databricks

Whitepaper

Databricks AI Governance Framework

Version 1.0



databricks

Table of Contents

Foreword 3

Introduction 5

1 Summary of the framework pillars 5
2 The Databricks AI Governance Framework pillars 6
3 Summary 6

Pillar I: AI Organizations 7

1 Business alignment 8
2 Governance model 10
3 Governance oversight 13
4 Guiding values 15
5 Strategy 16
6 Roles and responsibilities 18
7 Policies 20
8 Standards 22
9 Processes and procedures 25
10 Risk Management 29
i. Identify AI risks 29
ii. Assess, measure and prioritize AI risks 32
iii. Mitigate AI risks 35
iv. Continuous monitoring and evaluation 39
v. Internal assessment or audit 41
vi. AI incident management 42
11 Key performance indicators and monitoring 44
12 Reporting 46

Pillar II: Legal and Regulatory Compliance 48

1 Legal and regulatory compliance 49
i. Assess: Legal and regulatory considerations 49
ii. Prioritize: Liability and risk management 51
iii. Plan: Comprehensive legal planning for AI 52
iv. Deploy: Legal protections and safeguards 55
v. Monitor: Ongoing compliance and audits 56
vi. Prepare: Review emerging trends in AI ethics and regulation 57

Pillar III: Ethics, Transparency and Interpretability 58

1 AI Ethics 59
i. Accountability 59
ii. Fairness and nondiscrimination 62
iii. Human centricity and well-being 65
iv. Inclusivity 67
v. Cultural norms and sensitivity 68
2 Transparency 70
i. Transparency in AI development and design 71
ii. Transparency in AI operations 71
iii. Transparency in AI serving 71
iv. Challenges and trade-offs in AI transparency 72

Foreword

Introduction

Pillar I: AI Organizations

Pillar II: Legal and Regulatory Compliance

Pillar III: Ethics, Transparency and Interpretability

Pillar IV: Data, AIOps and Infrastructure

Pillar V: AI Security

References and Further Reading

Acknowledgements

Appendix A: Glossary

License

Pillar IV: Data, AIOps and Infrastructure 73

1 Data 75
- i. Data systems 76
- ii. Data classification standards 76
- iii. Data handling standards 78

2 AIOps 79
- i. AIOps and infrastructure 81

Pillar V: AI Security 92

1 Raw data 93
2 Data preparation 93
3 Datasets 93
4 Data catalog governance 94
5 Machine learning algorithms 94
6 Evaluation 94
7 Machine learning models 94
8 Model management 94
9 Model serving and inference requests 95
10 Model serving and inference response 95
11 Machine learning operations 95
12 Data and AI platform security 95

References and Further Reading 96

Acknowledgements 98

Appendix A: Glossary 99

License 112

AUTHORS



David Wells
Sr. Specialist Solutions Architect
databricks



Abhi Arikapudi
Sr. Director, Security Engineering
databricks

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword

As AI continues to transform and redefine the future of industries, the need for responsible and ethical AI adoption has never been more imperative. At Databricks, we recognize AI's immense potential to drive innovation and efficiency and the risks it poses if not governed properly. With the rapid pace of AI development, enterprises must navigate complex ethical, legal and societal challenges accompanying these powerful technologies. We are committed to developing a comprehensive AI governance framework to help organizations adopt AI responsibly and successfully.

The foundation of this framework is built on the principles of transparency, accountability, fairness and security. These elements are not just regulatory necessities but essential components for building trust in AI systems — trust that is critical for ensuring widespread acceptance and long-term success. As AI becomes more integrated into decision-making processes across industries, such as healthcare, finance and transportation, these systems must operate ethically, align with societal norms and respect individual rights.

As enterprises scale their AI initiatives, they face increasing challenges related to data governance, model transparency and regulatory compliance. Our framework addresses these challenges by providing structured guidelines for continuous monitoring, auditing and oversight throughout the entire lifecycle of AI models. By integrating governance into every stage of AI development — from data collection to model deployment — organizations can safeguard against risks while fostering innovation in a secure environment.

The Databricks AI Governance Framework (DAGF) will empower enterprises to harness the full potential of AI while upholding the highest standards of transparency, accountability and fairness.



Naveen Rao
VP, AI, Databricks
databricks

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Introduction

Generative AI (GenAI) and machine learning (ML) are fundamentally transforming business models and driving innovation across emerging and established industries. As organizations increasingly adopt AI tools to enhance their operations, the complexity of managing these initiatives requires a structured governance approach to align with business objectives, adhere to ethical standards and comply with evolving regulatory requirements. The Databricks AI Governance Framework (DAGF) offers organizations a comprehensive guide to effectively manage AI program development, integration and ongoing operations.

1 Summary of the framework pillars

This framework outlines five foundational components for building a responsible and resilient AI program, offering practical insights for decision-making and execution. It addresses core areas, such as AI governance, ethical compliance, risk management and operational oversight. This helps manage AI programs transparently, securely and effectively while fostering collaboration across people, processes and technology. Finally, the DAGF is a companion document to the Databricks AI Security Framework (DASF), which helps organizations secure their AI platforms.



[1]Gartner, Security Leader's Guide to Data Security, Andrew Bales. September 7, 2023.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

2 The Databricks AI Governance Framework pillars

- **Pillar I: AI organization**
  The AI organization pillar embeds AI governance within broader organizational structures. It underscores the foundation for an effective AI program through best practices like clearly defined business objectives and integrating the appropriate governance practices that oversee the organization's people, processes, technology and data. It explains how organizations can establish the oversight required to achieve their strategic goals while reducing risk.
- **Pillar II: Legal and regulatory compliance**
  The legal and regulatory compliance pillar helps organizations align AI initiatives with applicable laws and regulations. It guides managing legal risks, interpreting sector-specific requirements and adapting compliance strategies in response to evolving regulatory landscapes. This pillar ensures that AI programs are developed and deployed within a robust legal and regulatory framework.
- **Pillar III: Ethics, transparency and interpretability**
  The ethics, transparency and interpretability pillar supports organizations in building trustworthy and responsible AI systems. It emphasizes adherence to ethical principles, such as fairness, accountability and human oversight, while promoting explainability and stakeholder engagement. This pillar provides methodologies to ensure AI decisions are interpretable and aligned with evolving ethical standards, fostering long-term trust and societal acceptance.
- **Pillar IV: AI operations, data and infrastructure**
  The AI operations (AIOps), data and infrastructure pillar defines the foundation that supports organizations in fully deploying and maintaining AI. It provides guidelines for creating a scalable and reliable AI infrastructure, managing the ML lifecycle and ensuring data quality, security and compliance. This pillar also emphasizes best practices for AIOps, including model training, evaluation, deployment and monitoring, so AI systems are reliable, efficient and aligned with business goals.
- **Pillar V: AI security**
  The AI security pillar introduces the DASF, a comprehensive framework for understanding and mitigating security risks across the AI lifecycle. It covers critical areas, such as data protection, model management, secure model serving and the implementation of robust cybersecurity measures to protect AI assets.

3 Summary

The DAGF offers a holistic approach to managing AI programs, ensuring alignment with organizational goals, ethical integrity and regulatory compliance. By adopting this framework, organizations can effectively navigate AI's complexities while fostering a culture of transparency, accountability and continuous improvement. This approach empowers organizations to fully harness AI's potential, driving innovation, enhancing decision-making and achieving long-term strategic success in an increasingly AI-operated world.

AI Organizations

Organizations that embed AI governance within their overarching strategy and operational frameworks are best positioned to leverage AI's potential. This chapter outlines the elements for establishing a comprehensive AI organization by aligning AI initiatives with strategic objectives and integrating ethical, regulatory and operational standards across all levels.

Key areas of focus include:

- **Setting clear objectives**
- **Fostering cross-functional collaboration**
- **Managing risk**
- **Establishing robust monitoring and reporting systems**

This integrated approach empowers organizations to navigate AI's complexities, build stakeholder trust and support sustainable, responsible growth. All while aligning with strategic goals and operating transparently, ethically and in compliance with regulatory standards.

Section summary

Business requirements: Engage stakeholders to define precise requirements through needs assessment, goal setting, documentation and validation, ensuring AI projects align with enterprise objectives and deliver value.

Governance model: Select a governance model, such as centralized, distributed or hybrid, based on the organization's needs to manage AI practices responsibly, ethically and effectively.

Governance oversight: Establish dedicated oversight bodies, such as AI ethics committees and steering councils, so ethical, transparent and compliant AI practices align with organizational goals.

Resources: Identify the roles, responsibilities and skills required for the AI program and strategically integrate them into various parts of the organization.

Principles, policies and standards: Develop guiding principles, policies and standards for ethical AI use, ensuring fairness, transparency and accountability.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

1 Business alignment

Organizations that effectively align AI initiatives with strategic priorities establish a foundation for delivering measurable business value. This alignment balances innovation with acceptable business risk, promotes disciplined investment and supports scalable, long-term adoption of AI. Alignment is not a one-time activity — it requires ongoing coordination between business and technical teams, grounded in shared objectives and adaptive governance.

AI program alignment can be supported by establishing clear expectations, including:

- **Strategic contribution:** AI initiatives should directly support the organization's mission and strategic goals, embedding AI within the business's core value drivers.
- **Prioritized investment**: Initiatives should be evaluated and sequenced based on expected business impact, feasibility and alignment with broader transformation or operational goals.
- **Risk-aware framing:** Programs should reflect the organization's operational and strategic risk tolerance, enabling innovation within responsible boundaries.
- **Lifecycle ownership:** Business leaders should remain accountable for AI use cases throughout their lifecycle, from planning through deployment, monitoring and iteration.
- **Performance integration:** AI outcomes should be measured using business-relevant key performance indicators (KPIs) to ensure they contribute to enterprise performance goals, not just technical success.
- **Adaptive alignment:** AI priorities should be regularly reviewed and updated as business conditions, customer needs or market environments evolve.

Organizations should align AI programs with broader business objectives by creating a shared understanding of enterprise goals and AI's potential role across various business functions. This alignment helps to ensure that initiatives remain relevant, actionable and connected to the organization's overall outcomes.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Getting started

The process for establishing business requirements can include:

1. **Stakeholder engagement:** Engage with key organizational stakeholders to gather insights and understand business challenges and opportunities.
2. **Needs assessment:** Conduct a thorough needs assessment to identify critical business problems that AI can address. This includes analyzing current processes, pain points and performance metrics.
3. **Consider AI-specific challenges:** Design business requirements with an understanding of the unique constraints related to AI programs.

Examples include:

   a. **Cost:** AI programs can be expensive. The fiduciary benefits should justify the program's cost.
   b. **Risk:** AI programs introduce new risks that must be understood and addressed to meet the organization's risk tolerance.
   c. **Uncertainty and experimental nature:** AI initiatives require iterative research, design and trial to match business expectations.
   d. **Data dependency:** AI initiatives require high-quality data relevant to the business goals.
   e. **Retraining:** As businesses and circumstances evolve, models must evolve with them. Business requirements should define accuracy/predictability requirements to signal when retraining is required.
   f. **Process impact:** Net-new AI initiatives can change business processes. Business objectives should include the desired outcomes with updated processes.

4. **Set goals:** Define specific, measurable, achievable, relevant and time-bound (SMART) goals for AI initiatives that align with the organization's strategic objectives.
5. **Readiness:** Assess the organization's readiness for AI adoption by evaluating support, culture, headcount, skills, processes, infrastructure and alignment. Identify readiness gaps and enablement requirements. Ensure there's executive support and a culture conducive to innovation.
6. **Documentation:** Document the business requirements in detail, specifying the expected outcomes, performance metrics, risk treatments and success criteria.
7. **Validation and approval:** Validate the documented requirements with stakeholders and obtain formal approval to ensure alignment and commitment.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

2 Governance model

An AI governance model establishes a strategic framework to guide how organizations align AI programs and initiatives with their objectives and integrate them into enterprise-wide strategies and operations. It defines high-level principles that shape oversight, accountability and adaptability, supporting ethical standards, regulatory compliance and strategic goals.

Key challenges addressed by governance models include:

- **Accountability:** Defining clear frameworks for decision-making and responsibility, ensuring ethical and responsible outcomes from AI-driven processes.
- **Operational alignment:** Structuring AI programs to operate cohesively within the organization as independent entities or as integrated components of larger systems.
- **Adaptability:** Ensuring governance structures can evolve alongside advancements in AI technologies and shifts in organizational priorities.
- **Risk and compliance:** Establishing high-level guidance for integrating risk management and adherence to regulatory standards into AI-related decisions.

To meet these challenges, governance models will outline guiding values that inform the direction and decisions of the AI program.

Organizations establish a foundation for scalable, ethical and effective oversight of AI initiatives by articulating a governance model tailored to AI's complexities. This framework fosters alignment with organizational values and priorities, enabling adaptability to future technological advancements and evolving regulatory landscapes.

Getting started

Considerations when selecting and implementing a governance model can include:

1. **Clarify governance objectives:** Define the primary purpose of the AI governance model within the organization. Identify what the model aims to achieve regarding oversight, consistency and alignment with AI objectives.
2. **Assess organizational structure:** Examine existing organizational structures to understand how an AI governance model can integrate effectively. Consider how AI governance intersects current business units, operations and leadership hierarchies.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

3. **Choose a governance structure:** Based on the organization's operational needs, AI maturity and overall business strategy, determine which governance structure meets the organization's requirements:
    a. **Centralized structure:** A single governing body oversees AI policies and standards, ensuring consistency and compliance across the organization. This approach works well for regulated industries or organizations that prioritize uniformity in decision-making. However, it may slow decision-making and reduce flexibility for teams with specialized needs.
    b. **Distributed structure:** Departments or teams independently govern their AI initiatives, tailoring governance to their specific operational needs. This structure encourages flexibility and innovation but requires robust coordination to prevent fragmented standards and risk management inconsistencies.
    c. **Hybrid structure:** Hybrid structures combine centralized oversight with localized governance, balancing consistency and flexibility. They are particularly effective for organizations managing diverse AI programs or navigating complex transformations. However, hybrid structures require clear communication and efficient coordination to avoid duplicating efforts or accountability gaps.
4. **Outline key roles and authorities:** Identify essential roles within the governance model, including decision-making authorities and oversight responsibilities specific to AI initiatives. Define who will be accountable for strategic guidance, operational management and compliance within the AI ecosystem.
5. **Draft initial governance guidelines:** Develop foundational guidelines that will direct the organization's approach to AI governance. This includes setting initial parameters for decision-making authority, operational scope and role definitions.
6. **Identify stakeholder input needs:** Identify and engage initial stakeholders who will contribute to shaping the governance model. This can include executives, technical leaders, compliance officers and departmental heads who have a direct or indirect role in AI projects.
7. **Establish foundational reporting channels:** Set up initial reporting structures for the AI governance model to ensure early alignment and visibility. This includes identifying core metrics and information that must be communicated regularly within the governance framework.
8. **Develop a roadmap for model expansion:** Plan a high-level roadmap for maturing the governance model, starting with foundational elements and gradually expanding to address the full scope of AI operations as the organization's AI capabilities grow.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

How Databricks can help

- **Unified governance:** Databricks Unity Catalog offers a unified governance layer for data and AI within the Databricks Data Intelligence Platform. With Unity Catalog, organizations can seamlessly govern their structured and unstructured data, as well as ML models, notebooks, dashboards and files on any cloud or platform. Data scientists, analysts and engineers can use Unity Catalog to securely discover, access and collaborate on trusted data and AI assets, leveraging AI to boost productivity and unlock the full potential of lakehouse architecture. This unified approach to governance accelerates data and AI initiatives while simplifying regulatory compliance (Databricks).
- **Distributed use:** Databricks workspaces enable individual lines of business and teams to independently develop AI programs. Unity Catalog supports this distributed model by allowing teams to create and manage catalogs and schemas within a centralized governance framework. This ensures that departments have the autonomy to innovate and remain aligned with the organization's broader governance policies (Databricks documentation).

3 Governance oversight

Effective governance oversight establishes clear structures and accountability for AI programs, promoting ethical, transparent and aligned operations. Organizations can monitor AI initiatives, support structured decision-making and provide clear pathways for issue resolution through dedicated bodies like AI governance boards and ethics committees. This oversight framework helps manage risks, uphold ethical standards and comply with regulatory requirements.

Organizations should enact effective oversight mechanisms that support the efficient and trusted execution of their AI programs. Governance oversight can be strengthened by regularly updating these bodies to provide consistent guidance and strategic alignment across AI initiatives. By remaining responsive to regulatory changes, emerging risks and evolving organizational needs, these oversight structures support the governance model, ensuring that it is enforced and continuously adapted to meet the demands of responsible AI development.

Getting started

Considerations when selecting and implementing governance oversight can include:

1. **AI governance board:** A board of senior executives (e.g., chief AI/ML officer (CAIO), legal, compliance, security, business units) providing strategic oversight to align AI initiatives with organizational goals, risk tolerance and ethical standards. This board ensures high-level accountability, stakeholder trust and consistency in AI practices.
2. **AI ethics committee:** A group of AI ethicists, data scientists, legal experts and stakeholders that develop ethical principles, interpret those principles into actionable language and review new or developed systems for ethical adherence, transparency and accountability.
3. **AI center of excellence (CoE):** A cross-functional team offering expertise, support and best practices to drive innovation, ensure consistent AI practices and adapt governance policies. The CoE fosters collaboration, standardization and compliance.
4. **Department AI Leads:** Designated AI leads in each department ensure adherence to governance frameworks, align departmental initiatives with central policies and report to the CAIO or governance board to support effective risk management and coordination.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

How Databricks can help

- **Unified governance and compliance automation with Databricks:** The Databricks Data Intelligence Platform enables organizations to enforce unified governance and security policies across AI and business intelligence (BI) initiatives. This helps with consistently applying data handling practices, access controls and ethical guidelines, while automation reduces manual effort, minimizes errors and promotes robust adherence to governance standards (Databricks documentation).
- **Automated governance reviews and continuous integration and continuous delivery (CI/CD) best practices:** Databricks facilitates automated governance reviews to verify compliance with ethical standards and regulatory requirements by integrating seamlessly with data and AI lifecycle stages. Implementing automated checks, such as compliance or best practice verification, into your CI/CD process enhances the automated governance review process, streamlining development and deployment workflows (Databricks documentation, Databricks documentation).
- **Integration with BI tools for governance dashboards:** Databricks integrates with BI tools to create governance-specific dashboards, offering interactive visualizations and analytics. These dashboards enable real-time monitoring of key ethical adherence, risk management and regulatory compliance metrics, supporting informed decision-making and strategic oversight (Databricks documentation).
- **Real-time compliance monitoring and reporting:** Organizations can leverage Databricks real-time analytics and reporting capabilities to automate compliance report generation, track adherence to governance policies and gain visibility into deviations. Automated alerts notify governance boards and ethics committees of potential issues, enabling timely interventions (Databricks, Databricks).

4 Guiding values

Guiding values set an AI program's overarching direction and boundaries, influencing its ethical and operational approach to developing, deploying and managing AI systems. Integrating the AI program's guiding values into corporate structures can help align AI practices with broader ethical and operational standards.

Organizations should define and embed these guiding values within their core processes, creating adaptable guidelines that reflect their priorities and comply with regulatory requirements. Regularly reviewing these values in response to technological and societal changes will help maintain alignment with ethical standards, foster public and stakeholder trust and support resilient and accountable AI operations.

Getting started

Establishing guiding values for an AI program requires a deliberate and strategic approach. Organizations can consider the following steps when creating their guiding values:

1. **Articulate strategic intent:** Define the purpose of your AI guiding values in the context of your organization's goals. This ensures that guiding values are a strategic tool that supports business objectives, builds trust and manages risk.
2. **Champion cross-organizational alignment:** Facilitate collaboration among key stakeholders — including technical, legal, ethics and business teams — to ensure AI guiding values are practical and reflect the diverse priorities within the organization.
3. **Prioritize high-impact areas:** Focus on areas where AI adoption poses the most significant risks or opportunities. This includes fairness, transparency and accountability in decision-making or safeguarding against reputational and regulatory risk.
4. **Evaluate global trends and industry standards:** Stay ahead of evolving regulations and best practices by benchmarking against global frameworks and competitors. This strategic awareness positions your organization as a leader in AI governance.
5. **Integrate accountability mechanisms:** Ensure the guiding values are actionable by embedding them into governance structures, assigning clear ownership and establishing mechanisms for monitoring adherence and impact.
6. **Communicate and inspire action:** Lead the rollout of AI guiding values with clear, compelling messaging, emphasizing their importance to the organization's mission. Ensure ongoing visibility through executive sponsorship and regular reviews.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

5 Strategy

An effective AI strategy defines a structured path to adopt AI within an organization by directing initiatives to support short, medium and long-term business goals while remaining adaptable to change. By aligning AI objectives with broader business objectives, a well-designed strategy coordinates resources, supports cross-functional collaboration, scales AI solutions to the organization's AI appetite and prepares the organization to leverage advancements in AI and technology.

Organizations should implement an AI strategy that integrates AI goals with broader business objectives, maintaining flexibility to adapt to emerging insights or technologies. Regularly reviewing and updating the strategy can help organizations remain responsive to change, supporting a resilient and forward-looking approach to AI governance that maximizes AI's contribution to organizational success.

