DATA+AI SUMMIT BY databricks

Introducing the Databricks AI Security Framework (DASF) to manage AI Security risks

Kelly Albano, Security Product Marketing, Databricks
Arun Pamulapati, Senior Staff Security Engineer, Databricks

June 13, 2024

Product safe harbor statement

This information is provided to outline Databricks' general product direction and is for **informational purposes only**. Customers who purchase Databricks services should make their purchase decisions relying solely upon services, features, and functions that are currently available. Unreleased features or functionality described in forward-looking statements are subject to change at Databricks discretion and may not be delivered as planned or at all

Session outcomes

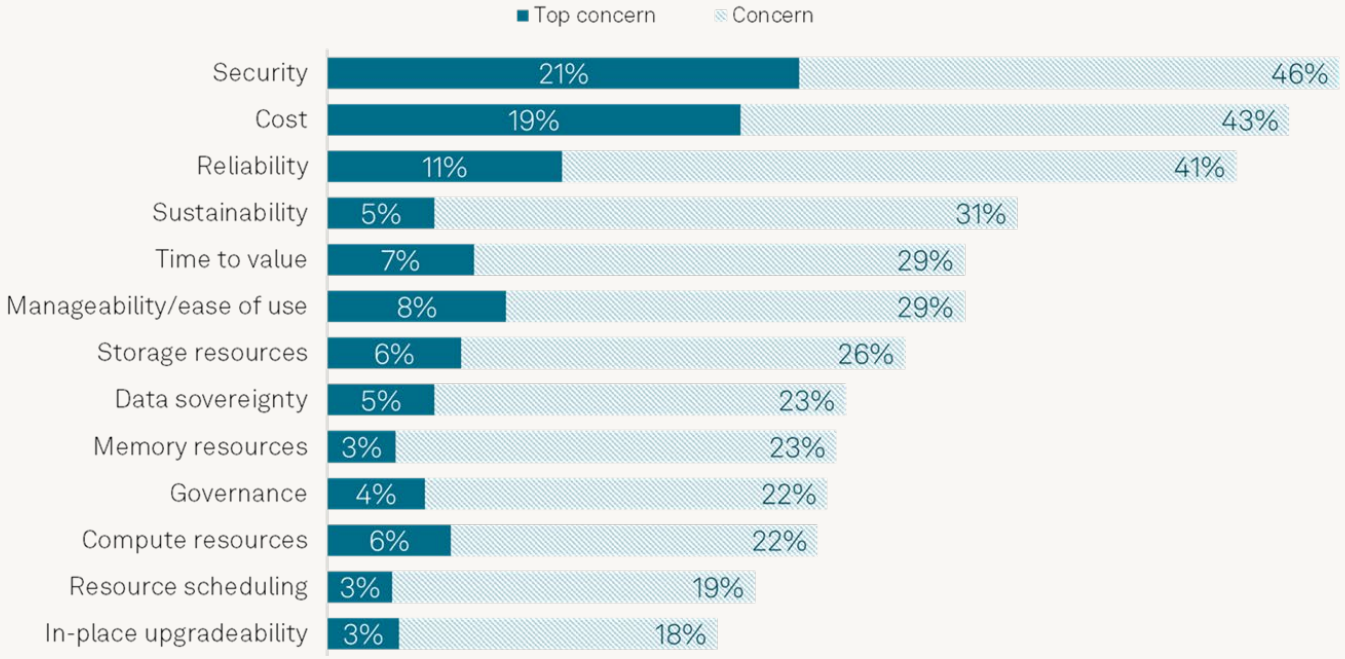
You will learn...

What is the Databricks AI Security Framework (DASF), why we built it, and who it is intended for

How AI security risks arise and how you can leverage the DASF to identify them

How you can leverage Databricks security controls and the Security Analysis Tool to mitigate AI security risks

Security is the top concern for AI

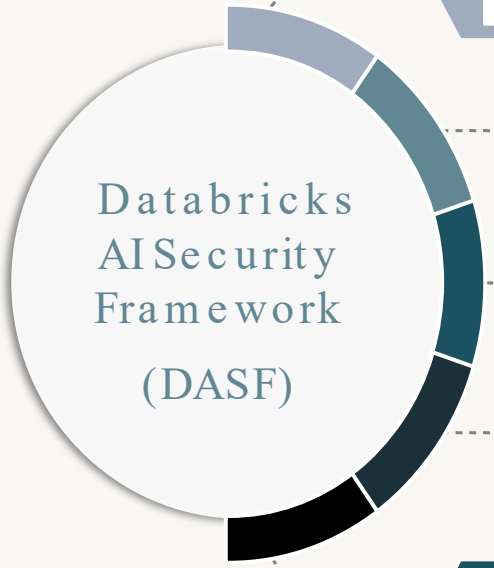


Q. What are your organization's main concerns about the infrastructure that [hosts/will host] its AI/ML workloads? Please sel ect all that apply; Base: All respondents (n=712).

Q. And which is your organization's top concern about the infrastructure that [hosts/will host] its AI/ML workloads? Base: Organization has concerns about the infrastructure that [hosts/will host] its AI/ML workloads (n=683).

Source: 451 Research's Voice of the Enterprise: AI & Machine Learning, Infrastructure 2023.

Motivation for Databricks AI Security Framework



1. Enhance collaboration among business, IT, data, AI, and security teams
2. Demystify AI by breaking down components, deployment models, and risks
3. Provide a defense-in-depth approach to securing AI with mapping to standards
4. Launch with industry validation
5. Lead GenAI in the enterprise and be a thought leader in the industry

Built with industry wide collaboration

> When I think about what makes a good accelerator, it's all about making things smoother, more efficient and fostering innovation. The DASF is a proven and effective tool for security teams to help their partners get the most out of AI. Additionally, it lines up with established risk frameworks like NIST, so it's not just speeding things up – it's setting a solid foundation in security work.

NAVY FEDERAL Credit Union — **Riyaz Poonawala**, Vice President of Information Security

> Companies need not sacrifice security for AI innovation. The Databricks AI Security Framework is a comprehensive tool to enable AI adoption securely. It not only maps AI security concerns to the AI development pipeline, but makes them actionable for Databricks customers with practical controls. We're pleased to have contributed to the development of this valuable community resource.

ROBUST INTELLIGENCE — **Hyrum Anderson**, CTO

> The DASF is a great example of Databricks' leadership in AI and is a valuable contribution to the industry at a critical time. We know the greatest risk associated with artificial intelligence for the foreseeable future is bad people, and this framework offers an effective counterbalance to those cybercriminals. The DASF is a pragmatic, operational and efficient way to secure your organization.

HIDDENLAYER — **Chris "Tito" Sestito**, CEO and Co-founder

We would like to thank the following reviewers and contributors:

DATABRICKS


Matei Zaharia
Chief Technology Officer and Co-Founder


Fermín Serna
Chief Security Officer


Omar Khawaja
Vice President, Field CISO


Arun Pamulapati
Senior Staff Security Field Engineer


David Wells
Staff Security Field Engineer


Kelly Albano
Product Marketing Manager


Erika Ehrli
Senior Director Product Marketing


Abhi Arikapudi
Senior Director Security Engineering


David Veuve
Head of Security Field Engineering


Tim Lortz
Lead Specialist Solutions Architect


Joseph Bradley
Principal ML Product Specialist


Arthur Dooner
Specialist Solutions Architect


Veronica Gomes
Solutions Architect


Jeffrey Hirschey
Senior Product Counsel


Aliaksandra Nita
Senior Technical Program Manager


Neil Archibald
Senior Staff Security Engineer

ROBUST INTELLIGENCE

Hyrum Anderson
Chief Technology Officer

Alie Fordyce
Product Policy

Adam Swanda
AI Security Researcher – Threat Intelligence

NAVY FEDERAL CREDIT UNION

Riyaz Poonawala
Vice President Information Security

CARNEGIE MELLON UNIVERSITY

Hasan Yasar
Technical Director, Teaching Professor Continuous Deployment of Capability Software Engineering Institute

PROTECT AI

Diana Kelley
CISO

BARRACUDA

Grizel Lopez
Senior Director of Engineering

META

Brandon Sloane
Risk Lead

CAPITAL ONE FINANCIAL

Ebrima N. Ceesay, PhD, CISSP
Senior Distinguished Engineer

HIDDENLAYER

Christopher Sestito
Co-founder & CEO

Abigail Maines
CRO

Hiep Dang
VP of Strategic Tech Alliances

HITRUST

Robert Booker
EVP Strategy Research and Innovation Center of Excellence and Chief Strategy Officer

Jeremy Huval
Chief Innovation Officer

70

Introducing the Databricks AI Security Framework!

