WHITEPAPER

AI Security Whitepaper for Matillion Data Productivity Cloud

Version 1.0 | 2025 - 06 - 26

matillion

Table of contents

01

AI in Matillion Data Productivity Cloud

Matillion Data Productivity Cloud leverages AI in two key ways. First, it offers Matillion's Maia, a digital agentic AI-powered data team helping the user at each and every step of the data journey, from understanding the data and bringing data pipeline authoring assistance, to operationalizing data integration and augmenting it with operational intelligence.

Second, Matillion provides AI-powered graphical components that can be added directly into data pipelines. These components facilitate integrations with Large Language Models (LLMs) and other AI services, allowing users to build AI-driven data workflows that include capabilities like data transformation, enrichment, and Retrieval Augmented Generation (RAG) pipelines.

Maia

1.1 Maia: Matillion’s Agentic Data Team

Matillion’s Maia, the digital data team in Matillion Data Productivity Cloud, empowers users to generate graphical pipelines using natural language and business prompts, as well as achieving all the tasks required during the data journey. Powered by Cloud AI platforms, Matillion’s Maia streamlines the creation of data pipelines, their operationalisation and their management, making it faster and easier to build even complex integration.

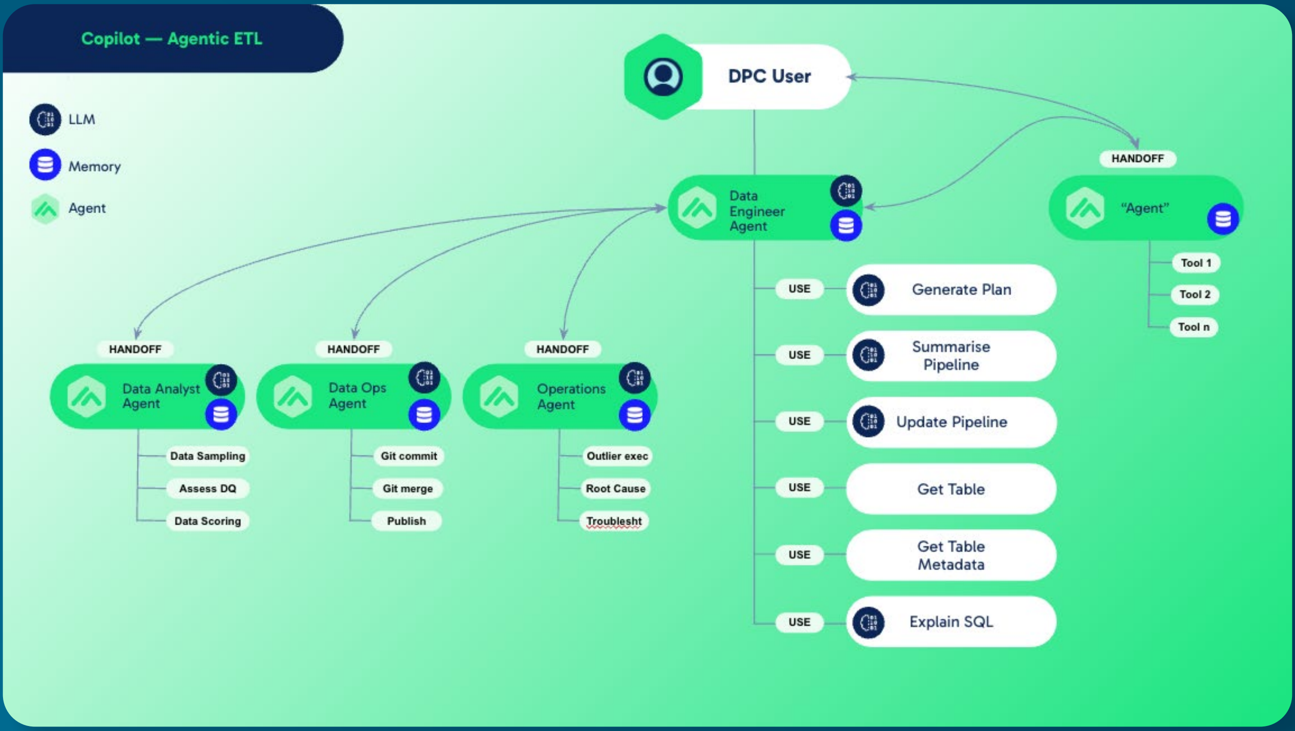
Matillion’s Maia employs an agentic AI architecture, with agents mapped to the personas of a data team (e.g., data analyst, data engineer, DataOps engineer, operations manager). Its tools replicate the tasks of the data team such as adding and editing components, sampling data, validating components, and discovering metadata and data assets.

