Abstract – This document provides an in-depth analysis of the DASF, exploring its structure, recommendations, and the practical applications it offers to organizations implementing AI solutions. This analysis not only serves as a quality examination but also highlights its significance and practical benefits for security experts and professionals across different sectors. By implementing the guidelines and controls recommended by the DASF, organizations can safeguard their AI assets against emerging threats and vulnerabilities.

I. INTRODUCTION

The Databricks AI Security Framework (DASF) is a comprehensive guide designed to address the evolving risks associated with the widespread integration of AI globally. The framework is created by the Databricks Security team and aims to provide actionable defensive control recommendations for AI systems, covering the entire AI lifecycle and facilitating collaboration between business, IT, data, AI, and security teams. The DASF is not limited to securing models or endpoints but adopts a holistic approach to mitigate cyber risks in AI systems, based on real-world evidence indicating that attackers employ simple tactics to compromise ML-driven systems.

The DASF identifies 55 technical security risks across 12 foundational components of a generic data-centric AI system, including raw data, data prep, datasets, data catalog governance, machine learning algorithms, evaluation, machine learning models, model management, model serving and inference, inference response, machine learning operations (MLOps), and data and AI platform security. Each risk is mapped to a set of mitigation controls that are ranked in prioritized order, starting with perimeter security to data security.

The Databricks Data Intelligence Platform is highlighted as a key component of the DASF, offering a unified foundation for all data and governance. The platform includes Mosaic AI, Databricks Unity Catalog, Databricks Platform Architecture, and Databricks Platform Security. Mosaic AI covers the end-to-end AI workflow, while Unity Catalog provides a unified governance solution for data and AI assets. The platform architecture is a hybrid PaaS that is data-agnostic, and the platform security is based on trust, technology, and transparency principles.

The DASF is intended for security teams, ML practitioners, governance officers, and DevSecOps engineering teams. It provides a structured conversation on new threats and mitigations without requiring deep expertise crossover. The DASF also includes a detailed guide for understanding the security and compliance of specific ML systems, offering insights into how ML impacts system security, applying security engineering principles to ML, and providing a detailed guide for understanding the security and compliance of specific ML systems.

The DASF concludes with Databricks' final recommendations on how to manage and deploy AI models safely and securely, consistent with the core tenets of machine learning adoption: identify the ML business use case, determine the ML deployment model, select the most pertinent risks, enumerate threats for each risk, and choose which controls to implement. It also provides further reading to enhance knowledge of the AI field and the frameworks reviewed as part of the analysis.

DASF document serves as a guide for how organizations can effectively utilize the framework to enhance the security of their AI systems, promoting a collaborative and comprehensive approach to AI security across various teams and AI model types:

- **Collaborative Use**: The DASF is designed for collaborative use by data and AI teams along with their security counterparts. It emphasizes the importance of these teams working together throughout the AI lifecycle to ensure the security and compliance of AI systems.
- **Applicability Across Teams**: The concepts in the DASF are applicable to all teams, regardless of whether they use Databricks to build their AI solutions. This inclusivity ensures that the framework can be utilized by a broad audience to enhance AI security.
- **Guidance on AI Model Types**: The document suggests that organizations first identify what types of AI models are being built or used. It categorizes models broadly into predictive ML models, state-of-the-art open models, and external models, providing a framework for understanding the specific security considerations for each type.
- **Understanding AI System Components**: Organizations are encouraged to review the 12 foundational components of a generic data-centric AI system as outlined in the document.
- **Risk Identification and Mitigation**: The DASF guides organizations to identify relevant risks and determine applicable controls from a comprehensive list provided in the document. This structured approach helps in

prioritizing security measures based on the specific needs of the organization.

- **Documentation and Features in Databricks Terminology**: While the document refers to documentation or features in Databricks terminology, it aims to be accessible to those who do not use Databricks. This approach helps in making the document useful for a wider audience while maintaining its practicality for Databricks users.