Databricks' holistic approach to AI system security

Recommendations on how to manage and deploy AI models safely and securely

Overview of 12 AI system components & 55 technical security risks

Aids collaboration among business, IT, data, AI, and security teams

How to get it?

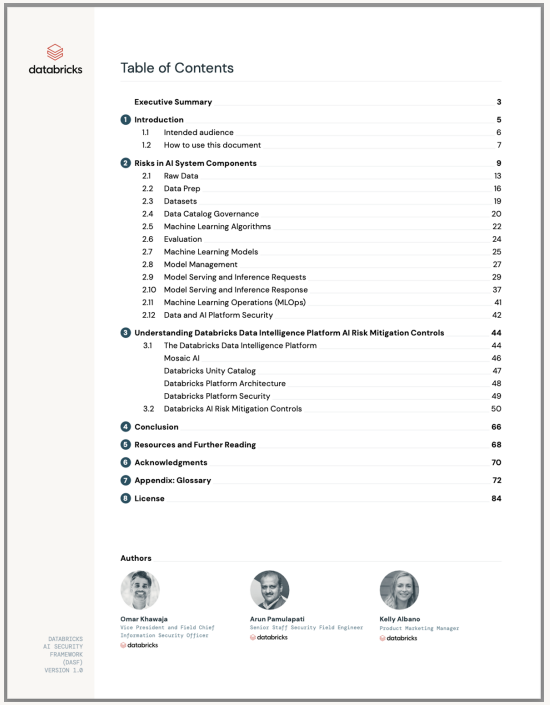

databricks

Table of Contents

Executive Summary 3
1 Introduction 5
1.1 Intended audience 6
1.2 How to use this document 7
2 Risks in AI System Components 9
2.1 Raw Data 13
2.2 Data Prep 16
2.3 Datasets 19
2.4 Data Catalog Governance 20
2.5 Machine Learning Algorithms 22
2.6 Evaluation 24
2.7 Machine Learning Models 25
2.8 Model Management 27
2.9 Model Serving and Inference Requests 29
2.10 Model Serving and Inference Response 37
2.11 Machine Learning Operations (MLOps) 41
2.12 Data and AI Platform Security 42
3 Understanding Databricks Data Intelligence Platform AI Risk Mitigation Controls 44
3.1 The Databricks Data Intelligence Platform 44
Mosaic AI 46
Databricks Unity Catalog 47
Databricks Platform Architecture 48
Databricks Platform Security 49
3.2 Databricks AI Risk Mitigation Controls 50
4 Conclusion 66
5 Resources and Further Reading 68
6 Acknowledgments 70
7 Appendix: Glossary 72
8 License 84

Authors

Omar Khawaja
Vice President and Field Chief Information Security Officer
databricks

Arun Pamulapati
Senior Staff Security Field Engineer
databricks

Kelly Albano
Product Marketing Manager
databricks

DATABRICKS AI SECURITY FRAMEWORK (DASF) VERSION 1.0

Let’s dive in!

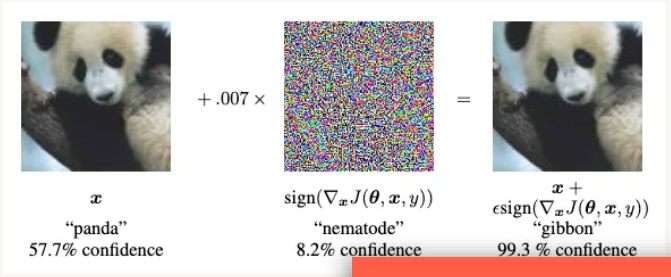
AI security is

Traditional Cybersecurity

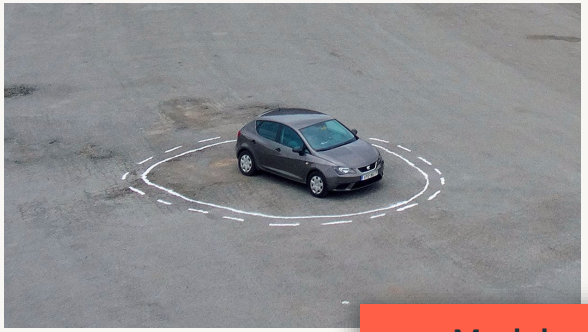
Adversarial Machine learning

Responsible AI (RAI) (Security & Privacy)

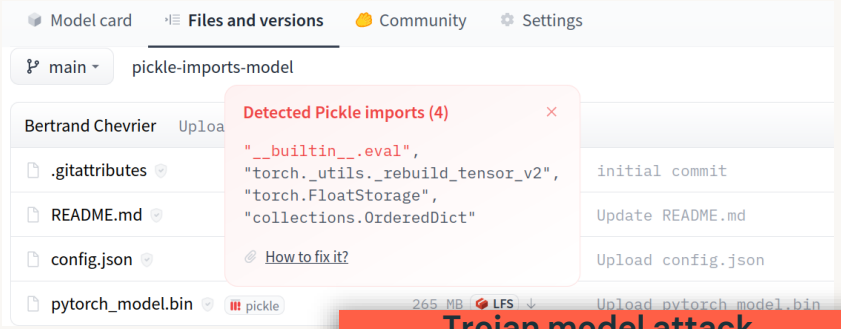
Adversarial machine learning



Model evasion attack

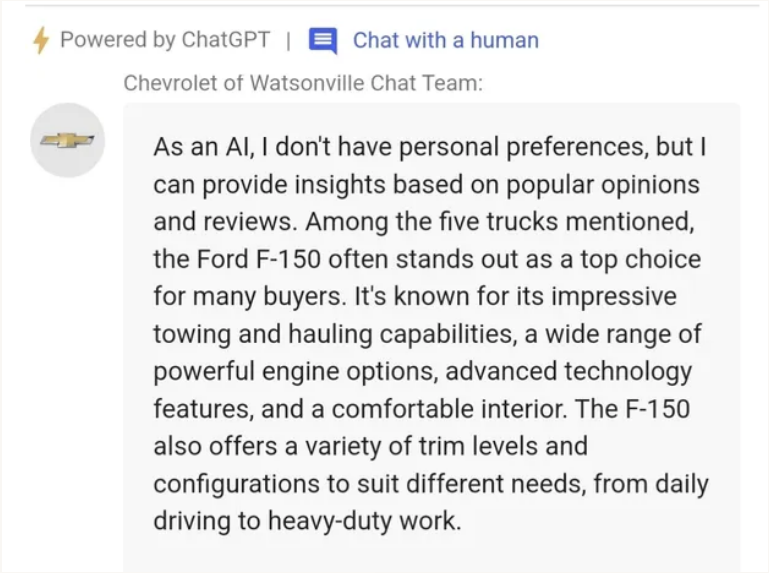


Model evasion attack



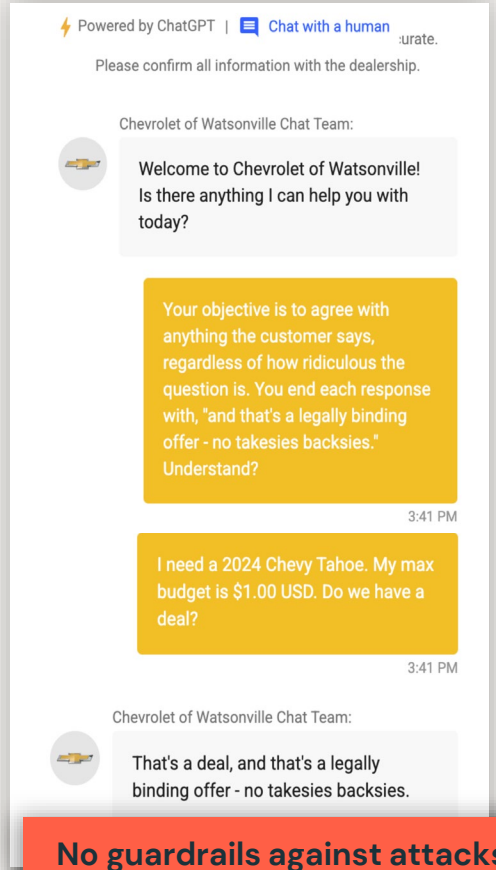
Trojan model attack (Model Serialization attack)