The main benefits of Matillion’s Maia include:

- ☑ Increased productivity
- ☑ Empowering less technically savvy users
- ☑ The ability to use business language to author technical data pipelines

These benefits and capabilities streamline data pipeline development and make it more accessible and faster to a wider range of users.

Matillion’s Maia Service & Agentic AI



Matillion’s Maia architecture is agentic, meaning it is composed of multiple intelligent agents that work collaboratively to achieve the goal of generating data pipelines. Each agent is designed to mimic the role and responsibilities of a specific persona within a data team. For example, there is a “Data Engineer Agent”, a “Data Analyst Agent”, a “DataOps Agent”, each with its own set of tools and capabilities.

This design is modular because the system is broken down into these discrete agents, each of which performs a specific set of functions. This modularity makes the system easier to understand, maintain, and extend. If a new capability is needed, a new agent or tool can be added to the system without requiring significant changes to the existing architecture.

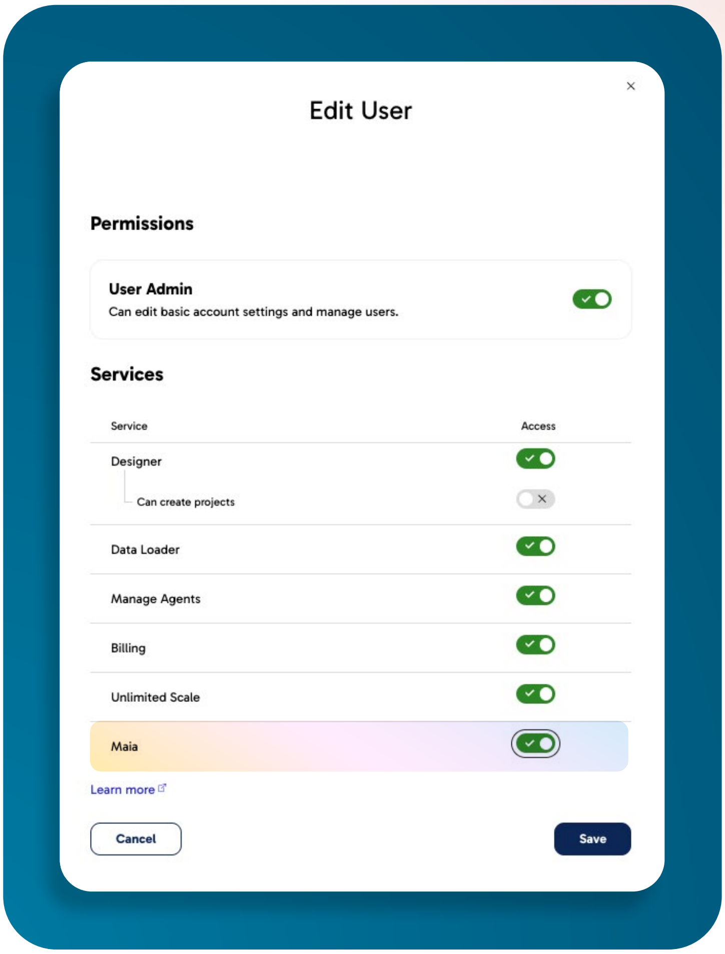
The agents use a variety of “tools” to perform their tasks. These tools represent the different actions that an agent can take, such as “Get Table Metadata,” “Explain SQL,” “Generate Plan,” and “Update Pipeline.” Each agent has access to the tools that are most relevant to its role. For instance, the Data Engineer Agent might use tools to generate an execution plan and update the pipeline, while the Data Analyst Agent might use tools to query data and explain SQL.

This agentic architecture enables Matillion’s Maia to break down complex tasks into smaller, more manageable steps, with each agent handling a specific part of the process. The agents communicate with each other, passing information and control as needed to achieve the overall goal of generating an effective data pipeline.

Matillion’s Maia Access

Customer’s Maia access can be granted or revoked on a per Matillion account basis, using an internal feature-flag that is only controlled by Matillion. This enables AI-sensitive customers to get the guarantee that not a single user can access the AI features. In case such a restriction is needed, it has to go through a special request to Matillion (via the customer’s Account Executive, or via a support ticket).