II. Audience

- **Security Teams**: This includes Chief Information Security Officers (CISOs), security leaders, DevSecOps, Site Reliability Engineers (SREs), and others responsible for the security of systems. They can use the DASF to understand how machine learning (ML) will impact system security and to grasp some of the basic mechanisms of ML.
- **ML Practitioners and Engineers**: This group comprises data engineers, data architects, ML engineers, and data scientists. The DASF helps them understand how security engineering and the "secure by design" mentality can be applied to ML.
- **Governance Officers**: These individuals are responsible for ensuring that data and AI practices within an organization comply with relevant laws, regulations, and policies. The DASF provides guidance on how ML impacts system security and compliance.
- **DevSecOps Engineering Teams**: These teams focus on integrating security into the development and operations processes. The DASF offers a structured way for these teams to have conversations about new threats and mitigations without requiring deep expertise crossover.

III. Benefits and Drawbacks

Databricks AI Security Framework (DASF) offers a comprehensive and actionable guide for organizations looking to understand and mitigate AI security risks. However, its complexity and Databricks-centric guidance may present challenges for some organizations.

A. Benefits

- **Holistic approach**: The DASF takes a holistic approach to AI security, addressing risks across the entire AI lifecycle and all components of a generic data-centric AI system. This comprehensive approach helps organizations identify and mitigate security risks more effectively.
- **Collaboration**: The framework is designed to facilitate collaboration between business, IT, data, AI, and security teams. This encourages a unified approach to AI security and helps bridge the gap between different disciplines.
- **Actionable recommendations**: The DASF provides actionable defensive control recommendations for each identified risk, which can be updated as new risks emerge and additional controls become available. This ensures that organizations can stay current with evolving AI security threats.
- **Applicability**: The DASF is applicable to organizations using various AI models, including predictive ML models, generative AI models, and external models. This broad applicability makes it a valuable resource for a wide range of organizations.
- **Integration with Databricks Data Intelligence Platform**: For organizations using the Databricks Data Intelligence Platform, the DASF offers specific guidance on leveraging the platform's AI risk mitigation controls. This helps organizations maximize the security benefits of the platform.

B. Drawbacks

- **Complexity**: The DASF covers a wide range of AI security risks and mitigation controls, which may be overwhelming for organizations new to AI security or with limited resources. Implementing the framework may require a significant investment of time and effort.
- **Databricks-centric guidance**: While the DASF offers valuable guidance for organizations using the Databricks Data Intelligence Platform, some of the recommendations may be less applicable or actionable for organizations using different AI platforms or tools.
- **Evolving landscape**: As the AI security landscape continues to evolve, organizations may need to continually update their security controls and practices to stay current.
- **Lack of specific examples**: The DASF provides a high-level overview of AI security risks and mitigation controls, but it may lack specific examples or case studies to illustrate how these risks and controls apply in real-world scenarios.
- **Focus on technical risks**: The DASF primarily focuses on technical security risks and mitigation controls. While this is an essential aspect of AI security, organizations should also consider non-technical risks, such as ethical, legal, and social implications of AI, which are not extensively covered in the DASF.

IV. Framework Alignment

The Databricks AI Security Framework (DASF) is designed to complement and integrate with other security frameworks, such as NIST, HITRUST, ISO/IEC 27001 and 27002, and CIS Critical Security Controls. The DASF takes a holistic approach to mitigating AI security risks instead of focusing only on the security of models or model endpoints. This approach aligns with the principles of these frameworks, which provide a structured process for identifying, assessing, and mitigating cybersecurity risks.

V. Databricks AI Security Framework

The framework categorizes the AI system into 12 primary components, each associated with specific security risks identified through extensive analysis. This analysis includes

predictive ML models, generative foundation models, and external models, informed by customer inquiries, security assessments, workshops with Chief Information Security Officers (CISOs), and surveys on AI risks. The identified risks are then mapped to corresponding mitigation controls within the Databricks Data Intelligence Platform, with links to detailed product documentation for each risk.

The document outlines the AI system components and their associated risks as follows:

- **Data Operations**: This stage encompasses the initial handling of raw data, including ingestion, transformation, and ensuring data security and governance. It is crucial for the development of reliable ML models and a secure DataOps infrastructure. A total of 19 specific risks are identified in this category, ranging from insufficient access controls to lack of end-to-end ML lifecycle management.
- **Model Operations**: This stage involves the creation of ML models, whether through building predictive models, acquiring models from marketplaces, or utilizing APIs like OpenAI. It requires a series of experiments and tracking mechanisms to compare various conditions and outcomes. There are 14 specific risks identified, including issues like lack of experiment reproducibility and model drift.
- **Model Deployment and Serving**: This stage focuses on securely deploying model images, serving models, and managing features such as automated scaling and rate limiting. It also includes the provision of high-availability services for structured data in RAG applications. A total of 15 specific risks are highlighted, including prompt injection and model inversion.
- **Operations and Platform**: This final stage includes platform vulnerability management, patching, model isolation, and ensuring authorized access to models with security built into the architecture. It also involves operational tooling for CI/CD to maintain secure MLOps across development, staging, and production environments. Seven specific risks are identified, such as lack of MLOps standards and vulnerability management.