Getting started

Examples of AI program strategy artifacts include:

1. **Long-term vision:** A strategic outlook aligning AI initiatives with the organization's goals to support business objectives and drive sustainable growth.
2. **Roadmap**: A detailed plan specifying:
   a. Short, medium and long-term goals
   b. Steps, timelines and resources for AI adoption
   c. Critical projects and priorities
3. **Milestones and deliverables:** Defined checkpoints and outputs to track progress and ensure accountability, using KPIs to evaluate success and make necessary adjustments.
4. **Scalability guidelines and requirements:** Documented standards and specifications for building a modular, interoperable infrastructure that supports the expansion of AI capabilities as the organization grows.
5. **Adaptability plans:** Documented mechanisms and resources to ensure the AI strategy remains flexible and responsive to evolving business needs and technology trends.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

6. **Stakeholder engagement plans:** A structured approach involving organizational stakeholders to build support, ensure alignment and encourage collaboration.
7. **Risk management framework:** A documented approach to proactively identifying and managing ethical, legal and operational risks associated with AI, including clear mitigation strategies.
8. **Continuous monitoring and evaluation processes:** Defined processes and tools for regular review, feedback and data-driven adjustments to optimize AI initiatives.

How Databricks can help

- **Thought leadership:** Databricks whitepapers, like the Big Book of Generative AI and the DASF, provide thought leadership to organizations considering deploying AI programs.
- **Scalability and adaptability:** The Databricks Platform's scalable and flexible architecture allows organizations to execute their AI strategies to match evolving business needs and technological advancements (Databricks).

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

6 Roles and responsibilities

AI programs require thoughtfully structured roles and responsibilities to meet immediate operational needs and future organizational priorities. By aligning roles with strategic goals, fostering interdisciplinary collaboration and embedding accountability, organizations can establish a foundation for effective and adaptive AI governance. This section outlines principles to guide defining roles that enable AI programs to deliver meaningful and sustainable outcomes. Guiding principles for structuring roles for an AI program include:

- **Strategic alignment:** Roles must be designed to directly support the AI program's objectives while aligning with the organization's goals. This involves tailoring responsibilities to address specific outcomes, such as operational efficiency, innovation or ethical compliance, and ensuring roles can evolve as AI capabilities advance.
- **Unique skill and competency requirements:** AI programs demand specialized technical expertise alongside ethics, business strategy and risk management competencies. Effective role design emphasizes integrating these skills while maintaining pathways for continuous upskilling to adapt to emerging technologies and standards.
- **Facilitating interdisciplinary collaboration:** AI initiatives often require contributions from diverse functions, such as data science, compliance, IT and business operations. Roles should be structured to enable seamless communication and shared accountability across these domains, ensuring the program benefits from a holistic perspective.
- **Governance and accountability:** Clearly defined responsibilities and decision-making authority are essential for effective oversight and operational efficiency. Roles must include mechanisms for accountability that support both innovation and adherence to governance standards, balancing flexibility with structure.
- **Embedding ethical and regulatory oversight:** AI programs operate within complex ethical and regulatory landscapes. Role definitions should explicitly include responsibilities for identifying and addressing risks related to bias, fairness, privacy and security, ensuring compliance and public trust.
- **Flexibility and scalability:** AI programs evolve rapidly, requiring roles that are adaptable to new challenges and scalable to support growth in complexity and scale. Role structures should anticipate and accommodate these changes to maintain program effectiveness.

By integrating these elements, organizations can bolster their AI capabilities, adapt effectively to emerging challenges and maintain long-term resilience in an evolving digital landscape.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Getting started

AI program role and responsibility considerations include:

1. **Identify key roles and skill requirements:** Start by defining roles that address the unique demands of AI programs, such as managing complex data systems or ensuring ethical compliance. Align these roles with strategic organizational goals and identify specialized skills needed in technical expertise, ethical oversight and strategic planning. For instance, a CAIO might lead the overall strategy, data engineers manage pipelines and AI ethicists address risks like bias or transparency.
2. **Create clear role definitions:** Develop descriptions that articulate specific responsibilities, objectives and decision-making authority. Highlight each role's contribution to AI initiatives and organizational goals, ensuring clarity and reducing overlap. Regularly reviewing these definitions will keep them aligned with evolving technologies and organizational needs.
3. **Establish training programs:** Establish training initiatives tailored to building technical and strategic competencies, focusing on areas like AI ethics and compliance. Ongoing education should anticipate technological changes and evolving regulations, helping teams remain effective as the program matures.
4. **Implement accountability mechanisms:** Introduce evaluation systems to measure role performance and program impact. Metrics and feedback loops can track how effectively roles meet AI objectives, fostering a culture of transparency and continuous improvement. Structured reviews can also identify opportunities to optimize cross-functional efforts.
5. **Foster cross-functional collaboration:** Establish structured communication channels and assign bridging roles to encourage collaboration across departments. Liaisons can connect technical teams with compliance officers or business strategists, helping to ensure cohesive decisions integrate diverse perspectives.
6. **Plan for scalability and adaptability:** Anticipate future demands by designing roles that can scale with the growth of AI initiatives. Incorporate flexibility into role structures to adapt to shifts in organizational priorities or external challenges, such as new regulations or technological disruptions.

Foreword
Introduction
AI Organizations
Legal and Regulatory Compliance
Ethics, Transparency and Interpretability
Data, AIOps and Infrastructure
AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

How Databricks can help

- **Education and training:** Databricks Academy is the Databricks educational ecosystem. It includes training, certification and continuous learning opportunities tailored to each role involved in an AI program. Databricks training programs equip teams with the latest knowledge and skills to drive successful AI initiatives (Databricks, Databricks).

7 Policies

AI policies provide high-level, formalized guidelines that set expectations for the responsible management and execution of AI programs. They address areas such as ethical considerations, data governance and compliance. By establishing overarching frameworks, policies ensure that AI initiatives align with organizational values and regulatory obligations, fostering accountability and consistency.

Organizations should develop effective AI policies that create adaptable rules reflecting internal standards and evolving external requirements. Regularly reviewing and updating these policies allows organizations to remain aligned with new ethical and regulatory insights, promoting stakeholder confidence and responsible AI use.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Getting started

Examples of AI program policies include:

1. **Acceptable AI use policy:** This policy establishes expectations for the organization's ethical, responsible and compliant use of AI. It ensures that AI applications align with organizational values and support beneficial, productive purposes while adhering to ethical and regulatory standards. This policy is commonly integrated with data policies, such as the acceptable data use policy.
2. **AI model governance and accountability policy:** This policy defines principles for responsible oversight of AI models throughout their lifecycle, ensuring models are designed, monitored and managed in alignment with organizational ethics, quality and transparency standards.
3. **Data governance and ethics policy for AI:** This policy establishes guiding principles for ethical data collection, use and management specific to AI applications. It ensures data practices meet privacy, security and quality requirements, supporting ethical and reliable AI outcomes aligned with organizational values.
4. **AI risk and impact management policy:** This policy provides a framework for identifying, assessing and addressing risks unique to AI, including ethical, operational and technical considerations. It supports the organization's commitment to safeguarding against potential AI risk and ensuring responsible oversight throughout the AI lifecycle.
5. **AI ethics, fairness and transparency policy:** This policy establishes standards for ethical AI practices and ensures fairness, transparency and accountability in all AI applications. It reflects the organization's commitment to ethical principles and supports equitable, understandable AI outcomes.
6. **Third-party AI vendor management policy:** This policy defines principles for engaging and managing third-party AI providers to ensure alignment with organizational ethics, security and compliance. It supports responsible collaboration with external partners while protecting organizational integrity and sensitive information.

How Databricks can help

- **Development and enforcement:** Databricks enables the creation and enforcement of comprehensive AI policies with features for data governance, model lifecycle management and compliance monitoring. This ensures adherence to best practices and regulatory requirements.
- **Regular review:** The platform's insights and monitoring tools support regular policy reviews and updates, helping organizations stay current with new challenges and regulatory changes (Databricks).

8 Standards

Technical and operational standards provide measurable benchmarks for AI model development, data management and operational procedures, ensuring consistency, quality and reliability across AI initiatives. These standards are foundational criteria aligning with industry best practices and regulatory requirements, supporting efficiency and compliance within the AI program.

Organizations should implement practical standards that focus on adopting and adapting industry best practices that reflect their strategic objectives and operational needs. Regularly reviewing and refining standards helps to ensure they remain relevant as industry practices and regulatory landscapes evolve, promoting robust and dependable AI systems.

Foreword
Introduction
AI Organizations
Legal and Regulatory Compliance
Ethics, Transparency and Interpretability
Data, AIOps and Infrastructure
AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Getting started

Implementing AI standards helps organizations ensure that their AI initiatives are practical, ethical, transparent and compliant with relevant regulations. AI program standards may include:

1. **Data preprocessing standard:** This standard defines specific requirements for data quality techniques, such as normalization, data cleaning and transformation, to ensure consistency across AI models. It supports interoperability and minimizes biases, establishing reliable and fair AI outcomes.
2. **Model performance testing standard:** Establishes minimum performance benchmarks to ensure AI models meet accuracy and reliability standards. Traditional ML models use metrics such as precision, recall and F1 score as requirements. For natural language processing (NLP) or GenAI models, include additional metrics (e.g., perplexity, bilingual evaluation understudy (BLEU) and BERTScore), reference internal evaluation tools (such as an Eval Gauntlet) and conduct red-teaming exercises to stress-test model responses.
3. **Security measures standard:** This standard specifies mandatory security protocols to safeguard AI systems and data from unauthorized access, breaches and attacks. Protections against threats like data poisoning, model extraction and privacy attacks ensure compliance with AI-specific security needs and relevant regulations.
4. **Bias mitigation and fairness standard:** This standard defines criteria for detecting, auditing and mitigating biases in AI models using fairness-enhancing algorithms. It includes requirements for routine bias audits and fairness checks to promote equitable AI outcomes and reinforce ethical standards.
5. **Privacy protection and data retention standard:** This standard mandates privacy-preserving techniques, such as data anonymization, encryption and access controls, to protect sensitive information in AI applications. It also includes AI-specific data retention guidelines, ensuring compliance with privacy regulations and secure data handling throughout the AI lifecycle.
6. **Explainability and interpretability standard:** Establishes requirements for making AI models explainable and interpretable, especially in high-impact applications. This standard ensures stakeholders understand and trust AI outputs, promoting transparency and accountability.

7. **Model documentation and traceability standard:** This standard sets requirements for detailed documentation of AI models, including assumptions, design choices, data sources and updates throughout the model lifecycle. It supports transparency, traceability and accountability, enabling thorough reviews and reproducibility.
8. **Data provenance and quality standard:** This standard defines criteria for verifying the sources, quality and diversity of data used in AI models. This standard ensures that data used for training and inference is accurate, relevant and ethically sourced to support reliable and fair AI outcomes.

How Databricks can help

- **Technical and operational standards:** Databricks helps define and enforce AI development and deployment standards, ensuring consistency with industry best practices and regulatory requirements. The platform's data and model governance capabilities facilitate adherence to these standards (Databricks).
- **Compliance and audits:** Robust Databricks governance features streamline regular audits and compliance checks, ensuring continuous improvement and adherence to established standards (Databricks).

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

9 Processes and procedures

Processes and procedures provide instructions for executing complex tasks, breaking them down into actionable outcomes or steps that are easy to understand and follow. In an AI program, processes and procedures guide critical tasks, such as model development, data quality management and system monitoring, supporting consistency, accuracy and accountability. Well-defined procedures enable team members to reliably carry out AI-related tasks in alignment with organizational objectives and standards.

Organizations should design and implement adequate procedures tailored to their AI operations by:

- Breaking down complex tasks into actionable steps that align with organizational standards, policies and ethical principles.
- Regularly reviewing and updating procedures to ensure they remain relevant to evolving tools, technologies and regulatory requirements.
- Ensuring procedures are clearly documented, easily accessible and effectively communicated to all relevant stakeholders.
- Monitoring adherence to procedures to drive efficiency, maintain compliance and support continuous improvement in AI operations.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Getting started

The following section provides a starting point for understanding the processes and procedures that can guide AI program management. By defining workflows and repeatable steps, organizations can explore a flexible framework to support governance tailored to their specific needs. Organizations should assess their AI program requirements and priorities to determine the relevant and impactful components.

1. **Model development and validation process:** Establish practices for developing AI models that align with organizational goals, prioritize fairness and address risks, such as bias or underperformance, fostering trust in AI outcomes.
    a. **Feasibility assessment:** Assess the feasibility of proposed AI models by identifying objectives, evaluating data readiness and recognizing potential risks to support decision-making.
    b. **Data preparation:** Develop processes for preparing datasets, including cleaning, formatting and validating, to support fairness and quality in AI development.
    c. **Model design:** Outline model architectures that align initiative objectives with program goals and risk obligations.
    d. **Testing and validation:** Apply systematic testing and auditing methods to evaluate model performance, fairness and compliance with established standards.
2. **Data management process:** Develop data practices that promote quality, compliance and traceability, enhancing the reliability of AI systems and alignment with organizational values and regulations.
    a. **Data sourcing:** Define criteria and practices for acquiring datasets that meet quality, ethical and compliance standards.
    b. **Data cleaning and validation:** Use techniques to address errors, gaps and inconsistencies in datasets, supporting data integrity and reliability.
    c. **Version control:** Design systems to track document data changes, facilitating reproducibility and traceability.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

3. **Deployment and monitoring process:** Provide pathways for the safe deployment of AI models, enabling monitoring to adapt to changes and address risks, such as model drift or performance issues.
    a. **Predeployment testing:** Conduct evaluations under simulated conditions to identify risks and optimize performance.
    b. **Incremental rollout:** Introduce models incrementally to monitor performance and gather feedback for iterative improvements.
    c. **Drift detection:** Establish monitoring processes to identify and address changes in data patterns or model performance over time.
4. **Compliance and risk management process:** Integrate compliance and risk management into AI operations to align with legal and ethical standards, fostering trust while minimizing liabilities.
    a. **Bias audits:** Periodically evaluate models against fairness criteria to identify and address bias or inequity risks.
    b. **Legal reviews:** Conduct expert reviews to verify alignment with applicable legal and regulatory frameworks.
    c. **Incident logging:** Build systems to document, track and resolve incidents, supporting accountability and learning.
5. **Security and privacy process:** Safeguard sensitive data and system integrity through security and privacy measures, addressing risks, such as unauthorized access and data breaches.
    a. **Access controls:** Develop mechanisms to manage and restrict access to sensitive data and systems based on roles and responsibilities.
    b. **Data encryption:** Encourage adequate encryption protocols to protect sensitive data during storage and transmission.
    c. **Breach response:** Create and validate response strategies to address data breaches and minimize potential impacts.
6. **Transparency and communication process:** Foster transparency and accountability in AI decision-making by educating stakeholders and maintaining clear documentation of critical processes.
    a. **User explainability tools:** Design tools and interfaces to enhance user understanding of AI-driven decisions, supporting transparency.
    b. **Stakeholder education:** Create initiatives to improve understanding of AI risks, opportunities and impacts.
    c. **Documentation of model decisions:** Maintain comprehensive records of decision-making processes to enable accountability and support audits.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

How Databricks can help

Built-in Databricks capabilities support consistent, repeatable, well-governed AI processes that are aligned with organizational standards and compliance requirements.

- **Unity Catalog for data governance:** Unity Catalog provides centralized, enterprise-grade governance with robust metadata management, data lineage tracking and fine-grained access controls. These capabilities enable secure, compliant and transparent data usage across the organization, aligning AI processes with regulatory and ethical standards (Databricks documentation).
- **MLflow for model lifecycle management:** Databricks integrates MLflow to manage the ML lifecycle, from experiment tracking to model versioning and deployment. MLflow ensures standardized, repeatable and auditable processes that improve consistency, reduce operational risks and support compliance across enterprise teams (Databricks documentation).
- **Delta Lake for data management:** Delta Lake enhances data reliability through atomicity, consistency, isolation and durability (ACID) transactions, schema enforcement and scalable metadata handling within Databricks. This ensures enterprise-wide data consistency and accuracy, which is essential for trusted AI model training, validation and decision-making processes (Databricks documentation).
- **Databricks Workflows for automation:** Databricks Workflows enable organizations to automate and orchestrate complex data processing and ML pipelines. Automating these processes ensures consistency, reduces manual errors and supports repeatable execution aligned with enterprise procedures and governance standards (Databricks documentation).

10 Risk management

AI program risk management involves systematically identifying, assessing and mitigating potential risks to support organizational integrity, strategic alignment and legal and regulatory compliance. Integrating risk management across the AI lifecycle — from development through deployment to ongoing operation — enables a structured approach that considers the organization's risk tolerance and adapts to changing conditions. Organizations should consider the following when establishing risk management for their AI programs:

i Identify AI risks

Identifying AI risks involves systematically recognizing potential challenges impacting organizational objectives, AI programs and initiatives and AI usage. These risks include direct challenges, such as model inaccuracy, bias and data security vulnerabilities, as well as broader risks, such as regulatory noncompliance, ethical concerns, reputational harm and the unintended consequences of AI deployment. Addressing these risks requires understanding immediate threats, such as adversarial attacks, and longer-term challenges, including evolving regulatory landscapes and technological advancements.

Organizations should establish robust processes for consistent risk identification that consider the full spectrum of risks associated with AI systems, from technical performance to alignment with organizational goals and external requirements. Grouping risks into categories — such as technical, operational, regulatory and societal — can facilitate clearer prioritization and strategy development. Regularly reviewing and refining these processes helps organizations stay aware of emerging risks and maintain readiness to address future challenges, fostering a proactive stance in AI governance.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Getting started

AI program risk considerations include:

1. **Strategic risks:**
   a. Misalignment of AI initiatives with organizational goals and values.
   b. Over-reliance on AI results in the erosion of human expertise or control.
   c. Failure to adopt AI leads to a competitive disadvantage.
2. **Financial risks:**
   a. Unexpected costs or budget overruns in AI initiatives.
   b. Financial losses arising from AI-driven decisions or recommendations.
   c. Overinvestment in AI technologies without a return on investment (ROI).
3. **Reputational risks:**
   a. Damage to brand image due to perceived unethical AI use.
   b. Loss of public trust stemming from AI-related controversies or failures.
   c. Negative media coverage surrounding AI initiatives.
4. **Legal and compliance risks:**
   a. Regulatory violations related to AI use (e.g., privacy laws and industry-specific regulations).
   b. Liability for damages caused by AI system decisions or actions.
   c. Intellectual property (IP) disputes connected to AI development or deployment.
5. **Operational risks:**
   a. Business disruptions due to critical AI system failures.
   b. Increased vulnerability to cyberattacks targeting AI systems.
   c. Dependence on third-party AI vendors or technologies.
6. **Ethical risks**:
   a. Reputational harm from biased or discriminatory AI outcomes.
   b. Ethical controversies arising from AI applications (e.g., surveillance or autonomy).
   c. Conflicts with organizational values or societal norms.

Foreword
Introduction
AI Organizations
Legal and Regulatory Compliance
Ethics, Transparency and Interpretability
Data, AIOps and Infrastructure
AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

How Databricks can help

Databricks tools assist organizations in proactively identifying and monitoring potential AI-related risks across the enterprise, including regulatory, ethical and operational concerns.

- **AI lifecycle-wide audit logging:** Databricks provides comprehensive audit logging through Unity Catalog, capturing detailed records of data access and model usage across the organization. These logs help identify risks, such as unauthorized access, regulatory violations or misuse of sensitive information, thereby mitigating legal, operational and reputational risks (Databricks documentation).
- **Bias and fairness assessments:** Databricks supports tools like SHapley Additive exPlanations (SHAP) (), which help to proactively identify fairness and ethical risks in AI models. Early detection of biased outcomes can prevent reputational harm, regulatory noncompliance and ethical controversies at the enterprise level (Databricks).
- **Data governance and access controls via Unity Catalog:** Unity Catalog provides centralized governance and fine-grained access controls, helping to identify risks related to data security, compliance breaches and unauthorized access. This capability safeguards against operational disruptions, regulatory noncompliance and potential financial liabilities (Databricks documentation).

ii Assess, measure and prioritize AI risks

Assessing, measuring and prioritizing AI risks enables organizations to evaluate how AI may impact the organization. Systematic evaluation using mechanisms such as impact and likelihood, risk scoring or ranking models allows organizations to make informed decisions and prioritize resource allocation and mitigation efforts based on the risks' potential impact on the enterprise.

Organizations should incorporate AI risk evaluation into their enterprise-level risk framework to align with governance structures and strategic goals. This approach provides consistency across risk management practices while including AI-specific considerations, such as model drift, bias and ethical implications. Organizations that cannot incorporate this into their enterprise-level risk framework can create a standalone risk evaluation framework. Regularly review risk practices to verify that they remain relevant and responsive while enabling organizations to adapt to emerging challenges, aligning with their core goals.