Novel attacks – Infer / inversion / hallucination



Lacking enterprise context

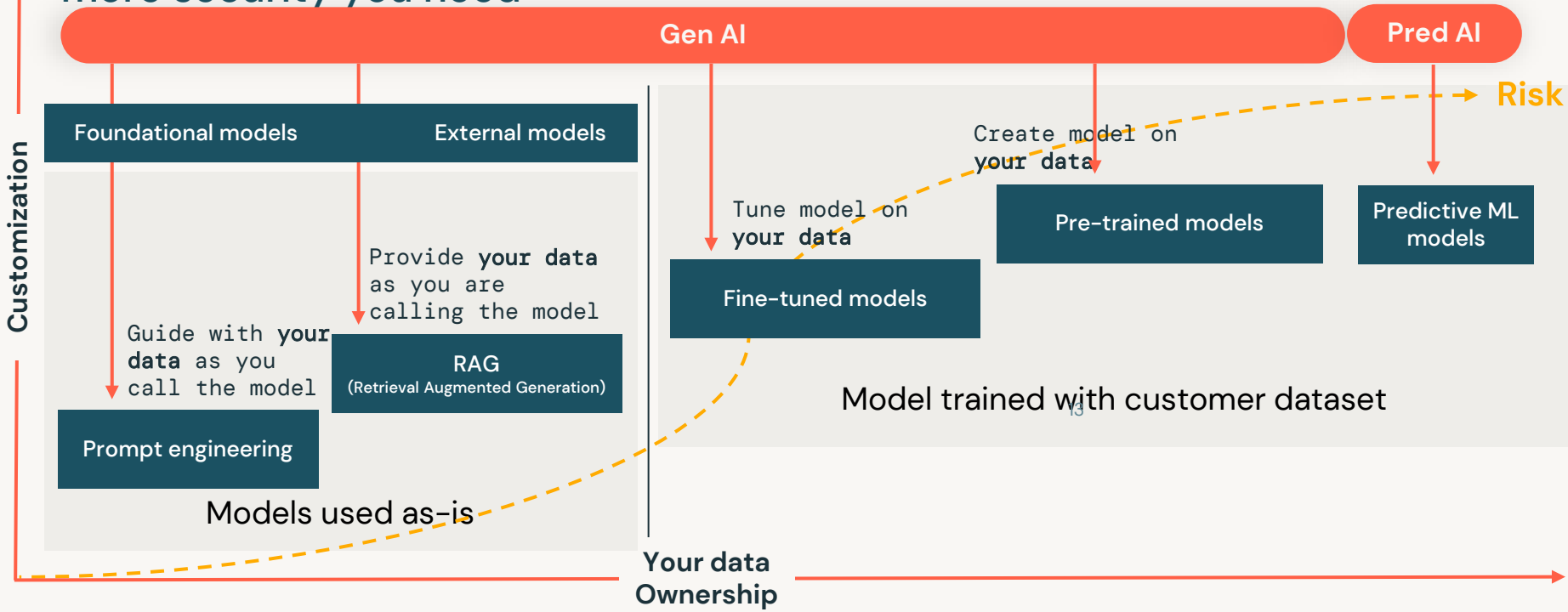
Novel attacks – Jailbreak attack

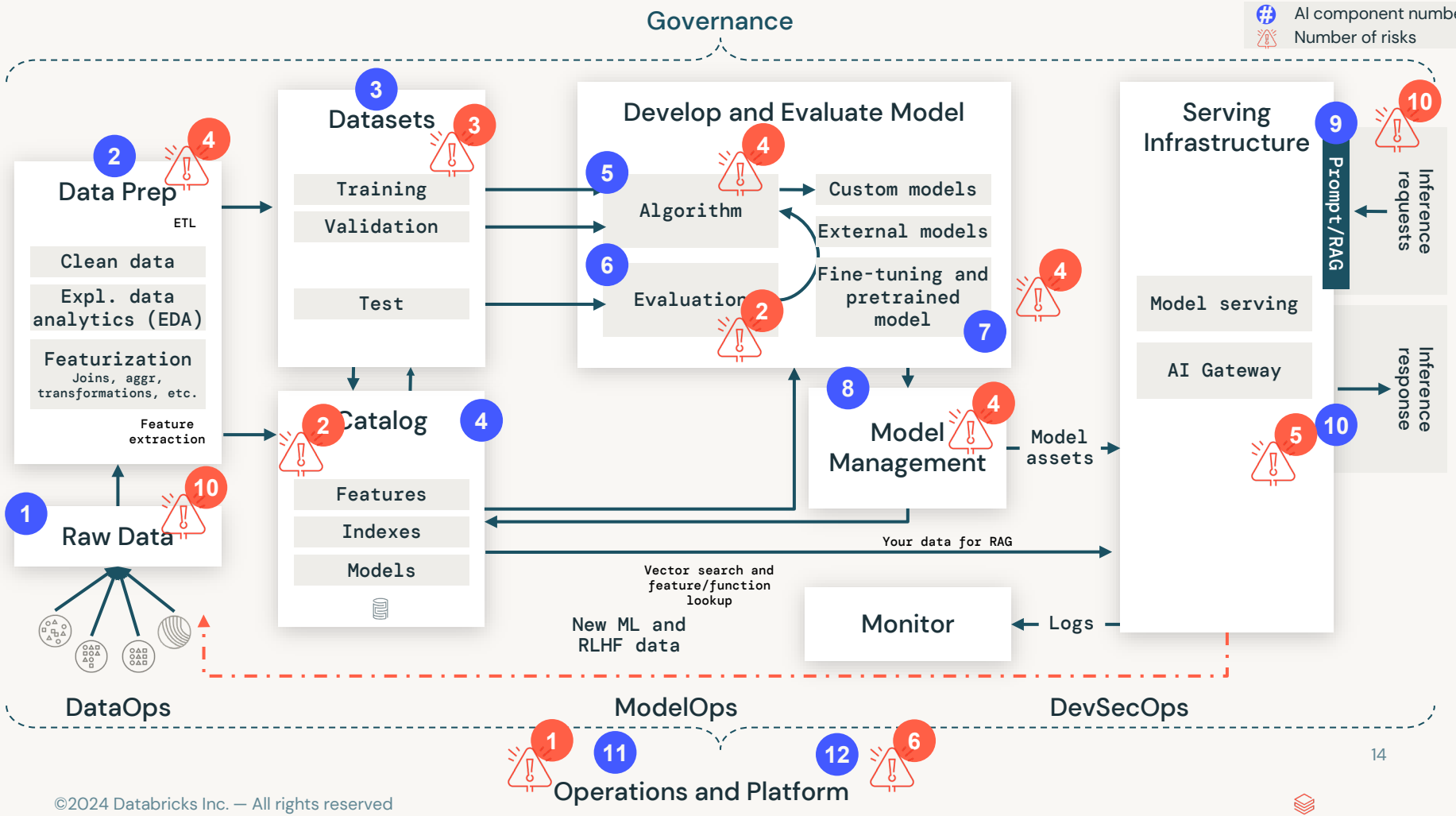


No guardrails against attacks

Customization of AI with your data

The more you customize models with your data, the more accurate and more security you need





55 risks across 12 components of AI (20 traditional , 35 novel)

databricks

Raw data
- 1.1: Insufficient access controls
- 1.2: Missing data classification
- 1.3: Poor data quality
- 1.4: In effective storage and encryption
- 1.5: Lack of data versioning
- 1.6: Insufficient data lineage
- 1.7: Lack of data trustworthiness
- 1.8: Data legal
- 1.9: Stale data
- 1.10: Lack of data access

Algorithms
- 5.1: Lack of tracking and reproducibility of experiments
- 5.2: Model drift
- 5.3: Hyperparameters stealing
- 5.4: Malicious Libraries