VI. Raw Data

- **Importance of Raw Data**: Raw data is the foundation of AI systems, encompassing enterprise data, metadata, and operational data in various forms such as semi-structured or unstructured data, batch data, or streaming data.
- **Data Security**: Securing raw data is paramount for the integrity of machine learning algorithms and any technical deployment particulars. It presents unique challenges, and all data collections in an AI system are subject to standard data security challenges as well as new ones.
- **Risk Mitigation Controls**: The document outlines specific risks associated with raw data and provides detailed mitigation controls for each. These controls include effective access management, data classification, data quality enforcement, storage and encryption, data versioning, data lineage, data trustworthiness, legal considerations, handling stale data, and data access logs.
- **Access Management**: Ensuring that only authorized individuals or groups can access specific datasets is fundamental to data security. This involves authentication, authorization, and finely tuned access controls.
- **Data Classification**: Classifying data is critical for governance, enabling organizations to sort and categorize data by sensitivity, importance, and criticality, which is essential for implementing appropriate security measures and governance policies.
- **Data Quality**: High data quality is crucial for reliable data-driven decisions and is a cornerstone of data governance. Organizations must rigorously evaluate key data attributes to ensure analytical accuracy and cost-effectiveness.
- **Storage and Encryption**: Encrypting data at rest and in transit is vital to protect against unauthorized access and to comply with industry-specific data security regulations.
- **Data Versioning and Lineage**: Versioning data and tracking change logs are important for rolling back or tracing back to the original data in case of corruption. Data lineage helps with compliance and audit-readiness by providing a clear understanding and traceability of data used for AI.
- **Trustworthiness and Legal Aspects**: Ensuring the trustworthiness of data and compliance with legal mandates such as GDPR and CCPA is essential. This includes the ability to "delete" specific data from machine learning systems and retrain models using clean and ownership-verified datasets.
- **Stale Data and Access Logs**: Addressing the risks of stale data and the lack of data access logs is important for maintaining the efficiency and security of business processes. Proper audit mechanisms are critical for data security and regulatory compliance.

VII. Data Prep

- **Definition and Importance**: Data preparation is defined as the process of transforming raw input data into a format that machine learning algorithms can interpret. This stage is crucial as it directly impacts the security and explainability of an ML system.
- **Security Risks and Mitigations**: The section outlines various security risks associated with data preparation and provides detailed mitigation controls for each. These risks include preprocessing integrity, feature manipulation, raw data criteria, and adversarial partitions.

- **Preprocessing Integrity**: Ensuring the integrity of preprocessing involves numerical transformations, data aggregation, text or image data encoding, and new feature creation. Mitigation controls include setting up Single Sign-On (SSO) with Identity Provider (IdP) and Multi-Factor Authentication (MFA), restricting access using IP access lists, and implementing private links to limit the source for inbound requests.
- **Feature Manipulation**: This risk involves the potential for attackers to manipulate how data is annotated into features, which can compromise the integrity and accuracy of the model. Controls include securing model features to prevent unauthorized updates and employing data-centric MLOps and LLMOps to promote models as code.
- **Raw Data Criteria**: Understanding the selection criteria for raw data is essential to prevent attackers from introducing malicious input that compromises system integrity. Controls include using access control lists and data-centric MLOps for unit and integration testing.
- **Adversarial Partitions**: This involves the risk of attackers influencing the partitioning of datasets used in training and evaluation, potentially controlling the ML system indirectly. Mitigation involves tracking and reproducing the training data used for ML model training and identifying ML models and runs derived from a particular dataset.
- **Comprehensive Mitigation Strategies**: The section emphasizes the importance of a comprehensive approach to securing the data preparation process, including the use of stringent security measures to safeguard against manipulations that can undermine the integrity and reliability of ML systems