Getting started

Strategies to assess, measure and prioritize AI risk include:

1. **Establish risk tolerance parameters:** Clearly outline acceptable thresholds for AI risks, focusing on operational, reputational, financial and compliance-related impacts. Use these parameters to guide decision-making across AI projects, ensuring risks are managed within acceptable limits while supporting business objectives. Tolerance parameters could include specifically defined circumstances or use cases where AI is *strictly and entirely unacceptable*.
2. **Establish risk identification processes:** Develop a comprehensive checklist for AI-specific risks, such as model drift, bias and security vulnerabilities. This checklist should be integrated seamlessly into enterprise risk management workflows to systematically address potential risks.
3. **Develop a risk scoring framework:** Create a scoring system that quantifies AI risks based on their impact — reputational, operational or financial — and their likelihood of occurrence. Use predefined scales, such as low, medium and high, with existing standardized departmental evaluations to facilitate a uniform understanding of risk levels.
4. **Integrate AI risks into enterprise-level risk management:** Align AI-specific risk assessments with the organization's overall governance frameworks. For those unable to fully integrate, establish a standalone AI risk management framework that considers ethical aspects, compliance with relevant regulations and alignment with corporate goals.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

5. **Categorize and rank risks:** Group risks into categories — including technical, operational, reputational and legal. Prioritize risks with high impact and likelihood to optimize resource allocation for mitigation efforts.
6. **Establish regular risk reviews:** Set a schedule for reviewing AI-related risks annually or semi-annually. This practice should include monitoring environmental changes, like new regulations or emerging threats, that may affect existing risk profiles.
7. **Use tools for monitoring and assessment:** Implement tools designed to track model performance, detect bias and measure data quality metrics. Additionally, deploy dashboards to visualize risk levels and trends, enhancing stakeholder communication.
8. **Engage cross-functional teams:** Involve stakeholders from IT, legal, compliance and operations in the risk evaluation process. Leverage diverse perspectives to support comprehensive risk identification and prioritization.
9. **Governance role:** Establish a dedicated governance role, such as a risk steward or assign a team to oversee AI risk assessment and reporting.
10. **Responsibilities and accountability:** Set clear expectations for accountability regarding risk responsibilities, such as identifying, treating and reporting risk.
11. **Record keeping:** Implement structured record-keeping practices to track consistency and transparency in risk-related decisions.
12. **Communicate risks to stakeholders:** Create a structured framework for clear and concise risk communication to executives, board members and external stakeholders. Visual aids, such as risk heatmaps, can help effectively convey prioritization and mitigation plans.
13. **Conduct risk awareness training:** Train employees to identify and escalate AI-specific risks. Emphasizing the ethical, compliance and operational implications of these risks in their daily tasks fosters a culture of awareness and responsibility within the organization.

Foreword
Introduction
AI Organizations
Legal and Regulatory Compliance
Ethics, Transparency and Interpretability
Data, AIOps and Infrastructure
AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

How Databricks can help

Databricks can systematically assess and prioritize AI risks, helping organizations align risk management with strategic objectives and resource allocation.

- **Enterprise risk dashboarding and reporting:** Databricks enables organizations to build comprehensive, enterprise-level dashboards for visualizing strategic, financial, operational and reputational risk indicators. These visualizations support executives and risk managers in effectively prioritizing AI-related risks based on their potential enterprise-wide impact (Databricks documenation).
- **Model registry and performance tracking (MLflow):** MLflow's centralized model registry allows organizations to systematically assess model risks by tracking performance metrics, regulatory compliance status and alignment with strategic objectives. This enables informed prioritization and effective resource allocation for risk mitigation (Databricks documentation).
- **Fine-grained access controls for risk management (Unity Catalog):** Unity Catalog's detailed access controls help measure and manage enterprise-level risks by ensuring that only authorized stakeholders access sensitive data, significantly reducing compliance and operational risks (Databricks documentation).

iii Mitigate AI risks

AI risk mitigation involves the strategic design and implementation of measures to address risks introduced by AI systems, ensuring alignment with organizational goals and societal expectations. These risks often stem from AI's dynamic nature, such as model drift, unintentional bias and autonomous actions. Mitigation strategies must be iterative, adaptable and tailored to specific organizational needs and operational contexts.

A comprehensive mitigation approach integrates multiple strategies to address these risks effectively.

For example:

- **Governance and policy:** Establish clear accountability structures, ethical guidelines and governance mechanisms to guide the responsible and ethical use of AI systems.
- **Technical measures:** Develop mechanisms like model monitoring, adversarial defenses and fairness auditing to maintain AI system performance, security and equity.
- **Cross-functional collaboration:** Facilitate collaboration among technical, legal, compliance and ethical experts to address risks that span organizational and societal domains.

This proactive approach enables organizations to manage risks effectively, foster trust in their AI systems and sustain alignment with broader goals and values.

Organizations should embed these strategies into their existing risk management workflows, tailoring them to reflect the scale and complexity of their AI initiatives. Regular reviews and iterative refinement help mitigation strategies remain relevant and responsive to changes in the AI landscape, including evolving regulatory, technological and operational conditions.

Getting started

Risk mitigation strategies can include:

1. **Strategic adjustment:** Align AI initiatives with organizational goals and risk tolerance. For example, prioritize AI projects that align with core business objectives and demonstrate manageable risk levels that consistently align with organizational strategy.
2. **Governance enhancement:** Define roles and responsibilities within AI governance structures to strengthen oversight and decision-making processes for AI programs. Establish clear accountability for risk assessment, project evaluation and model performance that supports comprehensive oversight.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

3. **Policy and procedural change:** Update organizational policies to address ethical, legal and compliance risks. This could include establishing clear guidelines for ethical AI usage, implementing processes for regular risk reviews and revising policies to accommodate evolving regulatory requirements.
4. **Technical solution:** Implement robust security measures, fail-safes and performance monitoring systems. For example, develop rollback mechanisms or restricted functionality that activate under specific conditions to minimize risks from unexpected AI behavior.
5. **Human oversight mechanism:** Maintain appropriate human involvement in critical AI-driven processes. Oversight mechanisms may include:
    a. **Human-in-the-loop (HITL):** Ensuring human review in high-stakes decisions, such as hiring or medical diagnosis.
    b. **Human-on-the-loop:** Providing ongoing human supervision with an override capability for dynamic AI applications, like automated trading.
    c. **Human-out-of-the-loop:** Limited human involvement in low-stakes or highly automated scenarios, such as content recommendation systems, while retaining accountability and audit trails.
6. **Stakeholder engagement:** Develop strategies to manage public perception and maintain trust, such as regular, transparent reporting on AI practices and outcomes. Proactively engaging stakeholders can help manage expectations and address concerns.
7. **Financial safeguards:** Implement financial controls and monitoring for AI investments, such as tracking AI project ROI and including early termination options for unprofitable projects. These safeguards help ensure that AI initiatives contribute positively to the organization's financial health.
8. **Workforce development:** Address potential workforce disruptions through training and change management. Initiatives might include upskilling programs focused on AI literacy, adaptability training and preparing employees for collaborative roles with AI systems.

9. **Ethical-by-design:** Embed ethical principles directly into AI development and deployment processes. This could involve:
    a. **Establishing ethical guidelines:** Define a set of ethical principles (e.g., fairness, transparency and accountability) that align with your organization's values.
    b. **Impact assessments:** Regularly assess the potential social, economic and ethical impacts of AI systems before deployment.
    c. **Bias mitigation:** Use bias detection and correction techniques to minimize potential biases in model predictions and ensure fairness across different groups.
    d. **Ethics committee:** Form a review committee that evaluates AI projects for ethical concerns, especially in high-impact applications.
10. **Privacy-by-design:** Ensure that AI systems respect and protect user privacy rights throughout their lifecycle. This can include:
    a. **Data minimization:** Collect only the data strictly necessary for the AI to function and implement strong anonymization techniques.
    b. **User consent and control:** Provide transparent ways for users to understand and control their data usage within AI systems.
    c. **Privacy impact assessments:** Regularly assess the AI's impact on user privacy to address potential risks.
    d. **Encryption and secure storage:** Implement encryption standards and secure data storage solutions to prevent unauthorized access.
11. **External specialists:** May assist with complex tasks, such as AI assurance, third-party monitoring and oversight and external bias assessments.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

How Databricks can help

Databricks includes features designed to automate and support risk mitigation strategies, addressing potential AI risks related to data governance, model performance and compliance.

- **Automated governance enforcement with Databricks Workflows:** Using Databricks Workflows, enterprises can automate processes, such as model retraining, auditing and checks, to meet governance requirements. Automation reduces operational risk by minimizing manual errors and ensuring consistent application of risk mitigation policies (Databricks documentation).
- **Real-time model monitoring and alerts:** Databricks real-time monitoring quickly identifies model performance degradation, data drift and anomalies, triggering automated alerts to stakeholders. This enables immediate enterprise-level response to prevent strategic or operational disruptions (Databricks documentation).
- **Role-based access control (RBAC) via Unity Catalog:** Unity Catalog's RBAC helps organizations mitigate enterprise-wide risks by controlling who can access sensitive AI models and data. This prevents unauthorized use that could lead to compliance issues or operational disruptions (Databricks documentation).

iv Continuous monitoring and evaluation

Continuous risk monitoring and evaluation involves tracking AI-related risks to ensure that mitigation efforts remain effective and adaptable. A proactive approach aligns risk management with evolving needs by regularly reviewing existing risks and actively monitoring for new ones. This process supports a dynamic AI risk management strategy that responds to technological changes, regulatory requirements and operational contexts.

Monitoring efforts should strategically assess and prioritize AI-specific risks by considering fundamental principles, such as transparency, to enhance accountability, bias and fairness. This contributes to ensuring equitable outcomes, the adaptability to manage changing data or environments and security against adversarial threats and AI applications' the ethical and societal implications of AI applications. These considerations provide a structured lens for evaluating risks and adjusting risk management strategies.

Incorporating structured feedback loops helps ensure that monitoring insights inform changes to mitigation measures, system performance and organizational policies. The frequency of evaluations should reflect the complexity and importance of AI systems, enabling timely and resource-efficient responses to evolving challenges. Maintaining compliance with dynamic regulatory landscapes requires robust governance mechanisms that inspire stakeholder confidence and ensure accountability.

By embedding these considerations into a dynamic monitoring strategy, organizations can address emerging risks, maintain compliance and build stakeholder trust. This approach ensures that AI initiatives remain reliable, ethical and impactful.

Getting started

Risk monitoring strategies include:

1. **Establish key risk indicators (KRIs):** Define and track key indicators aligning with the organization's risk appetite and AI objectives. These KRIs should be measurable, actionable and directly related to potential AI risks, such as model accuracy, bias detection rates and response times to anomalies.
2. **Regular reviews:** Periodically review AI initiatives to ensure alignment with the organization's strategy and risk tolerance. This involves assessing AI system objectives, performance and impact on strategic goals to confirm they remain within acceptable risk parameters.
3. **Periodic assessments:** Conduct periodic assessments of AI's impact on the organization's reputation, brand perception and stakeholder trust. Engage stakeholders, such as customers, partners and employees, to gather feedback and adjust AI practices to mitigate reputational risks.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

4. **Continuous tracking:** Establish systems and processes for the ongoing monitoring and review of identified risks. This can include regularly assessing changes impacting AI systems, such as model performance, data integrity and any operational or environmental factors. Assign ownership of remediation activities to the appropriate stakeholder teams. Automated alerts or dashboards can help ensure the timely identification of emerging risks.
5. **Update practices:** Regularly update risk management practices based on new insights, technological advancements and regulatory changes. Ensure that the risk management framework remains relevant by incorporating lessons learned from past incidents, adapting to new regulatory standards and leveraging emerging tools for risk mitigation.

How Databricks can help

Databricks supports ongoing monitoring and evaluation of AI systems, enabling organizations to respond efficiently to emerging issues or changes in operational conditions.

- **Streaming analytics and real-time risk detection:** The structured streaming capabilities of Databricks support continuous evaluation of data streams, identifying enterprise-level risks, such as evolving compliance threats, cybersecurity vulnerabilities or strategic shifts requiring rapid adaptation (Databricks documentation).
- **Enterprise alerting system:** Databricks admins can create alert systems that proactively warn stakeholders of emerging operational AI risks. These systems facilitate prompt decision-making and minimize the potential for strategic or reputational damage (Databricks documentation).

v Internal assessment or audit

AI programs can introduce unchecked risks, such as loss of stakeholder trust, regulatory noncompliance, operational failures or ethical lapses (e.g., bias or discrimination). This can result in significant organizational impacts, including reputational damage, regulatory penalties, financial losses and shareholder concerns. These risks are challenging to manage due to inherent complexities in AI systems, such as biases from poor data quality or incomplete datasets, lack of explainability in algorithmic decisions, model drift leading to unreliable outcomes and failure to adapt to varying legal, cultural or operational contexts. Audit programs provide critical oversight to help organizations effectively identify and manage these risks.

Organizations should consider adapting assessment, accountability and/or auditing frameworks to address the risks posed by their AI programs. These frameworks should incorporate AI-specific risks into existing processes and broader auditing practices based on the organization's specific needs and circumstances. Whether adapting established standards or designing new ones, organizations should integrate audit supportability into the AI lifecycle to proactively identify misalignments, strengthen governance and build trust with stakeholders.

Examples of potential AI Audit frameworks to adopt:

- U.S. Government Accountability Office (GAO) AI Accountability Framework
- European Data Protection Board (EDPB) AI Audit Framework
- UK Information Commissioner's Office (ICO) AI Audit Guidance

How Databricks can help

Databricks provides audit logging and tracking capabilities to facilitate internal assessments and audits, enhancing transparency and accountability in AI operations.

- **Detailed audit logging for enterprise transparency:** Databricks offers comprehensive logging of all data and model-related activities, which provides transparency for internal assessments and audits and ensures enterprise compliance with regulatory, ethical and internal standards (Databricks documentation).
- **Enterprise-wide reproducibility with MLflow:** MLflow records detailed model development metadata, ensuring auditors can verify model accuracy, compliance and alignment with enterprise risk management standards, significantly reducing audit risks (Databricks documentation).

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

vi AI incident management

Incident management in AI governance involves establishing protocols to address and mitigate issues in AI systems and services. A structured incident response capability enables organizations to respond promptly and effectively, minimizing operational disruptions, managing compliance risks and protecting organizational integrity.

Organizations should create or enhance incident response capabilities for AI-related risks, incorporating protocols for documenting and communicating incident management activities and outcomes to relevant stakeholders. Regularly reviewing and adapting these protocols supports continuous improvement, keeping the organization responsive to evolving risks and requirements.

Getting started

AI incident management strategies include:

1. **Establish incident response protocols:** Define clear roles, responsibilities and escalation paths for responding to AI-related incidents. Align these protocols with organizational priorities and address issues like data breaches, model failures and compliance violations.
2. **Document incident types and response plans:** Identify common AI incident scenarios, such as drops in model accuracy or data privacy breaches, and develop tailored response plans. For example, a plan for bias detection might include halting deployment, notifying stakeholders and conducting a model audit.
3. **Create communication guidelines:** Develop clear guidelines for documenting and sharing incident response activities with internal and external stakeholders. Include notification protocols for incidents involving customer-facing AI tools to help ensure transparency and timely updates.
4. **Conduct regular simulations and training:** Periodically simulate AI-related incidents, such as errors in financial predictions or data misclassification. Involve cross-functional teams to test response protocols, improve coordination and ensure readiness.
5. **Implement continuous improvement practices:** Regularly review incident reports with cross-functional stakeholders to identify patterns, root causes and opportunities for improvement. Use these findings to enhance AI systems, refine preventative measures and update operational processes.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
AI Organizations
Legal and Regulatory Compliance
Ethics, Transparency and Interpretability
Data, AIOps and Infrastructure
AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

How Databricks can help

Databricks provides functionality that supports structured incident management, including detection, response coordination and comprehensive documentation of AI incidents.

- **Incident response automation via Databricks Workflows:** Databricks Workflows can automate enterprise-level responses to incidents, such as pausing affected model deployments, initiating audit checks or notifying senior stakeholders, thus reducing operational and reputational impact (Databricks documentation).
- **Detailed incident logging and reporting:** Databricks provides robust documentation of incidents, system logs and remedial actions. This comprehensive logging supports incident investigation, regulatory reporting and continuous improvement initiatives at an enterprise level (Databricks documentation).
- **Collaborative incident resolution:** The Databricks collaborative notebooks and dashboards enable a coordinated cross-functional response to AI incidents, ensuring effective and transparent issue resolution and strengthening enterprise resilience (Databricks documentation).

11 Key performance indicators and monitoring

KPIs quantify the health, status and progress of AI programs, initiatives and operations. They provide a structured way to measure aspects such as technical performance, fairness, ethical compliance and stakeholder satisfaction. By offering measurable benchmarks, KPIs enable organizations to track whether their AI programs align with predefined goals and priorities.

Monitoring is the mechanism that powers KPIs, systematically collecting and analyzing data to populate these metrics. AI programs are unique because monitoring must address dynamic and evolving systems, focusing on technical performance and societal impacts like fairness and ethics. Unlike traditional programs, AI monitoring often involves real-time algorithm behavior, bias, compliance and stakeholder feedback tracking to ensure KPIs accurately reflect the program's current state.

Organizations should select KPIs that provide a comprehensive view of their program status, incorporating measures such as alignment with business goals, ethical standards, compliance requirements, model performance consistency, fairness outcomes and stakeholder satisfaction. Regularly reviewing and refining these metrics can help to keep pace with evolving organizational priorities and AI capabilities, ensuring an adaptive and effective AI governance framework.

Getting started

AI program and initiative metrics may include:

1. **Model development:** Monitor the efficiency of the development cycle and resource use across model iterations and updates. Track the effectiveness of model improvements using standardized evaluation protocols. For language models, measure the performance of fine-tuning and the success of adaptations across different business domains. Assess whether development efforts consistently deliver the expected performance gains while staying within planned resource constraints.
2. **Model performance:** Establish baselines and review thresholds that reflect specific accuracy and confidence score requirements for various business scenarios. Monitor prediction distributions and quality metrics to identify potential degradation before it affects operations. For language models, implement systematic sampling to assess consistency, factual accuracy and contextual appropriateness across different use cases. Additionally, track hallucination rates through representative output sampling.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

3. **Data quality:** Deploy monitoring systems to detect gradual pattern shifts and sudden anomalies in input data. Use this information to optimize retraining schedules and maintain performance standards. Regularly validate that the training data reflects the current business context and terminology for language models — track patterns of feature drift and data quality scores as early indicators of potential performance issues.
4. **System stability:** Monitor response times, throughput rates and resource usage, calibrating thresholds to the service's importance. Set alert thresholds for error rates and recovery times based on business impact. Operations involving language models require specific attention to optimizing token processing efficiency and memory usage to effectively manage computational costs.
5. **Output validation:** Implement automated validation pipelines to verify outputs against defined constraints and quality standards. Balance automated checks with targeted human reviews based on output risk and business impact. Develop structured evaluation frameworks for language models to assess response relevance, accuracy and adherence to business guidelines — track rates of constraint violations and correction patterns to identify systematic issues.
6. **Performance testing:** Regularly conduct tests across normal operations, edge cases and stress conditions using production-representative data. Define test scenarios based on actual usage patterns and observed production issues. For language models, emphasize testing across different prompt variations and business contexts to ensure consistent and reliable responses. Monitor performance across diverse inputs to validate generalization capabilities.
7. **Ethical performance:** Monitor model outputs for demographic fairness and potential bias across user groups. Track disparity metrics in predictions and automate the detection of potential fairness violations. Systematically monitor response patterns concerning sensitive topics and demographic references for language models. Validate that ethical guidelines and content filtering remain effective as usage patterns evolve.

Note: Appendix II lists examples of KPIs for AI programs.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

How Databricks can help

- **Performance indicators:** Databricks enables the tracking of KPIs with advanced analytics and reporting tools, measuring model performance, business impact and compliance.
- **Continuous improvement:** The platform's insights from KPI tracking support constant refinement and enhancement of AI practices, driving better outcomes over time (Databricks documentation, Databricks).

12 Reporting

The scope and sophistication of AI programs shape their reporting needs. As programs grow to influence critical business processes, reporting needs to capture performance across multiple dimensions — from technical metrics to business outcomes. This can involve translating complex indicators, like model behavior patterns and fairness measures, into a meaningful business context while maintaining necessary technical detail.

Beyond operational metrics, reporting may need to address broader governance considerations, such as bias mitigation and ethical implications. This comprehensive view enables stakeholders to assess how AI initiatives align with organizational objectives, risk parameters and compliance requirements as programs mature and scale.

Regular review and refinement of reporting practices support effective oversight as AI programs evolve, providing insights that inform governance decisions.

Getting started

GenAI program reports and metrics may include:

1. **Progress reports:** Provide a clear picture of where AI projects stand, tracking milestones, tasks and timelines. They help stakeholders understand what's been achieved, what's coming next and whether delays or challenges need attention.
2. **Performance reports**: Evaluate AI models' technical and operational performance and serving infrastructure. They provide insights into how well the system functions from end-to-end, ensuring reliability, efficiency and scalability in production environments.
3. **Impact reports:** Measure the tangible benefits of AI initiatives for the organization. They demonstrate how AI contributes to strategic objectives, such as increasing revenue, reducing costs or improving customer experiences.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

4. **Compliance reports:** Measure how AI systems operate within the bounds of regulations and organizational policies. They help demonstrate accountability and ethical practices to regulators, compliance teams and other stakeholders.
5. **Incident reports:** Document unexpected issues or failures in AI operations. These reports provide a detailed account of an event, its impact and the steps taken to resolve it, helping prevent future occurrences.
6. **Audit reports:** Provide an in-depth evaluation of AI systems and governance practices. They ensure processes, data usage and compliance efforts are consistent with organizational standards and regulatory requirements.

How Databricks can help

- **Transparency and accountability:** Databricks offers customizable dashboards and reporting tools that provide transparent and accountable reporting of AI initiatives. This ensures that progress and compliance are visible to all stakeholders.
- **Stakeholder communication:** The platform supports open communication with stakeholders through periodic updates and comprehensive reports, ensuring alignment and informed decision-making across the organization (Databricks).
- **AI-powered business intelligence (AI/BI):** Databricks AI/BI is an AI-assisted BI solution built into the Databricks Platform. It allows users to quickly create analytical datasets, interactive dashboards and visualizations. The AI/BI Genie feature offers a conversational interface, allowing users to ask questions in natural language and receive real-time insights, reducing reliance on expert practitioners (Databricks).
- **Model versioning and lifecycle management:** Databricks integrates MLflow Model Registry with Unity Catalog to provide centralized governance for ML models. This integration automatically tracks updated versions of registered models, ensuring a clear audit trail (Databricks documentation).