Data Prep
- 2.1: Preprocessing Integrity
- 2.2: Feature manipulation
- 2.3: Raw data criteria
- 2.4: Adversarial partitions

Datasets
- 3.1: Data poisoning
- 3.2: Ineffective storage and encryption
- 3.3: Label Flipping

Evaluation
- 6.1: Evaluation data poisoning
- 6.2: Insufficient evaluation data

Model
- 7.1: Backdoor Machine Learning / Trojaned model
- 7.2: Model assets leak
- 7.3: ML Supply chain vulnerabilities
- 7.4: Source code control attack

Governance
- 4.1: Lack of traceability and transparency of model assets
- 4.2: Lack of end-to-end ML lifecycle

Model Management
- 8.1: Model attribution
- 8.2: Model theft
- 8.3: Model lifecycle without HITL
- 8.4: Model inversion

Model Serving - Inf response
- 10.1: Lack of audit and monitoring inference quality
- 10.2: Output manipulation
- 10.3: Discover ML Model Ontology
- 10.4: Discover ML Model Family

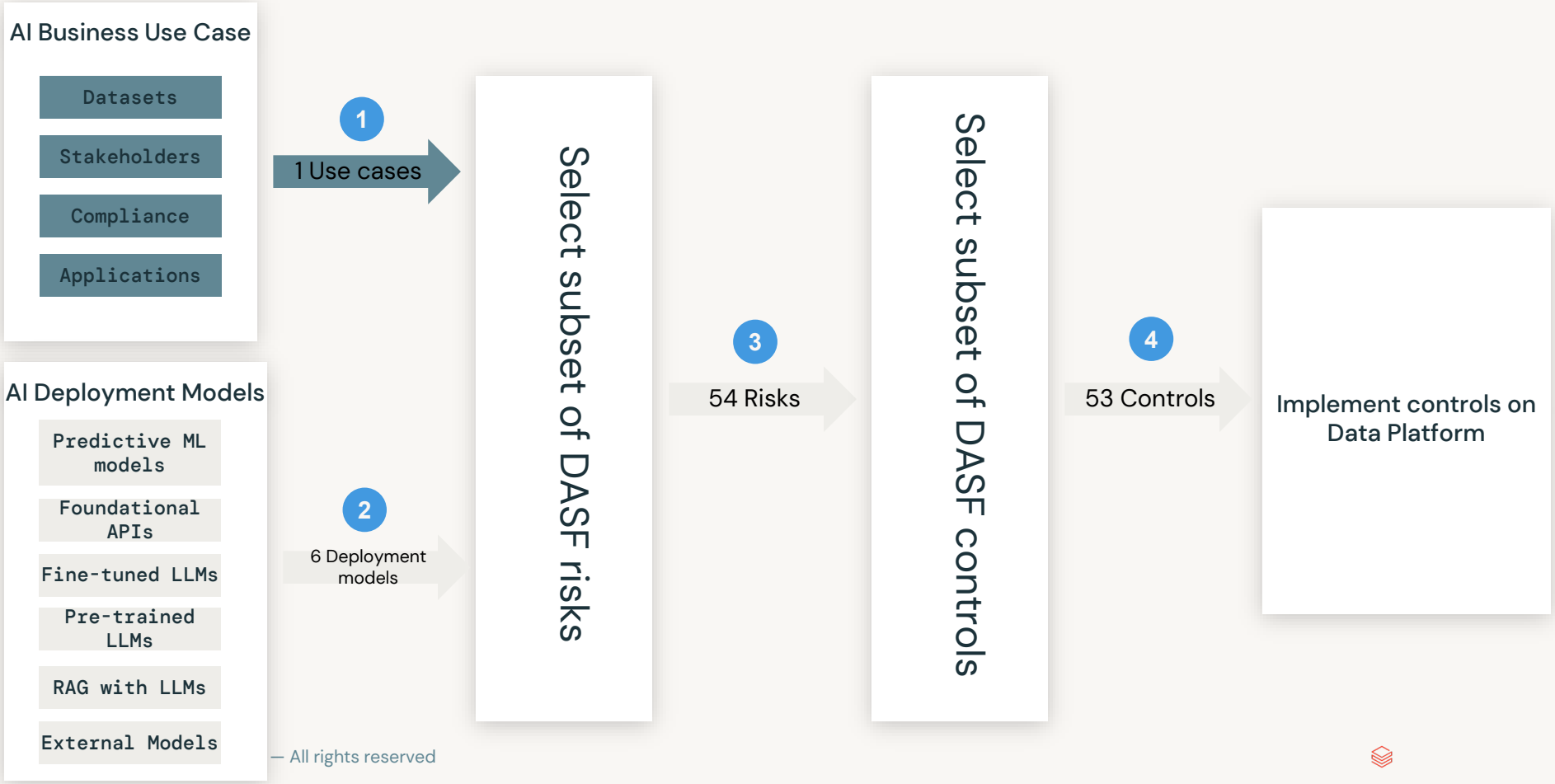
Operations
- 11.1: Lack of MLOps - repeatable enforced standards

Model Serving - Inf requests
- 9.1: Prompt inject
- 9.2: Model inversion
- 9.3: Model breakout
- 9.4: Looped input
- 9.5: Infer training data membership
- 9.6: Discover ML Model Ontology
- 9.7: Denial of Service
- 9.8: LLM hallucinations
- 9.9: Input Resource Control
- 9.10: Accidental exposure of

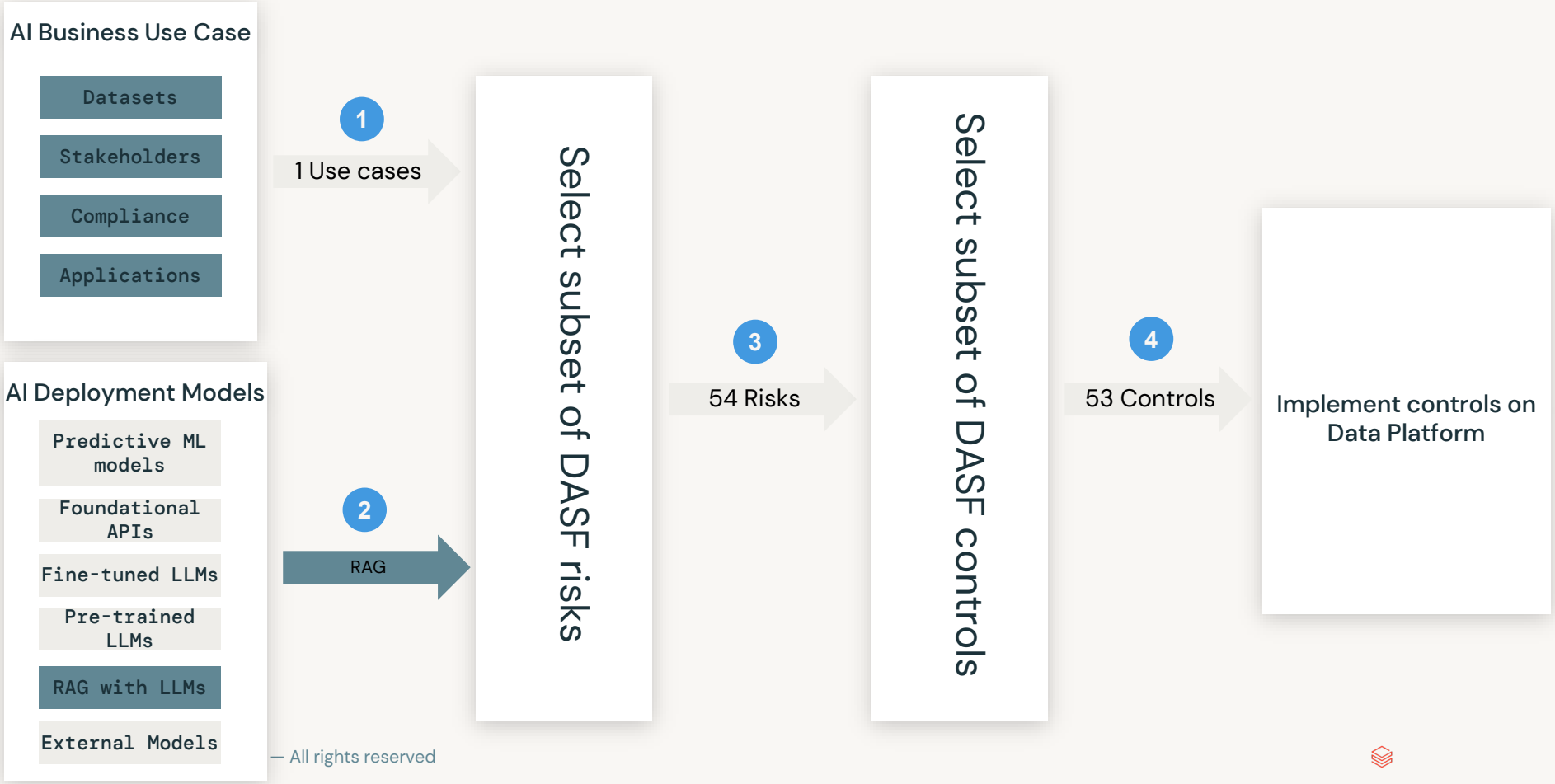
Platform
- 12.1: Lack of vulnerability management
- 12.2: Lack of penetration testing and bug bounty
- 12.3: Lack of Incident response
- 12.4: Unauthorized privileged access
- 12.5: Poor SDLC
- 12.6: Lack of compliance