VIII. Datasets

- **Significance of Datasets**: Datasets are crucial for training, validating, and testing machine learning models. They must be carefully managed to ensure the integrity and effectiveness of the AI systems.
- **Security Risks**: The section outlines various security risks associated with datasets, including data poisoning, ineffective storage and encryption, and label flipping. These risks can compromise the reliability and performance of machine learning models.
- **Data Poisoning**: This risk involves attackers manipulating training data to affect the model's output at the inference stage. Mitigation strategies include robust access controls, data quality checks, and monitoring data lineage to prevent unauthorized data manipulation.
- **Ineffective Storage and Encryption**: Proper data storage and encryption are critical to protect datasets from unauthorized access and breaches. The framework recommends encryption of data at rest and in transit, along with stringent access controls.
- **Label Flipping**: This specific type of data poisoning involves changing the labels in training data, which can mislead the model during training and degrade its performance. Encryption and secure access to datasets are recommended to mitigate this risk.
- **Mitigation Controls**: For each identified risk, the DASF provides detailed mitigation controls. These controls include the use of Single Sign-On (SSO) with Identity Providers (IdP), Multi-Factor Authentication (MFA), IP access lists, private links, and data encryption to enhance the security of datasets.
- **Comprehensive Risk Management**: The section emphasizes the importance of a comprehensive approach to managing dataset security, from the initial data collection to the deployment of machine learning models. This includes regular audits, updates to security protocols, and continuous monitoring of data integrity.

IX. Data Catalog Governance

- **Comprehensive Governance Approach**: Data catalog and governance involve managing an organization's data assets throughout their lifecycle, which includes principles, practices, and tools for effective management.
- **Centralized Access Control**: Managing governance for data and AI assets enables centralized access control, auditing, lineage, data, and model discovery capabilities, which limits the risk of data or model duplication, improper use of classified data for training, loss of provenance, and model theft.
- **Data Privacy and Security**: When dealing with datasets that may contain sensitive information, it is crucial to ensure that personally identifiable information (PII) and other sensitive data are adequately secured to prevent breaches and leaks. This is particularly important in sectors with stringent regulatory requirements.
- **Audit Trails and Transparency**: Proper data catalog governance allows for audit trails and tracing the origin and transformations of data used to train AI models. This transparency encourages trust and accountability, reduces the risk of biases, and improves AI outcomes.
- **Regulatory Compliance**: Ensuring that sensitive information in datasets is adequately secured is essential for compliance with regulations such as GDPR and CCPA. This includes the ability to demonstrate data security and maintain audit trails.
- **Collaborative Dashboard**: For computer vision projects involving multiple stakeholders, having an easy-to-use labeling tool with a collaborative dashboard is essential to keep everyone on the same page in real-time and avoid mission creep.
- **Automated Data Pipelines**: For projects with large volumes of data, automating data pipelines by connecting datasets and models using APIs can streamline the process and make it faster to train ML models.

- **Quality Control Workflows**: It is important to have customizable and manageable quality control workflows to validate labels and annotations, reduce errors and bias, and fix bugs in datasets. Automated annotation tools can help in this process

X. Machine Learning Algorithms

- **Technical Core of ML Systems**: Machine learning algorithms are described as the technical core of any ML system, crucial for the functionality and security of the system.
- **Lesser Security Risk**: It is noted that attacks against machine learning algorithms generally present significantly less security risk compared to the data used for training, testing, and eventual operation.
- **Offline and Online Systems**: The section distinguishes between offline and online machine learning algorithms. Offline systems are trained on a fixed dataset and then used for predictions, while online systems continuously learn and adapt through iterative training with new data.
- **Security Advantages of Offline Systems**: Offline systems are said to have certain security advantages due to their fixed, static nature, which reduces the attack surface and minimizes exposure to data-borne vulnerabilities over time.
- **Vulnerabilities of Online Systems**: Online systems are constantly exposed to new data, which increases their susceptibility to poisoning attacks, adversarial inputs, and manipulation of learning processes.
- **Careful Selection of Algorithms**: The document emphasizes the importance of carefully considering the choice between offline and online learning algorithms based on the specific security requirements and operating environment of the ML system