Legal and Regulatory Compliance

Legal and regulatory compliance is paramount in AI to ensure that AI technologies are developed, deployed and managed responsibly and transparently. This pillar outlines legal and regulatory governance components, emphasizing the importance of accountability, fairness, legal adherence and proactive risk management. By establishing legal and regulatory AI frameworks and practices, organizations can ensure that their AI initiatives comply with existing regulations and anticipate future challenges and opportunities.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Section summary

Establish robust legal and regulatory frameworks to comprehensively govern AI systems, ensuring compliance, reducing liability risks and fostering responsible AI use. Core components include legal adherence, proactive risk management, comprehensive documentation, contractual safeguards and dynamic adaptability to evolving legal standards and regulations.

- **Assessment of legal and regulatory considerations:** Evaluate and respond to obligations stemming from AI-specific regulations (e.g., EU AI Act), privacy laws (e.g., GDPR), data residency requirements and industry-specific regulatory mandates.
- **Operational liability and governance:** Clarify accountability structures, liability distribution, IP protections and contractual obligations to mitigate risks inherent to AI's autonomous capabilities.
- **Comprehensive legal planning:** Implement thorough documentation and robust contractual frameworks addressing AI-specific complexities, performance obligations, data handling, IP rights and third-party compliance.
- **Proactive risk and liability management:** Prioritize legal risk management strategies, restrict or guide AI applications in high-risk contexts and emphasize explainability to reduce liability exposure.
- **Ongoing monitoring and compliance:** Monitor regulatory developments, maintain audit readiness, promptly adapt governance frameworks and ensure compliance throughout the AI lifecycle.
- **Emerging regulatory trends:** Actively track and respond to global legal and regulatory changes, ensuring the adaptability of governance frameworks and practices to sustain compliance and stakeholder trust.

1 Legal and regulatory compliance

AI presents unique legal and regulatory challenges due to its complex and dynamic nature. Characteristics, such as autonomy, dependence on extensive datasets and the capacity to generate novel outputs, lead to ambiguities in legal risk areas, including accountability, compliance and IP. Furthermore, issues like system opacity, potential bias, cross-jurisdictional operations and evolving regulatory frameworks further complicate the landscape.

To navigate these complexities, organizations should consider establishing an adaptable legal framework that mitigates risk, aligns explicitly with applicable regulatory requirements and legally mandated ethical standards, adapts to shifting requirements and fosters stakeholder trust. This section outlines considerations for legal programs that aim to reduce risks and support responsible AI development and deployment.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

i Assess: Legal and regulatory considerations

AI programs introduce unique capabilities, like suggested or autonomous actions, decision-making and inference from vast data sources, creating a complex legal landscape. Organizations should assess and develop an understanding of their legal obligations based on their AI program's goals. Legal considerations include, but are not limited to:

a. **Legal and regulatory compliance**

- **AI-specific regulations (e.g., EU AI Act):** Emerging AI regulations emphasize transparency, safety and bias prevention. Aligning with these frameworks helps organizations mitigate risks and promote responsible deployment.
- **Privacy laws (e.g., GDPR, CCPA, PDPA):** Privacy regulations are a key aspect of the broader legal landscape. They ensure compliance with data handling, transparency and informed consent requirements. Adhering to privacy laws helps organizations protect individual rights and reduce compliance risks.
- **Data residency laws (e.g., CSL, BDSG):** Data residency laws impact the physical location where data can be processed and stored.
- **Industry-specific regulations (e.g., HIPAA, ECOA, FHA):** Specific industries impose distinct legal requirements for AI use. For example, healthcare AI should comply with patient privacy laws, while financial AI applications should address anti-discrimination practices.
- **Local, regional and cross-border compliance:** AI programs operating locally, regionally or across borders should evaluate each region's regulatory requirements.
- **Cross-jurisdictional harmonization:** Develop and maintain a compliance matrix mapping specific AI legal obligations across key jurisdictions, clearly highlighting differences, overlaps or conflicts. Use this matrix as a reference for governance decisions on system design, data residency and operational protocols.

b. Operational and stakeholder considerations

- **Terms, conditions and contractual obligations:** To mitigate legal risks, organizations should consider crafting contractual frameworks explicitly detailing ownership of AI-generated outputs, data usage rights, IP protections, liability distributions, vendor audit rights and indemnification clauses.
- **Operational liability and governance:** AI's autonomous decision-making introduces complexities in assigning liability. Organizations should consider establishing governance structures and legal controls to clarify accountability, manage risks and define responsibilities.
- **Intellectual property rights:** AI's transformative nature raises challenges around IP ownership. Organizations are encouraged to safeguard AI-generated works, proprietary algorithms and trade secrets to maintain competitive advantages.
- **Customer and employee protections:** AI impacts employees and consumers, necessitating compliance with customer-protection laws, such as anti-discrimination practices and fairness in automated decision-making processes.

A strategic legal assessment should guide governance, risk and compliance strategies throughout the AI lifecycle. By regularly revisiting these assessments, organizations can stay aligned with emerging themes, such as transparency, fairness and accountability, ultimately remaining proactive and compliant.

Getting started

Steps organizations can take to assess GenAI programs may include:

1. **Assemble a cross-functional compliance team:** Form a dedicated team comprising legal experts, data privacy officers, AI/ML engineers, risk managers and compliance officers to oversee AI compliance initiatives and ensure alignment with legal obligations and ethical standards.
2. **Conduct a comprehensive AI compliance review:** Initiate a thorough review of existing and planned AI systems to assess adherence to relevant regulations, such as the EU AI Act, GDPR, CCPA, HIPAA and industry-specific standards and identify gaps.
3. **Develop a regulatory compliance matrix:** Create a matrix mapping applicable legal requirements across jurisdictions where AI systems operate. Matrix development should consider data residency laws, cross-border data transfer regulations and sector-specific mandates.
4. **Integrate privacy by design principles:** Embed privacy considerations into the AI development lifecycle. Implement data minimization, purpose limitation and user consent mechanisms.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

5. **Establish clear contractual frameworks:** Establish preapproved legal language and contractual clauses for various AI program types. Review and update contracts with vendors, partners and stakeholders to ensure that AI initiative and service contracts reflect the organization's risk tolerance and compliance obligations.
6. **Engage with regulatory bodies and industry groups:** Maintain active communication with regulators and participate in industry forums to stay abreast of regulatory developments.
7. **Schedule regular compliance reviews:** Establish a cadence for periodic reviews of AI systems and compliance measures. These reviews should assess the effectiveness of current practices and identify areas for improvement, ensuring that the organization remains proactive in its compliance efforts.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

ii Prioritize: Liability and risk management

Liability and risk management are essential for AI program risk reduction, particularly given AI's autonomous nature and potential unintended outcomes. This section focuses on identifying and prioritizing key legal risks.

For example, effectively managing legal risks can include:

- Assessing potential liabilities across different AI applications using a legal risk scoring system. This approach can guide prioritization and resource allocation while ensuring alignment with the organization's broader legal strategy.
- Establish internal policies restricting or prohibiting AI use in high-risk areas or processes. For instance, some organizations avoid AI-driven employee evaluation and management systems due to potential legal implications and reputational harm arising from unintended bias or inaccuracies.
- Promoting explainable model outcomes to manage liability, particularly in high-risk domains like healthcare, finance or autonomous vehicles. Addressing risks specific to AI, such as privacy loss, bias and unintended harm, requires a robust combination of accountability structures, legal oversight and audit trails.

Organizations should evaluate liability distribution when engaging third-party AI services, particularly for damages from errors or misuse. Contracts should define liability boundaries and include safeguards like indemnification clauses.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Getting started

Steps organizations can take to prioritze GenAI program liability and risk may include:

1. **Implement a legal risk scoring framework:** Develop a structured framework to assess AI applications based on potential legal exposure, operational impact and reputational risk. Align this framework with regulatory classifications, such as the EU AI Act's risk tiers, to prioritize oversight and resource allocation.
2. **Define expectations for high-risk AI use cases:** Establish clear expectations that restrict or prohibit AI deployment in sensitive areas, such as employee evaluations or autonomous decision-making in critical systems, based on the organization's risk appetite and legal obligations.
3. **Integrate explainability into AI systems:** Assess the organization's requirements for AI models to provide transparent and interpretable outputs, especially in high-stakes domains like healthcare and finance. Implement tools and methodologies that facilitate understanding of AI decision-making processes to support accountability and compliance where required.
4. **Clarify liability in third-party AI agreements:** When engaging external AI vendors, delineate liability boundaries within contracts. Include indemnification clauses and specify responsibilities for errors or misuse to mitigate legal risks.

iii Plan: Comprehensive legal planning for AI

a. Legal documentation

Legal documentation supports oversight, compliance and transparency by setting requirements for clear records of the use, deployment and governance of AI systems. It should address AI-specific complexities like dynamic models, data provenance and algorithmic rationale.

Well-maintained records facilitate regulatory compliance and help organizations manage risks effectively. Governance frameworks should incorporate processes for regularly reviewing and updating documentation, helping it remain comprehensive and adaptable to evolving legal requirements.

b. Contractual frameworks for AI products and services

Contracts governing AI products and services should consider the complexities introduced by AI systems. For example, they should define performance and reliability standards, including protocols for retraining, errors and outages. When using external datasets, data ownership and permitted uses should be addressed, along with compliance with data protection and IP laws.

Vendor compliance requirements should be included to help ensure third-party compliance with privacy and ethical standards. Vendor contracts could include audit rights and require adherence to specified AI-transparency requirements that apply both to vendors and their underlying supply chains.

Organizations should develop and maintain preapproved legal clauses for AI-related contracts. Preapproved clauses can improve contract negotiation efficiency and consistency and standardize legal risk mitigation approaches that address key legal issues, such as liability, privacy, IP and compliance.

c. Legal change management strategy

Given the rapid pace of AI regulation development globally, organizations should consider a legal change management strategy. This would include:

- **Monitoring regulatory developments:** Establish a process to monitor changes in relevant AI laws and regulations.
- **Adapting governance processes:** Enable processes for governance frameworks to adapt to new or revised regulations.
- **Evaluating open source model license suitability:** Commercial deployment of open source AI models requires careful license evaluation since each model will likely impose specific usage restrictions. License terms may also vary for different model sub-components (such as weights, training code, deployment, etc.).
- **Regulatory impact assessment:** Assess how new regulations impact AI operations, contractual obligations and compliance strategies.
- **Regulatory foresight reviews:** Conduct bi-annual regulatory foresight reviews, proactively assessing emerging AI regulations and identifying potential legal impacts. Incorporate findings directly into strategic governance updates and internal training programs.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Getting started

Steps organizations can take to support comprehensive legal planning for AI may include:

1. **Establish structured AI documentation practices:** Implement standardized documentation frameworks to capture essential details about AI systems, such as data sources, model architectures, training methodologies and decision-making processes,to meet internal and regulatory requirements.
2. **Develop comprehensive contractual frameworks for AI products and services:** Draft contracts that address AI-specific considerations, including data ownership, usage rights and compliance obligations related to data protection and IP laws. Incorporate clauses that require third-party vendors to adhere to specified ethical standards and data protection regulations.
3. **Implement preapproved legal clauses for AI agreements:** Create a repository of standardized contractual clauses addressing common AI-related legal issues, such as liability, privacy, IP and compliance. Preapproved language streamlines contract negotiations, ensures consistency and mitigates legal risks.
4. **Establish a legal change management strategy:** Develop a proactive approach to monitor and respond to evolving AI regulations. This includes setting up processes to track regulatory developments, assess their impact on AI operations and contracts and update governance frameworks accordingly. Regularly conduct regulatory foresight reviews to anticipate changes and adjust internal policies and training programs.
5. **Evaluate open source AI model licenses for commercial use:** When deploying open source AI models, thoroughly assess license terms to understand usage restrictions and compliance obligations associated with different model components, such as weights, training code and deployment tools. This evaluation ensures lawful use and mitigates potential legal risks.

iv Deploy: Legal protections and safeguards

Proactive legal protections help manage AI deployment risks. Adopt best practices and regulatory guidelines to prepare for upcoming requirements and minimize legal exposure. Establish internal safeguards for AI usage, data handling and processes and external protections for third-party relationships, data exchange and agreements.

Verify that protections cover IP rights, liability limits and jurisdictional compliance. Including these safeguards in the governance framework strengthens resilience and helps navigate legal challenges.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Getting started

Steps organizations can take to establish legal protections and safeguards during AI deployment may include:

1. **Implement deployment risk assessments:** Conduct structured evaluations of AI systems before deployment to identify potential legal exposures, focusing on areas such as data privacy, IP rights and jurisdictional compliance.
2. **Establish internal safeguards:** Develop and enforce policies governing AI usage, data handling and decision-making processes to ensure alignment with legal requirements and ethical standards.
3. **Define external protection measures:** To mitigate legal risks associated with external partnerships, incorporate clauses in third-party agreements that address liability limitations, data exchange protocols and compliance obligations.
4. **Integrate IP protections:** Ensure that AI-generated outputs and underlying algorithms are protected under appropriate IP laws and that ownership rights are clearly defined and documented.
5. **Align with jurisdictional compliance requirements:** Map AI deployment strategies to each operating jurisdiction's legal and regulatory frameworks, ensuring adherence to local laws and international standards.

v Monitor: Ongoing compliance and audits

Regular monitoring helps organizations maintain adherence to established protocols, proactively mitigate risks and address emerging challenges. Compliance should also be reassessed when trigger events occur, such as releasing new legislation, significant modifications to AI models or integrating new data sources. Monitoring should be ongoing to align with evolving standards and enable rapid response to legal developments.

Getting started

Steps organizations can take to monitor AI compliance and conduct audits may include:

1. **Establish continuous monitoring protocols:** Implement systems to regularly track AI performance, data usage and decision-making outcomes, enabling prompt identification of compliance issues.
2. **Conduct regular compliance audits:** Schedule periodic audits to assess adherence to legal requirements, internal policies and ethical standards, adjusting practices as necessary based on findings.
3. **Define trigger events for reassessment:** Identify specific events, such as significant model updates, integration of new data sources or changes in legislation, that necessitate immediate compliance reviews.
4. **Use audit trails and documentation:** Maintain comprehensive records of AI system operations and decision-making processes to support transparency and facilitate regulatory inspections.
5. **Engage independent auditors:** Consider involving third-party experts to provide objective assessments of AI systems, which would enhance credibility and uncover potential blind spots in internal evaluations.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

vi Prepare: Review emerging trends in AI ethics and regulation

Monitor AI ethics and regulations trends across relevant jurisdictions to stay informed of emerging standards and best practices. Establish mechanisms to incorporate evolving ethical guidelines and legal requirements into your governance framework, clearly assigning internal monitoring, assessment and implementation responsibilities. Proactively engage with policymakers and industry stakeholders and maintain regular internal communication to support organizational adaptability and alignment with current and anticipated developments.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Getting started

Steps organizations can take to monitor AI compliance and conduct audits may include:

1. **Monitor regulatory developments:** Establish a process to track changes in AI-related laws and ethical guidelines across relevant jurisdictions, ensuring timely updates to compliance strategies.
2. **Engage with policymakers and industry stakeholders:** Participate in industry forums, regulatory consultations and collaborative initiatives to influence and stay informed about emerging standards.
3. **Assign internal responsibilities:** Designate specific roles or teams responsible for monitoring, assessing and implementing changes related to AI ethics and regulations within the organization.
4. **Incorporate findings into governance frameworks:** Regularly update internal policies, training programs and operational procedures to reflect evolving ethical considerations and legal requirements.
5. **Conduct scenario planning exercises:** Simulate potential regulatory changes and ethical dilemmas to assess organizational readiness and develop proactive response strategies.

Ethics, Transparency and Interpretability

Ethics transparency in AI refers to the openness and clarity of an AI system's design, operational processes, decision-making criteria and communicated details, such as model logic, data usage and limitations. This transparency enables stakeholders, including users, regulators and impacted individuals, to understand, verify and challenge AI-driven outcomes.

Transparency is fundamental to effective AI governance, as it directly supports organizational trust, strengthens accountability and supports compliance with evolving global regulatory requirements, such as the EU AI Act or OECD AI principles. However, achieving transparency in an AI program is an incremental process influenced by factors such as model complexity, industry standards, regulatory expectations and specific organizational contexts.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Section summary

Integrate ethical principles and transparency practices across AI governance, covering both development and operations. This fosters responsible use, builds stakeholder trust and supports regulatory alignment.

Ethics provides foundational principles, such as accountability, fairness, human-centricity, inclusivity and cultural sensitivity. Transparency operationalizes these principles through interpretability, explainability, traceability and strategic information disclosure.

- **Core ethical principles:** Uphold accountability by clearly defining roles and responsibilities, promoting fairness through bias mitigation, emphasizing human-centric design to safeguard well-being, practicing inclusivity to ensure equitable outcomes and embedding cultural sensitivity to respect diverse norms and contexts.
- **Dimensions of transparency:** Implement interpretability (intuitive model logic), explainability (clear rationale behind AI decisions), traceability (systematic documentation and auditability) and strategic disclosability (balanced information sharing considering security and competitiveness).
- **Transparency and ethics in AI development:** Embed ethics and transparency early into AI lifecycle management via interpretable architectures, proactive bias and risk mitigation, inclusive and culturally aware design processes and lifecycle-long transparency objectives.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

- **Transparency and ethical oversight in operations:** Foster operational transparency through clear communication of AI roles, decision logic, confidence levels, uncertainties and active oversight to maintain ethical integrity and stakeholder trust.
- **Ethical transparency in AI serving:** Provide tailored, real-time explanations of AI outcomes and actively disclose known model limitations and ethical considerations. Integrate continuous stakeholder feedback to refine AI systems and uphold ethical standards.
- **Navigating ethical and transparency trade-offs:** Strategically balance ethical obligations and transparency requirements against practical constraints, including proprietary protection, security risks and competitive considerations, while preserving stakeholder confidence and regulatory compliance.

1 AI Ethics

AI ethics involves establishing principles that guide the responsible design, deployment and use of AI across governance structures and the broader organization. Ethics within governance underpins policies, accountability mechanisms and oversight processes, ensuring ethical considerations are integrated throughout the AI lifecycle. Building ethics into the organizational culture influences decision-making, fosters collaboration and aligns stakeholder engagement with ethical commitments. This integrated approach addresses challenges, such as bias, decision-making opacity and societal impacts, while helping organizations align AI initiatives with their missiosn, build trust and manage risks effectively. When designing an ethical AI program, consider the following guidelines.

i Accountability

Accountability establishes clear roles and responsibilities for overseeing AI systems, fostering transparency and enabling effective oversight. AI systems can make autonomous decisions or generate outputs that inform organizational actions. In both cases, organizations remain fully accountable for the resulting impacts and decisions. Addressing complexities, such as AI's inherent opacity, autonomy, potential biases and long-term effects, supports organizations in proactively managing risks, responding to ethical considerations and sustaining stakeholder trust. Regular evaluations and integration of clearly defined AI-related roles into broader governance frameworks ensure accountability remains consistent, actionable and firmly aligned with organizational objectives and responsibilities.

Getting started

Accountability considerations include:

1. **Boards and committees:** Establish an AI ethics board or committee to oversee AI ethics policies and decisions, such as:
    a. **Scope and authority:** Clearly define the committee's scope, authority and decision-making processes, ensuring effective integration with organizational strategy and operations.
    b. **Representation:** Ensure diverse stakeholder representation for comprehensive oversight.
    c. **Escalation paths:** Define explicit escalation pathways for promptly addressing ethical and accountability concerns.
2. **Roles:** Define clear roles and responsibilities to align AI accountability with organizational goals:
    a. **AI ethics officer:** Ensures compliance with ethical standards and best practices.
    b. **AI developers:** Create and maintain AI systems aligned with ethical guidelines.
    c. **AI auditors:** Conduct regular assessments to verify compliance with ethical and regulatory standards.
3. **Cross-functional coordinators:** Coordinate accountability practices consistently across departments.
    a. **Accountability mechanisms:** Implement mechanisms to hold individuals and teams accountable for AI-related decisions and actions.
    b. **Performance metrics:** Develop metrics such as fairness, accuracy and system uptime to evaluate AI system performance.
    c. **Accountability frameworks:** Outline clear steps and procedures for ensuring accountability in AI operations.
    d. **Incident response:** Establish escalation procedures and response plans for addressing unintended AI impacts or ethical breaches.
    e. **Impact assessments:** Conduct periodic assessments of AI system performance, including ethical, social and reputational dimensions.
    f. **External engagement:** Engage independent reviewers and external stakeholders to reinforce transparency and external accountability.
    g. **Training programs:** Implement regular training initiatives to reinforce team accountability and ethical practices.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

4. **Reporting and transparency:** Foster trust and transparency through transparent, consistent reporting practices:
    a. **Transparency reports:** Publish regular reports detailing AI performance, ethical practices, decision-making processes and compliance status.
    b. **Stakeholder communication:** Establish transparent processes for communicating accountability outcomes internally and externally.
    c. **Feedback mechanisms:** Provide structured channels for stakeholders to report ethical or accountability concerns beyond statistical metrics.
    d. **External transparency:** Consider proactive external disclosures to position ethical AI practices as a competitive differentiator.

How Databricks can help

- **RBAC:** Databricks provides robust RBAC features, allowing precise definition and assignment of roles such as AI ethics officer, AI developers and AI auditors. This ensures that responsibilities are delineated and adhered to, maintaining a structured approach to AI governance (Databricks documentation, Databricks documentation).
- **Audit trails:** Databricks offers comprehensive logging and audit trails that capture detailed information about AI system usage, model training and deployment activities. These logs are essential for conducting thorough internal and external audits to ensure compliance with ethical standards (Databricks documentation, Databricks).
- **Compliance monitoring:** Databricks facilitates continuous compliance monitoring with automated alerts and reports to ensure ongoing adherence to ethical standards and regulatory requirements (Databricks, Databricks).
- **Dashboard reporting:** Databricks supports the generation of detailed transparency dashboards that outline AI performance, decision-making processes and compliance with ethical guidelines. These reports can be shared with stakeholders to maintain transparency and accountability (Databricks documentation).