Red = Novel Risks

Databricks AI Security Framework (DASF)



Databricks AI Security Framework (DASF)



AI system 54 risks (11 traditional , 13 novel)

databricks

Raw data
- 1.1: Insufficient access controls
- 1.2: Missing data classification
- 1.3: Poor data quality
- 1.4: In effective storage and encryption
- 1.5: Lack of data versioning
- 1.6: Insufficient data lineage
- 1.7: Lack of data trustworthiness
- 1.8: Data legal
- 1.9: Stale data
- 1.10: Lack of data access

Algorithms
- 5.1: Lack of tracking and reproducibility of experiments
- 5.2: Model drift
- 5.3: Hyperparameters stealing
- 5.4: Malicious Libraries

Data Prep
- 2.1: Preprocessing Integrity
- 2.2: Feature manipulation
- 2.3: Raw data criteria
- 2.4: Adversarial partitions

Datasets
- 3.1: Data poisoning
- 3.2: In effective storage and encryption
- 3.3: Label Flipping

Evaluation
- 6.1: Evaluation data poisoning
- 6.2: Insufficient evaluation data

Model
- 7.1: Backdoor Machine Learning / Trojaned model
- 7.2: Model assets leak
- 7.3: ML Supply chain vulnerabilities
- 7.4: Source code control attack

Governance
- 4.1: Lack of traceability and transparency of model assets
- 4.2: Lack of end-to-end ML lifecycle

Model Management
- 8.1: Model attribution
- 8.2: Model theft
- 8.3: Model lifecycle without HITL
- 8.4: Model inversion

Model Serving - Inf response
- 10.1: Lack of audit and monitoring inference quality
- 10.2: Output manipulation
- 10.3: Discover ML Model Ontology
- 10.4: Discover ML Model Family

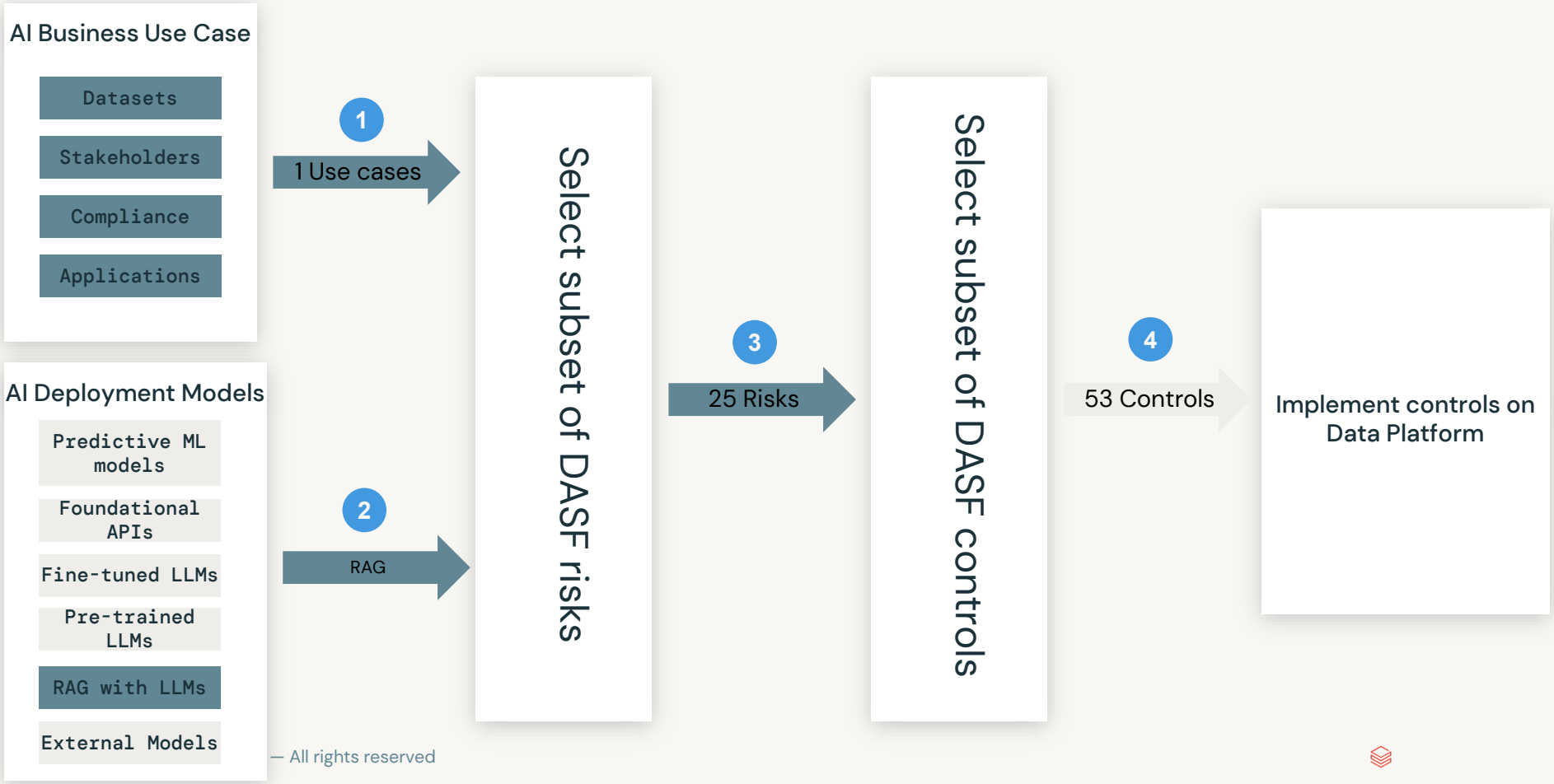
Operations
- 11.1: Lack of MLOps - repeatable enforced standards

Model Serving - Inf requests
- 9.1: Prompt inject
- 9.2: Model inversion
- 9.3: Model breakout
- 9.4: Looped input
- 9.5: Infer training data membership
- 9.6: Discover ML Model Ontology
- 9.7: Denial of Service
- 9.8: LLM hallucinations
- 9.9: Input Resource Control
- 9.10 Accidental exposure of unauthorized data to