XI. Evaluation

- **Critical Role of Evaluation**: Evaluation is essential for assessing the effectiveness of machine learning systems in achieving their intended functionalities. It involves using dedicated datasets to systematically analyze the performance of a trained model on its specific task.
- **Evaluation Data Poisoning**: There is a risk of upstream attacks against data, where the data is tampered with before it is used for machine learning, significantly complicating the training and evaluation of ML models. These attacks can corrupt or alter the data in a way that skews the training process, leading to unreliable models.
- **Insufficient Evaluation Data**: Evaluation datasets can also be too small or too similar to the training data to be useful. Poor evaluation data can lead to biases, hallucinations, and toxic output. It is difficult to effectively evaluate large language models (LLMs), as these models rarely have an objective ground truth labeled.
- **Mitigation Controls**:
    - Implementing Single Sign-On (SSO) with Identity Provider (IdP) and Multi-Factor Authentication (MFA) to limit who can access your data and AI platform.
    - Using IP access lists to restrict the IP addresses that can authenticate to Databricks.
    - Encrypting data at rest and in transit.
    - Monitoring data and AI system from a single pane of glass for changes and take action when changes occur.
- **Importance of Robust Evaluation**: Effective evaluation is crucial for ensuring the reliability and accuracy of machine learning models. It helps in identifying discrepancies or anomalies in the model's decision-making process and provides insights into the model's performance.

XII. Machine Learning Models

- **Model Security**: Machine learning models are the core of AI systems, and their security is crucial to ensure the integrity and reliability of the system. The section discusses various risks associated with machine learning models and provides mitigation controls for each risk.
- **Backdoor Machine Learning/Trojaned Model**: This risk involves an attacker embedding a backdoor in the model during training, which can be exploited later to manipulate the model's behavior. Mitigation controls include monitoring model performance, using robust training data, and implementing access controls.
- **Model Asset Leak**: This risk involves the unauthorized disclosure of model assets, such as model architecture, weights, and training data. Mitigation controls include encryption, access control, and monitoring for unauthorized access.
- **ML Supply Chain Vulnerabilities**: This risk arises from vulnerabilities in the ML supply chain, such as third-party libraries and dependencies. Mitigation controls include regular vulnerability assessments, using trusted sources, and implementing secure development practices.
- **Source Code Control Attack**: This risk involves an attacker gaining unauthorized access to the source code repository and modifying the code to introduce vulnerabilities or backdoors. Mitigation controls include access control, code review, and monitoring for unauthorized access.
- **Model Attribution**: This risk involves the unauthorized use of a model without proper attribution to its original creators. Mitigation controls include using digital watermarking, maintaining proper documentation, and enforcing licensing agreements.
- **Model Theft**: This risk involves an attacker stealing a model by reverse-engineering its behavior or directly accessing its code. Mitigation controls include

encryption, access control, and monitoring for unauthorized access.

- **Model Lifecycle without HITL**: This risk arises from the lack of human-in-the-loop (HITL) involvement in the model lifecycle, which can lead to biased or incorrect predictions. Mitigation controls include regular model validation, human review, and continuous monitoring.
- **Model Inversion**: This risk involves an attacker inferring sensitive information about the training data by analyzing the model's behavior. Mitigation controls include using differential privacy, access control, and monitoring for unauthorized access.

XIII. Model Management

- **Model Management Overview**: Model management is the process of organizing, tracking, and maintaining machine learning models throughout their lifecycle, from development to deployment and retirement.
- **Security Risks**: The section outlines various security risks associated with model management, including model attribution, model theft, model lifecycle without human-in-the-loop (HITL), and model inversion.
- **Model Attribution**: This risk involves the unauthorized use of a model without proper attribution to its original creators. Mitigation controls include using digital watermarking, maintaining proper documentation, and enforcing licensing agreements.
- **Model Theft**: This risk involves an attacker stealing a model by reverse-engineering its behavior or directly accessing its code. Mitigation controls include encryption, access control, and monitoring for unauthorized access.
- **Model Lifecycle without HITL**: This risk arises from the lack of human-in-the-loop (HITL) involvement in the model lifecycle, which can lead to biased or incorrect predictions. Mitigation controls include regular model validation, human review, and continuous monitoring.
- **Model Inversion**: This risk involves an attacker inferring sensitive information about the training data by analyzing the model's behavior. Mitigation controls include using differential privacy, access control, and monitoring for unauthorized access.
- **Mitigation Controls**: For each identified risk, the DASF provides detailed mitigation controls. These controls include the use of Single Sign-On (SSO) with Identity Providers (IdP), Multi-Factor Authentication (MFA), IP access lists, private links, and data encryption to enhance the security of model management.
- **Comprehensive Risk Management**: The section emphasizes the importance of a comprehensive approach to managing model security, from the initial development to the deployment and retirement of machine learning models. This includes regular audits, updates to security protocols, and continuous monitoring of model integrity.

