reduce manual oversight and enhance efficiency, ensuring consistent application of responsible AI principles.



Internal red team to probe for vulnerabilities

Establishing a dedicated team to test and tune models before they are moved to production helps eliminate errors, weed out biases, and validate models. This proactive approach ensures that only robust, compliant models are deployed, maintaining the integrity and trustworthiness of AI operations.



Centralized management and continuous monitoring

Implementing a centralized observability framework provides a unified view of all AI models across the organization, enhancing visibility and control. Continuous monitoring of models in production ensures they maintain accuracy and effectiveness over time, adapting to new data or conditions without compromising performance.



Collaboration and feedback loops

End-to-end observability facilitates structured feedback loops among data scientists, engineers, and stakeholders, aligning models with evolving business objectives, regulatory requirements, and ethical considerations. This collaboration promotes ongoing improvements and refinements to AI models.

By embedding these strategies within platform engineering solutions, enterprises can ensure their AI operations are efficient, innovative, and aligned with rigorous standards for responsible AI. This approach guarantees that AI systems are transparent, compliant, and ready to meet rapidly evolving technological and regulatory landscape challenges.

A centralized operating model for AI at scale

To scale AI operations efficiently and responsibly, enterprises should adopt an operational paradigm that centralizes the initial development and training of models while enabling localized fine-tuning and deployment (see Figure 1):

AI center of excellence

Enterprises concentrate their data science expertise in a centralized AI Center of Excellence. This hub serves as the core for developing and training AI models, leveraging the collective knowledge and resources of the organization's top data scientists.

Drawing on open-source AI models

Rather than developing new models from the ground up, the AI Center of Excellence data science team utilizes open-source AI models available in public registries as foundational elements. This approach is essential as enterprises scale their AI operations and simultaneously have scores of models in training and production.

Initial training on proprietary data

By training open-source models on proprietary data, enterprises create proprietary models that reflect the organization's unique insights and capabilities. This step ensures that the models are prepared to address the enterprise's specific business objectives.

Private model registries

All proprietary models are containerized and stored in a private registry to protect the intellectual property they now contain. This "walled garden" comprises a complete inventory of the enterprise's proprietary models, ensuring they are visible and available to localized data science teams distributed across the enterprise's various geographies.

Fine-tuning by localized data science teams

Data science teams working in different geographies leverage the enterprise's AI platform engineering solution to access the inventory of proprietary models in the private registry. Data science teams set up Kubernetes clusters in edge locations and deploy the containerized AI models to these edge clusters. Here the data scientists fine-tune the models on regional or local data to account for specific regional characteristics and anomalies while maintaining compliance with local data governance requirements.

Vector databases for retrieval augmented generation (RAG)

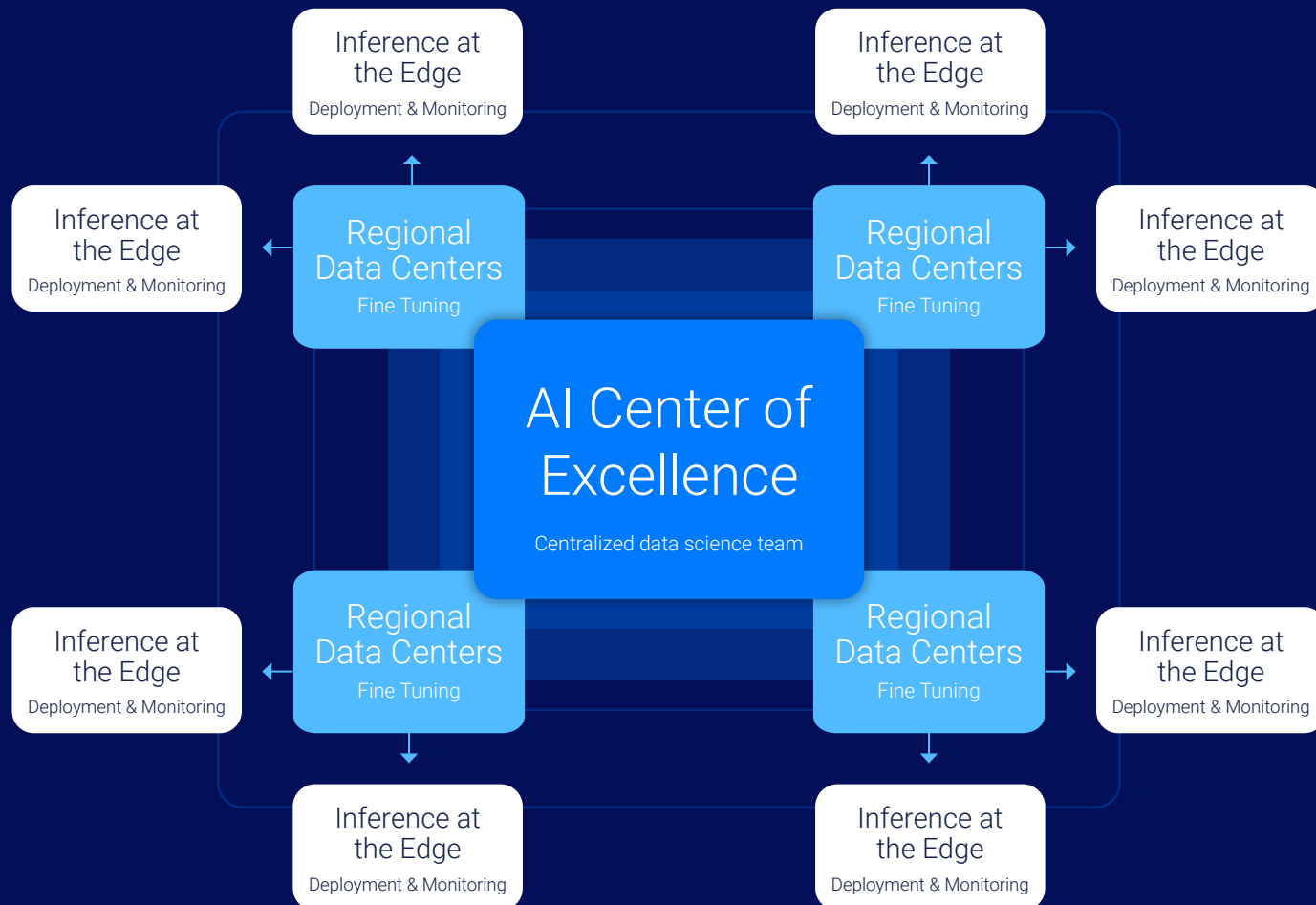
Data science teams store relevant (and often highly confidential) data they wish to exclude from the core training data as embeddings in vector databases, improving the quality, accuracy, and factuality of the model's outputs. Storing such data as embeddings offers three key benefits: it allows data scientists to incorporate up-to-date information from external sources that may not be present in the original training data; it makes the model's outputs more transparent by providing sources for the retrieved context; and it reduces the need to retrain the model as new data becomes available.