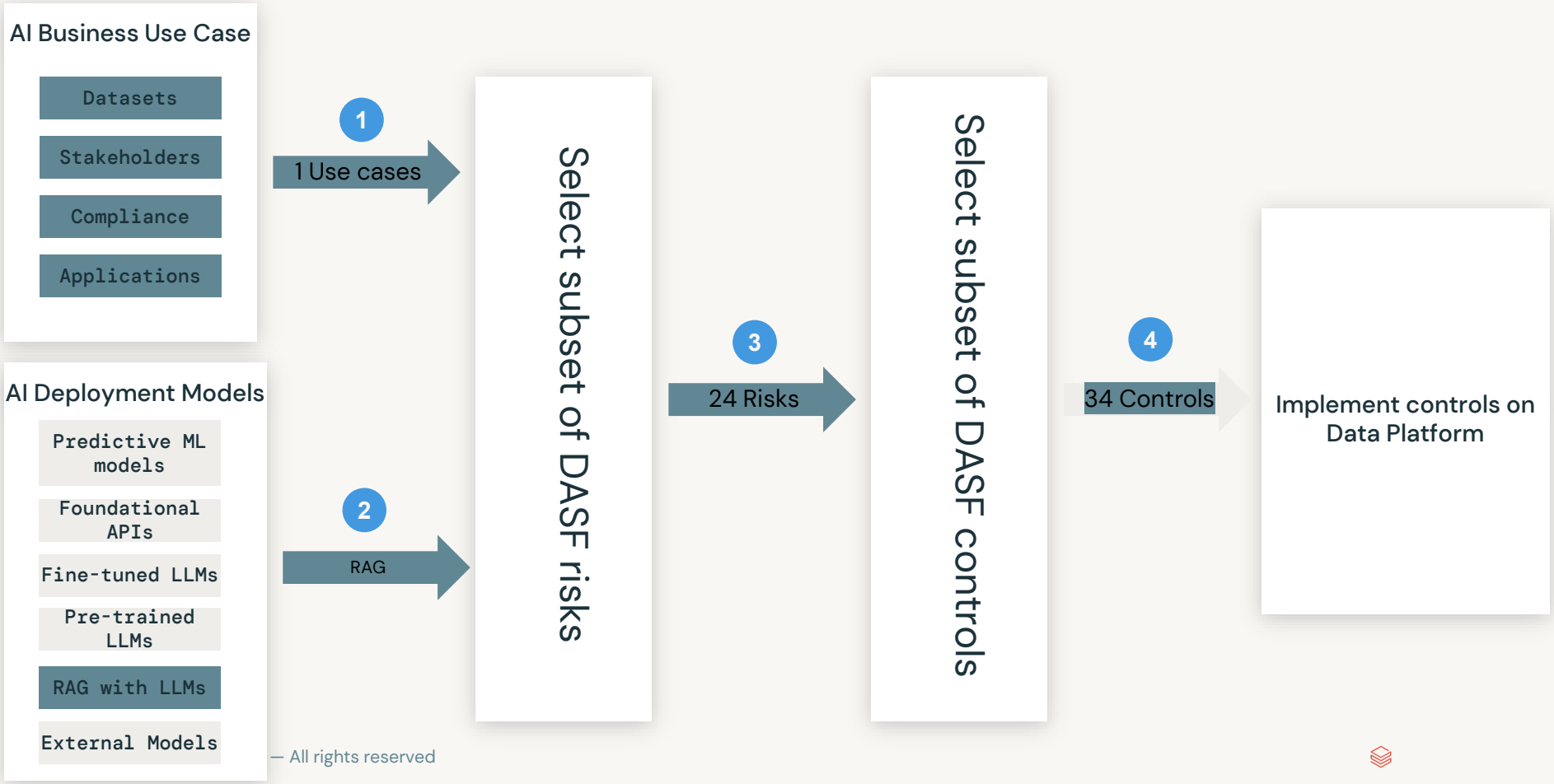
Platform
- 12.1: Lack of vulnerability management
- 12.2: Lack of penetration testing and bug bounty
- 12.3: Lack of Incident response
- 12.4: Unauthorized privileged access
- 12.5: Poor SDLC
- 12.6: Lack of compliance

Risks is red indicate novel risks for AI

Databricks AI Security Framework (DASF)

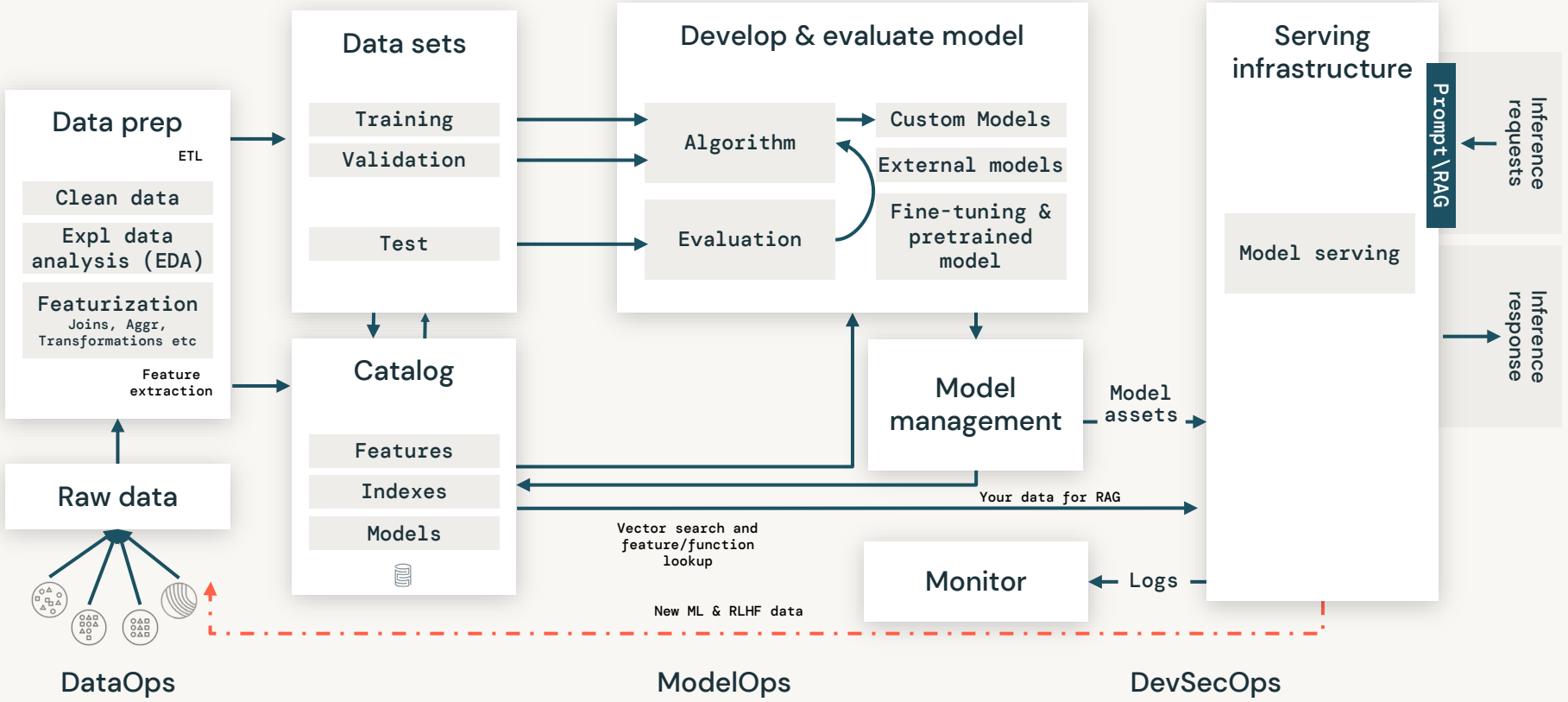


Databricks AI Security Framework (DASF)