XIV. Model Serving and Inference Requests

- **Model Serving**: Model serving is the process of deploying a trained machine learning model in a production environment to generate predictions on new data.
- **Inference Requests**: Inference requests are the input data sent to the deployed model for generating predictions.
- **Security Risks**: The section outlines various security risks associated with model serving and inference requests, including prompt injection, model inversion, model breakout, looped input, inferring training data membership, discovering ML model ontology, denial of service (DoS), LLM hallucinations, input resource control, and accidental exposure of unauthorized data to models.
- **Prompt Injection**: This risk involves an attacker injecting malicious input into the model to manipulate its behavior or extract sensitive information.
- **Model Inversion**: This risk involves an attacker attempting to reconstruct the original training data or sensitive features by observing the model's output.
- **Model Breakout**: This risk involves an attacker exploiting vulnerabilities in the model serving environment to gain unauthorized access to the underlying system or data.
- **Looped Input**: This risk involves an attacker submitting repeated or looped input to the model to cause resource exhaustion or degrade the system's performance.
- **Inferring Training Data Membership**: This risk involves an attacker attempting to determine whether a specific data point was used in the model's training data.
- **Discovering ML Model Ontology**: This risk involves an attacker attempting to extract information about the model's internal structure or functionality.
- **Denial of Service (DoS)**: This risk involves an attacker submitting a large volume of inference requests to overwhelm the model serving infrastructure and cause service disruption.
- **LLM Hallucinations**: This risk involves the model generating incorrect or misleading output due to the inherent uncertainty or limitations of the underlying algorithms.
- **Input Resource Control**: This risk involves an attacker manipulating the input data to consume excessive resources during the inference process.
- **Accidental Exposure of Unauthorized Data to Models**: This risk involves unintentionally exposing sensitive or unauthorized data to the model during the inference process.

XV. Model Serving and Inference Response

- **Model Serving**: Model serving is the process of deploying a trained machine learning model in a production environment to generate predictions on new data.
- **Inference Response**: Inference response refers to the output generated by the deployed model in response to the input data sent for prediction.
- **Security Risks**: The section outlines various security risks associated with model serving and inference response, including lack of audit and monitoring inference quality, output manipulation, discovering ML model ontology, discovering ML model family, and black-box attacks.
- **Lack of Audit and Monitoring Inference Quality**: This risk involves the absence of proper monitoring and auditing mechanisms to ensure the quality and accuracy of the model's predictions.
- **Output Manipulation**: This risk involves an attacker manipulating the model's output to cause incorrect or misleading predictions.
- **Discovering ML Model Ontology**: This risk involves an attacker attempting to extract information about the model's internal structure or functionality by analyzing the output.
- **Discovering ML Model Family**: This risk involves an attacker attempting to identify the specific type or family of the model used in the system by analyzing the output.
- **Black-Box Attacks**: This risk involves an attacker exploiting the model's vulnerabilities by treating it as a black box and manipulating the input data to generate desired outputs.
- **Mitigation Controls**: For each identified risk, the DASF provides detailed mitigation controls. These controls include the use of Single Sign-On (SSO) with Identity Providers (IdP), Multi-Factor Authentication (MFA), IP access lists, private links, and data encryption to enhance the security of model serving and inference response