Matillion’s Maia access can be granted or revoked on a per-user basis using Matillon Hub, by setting the related permission in the user’s RBAC setup.



Edit User

Permissions

User Admin ✓
Can edit basic account settings and manage users.

Services

Service	Access
Designer	✓
Can create projects	✕
Data Loader	✓
Manage Agents	✓
Billing	✓
Unlimited Scale	✓
Maia	✓

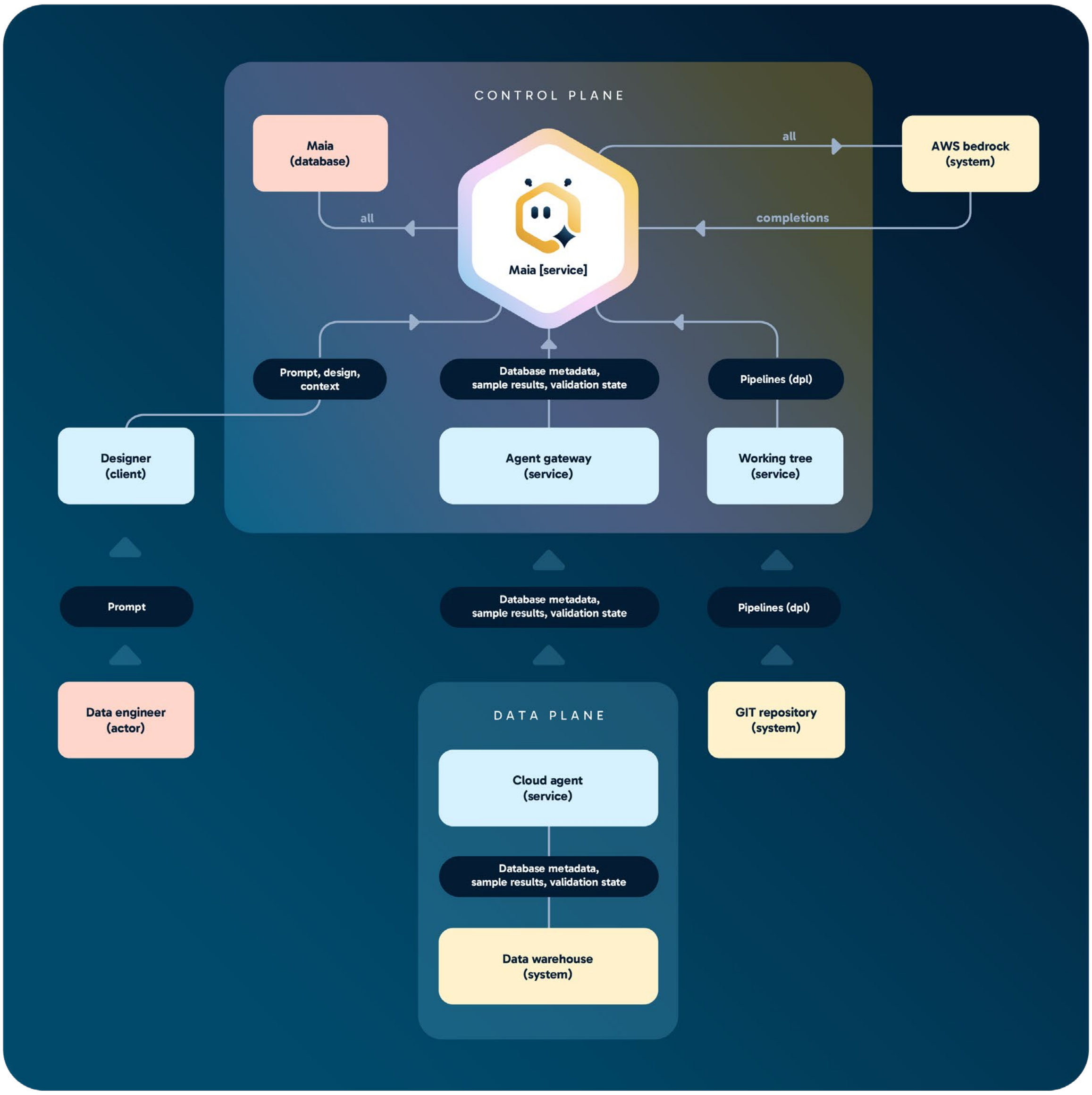
Matillion’s Maia Data Access

Matillion’s Maia requires data that is consistent with the requirements of a data engineer performing similar tasks. Matillion’s Maia uses Amazon Bedrock and its collection of LLMs to power its decision logic and reasoning capabilities. Data may be transmitted to AWS to generate content, but it is never stored or used to train any models. Also, this data never goes outside of Matillion’s Control Plane VPC.