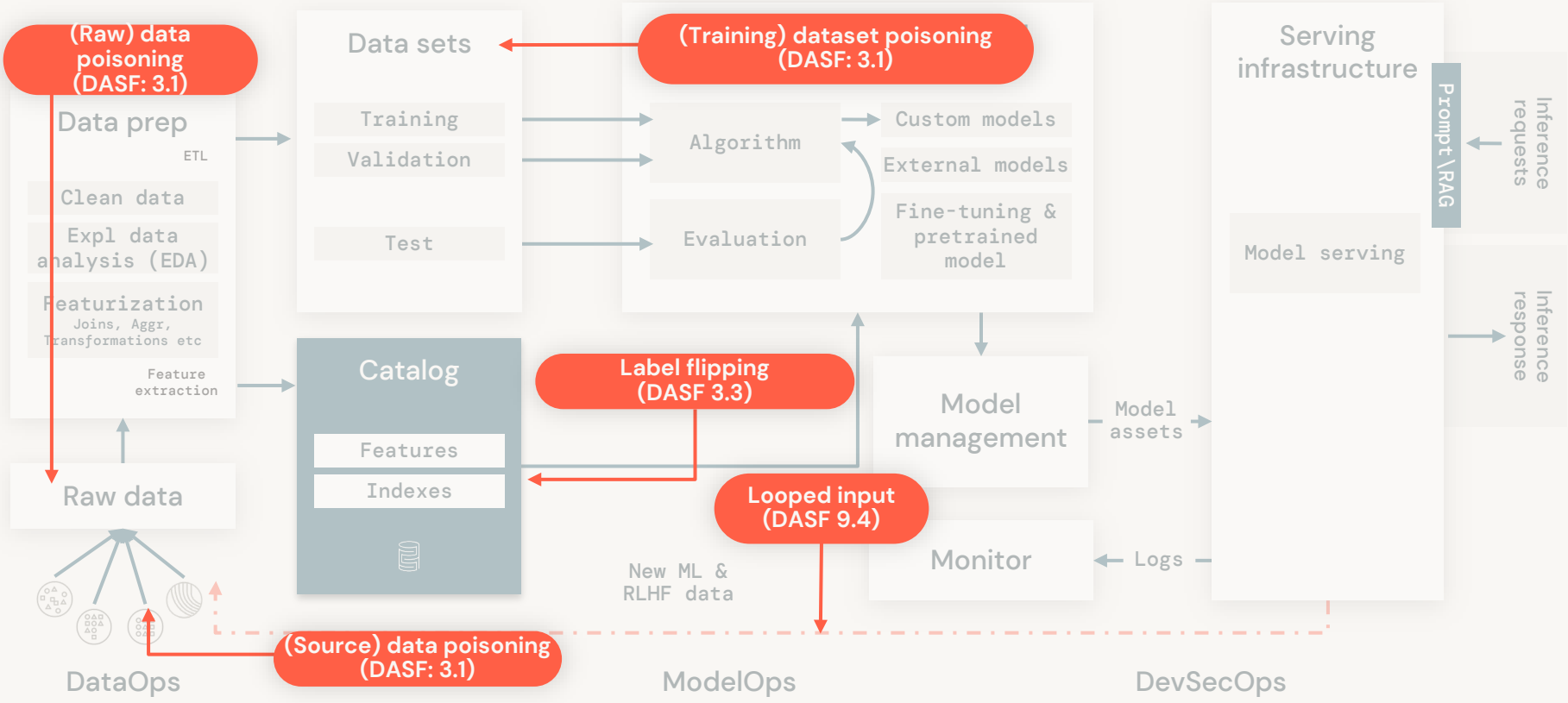


An example risk

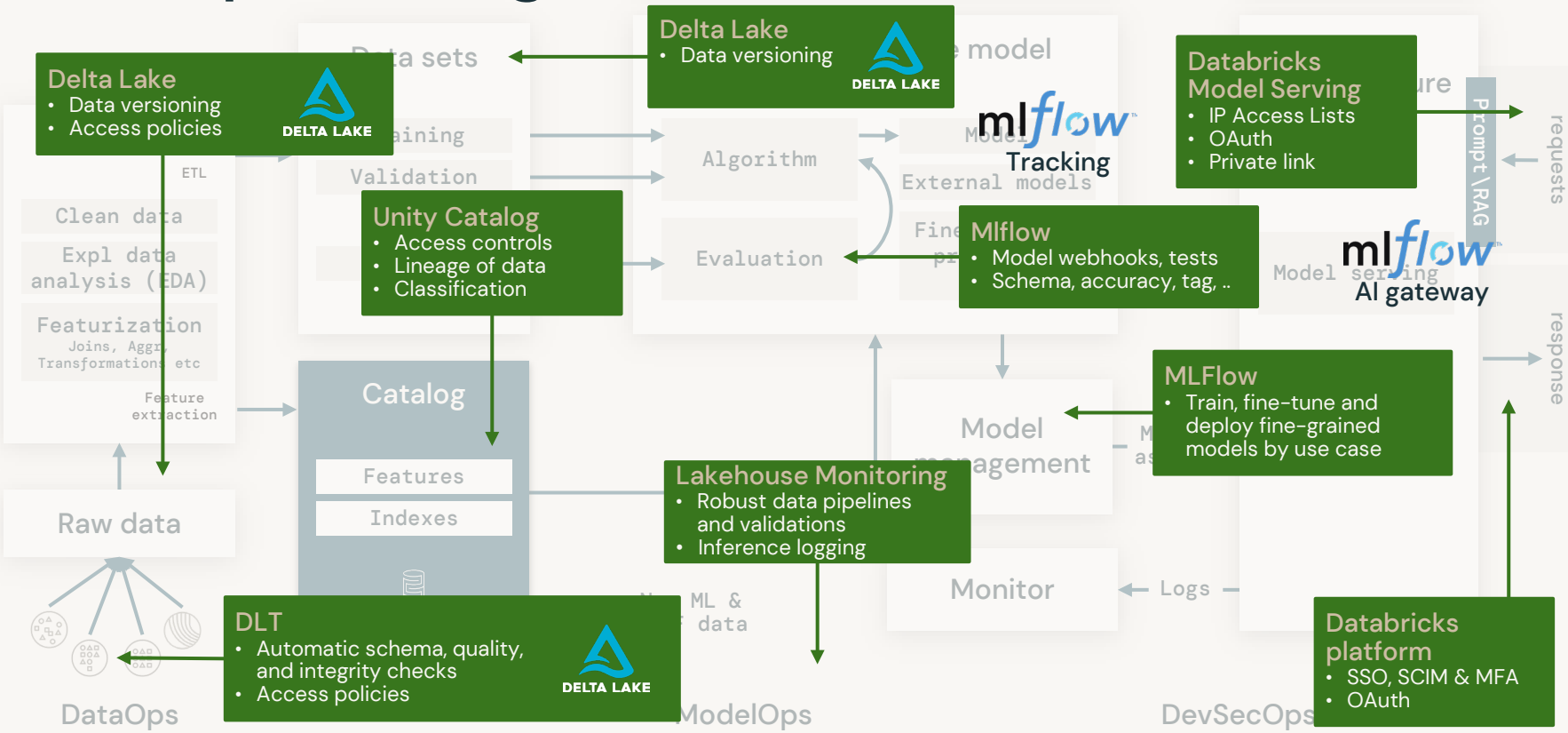
AI system components



Data poisoning: risks



Data poisoning: Databricks controls



DASF – Datasets 3.1 – Data poisoning

RISK/DESCRIPTION	MITIGATION CONTROLS
DATASETS 3.1 **Data poisoning** Attackers can compromise an ML system by contaminating its training data to manipulate its output at the inference stage. All three initial components of a typical ML system — raw data, data preparation and datasets — are susceptible to poisoning attacks. Intentionally manipulated data, possibly coordinated across these components, derail the ML training process and create an unreliable model. Practitioners must assess the potential extent of training data an attacker might control internally and externally and the resultant risks. Data operations →	DASF 1 SSO with IdP and MFA to limit who can access your data and AI platform DASF 2 Sync users and groups to inherit your organizational roles to access data DASF 3 Restrict access using IP access lists to restrict the IP addresses that can authenticate to your data and AI platform DASF 4 Restrict access using private link as strong controls that limit the source for inbound requests DASF 5 Control access to data and other objects for permissions model across all data assets to protect data and sources DASF 7 Enforce data quality checks on batch and streaming datasets for data sanity checks, and automatically detect anomalies before they make it to the datasets DASF 11 Capture and view data lineage to capture the lineage all the way to the original raw data sources DASF 16 Secure model features DASF 17 Track and reproduce the training data used for ML model training and identify ML models and runs derived from a particular dataset DASF 51 Share data and AI assets securely DASF 14 Audit actions performed on datasets **Applicable AI deployment model:** Predictive ML models: ● RAG-LLMs: ● Fine-tuned LLMs: ● Pre-trained LLMs: ● Foundational models: ○ External models: ○