XVI. Machine Learning Operations (MLOps)

- **MLOps Definition**: MLOps is the practice of combining Machine Learning (ML), DevOps, and Data Engineering to automate and standardize the process of deploying, maintaining, and updating ML models in production environments.
- **Security Risks**: The section outlines various security risks associated with MLOps, including lack of MLOps, repeatable enforced standards, and lack of compliance.
- **Lack of MLOps**: This risk involves the absence of a standardized and automated process for deploying, maintaining, and updating ML models, which can lead to inconsistencies, errors, and security vulnerabilities.
- **Repeatable Enforced Standards**: Enforcing repeatable standards is crucial for ensuring the security and reliability of ML models in production environments. This includes implementing version control, automated testing, and continuous integration and deployment (CI/CD) pipelines.
- **Lack of Compliance**: This risk involves the failure to comply with relevant regulations and industry standards, which can result in legal and financial consequences for the organization.
- **Mitigation Controls**: For each identified risk, the DASF provides detailed mitigation controls. These controls include the use of Single Sign-On (SSO) with Identity Providers (IdP), Multi-Factor Authentication (MFA), IP access lists, private links, and data encryption to enhance the security of MLOps

XVII. Data and AI Platform Security

- **Inherent Risks and Rewards**: The choice of platform used for building and deploying AI models can have inherent risks and rewards. Real-world evidence suggests that attackers often use simple tactics to compromise ML-driven systems.
- **Lack of Incident Response**: AI/ML applications are mission-critical for businesses, and platform vendors must address security issues quickly and effectively. A combination of automated monitoring and manual analysis is recommended to address general and ML-specific threats (DASF 39 Platform security — Incident Response Team).
- **Unauthorized Privileged Access**: Malicious internal actors, such as employees or contractors, can pose a significant security threat. They might gain unauthorized access to private training data or ML models, leading to data breaches, leakage of sensitive information, business process abuses, and potential sabotage of ML systems. Implementing stringent internal security measures and monitoring protocols is crucial to mitigate insider risks (DASF 40 Platform security — Internal access).
- **Poor Security in the Software Development Lifecycle (SDLC)**: Software platform security is an important part of any progressive security program. Hackers often exploit bugs in the platform where AI is built. The security of AI depends on the platform's security (DASF 41 Platform security — secure SDLC).
- **Lack of Compliance**: As AI applications become more prevalent, they are increasingly subject to scrutiny and regulations such as GDPR and CCPA. Utilizing a compliance-certified platform can be a significant advantage for organizations, as these platforms are specifically designed to meet regulatory standards and provide essential tools and resources to help organizations build and deploy AI applications that are compliant with these laws

XVIII. Databricks Data Intelligence Platform

The Databricks Data Intelligence Platform is a comprehensive solution for AI and data management.

- **Mosaic AI:** This component of the platform covers the end-to-end AI workflow, from data preparation to model deployment and monitoring.
- **Databricks Unity Catalog**: This is a unified governance solution for data and AI assets. It provides data discovery, data lineage, and fine-grained access control.
- **Databricks Platform Architecture**: The platform architecture is a hybrid PaaS that is data-agnostic, supporting a wide range of data types and sources.
- **Databricks Platform Security**: The security of the platform is based on trust, technology, and transparency principles. It includes features like encryption, access control, and monitoring.
- **AI Risk Mitigation Controls**: Databricks has identified 55 technical security risks across 12 foundational components of a generic data-centric AI system. For each risk, the platform provides a guide to the AI and ML mitigation control, its shared responsibility between Databricks and the organization, and the associated Databricks technical documentation.

XIX. Databricks AI Risk Mitigation Controls

- **Databricks AI Risk Mitigation Controls**: Databricks has identified 55 technical security risks across 12 foundational components of a generic data-centric AI system. For each risk, the DASF provides a guide to the AI and ML mitigation control, its shared responsibility between Databricks and the organization, and the associated Databricks technical documentation.
- **Shared Responsibility**: The responsibility for implementing the mitigation controls is shared between Databricks and the organization using the platform. Databricks provides the tools and resources needed to implement the controls, while the organization is responsible for configuring and managing them according to their specific needs.
- **Comprehensive Approach**: The Databricks AI risk mitigation controls cover a wide range of security risks, from data security and access control to model deployment and monitoring. This comprehensive approach helps organizations reduce overall risk in their AI system development and deployment processes.
- **Applicability**: The Databricks AI risk mitigation controls are applicable to all types of AI models, including predictive ML models, generative AI models, and external models. This ensures that organizations can implement the appropriate controls based on the specific AI models they are using.
- **Effort Estimation**: Each control is tagged as "Out-of-the-box," "Configuration," or "Implementation," helping teams estimate the effort involved in implementing the control on the Databricks Data Intelligence Platform. This allows organizations to prioritize their security efforts and allocate resources effectively