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

- **Tracing for GenAI:** Databricks enhances GenAI observability through MLflow Tracing, which captures detailed execution data of AI agents. This facilitates debugging, performance optimization and compliance auditing. This end-to-end observability is crucial for understanding agent behavior and ensuring alignment with organizational standards (Databricks documentation).
- **Agent evaluation:** Databricks supports the systematic evaluation of GenAI agents through integrated MLflow capabilities, capturing metrics related to accuracy, reliability and compliance. This enables comprehensive assessments, validation and reporting aligned with organizational and regulatory standards (Databricks documentation).

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

ii Fairness and nondiscrimination

Fairness and nondiscrimination guide the development of AI systems to promote equitable outcomes and reduce the risk of perpetuating biases. Because AI systems often rely on historical data and algorithms, they present unique challenges in avoiding systemic inequities and detecting deeply embedded biases. Governance frameworks should include regular assessments of AI system impacts on different demographic groups to support fair and equitable outcomes over time.

Getting started

Key fairness and nondiscrimination methodologies include:

1. **Bias identification:** Use statistical and computational methods to detect biases in data and models, considering factors such as race, gender, sexual orientation and socioeconomic status. Techniques might include:
    a. **Data audits:** Regularly audit datasets to identify potential biases.
    b. **Fairness metrics:** Implement metrics to measure bias, such as disparate impact analysis. Fundamental exploratory analyses (e.g., t-tests, Chi-square tests and ANOVA) can help identify biases across various data segments.
    c. **GenAI checks:** Bias assessments are increasingly being applied to GenAI systems, particularly to detect gender or racial bias when generating descriptions, images or content related to occupations.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

2. **Bias mitigation:** Apply techniques incorporating fairness constraints in algorithm design to ensure equitable outcomes during model training. Example techniques and constraints may include:
    a. **Resampling:** Adjust the data distribution to balance representation.
    b. **Re-weighting:** Assign weights to data points to reduce bias.
    c. **Adversarial debiasing:** Use adversarial techniques to minimize bias in the model.
    d. **Compensating controls:** Implement controls during model serving to address biases that cannot be removed during training.
    e. **Additional data collection:** Gather additional or more representative data to mitigate biases.
    f. **Label adjustments:** Modify labeling strategies to reduce unintended bias.
    g. **Counterfactual testing:** Use counterfactual scenarios to evaluate and adjust model fairness.
3. **Continuous monitoring:** Implement systems to track AI performance and fairness metrics over time. Regularly update models to address newly identified biases and ensure ongoing fairness. Specific actions include:
    a. **Performance tracking:** Continuously monitor AI outputs for fairness.
    b. **Model updates:** Periodically retrain models with new data to correct for any emerging biases.
    c. **Feedback loops:** Establish feedback mechanisms to incorporate stakeholder input and findings from bias audits.

How Databricks can help

- **Fairness metrics:** The platform provides predictive parity, equality, equal opportunity and statistical parity to measure and compare model performance across different demographic groups (Databricks documentation).
- **Data audits:** Databricks enables regular audits of datasets to identify and address potential biases. Tools like Fairlearn and SHAP can be integrated to detect biases across factors such as race, gender, sexual orientation and socioeconomic status (Databricks documentation, Databricks).

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

- **Re-sampling and re-weighting:** Databricks supports techniques like re-sampling and re-weighting to adjust the data distribution and reduce bias during model training. This ensures that underrepresented groups are adequately represented in the training process (Databricks documentation).
- **Adversarial debiasing:** Using adversarial debiasing techniques, Databricks helps minimize bias within models. This involves training models so they cannot distinguish between different demographic groups, leading to fairer outcomes (Databricks).
- **Performance tracking:** Databricks Lakehouse Monitoring allows organizations to continuously track AI model performance, capturing inference data and enabling local explainability. This facilitates comprehensive monitoring for data quality, model drift and bias detection and ensures sustained fairness, transparency and reliability of AI systems over time (Databricks documentation).
- **Feedback loops:** Databricks supports establishing feedback mechanisms to incorporate stakeholder input and findings from bias audits into the model development lifecycle. This helps continuously improve AI system fairness and accountability (Databricks).
- **Good governance and compliance:** The platform's Unity Catalog and other governance tools help enterprises track data lineage, ensure quality control and maintain compliance with global privacy and data protection regulations. This comprehensive governance framework is critical for responsible AI deployment (Databricks).

iii Human centricity and well-being

Human centricity prioritizes aligning AI systems with human values, rights and societal well-being. As AI increasingly makes autonomous decisions in areas such as healthcare, IT and public safety, robust safeguards and oversight mechanisms are essential to mitigate risks and protect individuals and communities. Building transparency and interpretability in these contexts is vital, as they are crucial to maintaining effective human oversight. Regular impact assessments and adaptable safeguards can ensure that AI systems continue to support societal well-being and align with ethical goals as they evolve.

Foreword

Introduction

Pillar I: AI Organizations

Pillar II: Legal and Regulatory Compliance

Pillar III: Ethics, Transparency and Interpretability

Pillar IV: Data, AIOps and Infrastructure

Pillar V: AI Security

References and Further Reading

Acknowledgements

Appendix A: Glossary

License

Getting started

Human-centric methodologies may include:

1. **User-centered design:** Conduct user research to understand needs and expectations. Both direct users and those who may be indirectly affected by the AI system should be involved through participatory design sessions and usability testing.
    a. **User research:** Gather insights through interviews, surveys and observational studies to inform design decisions, including secondary and broader societal impacts.
    b. **Participatory design:** Engage diverse user groups and stakeholders in the design process to co-create solutions that address user needs.
    c. **Usability testing:** Test AI systems with real users and stakeholders to identify and address usability and secondary impact issues.
2. **Safety protocols:** Develop human safety guidelines for AI application development and integrate them into fail-safes and HITL mechanisms for critical decision-making processes:
    a. **Human safety guidelines:** Establish comprehensive guidelines to ensure AI applications prioritize user safety.
    b. **Fail-safes:** Implement technical measures to prevent AI systems from causing harm in case of failure.
    c. **HITL:** Ensure critical decisions involve human oversight to maintain control and accountability.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

3. **Well-being monitoring:** Evaluate AI system impact on user well-being and adjust AI applications based on feedback and observed effects:
    a. **Impact assessment:** Regularly assess the impact of AI systems on user well-being through surveys, interviews and behavioral analysis.
    b. **Feedback integration:** Continuously collect and incorporate user feedback to improve AI applications.
    c. **Adaptation:** Modify AI systems to address any negative impacts on user well-being and enhance positive outcomes.
    d. **Policy-driven red teaming**: Perform red-teaming exercises focused on specific risk types (such as child safety, election integrity, etc.), potentially involving collaboration with external experts.

How Databricks can help

- **Participatory design:** Databricks facilitates participatory design sessions through collaborative workspaces and version-controlled notebooks. These features enable stakeholders to co-create solutions, ensuring AI systems meet user needs and preferences (Databricks).
- **Human safety guidelines:** Using Databricks, you can establish and integrate comprehensive human safety guidelines into AI application development. The platform supports ML operations (MLOps) capabilities to enforce these guidelines throughout the development lifecycle, prioritizing user safety (Databricks).
- **Feedback integration:** The platform supports continuous feedback loops, enabling the collection and incorporation of user feedback to refine AI applications based on real user experiences and needs (Databricks documentation).
- **Adaptation:** Databricks provides a flexible and scalable environment that allows for rapid modifications to AI systems. This adaptability helps address any negative impacts on user well-being and enhances positive outcomes, ensuring a consistently positive user experience (Databricks).
- **LLM-as-a-judge:** The Databricks AI agent evaluation incorporates built-in AI judges that leverage LLMs to assess various quality aspects of AI-generated outputs, such as correctness, groundedness and safety. These judges provide pass or fail ratings and rationales, facilitating systematic evaluation of GenAI applications without extensive human oversight (Databricks documentation).

- **AI gateway guardrails:** The Databricks AI gateway offers configurable guardrails at the model serving endpoint level, enabling enforcement of data compliance and reducing harmful content. Features include safety filtering to prevent unsafe interactions and personally identifiable information detection to protect sensitive information, ensuring secure and responsible AI deployments (Databricks documentation).

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

iv Inclusivity

Inclusivity in AI design emphasizes addressing the needs of diverse populations while minimizing the risks of marginalization for underrepresented groups. AI systems depend on training data and design choices — promoting inclusivity in these processes improves accessibility and equity, minimizes cognitive biases within design and development teams and leads to fairer and more effective outcomes. Scalable AI systems often encounter challenges in accurately reflecting local or regional contexts, adding complexity to inclusivity efforts. Organizations can manage these challenges by developing adaptable systems, collaborating meaningfully with diverse stakeholders and embedding continuous learning and improvement mechanisms.

Getting started

Inclusivity methodologies may include:

1. **Inclusive design practices:** Follow inclusive design principles to ensure user accessibility and design AI interfaces with adaptive features to cater to diverse needs, such as:
    a. **Universal design:** Implement design principles that make AI systems usable by the broadest range of people possible without needing adaptation.
    b. **Adaptive features:** Incorporate features that can be customized to meet users' individual needs, such as text-to-speech, adjustable font sizes and alternative input methods.
    c. **Accessibility standards:** Adhere to established accessibility standards (e.g., WCAG) to ensure AI systems are accessible to individuals with disabilities.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

2. **Community engagement:** Engage with community groups to understand diverse needs and perspectives and co-create AI solutions with input from underrepresented and marginalized communities:
    a. **Stakeholder consultations:** Consult various community groups to gather AI design and deployment feedback.
    b. **Co-creation workshops:** Organize workshops that bring together diverse stakeholders to collaborate on designing inclusive AI solutions.
    c. **Feedback mechanisms:** Establish channels for continuous feedback from underrepresented and marginalized communities to ensure their needs are addressed throughout the AI lifecycle.

v Cultural norms and sensitivity

Cultural norms and sensitivity emphasize the importance of designing AI systems that respect and adapt to the values and expectations of different societies. However, the global scalability of AI systems introduces challenges in aligning with local customs and cultural contexts. These challenges compound when systems prioritize efficiency or uniformity over cultural nuance, risking the erasure of local values. To address this, engaging cultural experts to incorporate diverse regional perspectives into AI design and leveraging diverse training datasets allows systems to remain appropriate and adaptable over time. All cultural sensitivity considerations included in systems design and deployment should be appropriately documented for future explainability.

Getting started

Cultural norms and sensitivity considerations may include:

1. **Cultural awareness:** Train AI developers and stakeholders to recognize and respect cultural differences, such as:
    a. **Cultural sensitivity training:** Provide training programs to educate AI developers and stakeholders on cultural awareness and sensitivity.
    b. **Diversity and inclusion workshops:** Organize workshops that focus on the importance of cultural diversity and inclusion in AI development.

2. **Localization:** Adapt AI models and applications to align with local cultural norms, values and languages:
    a. **Language support:** Ensure AI applications support multiple languages and dialects relevant to the target user base.
    b. **Cultural customization:** Customize AI features and interactions to reflect local customs, traditions and values.
    c. **Region-specific content:** Incorporate culturally relevant and appropriate content for different regions.
3. **Ethical review:** Include cultural experts in ethical review processes to evaluate the cultural implications of AI applications:
    a. **Cultural expert panels:** Establish panels of cultural experts to provide insights and recommendations during the ethical review of AI projects.
    b. **Impact assessments:** Conduct cultural impact assessments to proactively identify and address potential cultural sensitivities.
    c. **Continuous feedback:** Implement mechanisms for constant feedback from cultural experts to ensure ongoing alignment with cultural norms and sensitivities.

Foreword
Introduction
AI Organizations
Legal and Regulatory Compliance
Ethics, Transparency and Interpretability
Data, AIOps and Infrastructure
AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

2 Transparency

The effectiveness of transparency in AI programs relies on clearly defined and integrated dimensions. These dimensions described provide the foundational structure enabling stakeholders to understand AI system functionality, decision rationale, developmental processes and communication strategies.

- **Interpretability:** Interpretability emphasizes building AI systems with inherently understandable logic and internal processes. These systems, such as decision trees or linear regression models, are beneficial in high-stakes decision contexts.
- **Interpretability decision criteria:** Complex 'black-box' AI models should be considered acceptable when their predictive performance significantly surpasses interpretable alternatives — provided they incorporate robust transparency mechanisms, such as detailed documentation, rigorous validation, structured explainability tools (e.g., SHAP, LIME) and enhanced auditability. Leadership should explicitly approve these decisions, documenting justification aligned with the organization's strategic risk tolerance.
- **Explainability:** Explainability methods are chosen based on defined factors that help stakeholders gain actionable insights into AI outcomes, including stakeholder information needs, regulatory expectations and the importance of AI-driven decisions.
- **Traceability:** Traceability supports comprehensive documentation of AI system data sources, model iterations, version control and audit trails, facilitating reproducibility and strengthening accountability.
- **Disclosability:** Disclosability requires systematically determining what AI-related information to communicate, whom to communicate it to and under what conditions. Consider developing a formal disclosability framework, specifying clear criteria for disclosure aligned with stakeholder needs, competitive sensitivity, regulatory mandates and security risks.

Together, these dimensions support informed oversight, accountability and stakeholder trust.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword

Introduction

Pillar I: AI Organizations

Pillar II: Legal and Regulatory Compliance

Pillar III: Ethics, Transparency and Interpretability

Pillar IV: Data, AIOps and Infrastructure

Pillar V: AI Security

References and Further Reading

Acknowledgements

Appendix A: Glossary

License

i Transparency in AI development and design

Integrating transparency during AI development and design involves directly embedding clear and purposeful transparency objectives into the development lifecycle. At this stage, transparency considerations include:

- Explicitly designing system architectures for interpretability
- Embedding audibility and accountability considerations directly into system design
- Proactively identifying and mitigating potential biases
- Defining how transparency will be maintained throughout the model lifecycle

Proactive attention to these aspects during design positions organizations to build inherently transparent systems, thereby reducing later-stage risks related to accountability, compliance and stakeholder trust.

ii Transparency in AI operations

Embedding transparency into AI operations involves defining processes for communicating AI system outputs and decision roles. Organizations operationalize transparency by providing stakeholders, such as users, management and oversight bodies, with accessible information about AI's roles in organizational decisions, including how human oversight validates or reviews AI recommendations.

Operational transparency also involves model confidence levels, openly acknowledging model limitations and concisely explaining uncertainties to inform effective decision-making.

Organizations can systematically integrate transparency into routine operations, such as structured communication processes, standardized reporting formats and a consistent feedback loop, to enhance stakeholder understanding, support regulatory alignment and foster trust.

iii Transparency in AI serving

Integrating transparency into AI model serving involves communicating the reasoning behind AI-generated outputs directly at the point of user interaction. At this stage, transparency entails providing stakeholders with real-time or near-real-time explanations suitable to their specific context, level of expertise and decision needs.

Effective transparency practices during model serving include using explainability techniques (e.g., LIME, SHAP and counterfactual explanations) to provide an intuitive interface for users to query or challenge AI outputs, and disclosing the AI model's level of confidence, known limitations or potential biases in its recommendations. This supports informed decision-making, strengthens trust and enhances user acceptance precisely when stakeholders critically engage with AI.

iv Challenges and trade-offs in AI transparency

Transparency strategies should consider strategically balancing openness with necessary protections for proprietary methods and competitive advantages, employing techniques such as phased transparency initiatives, partial disclosures and tailored confidentiality protections. Organizations can manage these constraints by building stakeholder trust and enhancing credibility, such as partial disclosures, confidentiality protections or phased transparency initiatives, without inadvertently exposing sensitive IP or competitive advantages.

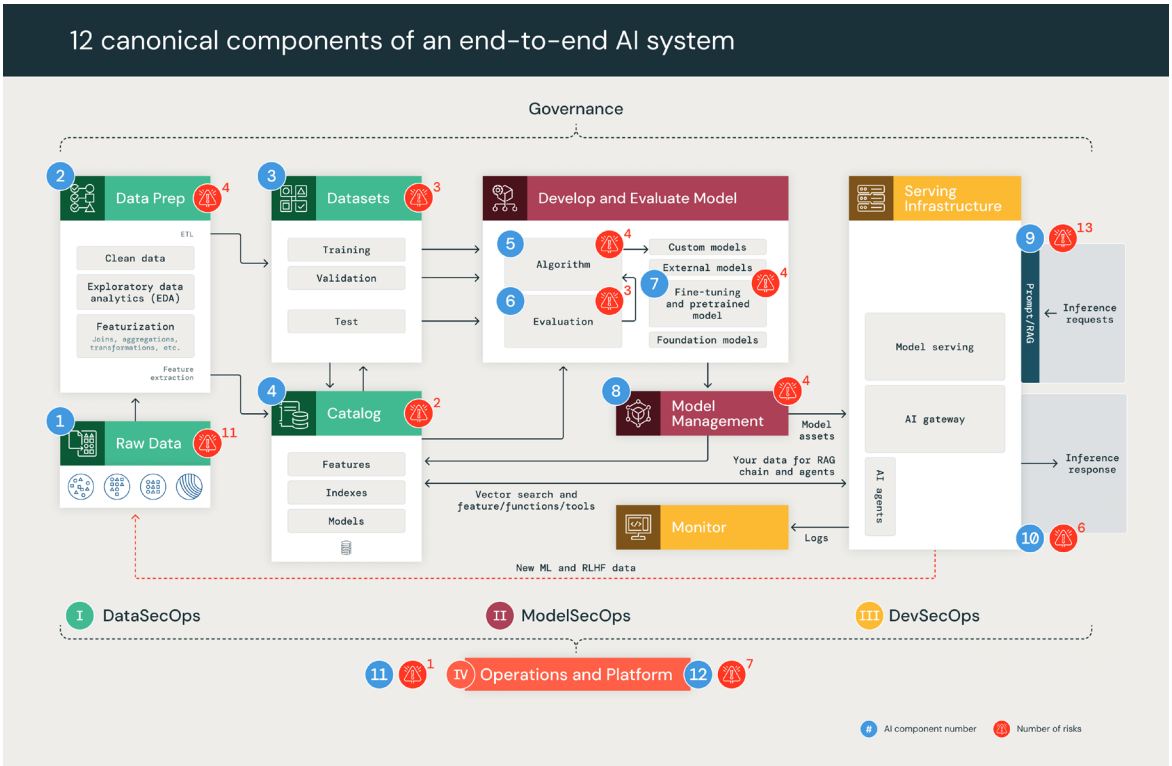
Organizations inevitably face trade-offs between leveraging 'black-box' or complex AI models that may provide improved accuracy and adopting simpler, more interpretable models. Leaders should thoughtfully evaluate these trade-offs in alignment with organizational goals, stakeholder expectations and regulatory requirements.

Finally, excessive transparency carries risks, including exposing systems to vulnerabilities or enabling exploitation by adversaries. For example, detailed disclosures about fraud detection criteria could enable system manipulation. Organizations may consider practices that maintain openness without compromising security and system integrity, such as segmented information disclosures or tailored transparency guidelines.

Foreword
Introduction
AI Organizations
Legal and Regulatory Compliance
Ethics, Transparency and Interpretability
Data, AIOps and Infrastructure
AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Data, AIOps and Infrastructure

Over the past decade, leading organizations have increasingly leveraged AI to enhance business operations, drive innovation and strengthen their competitive position in traditional and evolving markets. Rather than simply augmenting existing services and operations, AI enables the creation of entirely new systems and business models that surpass the capabilities of traditional approaches. These advancements allow organizations to automate intelligent workflows, optimize decision-making at scale and dynamically adapt to shifting market conditions — fundamentally redefining how they compete and deliver value.



The Databricks AI development lifecycle, courtesy of the DASF v2.0.

Achieving these outcomes is not simple. AI program success depends on multiple factors, including high-quality data, robust algorithm design, scalable and reliable infrastructure, effective governance and continuous adaptability through human oversight, monitoring and dynamic feedback loops.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Organizations that have successfully developed mature AI programs often establish governance frameworks that balance innovation, productivity, accountability, transparency and risk throughout the entire AI lifecycle. Based on experience Databricks has gained from leading AI programs, success hinges on focusing on four key areas of AI maturity.

- **Data operations and integrity:** The foundation of AI reliability, encompassing data ingestion, transformation and serving. Data influences every stage of the AI lifecycle and affects the health of ML features, including accuracy, bias mitigation and compliance.
- **Model architecture, development and oversight:** Encompasses the iterative design, training, evaluation and preparation of production models. This stage aligns development with business objectives and defines the AI's capabilities, such as transparency, explainability, interpretability and constraints.
- **AI lifecycle and AIOps:** Focuses on monitoring, scaling and automating the performance of AI systems. It includes MLOps, DevSecOps and continuous optimization to maintain AI effectiveness in production.
- **Infrastructure and platform resilience:** The technical foundation of the AI lifecycle, supporting AI scalability, security and adaptability across enterprise environments.

Most organizations do not deploy the full AI development lifecycle but adopt the components that align with their strategic and operational needs. These components should be thoughtfully designed to support mature, reliable, scalable and responsible AI capabilities across data integrity, model development, AI lifecycle operations and infrastructure resilience. This approach fosters innovation while maintaining accountability, transparency and risk awareness.

