

The Executive Guide to Responsible AI on Databricks

A unified approach to AI can help you deliver model quality, provide more secure applications and maintain compliance with regulatory standards

Artificial Intelligence (AI) has the potential to transform business, but its advancement relies on building trust in intelligent applications by following responsible AI development and use practices. Leaders may be in a rush to innovate with AI, but they must first consider safety and ethical issues. A careful balance should be struck between exercising caution and embracing experimentation. The most effective leaders will ensure they have strong policies and security in place before they roll out generative AI across the business. Failing to adhere to responsible AI use can expose organizations to three common challenges:

Challenges

1

Lack of visibility into model quality

Generative AI tools are known to produce “hallucinations.” This is when the model generates inaccurate or nonsensical responses as a result of limitations in its understanding but presents it as fact. Hallucinations can be dangerous when models generate unsafe or biased recommendations or disclose private or sensitive information. This can lead to a lack of trust in the reliability of AI models to consistently deliver outcomes that are safe and fair. Leaders need to ensure staff is adequately trained on the limitations of generative AI so they don’t inadvertently create content or develop products and solutions based on false information.

2

Inadequate security safeguards

Interactions with AI models expand the attack surface of an organization by providing a new way for bad actors to interact with data. A lack of security safeguards can allow applications like chatbots to reveal sensitive data and proprietary intellectual property.

3

Siloed governance

Governance for data and AI is complex. Organizations frequently deploy separate data and AI platforms, creating governance silos that result in limited visibility and explainability of AI models. This disjointed approach leads to inadequate cataloging, monitoring and auditing of AI models, which can impede the ability to guarantee their appropriate use. Furthermore, a lack of data lineage complicates understanding which data is being utilized for AI models and obstructs effective oversight.

These challenges can lead to multiple pain points, most notably:

- Legal risk from inappropriate model decisions
- Risk of data breaches
- Misuse and regulatory non-compliance

“

80% of companies plan to increase investment in responsible AI

accenture

The Solution

A unified approach to AI empowers organizations to meet responsible AI objectives and:



Improve confidence in model decisions

Organizations have confidence in the development of models that adhere to their ethical standards and policies. By proactively addressing risks, these models can be confidently deployed into production.



Reduce legal risk from inappropriate model decisions

A common approach to security protects the data, models and applications that make up the AI system. The associated components are protected end to end by detecting, preventing and logging suspicious behaviors throughout model training, development and inference. Vulnerabilities and breaches are more easily identified and mitigated to avoid business disruption and financial risks.



Enable better control and regulatory adherence

A unified governance framework improves auditability and accountability throughout the AI development and deployment process. Organizations can trace every step and decision of the model development journey and precisely assign responsibility to distinct roles to reinforce accountability at every phase. Compliance is easy with much of the necessary information available automatically.

The Databricks Solution

Delivering AI you can trust

The advancement of AI requires that every organization has ownership and control over their data and models with comprehensive monitoring, privacy controls and governance throughout development and deployment. The Databricks Data Intelligence Platform allows you to unify data, model training, management, monitoring and governance of the entire AI lifecycle. This unified approach enables:



End-to-end AI monitoring

Monitor the entire data workflow — from data and features to ML models — without additional tools and complexity. Developers benefit from auto-generated metrics, root-cause tracing, experiment logging and auto-scanning for toxic and unsafe content.



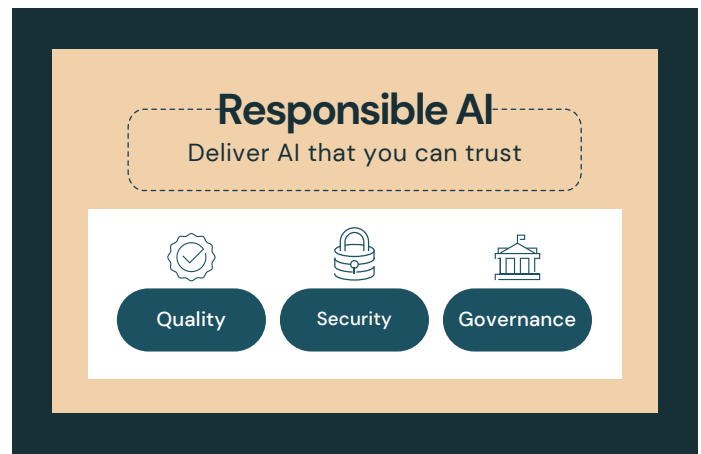
Unified security

Safely deploy AI systems with strong security controls, including encryption, network controls, data governance and auditing. These measures ensure the protection of your entire AI lifecycle, covering data operations, model operations and model serving.



Unified governance

Discover, automatically classify and secure all your data and AI assets, including models and features, with one centralized platform. Employ a user-friendly interface for access control and audit management with automated data lineage to trace the end-to-end flow of data. Securely share AI models and collaborate across business units, data platforms and clouds.



To learn more, visit:
databricks.com/trust/responsibleAI