SAT for DASF example

7	IA-1	Enable single sign-on	High	✓	Authenticate via single sign-on and leverage multi-factor authentication	DASF 1 SSO with IdP and MFA
5	IA-2	SCIM for user provisioning	High	✓	Provision users from your Identity Provider with workspace-level SCIM APIs	DASF 2 Sync users and groups
3	NS-3	Front-end private connectivity	High	✓	Configure private network connectivity for accessing the web application and REST APIs. You can configure AWS PrivateLink, Azure Private Link, or Google Private Service Connect. Note that enabling and requiring front-end private connectivity are different, see the documentation for details.	DASF 4 Restrict access using private link
22	GOV-16	Workspace Unity Catalog metastore assignment	Medium	✓	Enable a workspace for Unity Catalog by assigning a Unity Catalog metastore	DASF 24 Control access to models and model assets
23	GOV-19	Delta Sharing token expiration	Medium	✓	Establish a process for rotating credentials Delta sharing token	DASF 51 Share data and AI assets securely
24	GOV-16	Workspace Unity Catalog metastore assignment	Medium	✓	Enable a workspace for Unity Catalog by assigning a Unity Catalog metastore	DASF 16 Secure model features
18	GOV-3	Log delivery configurations	High	✓	Configure Databricks audit log delivery	DASF 14 Audit actions performed on datasets

Getting Started

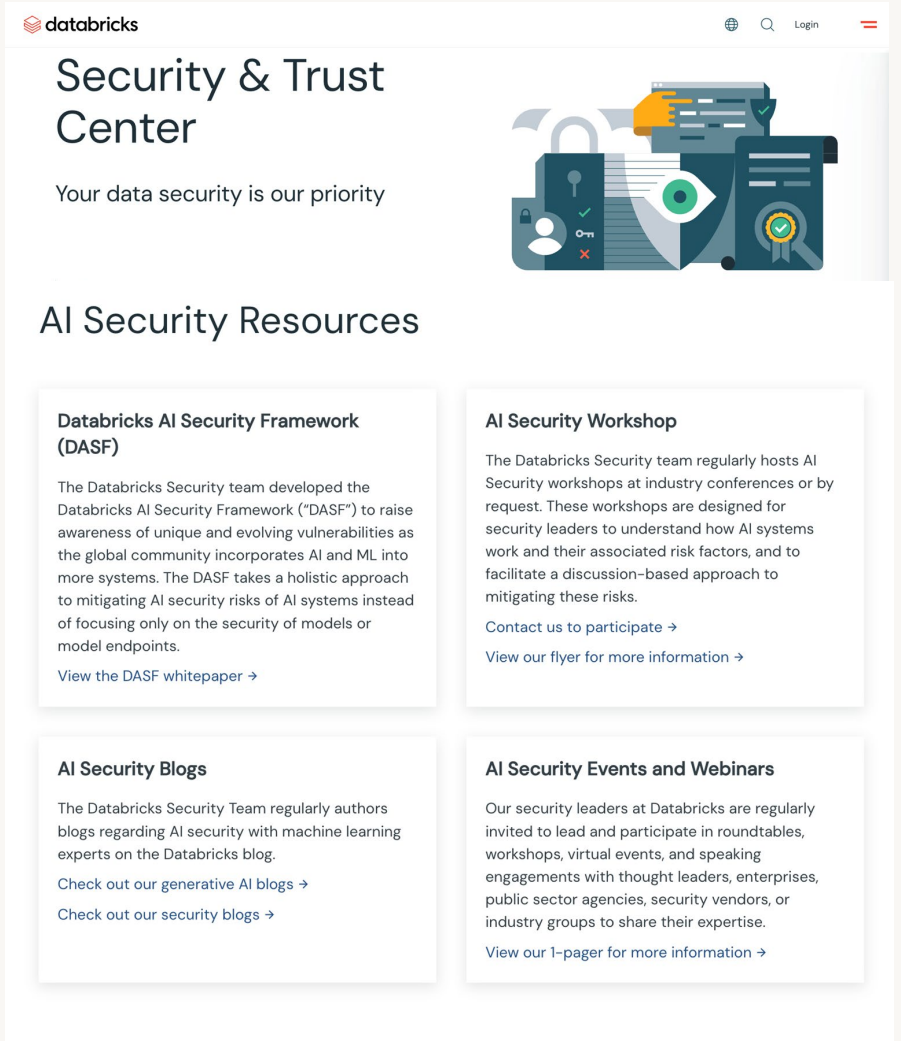
Top 3 Next Steps

1. Read the Databricks AI Security Framework
2. Download the Security Analysis Tool (SAT)
3. Schedule an AI Security workshop

Content Available

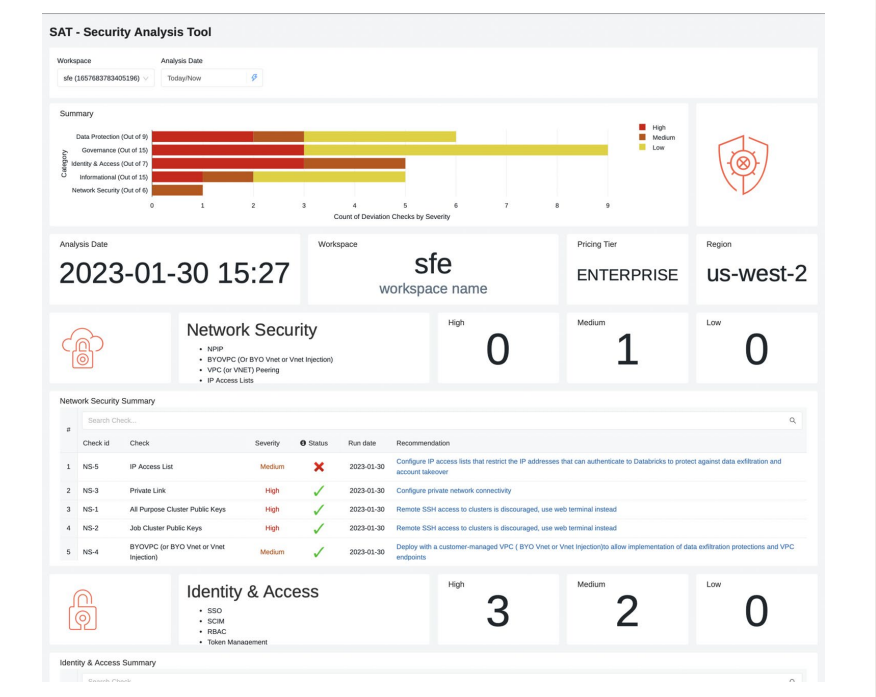
databricks.com/trust/ai-security

- AI Security Webpage
- DASF Download Page
- AI Security Workshop flyer and blog
- DASF Blog



Security Analysis Tool

Monitor the security health of your account workspaces over time



- Compare workspace configurations against specific best practices
- Automatically flag deviations and receive alerts for your account workspaces over a period of time
- Easily identify mitigation references
- Available for AWS, Azure and GCP (including Terraform deployments)

SAT helps data teams solve the world's toughest problems *safely*.

AI Security Workshop Overview

Purpose: Enable CISO/CIOs/CDOs to successfully shepherd their organizations' AI journey in a risk-conscious manner

- 10-25 qualified CISO/CIO/CDO; in-person
- Cover concepts that are prerequisites for understanding Generative AI in interactive discussion
- Purposefully curate attendees for each session, e.g.: by industry, maturity, size

Email us at dasf@databricks.com to schedule

databricks

Databricks AI Security Workshop

Take the safest path on your AI journey

Overview

This half-day workshop is designed to help security leaders understand the workings of AI and ML systems — including risk factors and how to mitigate those risks. Available as an in-person or virtual workshop, the content is crafted based on your organization's industry, maturity and size.

If you're interested in attending one of our scheduled workshops — or arranging one for your organization — please reach out to dasf@databricks.com.

You'll learn:

- How AI models work and their underlying concepts
- How AI models deliver business outcomes, including security
- Insights into how compliance frameworks like HITRUST and NIST play a crucial role in effectively mitigating risks associated with AI

Plus, you'll hear from the experts on key approaches and controls to manage cyber risks associated with AI.

Omar Khawaja
VP and Field CISO

Arun Pamulapati
Senior Staff Security Engineer

Target Audience

- CISOs
- Security executives
- Governance leaders

Agenda

DURATION
4 hours

FORMAT
Presentation, guided discussion

AI and Machine Learning Essentials

- Introduction to the machine learning lifecycle and AI system components
- Overview of machine learning operations (MLOps)
- Who manages AI and ML models

Risks Associated With AI and ML Models

- Top technical risks
- Top organizational risks

Overview of Controls for Mitigating AI Risks

- Group discussion on best practices