1 Data

AI systems are inherently data driven, especially when powered by LLMs and ML models. Their capabilities, adaptability and risks are shaped by the quality, representativeness and evolution of data throughout its lifecycle. Unlike traditional software that follows predefined instructions, AI systems can continuously refine their understanding, adjusting to changing conditions through ongoing data interactions. The role of data extends beyond model training — it dictates how AI systems generalize, respond and sustain long-term effectiveness.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Section summary

Integrate AI requirements into data governance programs:

- **Data classification:** Classify data based on sensitivity. Exclude or anonymize sensitive information in training data.
- **Data handling standards:** Implement protocols for accurate, unbiased, secure data sourcing, usage and storage.
- **Data catalog:** Organize and document datasets with metadata. Enable secure stakeholder access.
- **Data lineage:** Track preprocessing steps and transformations for transparency and accountability.
- **Data processing:** Set standards for consistent, accurate and secure data preparation and transformation.
- **Data quality:** Monitor and maintain data quality with automated checks and anomaly detection.
- **Model training and evaluation:** Use high-quality data with bias mitigation techniques and robust metrics for training and evaluation.
- **Model serving and use:** Set guidelines for responsible, compliant data use in internal and external models.
- **Data access and entitlement:** Apply fine-grained, role-based access controls for secure and compliant data management.

The way that AI systems use data differs vastly across the AI lifecycle depending on their type, function, objectives and operational constraints. This creates high complexity for governance programs, creating the potential for governance to impact the goals of AI programs.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

i Data systems

AI systems use data differently depending on their function, objectives and operational constraints.

For example:

- Real-time AI systems (e.g., fraud detection and recommendation engines) require low-latency data streaming and robust monitoring for drift detection.
- Batch-driven AI systems (e.g., predictive maintenance and churn modeling) rely on scalable storage, efficient data pipelines and scheduled retraining workflows.
- Regulated AI systems (e.g., healthcare, finance and legal AI) must embed data governance, compliance and explainability frameworks from inception.
- Adaptive AI systems (e.g., continuous learning models and reinforcement learning) demand active feedback loops and automated retraining mechanisms.

ii Data classification standards

Effective data classification supports transparency, accountability and ethical use, enhancing AI program overall quality and reliability. Establish robust data classification standards that address the unique requirements of AI programs at different lifecycle stages.

Getting started

Data classification considerations for AI datasets include:

1. **Data categories:** Define detailed data categories and subcategories relevant to AI models, considering specific features and variables, such as granularity, diversity and context.
2. **Privacy:** Implement privacy-preserving techniques, like data anonymization and differential privacy, to protect individual privacy and IP while maintaining data utility.
3. **Compliance:** Tag and align classifications with relevant regulations (e.g., GDPR, CCPA, HIPA and FERPA).
4. **Data provenance and lineage:** Document data sources, preprocessing steps, transformations and collection methodologies to ensure transparency and accountability.
5. **Data diversity and representativeness:** Classify data using categories that support diversity (e.g., demographic, geographic and behavioral attributes). Regularly review and update classifications to include underrepresented groups.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

6. **Data lifespan and retention:** Define retention periods for different data types (e.g., data, features and models) to ensure compliance with legal requirements.
7. **Ethical and responsible use:** Classify data based on potential misuse or harm. Establish guidelines for ethical data use, including considerations for fairness, accountability and transparency.
8. **Assess data sensitivity:** Evaluate data sensitivity in AI initiatives (model preparation, training, inference). Identify sensitive or personally identifiable information.
9. **Establish classification criteria:** Define clear criteria for data sensitivity levels (public, internal, confidential and highly confidential), considering privacy regulations, ownership and potential misuse impact.
10. **Document decisions:** Maintain thorough documentation of classification decisions and rationale for transparency and future audits.
11. **Review and update:** Regularly review and update classifications to reflect data use, sensitivity or regulation changes.

How Databricks can help

- **Data classification, diversity and representative tagging:** Databricks provides comprehensive data governance through Unity Catalog, which helps define detailed data categories and subcategories relevant to AI models. Unity Catalog supports tagging and annotating datasets with specific features and variables, allowing for granular classification based on AI needs, such as granularity, diversity and context. This aids in better data organization and retrieval, which is crucial for AI model training and analysis (Databricks documentation).
- **Compliance:** Databricks helps customers align with major data privacy regulations, like GDPR, CCPA and HIPAA, through fine-grained access controls, centralized governance, auditability and data encryption (Databricks).
- **Data lineage:** Unity Catalog integrated data lineage features capture data preprocessing steps, transformations and feature engineering processes. This transparency aids in troubleshooting and ensures accountability, which is vital for managing AI data workflows and maintaining data integrity across different stages (Databricks, Databricks documentation).

- **Data versioning:** The unified data platform from Databricks supports defining and managing data retention policies. Users can specify retention periods for different data types, ensuring compliance with legal requirements. Databricks also supports automated data archiving and deletion processes to manage the data lifecycle effectively (Databricks documentation).
- **Responsible AI:** Databricks promotes ethical data use by providing robust governance frameworks that classify data based on potential misuse or harm. The platform incorporates comprehensive security measures, access controls and audit capabilities to ensure data usage aligns with ethical standards and organizational policies. This includes considerations for fairness, accountability and transparency in AI development and deployment (Databricks).

iii Data handling standards

Define and enforce data handling standards that ensure reliable, secure and ethical data use in AI programs. Data classification processes must be integrated to associate datasets with appropriate usage policies, ensuring alignment with privacy and security requirements.

Getting started

Data handling standards for AI programs considerations include:

1. **Access and entitlement:** Implement access control measures to ensure security and privacy requirements compliance.
2. **Data sourcing, usage and storage:** Enforce strict data sourcing and storage protocols to ensure compliance with regulations.
3. **Data catalog:** Maintain a detailed catalog of datasets, including sources, usage rights and quality metrics.
4. **Data lineage:** Track data preprocessing steps and maintain transparency in all data transformations.
5. **Data processing:** Define clear standards for data processing to ensure consistency and privacy.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

6. **Data quality:** Continuously monitor data quality and address issues proactively.
7. **Model training and evaluation:** Establish standards for model training, ensuring unbiased, high-quality data use.
8. **Model serving and use:** To ensure regulatory compliance, define acceptable data use for internal and external models.

Foreword

Introduction

Pillar I: AI Organizations

Pillar II: Legal and Regulatory Compliance

Pillar III: Ethics, Transparency and Interpretability

Pillar IV: Data, AIOps and Infrastructure

Pillar V: AI Security

References and Further Reading

Acknowledgements

Appendix A: Glossary

License

2 AIOps

AI programs, while powerful, can become inefficient and untenable, leading to poor outcomes when inadequately managed. AIOps provides a structured approach to managing the development, deployment and maintenance of AI initiatives — ensuring these programs are scalable, effective and reliable. This section outlines the critical components and best practices for successfully implementing AIOps and achieving consistent, high-quality results.

Section summary

Integrate AI requirements into data governance programs:

- **AIOps overview:** AIOps streamlines the development, deployment and maintenance of AI initiatives to ensure efficient, scalable and reliable AI programs. AIOps components include a well-defined infrastructure, a comprehensive ML and AI lifecycle and best practices for effective AIOps.
- **AIOps infrastructure:** A well-defined infrastructure supports scalable, reliable and efficient AIOps. It includes the hardware, software and services required for successful AI program initiatives. Essential elements are computing resources, storage systems, networking and cloud services.
- **ML and AI lifecycle:** Establishing a comprehensive lifecycle is crucial for managing AI projects effectively. Key stages include:
  - **Raw data collection and storage:** Implement robust processes for data collection, access, integrity, quality and security.
  - **Data processing:** Clean, transform and prepare data for model training.
  - **Model pipelines:** Automate data processing, model training, evaluation and deployment.
  - **Model training:** Optimize model performance through appropriate algorithm selection and hyperparameter tuning.

Foreword
Introduction
AI Organizations
Legal and Regulatory Compliance
Ethics, Transparency and Interpretability
Data, AIOps and Infrastructure
AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

- **Model evaluation:** Assess model performance using relevant metrics and validation techniques. Verify model objectives, such as accuracy, transparency, consistency and explainability.
- **Model monitoring:** Track model performance and detect issues in production.
- **Experimentation:** Test and evaluate different model versions using techniques like A/B testing.
- **Model registry**: Store, version and manage ML models.
- **Dataset and feature management:** Manage datasets and features to ensure data consistency and quality.
- **ML metadata and artifact tracking:** Record and manage metadata and artifacts for reproducibility and compliance.
- **Model serving:** Deploy models into production for real-time predictions or batch inference.
- **Model gateway:** Manage the transition of models from development to production.

See the Big Book of MLOps.

- **Roadmap and infrastructure:** A technical roadmap for AIOps includes seamless adoption, implementation and continuous improvement. Infrastructure components such as computing resources, storage systems, networking and cloud services support scalable, efficient and reliable AIOps.

i AIOps and infrastructure

Well-defined infrastructure and operations support scalable, reliable and efficient AI programs, enabling organizations to achieve their strategic goals and maintain a competitive advantage. Identify the necessary infrastructure and AIOps services required for the successful deployment and execution of AI program initiatives.

a. AI architecture and roadmap

Define a target state architecture or technical roadmap for AIOps that provides a structured and strategic plan to integrate AI-driven solutions into IT operations. This roadmap should consider seamless adoption, implementation and continuous improvement of AIOps, fostering a proactive, efficient and resilient IT environment. Technical roadmaps can be integrated into target-state enterprise and solution architectures for projects.

b. ML and AI lifecycle

Establish a comprehensive ML and AI lifecycle to develop, deploy and manage AI programs effectively. This lifecycle should address all relevant stages of AI projects, ensuring that each step is systematically planned and executed. Lifecycle considerations include:

1. **Raw data collection and storage:** Implement robust processes for collecting, accessing, ensuring integrity, maintaining quality, preventing poisoning and securing and protecting the privacy of raw data to ensure its accuracy and completeness.

Best practices: Establish data collection and handling standards, secure storage solutions and implement data validation procedures.

How Databricks can help

- **Databricks Unity Catalog:** Databricks Unity Catalog provides a robust data governance platform, including features to manage and audit access to data and AI assets. This helps classify data and ensure compliance with privacy standards by enforcing fine-grained access controls and maintaining data integrity. Unity Catalog also supports data lineage tracking, crucial for understanding data transformations and maintaining data classification throughout the ML lifecycle (Databricks, Databricks documentation).

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

- **Secure data sourcing, usage and storage:** Databricks provides comprehensive governance and security features that help customers ensure data reliability, security and compliance. This includes Databricks Marketplace for secure data sourcing, encryption, access controls and audit logging, all managed centrally via Unity Catalog. Databricks also supports Delta Sharing, an open protocol for secure data sharing across organizations, ensuring data sourcing and usage remain compliant and ethical (Databricks, Databricks, Databricks documentation).
- **Databricks Auto Loader:** Databricks Auto Loader simplifies raw data ingestion by automatically detecting and processing new files in cloud storage. Unity Catalog helps manage data assets with fine-grained access controls and lineage tracking, ensuring data integrity and privacy (Databricks documentation, Databricks documentation).

2. **Data processing:** Clean, transform and prepare raw data for model training by addressing missing values, applying quality controls, normalizing data and performing feature engineering and augmentation.

Best practices: Establish data quality standards; automate procedures for data validation, usage and sharing; and use version control for datasets.

How Databricks can help

- **Databricks DLT** automates data preparation and transformation, ensuring consistency and accuracy. The platform supports various data processing tasks, from extract, transform, load (ETL) to complex transformations, all managed through a single interface (Databricks).
- **Feature store:** The Databricks Feature Store facilitates creating, managing and reusing ML features. It ensures that features used in training are consistent with those used in production, reducing online/offline skew and improving model performance (Databricks).

3. **Model pipelines:** Define interconnected steps to automate the end-to-end data processing, model training, evaluation and deployment processes. This will ensure reproducibility, scalability and efficiency.

Best practices: Design modular and reusable pipeline components and implement robust error handling, monitoring and service level agreements (SLAs to ensure pipeline stability and scalability.)

How Databricks can help

- **Databricks Workflows:** Databricks Workflows orchestrate complex data pipelines, integrating with Auto Loader and DLT for scalable and efficient data processing. The platform's managed service simplifies pipeline management and monitoring (Databricks, Databricks documentation).

Note: Based on their requirements, organizations can approach this step from an enterprise, department/line of business or initiative/project perspective.

4. **Model training:** Feed prepared data into ML algorithms to learn patterns and relationships. Optimize model performance by selecting appropriate algorithms, tuning hyperparameters and iterating configurations.

Best practices: Use automated hyperparameter tuning, track experiments and leverage distributed training for large datasets.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

How Databricks can help

- **Mosaic AI:** Databricks Mosaic AI provides unified tooling to build, deploy and monitor AI and ML solutions — from building predictive models to the latest GenAI and LLMs. Built on the Databricks Data Intelligence Platform, Mosaic AI enables organizations to securely and cost-effectively integrate their enterprise data into the AI lifecycle (Databricks).
- **Databricks run time:** Databricks supports distributed model training with scalable compute resources and integrated MLflow for experiment tracking. The platform also provides optimized machine learning libraries and frameworks like TensorFlow and PyTorch (Databricks documentation).

5. **Model evaluation:** Assess the performance of trained models using metrics and validation techniques relevant to business and AI program objectives. Verify model objectives, such as accuracy, transparency, consistency and explainability.

Best practices: Perform thorough cross-validation, use multiple evaluation metrics and assess model fairness and bias. Ensure models generalize well to unseen data and meet performance criteria before deployment.

How Databricks can help

- **Databricks Managed MLflow:** Managed MLflow extends the functionality of MLflow, an open source platform developed by Databricks for building better models and GenAI apps. It focuses on enterprise reliability, security and scalability (Databricks).
- **Databricks AutoML:** Databricks AutoML simplifies the evaluation process by automatically generating and assessing multiple baseline models. The platform integrates these evaluations into CI/CD pipelines for rapid iteration and deployment (Databricks).

6. **Model monitoring:** Implement monitoring tools to track model performance and behavior in production and detect issues like drift, bias or performance degradation.

Best practices: Set up alerts, dashboards and metrics to monitor key indicators such as accuracy, precision, recall and business-specific KPIs. Implement drift detection mechanisms to identify data, concept and label drift. Automate alerts to notify stakeholders of significant deviations in model performance. Schedule regular retraining based on performance thresholds or time intervals. Maintain comprehensive logs of model predictions, input data and performance metrics to facilitate audits and compliance checks.

How Databricks can help

- **Lakehouse Monitoring:** Databricks Lakehouse Monitoring offers comprehensive tools for tracking model performance and detecting anomalies. Integrated dashboards and automated alerts help maintain model accuracy and compliance (Databricks).

7. **Experimentation:** Thoroughly test and evaluate different model versions using techniques like A/B testing, multi-armed bandits and canary releases to identify the best-performing models under real-world conditions.

Best practices: Conduct A/B testing by randomly assigning users to different model variants to compare performance metrics and identify the best model.

- Use multi-armed bandits to allocate more traffic to better-performing models, optimizing outcomes over time.
- Implement canary releases to gradually roll out new models to a subset of users, monitoring performance and mitigating risks before full-scale deployment.
- Perform shadow testing by running new models in parallel with existing ones without affecting end users, allowing for performance comparison in a live environment.

Document experimentation processes to ensure reproducibility and facilitate knowledge sharing.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

How Databricks can help

- **MLflow and Model Serving automation:** Databricks Model Serving with Managed MLflow simplifies A/B testing and model quality monitoring in live production environments, allowing for efficient experimentation and performance optimization (Databricks).

8. **Model registry:** Adopt a model registry to store, version and manage ML models, to help enhance collaboration, reproducibility and governance by tracking model metadata, lineage and deployment status.

Best practices: Implement version control for models, track metadata and manage model lifecycle stages effectively.

How Databricks can help

- **MLflow Model Registry:** The MLflow Model Registry on Databricks centralizes model management, tracking versions and metadata to ensure reproducibility and governance. It supports stage transitions and comprehensive lineage tracking (Databricks documentation).

9. **Dataset and feature management:** Use repositories for tracking, storing and managing datasets and engineered features to ensure consistent access to high-quality data, promote feature reuse and support data versioning and lineage tracking.

Best practices: Use data versioning, maintain consistency and promote team feature reuse.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

How Databricks can help

- **Databricks Online Tables:** Databricks Online Tables facilitate the creation, management and reuse of ML features, ensuring consistency between training and production environments (Databricks documentation).

10. **ML metadata and artifact tracking:** Record and manage metadata and artifacts generated during the ML lifecycle to support reproducibility, accountability and compliance, and to facilitate better team collaboration.

Best practices: Track all relevant metadata, maintain comprehensive logs and ensure transparency and reproducibility.

How Databricks can help

- **MLflow and Unity Catalog:** Databricks Unity Catalog offers a unified governance layer for data and AI assets, providing detailed lineage tracking and metadata management to support compliance and collaboration (Databricks documentation).

11. **Model serving:** Deploy approved models into production environments for real-time predictions or batch inference. Set up infrastructure to handle inference requests and ensure scalability, low latency, cost-effectiveness and high availability.

Best practices: Implement robust API management, use containerization for portability and adopt deployment strategies like canary releases.

How Databricks can help

- **Model serving:** Databricks Model Serving provides a unified service for deploying and monitoring models, ensuring low latency and high availability. The platform supports containerization and scalable deployment strategies (Databricks).

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

c. Infrastructure

Infrastructure refers to the hardware and software resources required to support the entire ML lifecycle, from development to deployment and ongoing operations. This includes computing resources (CPUs and GPUs), storage systems, networking components and cloud services. A well-architected infrastructure can ensure ML workflow scalability, performance and reliability.

Getting started

Ensure that AI and ML projects are managed efficiently and effectively from development through deployment and ongoing maintenance. Consider monitoring and measurement strategies for the following key components.

1. **Scalability:** Enables the seamless scaling of computing resources to handle varying workloads and large datasets.
2. **Performance:** Ensures AI tasks are executed efficiently, from data processing to model inference.
3. **Reliability:** Provides a stable environment that minimizes downtime and supports high availability.
4. **Security:** Incorporates measures to protect data and models from unauthorized access and breaches.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

How Databricks can help

- **Simplified workflows:** Databricks Workflows is a managed orchestration service fully integrated with the Databricks Data Intelligence Platform. It lets you easily define, manage and monitor multitask workflows for ETL, analytics and ML pipelines. With a wide range of supported task types, deep observability capabilities and high reliability, your data teams can better automate and orchestrate any pipeline and become more productive (Databricks).
- **Databricks jobs and DLT:** DLT simplifies the creation and management of data pipelines and automates data preparation and transformation enforcement, helping to ensure consistency, accuracy, privacy and security (Databricks).
- **Automated model training and artifact tracking:** Databricks supports scalable model training with distributed computing capabilities. You can use MLflow, an open source platform integrated with Databricks, to manage the complete machine learning lifecycle, including model versioning and experiment tracking, ensuring reproducibility and transparency. Databricks ML Runtime also offers optimized machine learning libraries and frameworks like TensorFlow, PyTorch and XGBoost. It provides a scalable environment for training models on large datasets, leveraging the power of distributed computing on Apache Spark™ (Databricks, Databricks).
- **Automated ML, AI and GenAI development:** Databricks AutoML simplifies the model evaluation process by automatically generating and evaluating multiple baseline models using preset metrics. AutoML experiments can be auto-triggered when an existing model is identified as inefficient and they can integrate into CI/CD pipelines for fast time-to-production lifecycles (Databricks).
- **Model serving:** Databricks Model Serving is a unified service for deploying, governing, querying and monitoring models fine-tuned or predeployed by Databricks, like Meta Llama 3, DBRX or BGE, or from other model providers like Azure OpenAI, AWS Bedrock, AWS SageMaker and Anthropic. Model Serving also has predeployed models such as Llama 2 70B, allowing you to jump-start developing AI applications like retrieval augmented generation (RAG) and provide pay-per-token access or pay-for-provisioned compute for throughput guarantees (Databricks).
- **A/B Testing:** Databricks Model Serving with Managed MLflow and Lakehouse Monitoring simplifies A/B testing of different models and monitors model quality on live production data. Our unified approach makes it easy to experiment with cloud or provider production models to find the best candidate for your real-time application. Once deployed, you can do A/B testing of different models and monitor model quality on live production data (Databricks, Databricks, Databricks).

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

- **Monitoring:** Databricks Lakehouse Monitoring allows teams to monitor their entire data pipelines — from data and features to ML models — without additional tools and complexity. Powered by Unity Catalog, it lets users uniquely ensure that their data and AI assets are high-quality, accurate and reliable through deep insight into their data and AI asset lineage. The single, unified approach to monitoring enabled by lakehouse architecture makes diagnosing errors, performing root cause analysis and finding solutions simple (Databricks).
- **Model registry:** The MLflow Model Registry in Databricks provides a centralized repository for managing model versions. It tracks model lineage, supports stage transitions (e.g., staging to production) and ensures that the best models are deployed in production environments (Databricks).
- **Feature Store:** The Databricks Feature Store facilitates creating, managing and reusing ML features. It ensures that features used in training are consistent with those used in production, reducing online/offline skew and improving model performance (Databricks).
- **Comprehensive security and governance:** Databricks provides robust security and governance features, including the Unity Catalog, which offers fine-grained access controls and data lineage tracking. This ensures your data is secure and compliant with regulatory requirements, supporting enterprise-grade deployments (Databricks, Databricks).
- **Scalable and cost-efficient infrastructure:** Databricks provides a highly scalable infrastructure that can handle workloads ranging from gigabytes to petabytes. By using the Databricks Data Intelligence Platform, you can achieve better cost-efficiency than with other cloud data warehouses. The platform's next-generation vectorized query engine, Photon, delivers a price to performance ratio up to 12 times better, allowing you to optimize your infrastructure costs while maintaining high performance (Databricks documentation).
- **Multicloud flexibility:** The Databricks Platform supports deployment across multiple cloud providers, including AWS, Azure and Google Cloud. This multicloud capability ensures you are not locked into a single vendor, allowing you to adapt your infrastructure as your business grows and evolves. This is particularly beneficial for maintaining infrastructure flexibility and avoiding vendor lock-in (Databricks, Databricks documentation).
- **Unified data and AI platform:** Databricks integrates various data management and processing tools into a unified platform. This includes support for ETL processes via DLT, data warehousing via Databricks SQL, ML and real-time analytics. Lakehouse architecture combines the strengths of data lakes and data warehouses, allowing you to store all your data in one place and access it seamlessly for different use cases (Databricks, Databricks, Databricks, Databricks, Databricks).