This data may be stored within our internal database, as it is essential for driving the conversation with the AI. The AI cannot effectively reason and generate accurate content to solve a diverse set of problems without the conversation history and knowledge of the context.

In accordance with our security principles, all data is tenancy-bound and accessible only to the specific account and user that initiated a request with Matillion’s Maia.

Matillion's Maia Technical Architecture



The Matillion's Maia architecture is designed to facilitate the generation of data pipelines from natural language prompts. It comprises several key components that work together to understand user intent, design pipelines, and interact with data systems. Here's a breakdown of each section, in line with the representation in the architecture diagram:

Control Plane

- **Designer (Client):**
  This is the user interface where the data engineer interacts with Matillion's Maia, providing prompts and viewing the generated graphical pipeline. Matillion's Maia surfaces as a chat interface in Designer, as well as in other key moments of the data journey (documentation generation, git commit messages generation...).

- **Agent Gateway (Service):**
  The Agent Gateway gathers all resources that are relevant in order to fulfil the generation (getting metadata, sample, validating the individual graphical components).

- **Matillion's Maia Service:**
  This is the core of Matillion's Maia, which uses an agentic AI architecture powered by Amazon Bedrock LLMs to interpret the prompt, generate a pipeline design, and manage the interaction flow. It uses the following information:

  – Database metadata, sample results, and validation: This information is retrieved to provide context for pipeline generation.

  – State: Maintains the conversation history and current state of the pipeline design.

- **Matillion's Maia Database:**
  This database stores all the data required to drive the conversation with the AI, including prompt, design context, sample results, validation of pipelines, and state. The current retention period is 3 days / 72 hours.

- **LLM provider (AWS Bedrock):**
  Matillion's Maia leverages Amazon Bedrock and its collection of LLMs to power its decision logic and reasoning capabilities. Data may be transmitted to AWS to generate content, but it is never stored or used to train any models. While it is currently leveraging Amazon Bedrock, this could change over time.

- **Working Tree (Service):**
  This service represents the current state of the pipeline being designed, including the pipeline definition (DPL), database metadata, sample results, and validation state.

Data Plane

- **Cloud Agent (Service):**
  The Cloud Agent interacts with the data warehouse, retrieving metadata and data samples, and executing validation pipelines.

- **Data Warehouse (System):**
  The data warehouse stores the actual data being processed by the pipelines.

3rd-party services

- **Git Repository (System):**
  The Git repository is used to store and version-control the pipeline definitions, in DPL.

Communication between services

- Connections between the client and the designer backend as well as the designer backend and Bedrock both use HTTPS which ensures message integrity is preserved and any tampering is easily detected by either side.