Deployment and monitoring

Fine-tuned models are moved to production, where the local data science teams leverage observability tools in the platform engineering solution to continuously monitor model performance. This localized monitoring ensures that data scientists can quickly adapt the models to account for any changes or anomalies in the local environment and correct any instances of drift or bias.

FIGURE 1

A centralized operating model for AI at scale



Model development starts in this centralized hub housing the organization's top data science team.

- > Open-source models from public registries form the foundation of the enterprise's AI model inventory.
- > These models are trained on proprietary company data, thereby creating proprietary models.
- > Proprietary models are containerized and stored in a private registry housing the full inventory of the enterprise's models.

Model development continues with fine-tuning on localized data to account for regional characteristics and data governance requirements.

- > Data science teams set up Kubernetes clusters in edge locations to deploy the containerized AI models.
- > Relevant data they wish to exclude from the core training data is stored as embeddings in vector databases.

AI operations culminate in model deployment, monitoring, and inference in edge environments.

- > Data science teams leverage observability tools to continuously monitor model performance and correct any instances of drift or bias.

Responsible AI as a competitive advantage

Many executive leaders may consider regulation to be a business inhibitor. While this notion may be accurate in some contexts, when it comes to AI regulation, executive leaders should view their approach as a business opportunity.

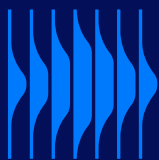
Assuming a proactive posture toward AI regulation by instituting responsible AI and providing transparency across all AI operations helps build trust in AI and, by extension, trust in the brand. Positioning the company as operating ahead of the coming regulation demonstrates wisdom and foresight, further reinforcing the brand's (and its executives') identity as a leader.

Embracing strong data governance and compliance can give those wary of AI a reason to believe. Brands that exhibit unflappable adherence to responsible AI provide an answer to the question, "What's your commitment to ethical AI?" without explicitly being asked. Responsible AI becomes a formidable differentiator for brands that can stake this claim.

Looking ahead to the new reality of increasing AI regulation

As new AI regulations continue to emerge in different parts of the world, monitoring the tangled web of AI legislation grows increasingly complex. The only way for enterprises to proceed in such a challenging global environment is to adopt a proactive rather than a reactive posture toward responsible AI. Setting the highest standards for responsible AI across the organization is the winning formula that will keep enterprises in good standing with all AI regulations while safeguarding the employees and customers who use their AI models.

Leveraging the proper infrastructure, tools, and processes is crucial for scaling AI operations responsibly. This includes utilizing global and composable AI infrastructure, platform engineering solutions tailored for AI, and adopting a centralized operating model to manage and deploy AI models across geographies efficiently.



By taking these proactive steps, enterprises can maintain the agility needed to keep their AI operations relevant and productive amid an evolving regulatory landscape. This approach ensures compliance and drives innovation and efficiency, positioning enterprises to lead in the AI-driven future.



A final word

The path to responsible AI at scale is clear: by setting high standards and leveraging the proper infrastructure, tools, and processes, enterprises can scale their AI operations responsibly and effectively. This strategic approach will enable them to navigate the complexities of increasing AI regulation, build brand affinity among employees and customers, and ultimately drive sustained innovation and excellence in business operations.

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

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

[CONTACT](#)