- **Managed open source integrations:** Databricks is strongly committed to open source technologies, managing integrations and updates within the Databricks Runtime. This includes projects like Delta Lake, MLflow and Apache Spark™, which are critical for building and managing robust data and AI infrastructure (Databricks documentation, Databricks documentation).
- **MLflow AI Gateway:** MLflow AI Gateway enables organizations to centrally manage credentials for SaaS models or model APIs and provide access-controlled routes for querying. MLflow AI Gateway will also enable prediction caching to track repeated prompts and rate limiting to manage costs (Databricks, MLflow documentation).
- **Databricks Clean Rooms:** Databricks offers a robust clean room solution that supports secure data collaboration across multiple clouds and regions without requiring data movement. Powered by Delta Sharing, it ensures data privacy and integrity while enabling seamless collaboration. The platform supports various data formats and workloads, providing flexibility to run complex computations, including ML tasks, in a privacy-safe manner (Databricks, Databricks).

Foreword
Introduction
AI Organizations
Legal and Regulatory Compliance
Ethics, Transparency and Interpretability
Data, AIOps and Infrastructure
AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

AI Security

Databricks has published the DASF, which provides a comprehensive guide to understanding and mitigating security risks associated with AI and ML systems. It is designed to assist organizations in securely adopting AI technologies while managing potential risks like data breaches and regulatory compliance issues. The DASF mapped AI risks and controls to 10 industry standards and frameworks, and goes into far more detail than is in the scope of this framework, so a summary of the DASF is described below. Please review the whitepaper for the complete threat model, associated risks and appropriate controls.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Section summary

- **Raw data:** Implement robust access controls, ensure data quality and integrity checks, use modern encryption practices and protect against data manipulation.
- **Data preparation:** Ensure integrity in preprocessing steps, secure feature engineering processes and safeguard data selection criteria to prevent adversarial inputs.
- **Datasets:** Obtain secure training data from reliable sources, use secure storage and encryption and maintain label integrity to ensure model accuracy.
- **Data catalog governance:** Ensure traceability and transparency of data and model assets and manage the AI lifecycle effectively to maintain security and compliance.
- **Machine learning algorithms:** Thoroughly document experiments, monitor changes in data and performance and protect the confidentiality of hyperparameters.
- **Evaluation:** Secure evaluation data and use comprehensive, representative datasets to avoid biases and inaccuracies.
- **ML models:** Rigorously verify models to prevent malicious alterations and implement controls to prevent unauthorized access and theft of models.
- **Model management:** Secure the entire lifecycle of AI models and protect against inversion attacks to prevent sensitive data exposure.

- **Model serving and inference requests:** Secure inputs to prevent malicious data injection and ensure requests are authorized and secure.
- **Model serving and inference responses:** Secure outputs to prevent manipulation and protect against attacks that infer sensitive information from outputs.
- **MLOps:** Standardize operations for security, regularly assess vulnerabilities and apply timely patches to maintain a secure environment.
- **Data and AI platform security:** Implement robust access control mechanisms, continuously monitor for threats and maintain effective incident response capabilities to ensure security.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

1 Raw data

The DASF emphasizes the importance of securing raw data. Implementing robust access controls (DASF 5) is essential to prevent unauthorized access and mitigate significant security risks. Ensuring high data quality and performing integrity checks (DASF 7) are crucial to avoid compromising AI models. The framework also advocates adopting modern storage and encryption practices (DASF 8, DASF 9) to reduce vulnerability to breaches. Additionally, protecting raw data from manipulation by attackers (DASF 1, DASF 7) is necessary to maintain the integrity of AI systems.

2 Data preparation

In the data preparation phase, the framework underscores the importance of maintaining the integrity of data preprocessing steps (DASF 1, DASF 2, DASF 3, DASF 4, DASF 7) to prevent tampering. Securing feature engineering processes (DASF 16, DASF 42) is crucial to avoid introducing biased or malicious inputs into the system. Establishing and safeguarding criteria for selecting raw data (DASF 1, DASF 3, DASF 4) is also essential to prevent adversarial inputs that could compromise the AI system's functionality.

3 Datasets

When managing datasets, the framework advises obtaining secure training data from reliable sources (DASF 1, DASF 2, DASF 3, DASF 4, DASF 7, DASF 11) to avoid malicious manipulation. Secure storage and encryption practices (DASF 8, DASF 9) are recommended to protect the datasets from unauthorized access. The framework also stresses the importance of maintaining label integrity (DASF 8, DASF 9, DASF 5) to ensure that the models remain accurate and reliable.

4 Data catalog governance

The framework highlights the significance of data catalog governance by ensuring the traceability and transparency of data and model assets (DASF 5, DASF 11, DASF 16, DASF 17). This approach supports both security and compliance efforts. Effective management of the AI lifecycle (DASF 19, DASF 21) is critical to mitigate risks and ensure that security standards are consistently met throughout the development and deployment of AI systems.

5 Machine learning algorithms

For ML algorithms, the framework recommends thorough documentation of experiments (DASF 7, DASF 45) to ensure reproducibility and detect potential security issues early. Monitoring for changes in data or model performance (DASF 21) is essential for identifying security vulnerabilities. The framework also emphasizes the importance of protecting the confidentiality and integrity of hyperparameters (DASF 5) to safeguard the proprietary aspects of AI models.

6 Evaluation

The framework advises securing evaluation data (DASF 6.1) during the evaluation phase to prevent manipulation and ensure reliable model validation. Using comprehensive and representative evaluation datasets (DASF 6.2) is crucial to avoid biases and inaccuracies, which could otherwise lead to flawed AI models.

7 Machine learning models

The framework stresses the need for rigorous verification of ML models (DASF 7.1) to ensure they are free from malicious alterations, thus protecting against backdoor and Trojan models. Implementing controls to prevent unauthorized access and theft of models (DASF 5, DASF 24, DASF 30) is equally essential to maintain the security and integrity of AI systems.

8 Model management

In the model management phase, the framework recommends securing the entire lifecycle of AI models (DASF 29) from development to deployment. Protecting models from inversion attacks (DASF 24, DASF 31, DASF 33) is particularly important to prevent the disclosure of sensitive information that malicious actors could exploit.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

9 Model serving and inference requests

The framework emphasizes the importance of securing inputs for model serving and inference requests to prevent the injection of malicious data (DASF 1, DASF 3). Ensuring that requests to models are secure and authorized (DASF 4) is crucial for maintaining the integrity of AI systems and preventing unauthorized use or manipulation.

10 Model serving and inference response

Securing the outputs of AI models (DASF 10.3) is necessary to prevent manipulation and to ensure their accuracy. The framework also highlights the importance of protecting against attacks that exploit model outputs to infer sensitive information (DASF 10.4), which could compromise the security of the AI system.

11 Machine learning operations

In MLOps, the framework advocates for implementing standardized operations (DASF 42) to enhance security and consistency across the board. Regular vulnerability assessments and timely patching (DASF 21) are recommended to secure the AI environment from potential threats and maintain the overall security posture.

12 Data and AI platform security

The framework concludes by emphasizing the need for robust access control mechanisms (DASF 5) to protect data and AI assets. Continuous monitoring and effective incident response capabilities (DASF 21, DASF 36) are essential for detecting and mitigating security threats, ensuring the AI platform remains secure and resilient against potential attacks.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

References and Further Reading

Industry best-practice references

- NIST AI Risk Management Framework →
- ISO/IEC 42001:2023 →
- ISO/IEC 38507:2022 →
- Guidelines for Secure AI System Development →
- AI Playbook for UK Government →
- The EU AI Act →
- Model AI Governance Framework (Singapore) →
- OWASP AI Exchange →
- UNESCO AI Ethics and Governance Observatory →
- NIST Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations →
- Secure by Design — Shifting the Balance of Cybersecurity Risk: Principles and Approaches for Secure by Design Software →
- MITRE ATLAS Adversarial ML →
- OWASP Top 10 for LLM Applications Cybersecurity and Governance →

AI and machine learning on Databricks (Mosaic AI)

Training course: Generative AI Fundamentals →

Web page: AI and Machine Learning on Databricks →

Industry solutions: Solution Accelerators →

Blogs: AI/ML Blogs →

eBooks: Big Book of Generative AI → | Big Book of MLOps: 2nd Edition → | Data, Analytics and AI Governance →

Learning library: Generative AI Engineering With Databricks →

DASF video: Introducing the Databricks AI Security Framework (DASF) to Manage AI Security Risks →

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Databricks Unity Catalog

Web pages: Databricks Unity Catalog → | AI Governance →

eBook: Data and AI Governance →

Blogs: What's New in Unity Catalog → | Open Sourcing Unity Catalog → | Row and Column Level Security →

Databricks Platform security

Review the security features in the Security and Trust Center, along with the overall documentation about the Databricks security and compliance programs.

The Security and Trust Overview Whitepaper provides an outline of the Databricks architecture and platform security practices.

Databricks Platform Security Best Practices | AWS | Azure | GCP

Security Analysis Tool (SAT)

Databricks Security and Trust Blog

Databricks Security Best Practices YouTube channel

Databricks Delta UniForm

Web page: Delta Lake UniForm →

eBook: O'Reilly Delta Lake: The Definitive Guide →

Blog: General Availability of Delta Lake UniForm →

Responsible AI on the Databricks Data Intelligence Platform

Web page: Responsible AI →

Blogs: Responsible AI with the Databricks Data Intelligence Platform →
Helping Enterprises Responsibly Deploy AI →
Partnerships With Industry and Government Organizations →
AI Regulations and the Databricks Data Intelligence Platform →

The information in this document does not constitute or imply endorsement or recommendation of any third-party organization, product or service by Databricks. Links and references to websites and third-party materials are provided for informational purposes only and do not represent endorsement or recommendation of such resources over others.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Acknowledgements

The authors would like to thank the following contributors and reviewers, whose invaluable contributions made this framework possible.

Special Contributors

DATABRICKS



Omar Khawaja
Vice President, Field CISO



Arun Pamulapati
Senior Staff Security Field Engineer



Kelly Albano
Product Marketing Manager



Robin Sutara
Field Chief Technology Officer



Lexy Kassan
Lead Data and AI Strategist



Suchismita Pahi
Lead Counsel, Product



Sachin Thakur
Principal Product Marketing Manager

Implementation Partners

TRUSTIBLE

Gerald Kierce-Iturrioz,
CEO

Andrew Gamino-Cheong,
CTO

Reviewers

GLEAN

Sunil Agrawal,
CISO

CAPITAL ONE FINANCIAL

Stephen Jou,
Sr. Director and Head of Security of Enterprise Data

META

Craig Thiesen,
Privacy Specialist

GRAMMARLY

Alan Luk,
Head of GRC

Nicholas Ray,
Program Leader GRC

NETFLIX

Mosi Platt,
Compliance Engineering

McKenna Rivera-Yeakey,
Security Engineering

SCHELLMAN

Danny Manimbo,
Principal

COMPLYLEFT

Ben Johns,
Cybersecurity Specialist

FAIR INSTITUTE

Jacqueline Lebo,
Risk Advisory Manager

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Appendix A: Glossary

A

ACID transaction: A set of properties — Atomicity, Consistency, Isolation and Durability — that ensure reliable processing of database transactions.

Adversarial examples: Modified testing samples that induce misclassification of a machine learning model at deployment time.

Agentic AI: Autonomous artificial intelligence systems that can make decisions and take actions independently to achieve specific goals. These systems utilize learning capabilities to adapt to their environments and operate with a level of self-direction, reducing the need for constant human intervention.

AI agent: While the industry is refining the definition of AI agents, generally, an AI agent is an application capable of making decisions based on data, learning from experience and adapting to new situations over time. Unlike traditional rule-based systems like robotic process automation (RPA), AI agents can manage structured, unstructured and other data types, analyze them and make informed decisions in dynamic or uncertain environments.

AI augmentation: The use of artificial intelligence to enhance human capabilities and decision-making rather than replacing them. By automating routine tasks and providing data-driven insights, AI augmentation allows individuals to focus on more strategic activities, improving efficiency and overall outcomes.

AI gateway: A centralized system that acts as an intermediary between users or applications and various artificial intelligence services, facilitating access through standardized APIs. It manages data routing, ensures security and access control and provides monitoring and analytics to optimize the use of AI capabilities.

AI governance: The actions to ensure stakeholder needs, conditions and options are evaluated to determine balanced, agreed-upon enterprise objectives; setting direction through prioritization and decision-making; and monitoring performance and compliance against agreed-upon directions and objectives. AI governance may include policies on the nature of AI applications developed and deployed versus those limited or withheld.

AI guardrail: A safeguard that is put in place to prevent artificial intelligence from causing harm. AI guardrails are a lot like highway guardrails – they are both created to keep people safe and guide positive outcomes.

AI red teaming: The practice of rigorously testing AI systems by simulating adversarial attacks and identifying vulnerabilities. This approach helps organizations improve the security, robustness and ethical safeguards of their AI models by assessing how they respond to malicious or unexpected behaviors.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

Artificial intelligence (AI): A multidisciplinary field of computer science that aims to create systems capable of emulating and surpassing human-level intelligence.

Attribute based access control (ABAC): A security model that grants or denies access to resources based on the attributes of users, resources and the environment. This approach allows for fine-grained access control by evaluating various characteristics such as user roles, resource classifications and contextual factors to determine permissions dynamically.

Autonomous agents: AI systems capable of performing tasks or making decisions independently, without human intervention, by perceiving their environment and acting based on predefined goals or learned behaviors.

B

Bug bounty program: A program that offers monetary rewards to ethical hackers for successfully discovering and reporting a vulnerability or bug to the application's developer. Bug bounty programs allow companies to leverage the hacker community to improve their systems' security posture over time.

C

Compound AI system: An advanced artificial intelligence architecture that integrates multiple AI models or components to achieve more complex and capable functionalities than a single model can provide.

Compute plane: Where your data is processed in Databricks Platform architecture.

Concept drift: A situation where statistical properties of the target variable change and the very concept of what you are trying to predict changes as well. For example, the definition of what is considered a fraudulent transaction could change over time as new ways are developed to conduct such illegal transactions. This type of change will result in concept drift.

Continuous integration and continuous delivery/continuous deployment (CI/CD): CI is a modern software development practice in which incremental code changes are made frequently and reliably. CI/CD is common to software development, but it is becoming increasingly necessary to data engineering and data science. By automating the building, testing and deployment of code, development teams are able to deliver releases more frequently and reliably than with the manual processes still common to data engineering and data science teams.

Control plane: The back-end services that Databricks manages in your Databricks account. Notebook commands and many other workspace configurations are stored in the control plane and encrypted at rest.

D

Data classification: A crucial part of data governance that involves organizing and categorizing data based on its sensitivity, value and criticality.

Data clean room: A secure environment where multiple parties can share and analyze their data without directly exposing the underlying raw data to each other.

Foreword

Introduction

Pillar I:
AI Organizations

Pillar II:
Legal and Regulatory
Compliance

Pillar III:
Ethics, Transparency
and Interpretability

Pillar IV:
Data, AIOps and
Infrastructure

Pillar V:
AI Security

References and
Further Reading

Acknowledgements

Appendix A:
Glossary

License

Data drift: The features used to train a model are selected from the input data. When statistical properties of this input data change, it will have a downstream impact on the model's quality. For example, data changes due to seasonality, personal preference changes, trends, etc., will lead to incoming data drift.

Data governance: Data governance is a comprehensive approach that comprises the principles, practices and tools to manage an organization's data assets throughout their lifecycle. By aligning data-related requirements with business strategy, data governance provides superior data management, quality, visibility, security and compliance capabilities across the organization. Implementing an effective data governance strategy allows companies to make data easily available for data-driven decision-making while safeguarding their data from unauthorized access and ensuring compliance with regulatory requirements.

Data Intelligence Platform: A new era of data platform that employs AI models to deeply understand the semantics of enterprise data. It builds on the foundation of the data lakehouse — a unified system to query and manage all data across the enterprise — but automatically analyzes both the data (contents and metadata) and how it is used (queries, reports, lineage, etc.) to add new capabilities.

Data lake: A central location that holds a large amount of data in its native, raw format. Compared to a hierarchical data warehouse, which stores data in files or folders, a data lake uses a flat architecture and object storage to store the data. With object storage, data is stored with metadata tags and a unique identifier, which makes it easier to locate and retrieve data across regions and improves performance. By leveraging inexpensive object storage and open formats, data lakes enable many applications to take advantage of the data.

Data lakehouse: A new, open data management architecture that combines the flexibility, cost-efficiency and scale of data lakes with the data management and ACID transactions of data warehouses, enabling business intelligence (BI) and machine learning (ML) on all data.

Data lineage: A powerful tool that helps organizations ensure data quality and trustworthiness by providing a better understanding of data sources and consumption. It captures relevant metadata and events throughout the data's lifecycle, providing an end-to-end view of how data flows across an organization's data estate.

Data partitioning: A partition is composed of a subset of rows in a table that share the same value for a predefined subset of columns called the partitioning columns. Data partitioning can speed up queries against the table as well as data manipulation.

Data pipeline: A data pipeline implements the steps required to move data from source systems, transform that data based on requirements and store the data in a target system. A data pipeline includes all the processes necessary to turn raw data into prepared data that users can consume. For example, a data pipeline might prepare data so data analysts and data scientists can extract value from the data through analysis and reporting. An extract, transform and load (ETL) workflow is a common example of a data pipeline.

Data poisoning: Attacks in which a part of the training data is under the control of the adversary.

Data preparation (data prep): The set of preprocessing operations performed in the early stages of a data processing pipeline (i.e., data transformations at the structural and syntactical levels).

Data privacy: Attacks against machine learning models to extract sensitive information about training data.

Foreword

Introduction

Pillar I:
AI Organizations

Pillar II:
Legal and Regulatory
Compliance

Pillar III:
Ethics, Transparency
and Interpretability

Pillar IV:
Data, AIOps and
Infrastructure

Pillar V:
AI Security

References and
Further Reading

Acknowledgements

Appendix A:
Glossary

License

Data streaming: Data that is continuously and/or incrementally flowing from a variety of sources to a destination to be processed and analyzed in near real-time. This unlocks a new world of use cases around real-time ETL, real-time analytics, real-time ML and real-time operational applications that in turn enable faster decision-making.

Databricks Delta Live Tables: A declarative framework for building reliable, maintainable and testable data processing pipelines. You define the transformations to perform on your data and Delta Live Tables manages task orchestration, cluster management, monitoring, data quality and error handling.

Databricks Feature Store: A centralized repository that enables data scientists to find and share features and also ensures that the same code used to compute the feature values is used for model training and inference.

Databricks IQ: The data intelligence engine powering the Databricks Platform. It is a compound AI system that combines the use of AI models, retrieval, ranking and personalization systems to understand the semantics of your organization's data and usage patterns.

Databricks Secrets: Sometimes accessing data requires that you authenticate to external data sources through Java Database Connectivity (JDBC). Databricks Secrets stores your credentials so you can reference them in notebooks and jobs instead of directly entering your credentials into a notebook.

Databricks SQL: The collection of services that bring data warehousing capabilities and performance to your existing data lakes. Databricks SQL supports open formats and standard ANSI SQL. An in-platform SQL editor and dashboarding tools allow team members to collaborate with other Databricks users directly in the workspace. Databricks SQL also integrates with a variety of tools so that analysts can author queries and dashboards in their favorite environments without adjusting to a new platform.

Databricks Workflows: Orchestrates data processing, machine learning and analytics pipelines on the Databricks Data Intelligence Platform. Workflows has fully managed orchestration services integrated with the Databricks Platform, including Databricks Jobs to run noninteractive code in your Databricks workspace and Delta Live Tables to build reliable and maintainable ETL pipelines.

Datasets: A dataset in machine learning and artificial intelligence refers to a collection of data that is used to train and test algorithms and models.

Delta Lake: The optimized storage layer that provides the foundation for storing data and tables in the Databricks lakehouse. Delta Lake is open source software that extends Parquet data files with a file-based transaction log for ACID transactions and scalable metadata handling. Delta Lake is fully compatible with Apache Spark™ APIs, and was developed for tight integration with Structured Streaming, allowing you to easily use a single copy of data for both batch and streaming operations and providing incremental processing at scale.

Denial of service (DoS): An attack meant to shut down access to information systems, devices or other network resources, making them inaccessible to their intended users. DoS attacks accomplish this by flooding the target with traffic, or sending it information that triggers a crash. In both instances, the DoS attack deprives legitimate users (i.e., employees, members or account holders) of the service or resource they expected due to the actions of a malicious cyberthreat actor.

DevSecOps: Stands for development, security and operations. It's an approach to culture, automation and platform design that integrates security as a shared responsibility throughout the entire IT lifecycle.

Foreword

Introduction

Pillar I: AI Organizations

Pillar II: Legal and Regulatory Compliance

Pillar III: Ethics, Transparency and Interpretability

Pillar IV: Data, AIOps and Infrastructure

Pillar V: AI Security

References and Further Reading

Acknowledgements

Appendix A: Glossary

License

E

Egress control: Security measures and protocols designed to manage and monitor the exit of individuals and data from a secure environment.

Embeddings: Mathematical representations of the semantic content of data, typically text or image data. Embeddings are generated by a large language model and are a key component of many GenAI applications that depend on finding documents or images that are similar to each other. Examples are RAG systems, recommender systems and image and video recognition.

Explainable AI: Artificial intelligence systems designed to provide clear, understandable justifications for their predictions or decisions, making it easier for humans to interpret how and why the AI reached its outcomes.

Exploratory data analysis (EDA): Methods for exploring datasets to summarize their main characteristics and identify any problems with the data. Using statistical methods and visualizations, you can learn about a dataset to determine its readiness for analysis and inform what techniques to apply for data preparation. EDA can also influence which algorithms you choose to apply for training ML models.

External models: Third-party models hosted outside of Databricks. Supported by Model Serving, external models allow you to streamline the usage and management of various large language model (LLM) providers, such as OpenAI and Anthropic, within an organization.

Extract, transform and load (ETL): The foundational process in data engineering of combining data from multiple sources into a large, central repository called a data warehouse. ETL uses a set of business rules to clean and organize raw data and prepare it for storage, data analytics and machine learning (ML).

F

Feature engineering: The process of extracting features (characteristics, properties, attributes) from raw data to develop machine learning models.

Fine-tuned LLM: Adapting a pretrained LLM to specific datasets or domains.

Foundation model: A general purpose machine learning model trained on vast quantities of data and fine-tuned for more specific language understanding and generation tasks.

G

Generative: Type of machine learning methods that learn the data distribution and can generate new examples from distribution.

Generative AI: Also known as GenAI, this is a form of machine learning that uses large quantities of data to train models to produce content.

Foreword

Introduction

Pillar I: AI Organizations

Pillar II: Legal and Regulatory Compliance

Pillar III: Ethics, Transparency and Interpretability

Pillar IV: Data, AIOps and Infrastructure

Pillar V: AI Security

References and Further Reading

Acknowledgements

Appendix A: Glossary

License

H

Hallucination: A response generated by AI which contains false or misleading information presented as fact. This term draws a loose analogy with human psychology, where hallucination typically involves false perceptions. However, there is a key difference: AI hallucination is associated with erroneous responses or beliefs rather than perceptual experiences.

Hardened runtime: Databricks handles the actual base system image (e.g., AMI) by leveraging Ubuntu with a hardening configuration based on CIS. As a part of the Databricks Threat and Vulnerability Management program, we perform weekly scanning of the AMIs as they are making their way from dev to production.

Human-in-the-loop (HITL): The process of machine learning that allows people to validate a machine learning model's predictions as right or wrong at the time of training and inference with intervention.

Human-at-the-helm: Cyberattacks that exploit the reliance on human oversight in decision-making, often through misinformation, social engineering, or psychological pressure to mislead or manipulate the person in control.

Hybrid models: These models combine multiple methodologies or techniques, often integrating both traditional statistical approaches and modern machine learning methods, to leverage the strengths of each for improved performance and accuracy.

Hyperparameter: A parameter whose value is set before the machine learning process begins. In contrast, the values of other parameters are derived via training.

I

Identity provider (IdP): A service that stores and manages digital identities. Companies use these services to allow their employees or users to connect with the resources they need. They provide a way to manage access, adding or removing privileges, while security remains tight.

Incident response plan: A documented strategy outlining the processes and procedures an organization follows to prepare for, detect, respond to and recover from cybersecurity incidents.

Inference: The stage of ML in which a model is applied to a task by running data points into a machine learning model to calculate an output such as a single numerical score. For example, a classifier model produces the classification of a test sample.

Inference queries: Inputs sent to an AI model to generate predictions or outputs based on learned patterns and knowledge. In machine learning, inference queries are used after the model is trained, enabling it to make decisions or provide insights in real-world applications without further updating its internal parameters.

Inference tables: A table that automatically captures incoming requests and outgoing responses for a model serving endpoint and logs them as a table.

Initial access: Techniques that adversaries use with various entry vectors to gain their initial foothold within the system.

Insider risk: An insider is any person who has or had authorized access to or knowledge of an organization's resources, including personnel, facilities, information, equipment, networks and

Foreword

Introduction

Pillar I:
AI Organizations

Pillar II:
Legal and Regulatory
Compliance

Pillar III:
Ethics, Transparency
and Interpretability

Pillar IV:
Data, AIOps and
Infrastructure

Pillar V:
AI Security

References and
Further Reading

Acknowledgements

Appendix A:
Glossary

License

systems. Should an individual choose to act against the organization, with their privileged access and their extensive knowledge, they are well positioned to cause serious damage.

Interactive agents: AI systems designed to engage in dynamic, two-way communication with users, often using natural language processing and machine learning to understand and respond to queries or commands.

IP access list (IP ACL): Enables you to restrict access to your AI system based on a user's IP address. For example, you can configure IP access lists to allow users to connect only through existing corporate networks with a secure perimeter. If the internal VPN network is authorized, users who are remote or traveling can use the VPN to connect to the corporate network. If a user attempts to connect to the AI system from an insecure network, like from a coffee shop, access is blocked.

J

Jailbreaking: An attack that employs prompt injection to specifically circumvent the safety and moderation features placed on LLMs by their creators.

L

Label-flipping (LF) attacks: A targeted poisoning attack where the attackers poison their training data by flipping the labels of some examples from one class (i.e., the source class) to another (i.e., the target class).

Lakehouse Monitoring: Databricks Lakehouse Monitoring lets you monitor the statistical properties and quality of the data in all of the tables in your account. You can also use it to track the performance of machine learning models and model serving endpoints by monitoring inference tables that contain model inputs and predictions.

Large language model (LLM): A model trained on massive datasets to achieve advanced language processing capabilities based on deep learning neural networks.

Liquid clustering: A dynamic data clustering approach that adapts to changes in data over time, allowing clusters to evolve and reconfigure as new information becomes available. This method enables more accurate and timely insights by continuously refining groupings based on real-time data inputs, making it suitable for applications where data characteristics frequently shift.

LLM-as-a-judge: A scalable and explainable way to approximate human preferences, which are otherwise very expensive to obtain. Evaluating large language model (LLM) based chat assistants is challenging due to their broad capabilities and the inadequacy of existing benchmarks in measuring human preferences. LLMs as judges to evaluate these models on more open-ended questions.

LLM hallucination: A phenomenon wherein a large language model (LLM) — often a generative AI chatbot or computer vision tool — perceives patterns or objects that are nonexistent or imperceptible to human observers, creating outputs that are nonsensical or altogether inaccurate.

Foreword

Introduction

Pillar I: AI Organizations

Pillar II: Legal and Regulatory Compliance

Pillar III: Ethics, Transparency and Interpretability

Pillar IV: Data, AIOps and Infrastructure

Pillar V: AI Security

References and Further Reading

Acknowledgements

Appendix A: Glossary

License

M

Machine learning (ML): A form of AI that learns from existing data and makes predictions without being explicitly programmed.

Machine learning algorithms: Pieces of code that help people explore, analyze and find meaning in complex datasets. Each algorithm is a finite set of unambiguous step-by-step instructions that a machine can follow to achieve a certain goal. In a machine learning model, the goal is to establish or discover patterns that people can use to make predictions or categorize information.

Machine learning models: Process of using mathematical models of data to help a computer learn without direct instruction. Machine learning uses algorithms to identify patterns within data and those patterns are then used to create a data model that can make predictions. For example, in natural language processing, machine learning models can parse and correctly recognize the intent behind previously unheard sentences or combinations of words. In image recognition, a machine learning model can be taught to recognize objects — such as cars or dogs. A machine learning model can perform such tasks by having it "trained" with a large dataset. During training, the machine learning algorithm is optimized to find certain patterns or outputs from the dataset, depending on the task. The output of this process — often a computer program with specific rules and data structures — is called a machine learning model.

Machine learning operations (MLOps): The practice of creating new machine learning (ML) models and running them through a repeatable, automated workflow that deploys them to production. An MLOps pipeline provides a variety of services to data science processes, including model version control, continuous integration and continuous delivery (CI/CD), model catalogs for models in production, infrastructure management, monitoring of live model performance, security and governance. MLOps is a collaborative function, often comprising data scientists, devops engineers, security teams and IT.

Malicious libraries: Software components that were intentionally designed to cause harm to computer systems or the data they process. Such packages can be distributed through various means, including phishing emails, compromised websites or even legitimate software repositories.

Metadata: Data that annotates other data and AI assets. It generally includes the permissions that govern access to them with descriptive information, possibly including their data descriptions, data about data ownership, access paths, access rights and data volatility.

MLflow Model Registry: A centralized model store, set of APIs and UI to collaboratively manage the full lifecycle of an MLflow model. It provides model lineage (i.e., information about which MLflow experiment and run produced the model), model versioning, model aliasing, model tagging and annotations.

MLSecOps: The integration of security practices and considerations into the ML development and deployment process. This includes ensuring the security and privacy of data used to train and test models, as well as protecting deployed models and the infrastructure they run on from malicious attacks.

Model cards: Standardized documentation for machine learning models that provide essential details about a model's purpose, performance and limitations.

Model drift: The decay of models' predictive power as a result of the changes in real-world environments.

Foreword

Introduction

Pillar I:
AI Organizations

Pillar II:
Legal and Regulatory
Compliance

Pillar III:
Ethics, Transparency
and Interpretability

Pillar IV:
Data, AIOps and
Infrastructure

Pillar V:
AI Security

References and
Further Reading

Acknowledgements

Appendix A:
Glossary

License

Model inference: The use of a trained model on new data to create a result.

Model inversion: In machine learning models, private assets like training data, features and hyperparameters, which are typically confidential, can potentially be recovered by attackers through a process known as model inversion. This technique involves reconstructing private elements without direct access, compromising the model's security.

Model management: A single place for development, tracking, discovering, governing, encrypting and accessing models with proper security controls.

Model operations: The building of predictive ML models, the acquisition of models from a model marketplace or the use of LLMs like OpenAI or Foundation Models APIs. Developing a model requires a series of experiments and a way to track and compare the conditions and results of those experiments.

Mosaic AI AutoML: Helps you automatically apply machine learning to a dataset. You provide the dataset and identify the prediction target, while AutoML prepares the dataset for model training. AutoML then performs and records a set of trials that creates, tunes and evaluates multiple models. After model evaluation, AutoML displays the results and provides a Python notebook with the source code for each trial run so you can review, reproduce and modify the code. AutoML also calculates summary statistics on your dataset and saves this information in a notebook that you can review later.

Mosaic AI Model Serving: A unified service for deploying, governing, querying and monitoring models fine-tuned or predeployed by Databricks like Llama 3, MosaicML MPT or BGE, or from any other model provider like Azure OpenAI, AWS Bedrock, AWS SageMaker and Anthropic. Model Serving provides a highly available and low-latency service for deploying models. The service automatically scales up or down to meet demand changes, saving infrastructure costs while optimizing latency performance.

Mosaic AI Vector Search: A vector database that is built into the Databricks Data Intelligence Platform and integrated with its governance and productivity tools. A vector database is a database that is optimized to store and retrieve embeddings. Embeddings are mathematical representations of the semantic content of data, typically text or image data. Embeddings are generated by a large language model and are a key component of many GenAI applications that depend on finding documents or images that are similar to each other. Examples are RAG systems, recommender systems and image and video recognition.

Model theft: Theft of a system's knowledge through direct observation of its input and output observations, akin to reverse engineering. This can lead to unauthorized access, copying or exfiltration of proprietary models, resulting in economic losses, eroded competitive advantage and exposure of sensitive information.

Model Zoo: A repository or library that contains pretrained models for various machine learning tasks. These models are trained on large datasets and are ready to be deployed or fine-tuned for specific tasks.

N

Notebook: A common tool in data science and machine learning for developing code and presenting results.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

O

Observability: The ability to monitor, understand and gain insights into the health and performance of data pipelines and systems. It involves tracking metrics like data quality, lineage, latency and dependencies, helping teams proactively detect and address issues in data workflows to ensure reliability and trustworthiness across the data lifecycle.

Offline system: ML systems that are trained up, "frozen," and then operated using new data on the frozen trained system.

Online system: An ML system is said to be "online" when it continues to learn during operational use, modifying its behavior over time.

Ontology: A formally defined vocabulary for a particular domain of interest used to capture knowledge about that (restricted) domain of interest. Adversaries may discover the ontology of a machine learning model's output space — for example, the types of objects a model can detect. The adversary may discover the ontology by repeated queries to the model, forcing it to enumerate its output space. Or the ontology may be discovered in a configuration file or in documentation about the model.

P

Penetration testing (pen testing): A security exercise where a cybersecurity expert attempts to find and exploit vulnerabilities in a computer system through a combination of an in-house offensive security team, qualified third-party penetration testers and a year-round public bug bounty program. The purpose of this simulated attack is to identify any weak spots in a system's defenses that attackers could take advantage of.

Predictive optimization: A process that uses predictive analytics and machine learning algorithms to analyze historical data and forecast future outcomes, enabling organizations to make data-driven decisions that enhance performance and efficiency.

Pretrained LLM: Training an LLM from scratch using your own data for better domain performance.

Private link: Enables private connectivity between users and their Databricks workspaces and between clusters on the compute plane and core services on the control plane within the Databricks workspace infrastructure.

Prompt injection

- **Direct:** A direct prompt injection occurs when a user injects text that is intended to alter the behavior of the LLM
- **Indirect:** When a user might modify or exfiltrate resources (e.g., documents, web pages) that will be ingested by the GenAI model at runtime via the RAG process

Q

Query federation: A technique that allows users to execute a single query across multiple data sources, like databases or data lakes, without consolidating the data into one place.

Foreword

Introduction

Pillar I: AI Organizations

Pillar II: Legal and Regulatory Compliance

Pillar III: Ethics, Transparency and Interpretability

Pillar IV: Data, AIOps and Infrastructure

Pillar V: AI Security

References and Further Reading

Acknowledgements

Appendix A: Glossary

License

R

Red teaming: NIST defines cybersecurity red teaming as "a group of people authorized and organized to emulate a potential adversary's attack or exploitation capabilities against an enterprise's security posture. The Red Team's objective is to improve enterprise cybersecurity by demonstrating the impacts of successful attacks and by demonstrating what works for the defenders (i.e., the Blue Team) in an operational environment." (CNSS 2015 [80]) Traditional red teaming might combine physical and cyberattack elements, attack multiple systems and aim to evaluate the overall security posture of an organization. Penetration testing (pen testing), in contrast, tests the security of a specific application or system. In AI discourse, red teaming has come to mean something closer to pen testing, where the model may be rapidly or continuously tested by a set of evaluators and under conditions other than normal operation.

Reinforcement learning from human feedback (RLHF): A method of training AI models where human feedback is used as a source of reinforcement signals. Instead of relying solely on predefined reward functions, RLHF incorporates feedback from humans to guide the learning process.

Resource control: A capability in which the attacker has control over the resources consumed by an ML model, particularly for LLMs and RAG applications.

Responsible AI: Responsible Artificial Intelligence (Responsible AI) is an approach to developing, assessing and deploying AI systems in a safe, trustworthy and ethical way. Characteristics of trustworthy AI systems include: valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced and fair with harmful bias managed.

Retrieval augmented generation (RAG): An architectural approach that can improve the efficacy of large language model (LLM) applications by leveraging custom data. This is done by retrieving data/documents relevant to a question or task and providing them as context for the LLM.

S

Serverless compute: An architectural design that follows infrastructure as a service (IaaS) and platform as a service (PaaS), and which primarily requires the customer to provide the necessary business logic for execution. Meanwhile, the service provider takes care of infrastructure management. Compared to other platform architectures like PaaS, Serverless provides a considerably quicker path to realizing value and typically offers better cost efficiency and performance.

Single-sign on (SSO): A user authentication tool that enables users to securely access multiple applications and services using just one set of credentials.

Software development lifecycle (SDLC): A structured process that enables the production of high-quality, low-cost software, in the shortest possible production time. The goal of the SDLC is to produce superior software that meets and exceeds all customer expectations and demands. The SDLC defines and outlines a detailed plan with stages or phases, that each encompasses their own process and deliverables. Adherence to the SDLC enhances development speed and minimizes project risks and costs associated with alternative methods of production.

Source code control: A capability in which the attacker has control over the source code of the machine learning algorithm.

Foreword

Introduction

Pillar I: AI Organizations

Pillar II: Legal and Regulatory Compliance

Pillar III: Ethics, Transparency and Interpretability

Pillar IV: Data, AIOps and Infrastructure

Pillar V: AI Security

References and Further Reading

Acknowledgements

Appendix A: Glossary

License

Synthetic data generation: Creation of artificial data that mimics real-world data characteristics and structures, typically using algorithms or simulation techniques. This approach is often used to augment training datasets for machine learning models, enhance privacy by obfuscating sensitive information and facilitate testing and validation processes without relying on real data.

System for cross-domain identity management (SCIM): An open standard designed to manage user identity information. SCIM provides a defined schema for representing users and groups, and a RESTful API to run CRUD operations on those user and group resources. The goal of SCIM is to securely automate the exchange of user identity data between your company's cloud applications and any service providers, such as enterprise SaaS applications.

T

Table tops: Simulated, discussion-based sessions designed to help organizations prepare for and respond to potential incidents, typically in the context of emergency management or cybersecurity.

Train proxy: The ability of an attacker to extract training data of a generative model by prompting the model on specific inputs.

Train proxy via replication: Adversaries may replicate a private model. By repeatedly querying the victim's ML Model Inference API Access, the adversary can collect the target model's inferences into a dataset. The inferences are used as labels for training a separate model offline that will mimic the behavior and performance of the target model.

Trojan: A malicious code/logic inserted into the code of a software or hardware system, typically without the knowledge and consent of the organization that owns/develops the system and which is difficult to detect and may appear harmless, but can alter the intended function of the system upon a signal from an attacker to cause a malicious behavior desired by the attacker. For Trojan attacks to be effective, the trigger must be rare in the normal operating environment so that it does not affect the normal effectiveness of the AI and raise the suspicions of human users.

Trojan horse backdoor: In the context of adversarial machine learning, the term "backdoor" describes a malicious module injected into the ML model that introduces some secret and unwanted behavior. This behavior can then be triggered by specific inputs, as defined by the attacker.

U

Unity Catalog (UC): A unified governance solution for data and AI assets on the Databricks Data Intelligence Platform. It provides centralized access control, auditing, lineage and data discovery capabilities across Databricks workspaces.

Foreword

Introduction

Pillar I:
AI Organizations

Pillar II:
Legal and Regulatory
Compliance

Pillar III:
Ethics, Transparency
and Interpretability

Pillar IV:
Data, AIOps and
Infrastructure

Pillar V:
AI Security

References and
Further Reading

Acknowledgements

Appendix A:
Glossary

License

V

Vector database: A specialized type of database designed to store and manage vector embeddings, which are high-dimensional representations of data such as text, images or audio. These databases enable efficient similarity searches and retrieval based on the geometric proximity of vectors, making them particularly useful for applications in machine learning, natural language processing and recommendation systems.

Vision models: AI systems that analyze and interpret visual data from images or videos, often using techniques from computer vision and deep learning. These models can perform tasks such as object detection, image classification and facial recognition, enabling applications in fields like autonomous vehicles, medical imaging and security surveillance.

Vulnerability management: An information security continuous monitoring (ISCM) process of identifying, evaluating, treating and reporting on security vulnerabilities in systems and the software that runs on them. This, implemented alongside other security tactics, is vital for organizations to prioritize possible threats and minimizing their "attack surface."

W

Watering hole attacks: A form of cyberattack that targets groups of users by infecting websites that they commonly visit to gain access to the victim's computer and network.

Webhooks: Enable you to listen for Model Registry events so your integrations can automatically trigger actions. You can use webhooks to automate and integrate your machine learning pipeline with existing CI/CD tools and workflows. For example, you can trigger CI builds when a new model version is created or notify your team members through Slack each time a model transition to production is requested.

Foreword
Introduction
Pillar I: AI Organizations
Pillar II: Legal and Regulatory Compliance
Pillar III: Ethics, Transparency and Interpretability
Pillar IV: Data, AIOps and Infrastructure
Pillar V: AI Security
References and Further Reading
Acknowledgements
Appendix A: Glossary
License

License

This work is licensed under the Creative Commons Attribution-Share Alike 4.0 License.

To view a copy of this license, visit https://creativecommons.org/licenses/by-sa/4.0/ or send a letter to:

Creative Commons
171 Second Street, Suite 300
San Francisco, California 94105
USA

Databricks is the data and AI company. More than 10,000 organizations worldwide — including Block, Comcast, Condé Nast, Rivian, Shell and over 60% of the Fortune 500 — rely on the Databricks Data Intelligence Platform to take control of their data and put it to work with AI.

Databricks is headquartered in San Francisco, with offices around the globe, and was founded by the original creators of Lakehouse, Apache Spark™, Delta Lake and MLflow.

To learn more, follow Databricks on LinkedIn, X and Facebook.

Evaluate Databricks for yourself. Visit us at databricks.com and try Databricks free!

© Databricks 2025. All rights reserved. Apache, Apache Spark, Spark and the Spark logo are trademarks of the Apache Software Foundation.
Privacy Policy | Terms of Use