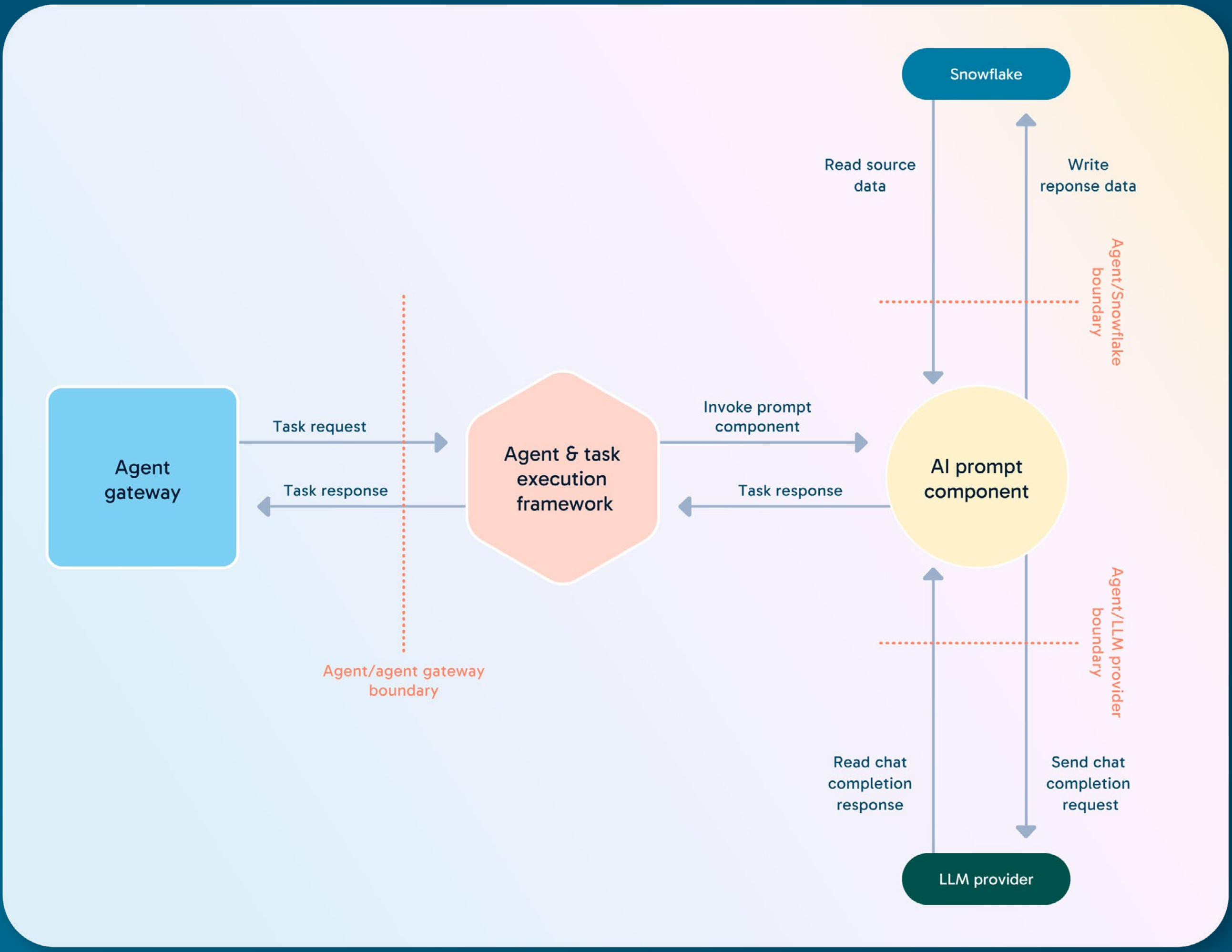
Matillion's Maia uses a combination of LLMs and access to data system information to understand user requests and generate executable data pipelines. The architecture is designed to be modular and extensible, allowing for integration with various data sources and systems.

1.2
Large Language Model Integrations & AI components

Matillion Data Productivity Cloud also provides a set of graphical components that users (or Matillion's Maia) can add to data pipelines. These components facilitate loading unstructured data, integrating with Large Language Models (LLMs), building Retrieval Augmented Generation (RAG) pipelines, and writing to vector stores. They enable seamless API integration with third-party AI services, including Amazon Bedrock, Azure OpenAI, Google Gemini, Pinecone, pg_vector, and pushdown AI transformations with Snowflake Cortex and Databricks AI. Note that this list isn't final and is subject to change in the future.

AI Prompt Components (API integration to LLM services)

The diagram illustrates the data flow and system interactions involved in utilizing an AI Prompt Component within a data processing architecture.



Component Interaction and Security Boundaries

- **Data Source (Cloud Data Platform):**
  Represents the repository of potentially sensitive customer data. Security here hinges on robust access controls to ensure that only authorized users can extract data. This includes evaluating authentication mechanisms, authorization policies, and the principle of least privilege.

- **LLM Provider**
  A third-party service providing AI model capabilities. Data transmitted to this provider falls outside our direct control, necessitating a thorough assessment of the provider's security posture. This assessment should include:

  - Evaluation of their data handling policies.
  - Verification of compliance with relevant regulations (e.g., GDPR, HIPAA).
  - Review of their security certifications and audit reports.
  - Analysis of their vulnerability management practices.

- **AI Prompt Component:**
  This component mediates between the data source and the LLM, forwarding data for processing and receiving the LLM's output. As such, it is a critical security control point. Security considerations include:

  - Input validation to prevent prompt injection attacks.
  - Secure data handling to protect data in transit.
  - Strict access control to the component itself.

02

Security Considerations for AI and LLMs in Matillion Data Productivity Cloud

The integration of Artificial Intelligence (AI) and Large Language Models (LLMs) into the Matillion Data Productivity Cloud introduces a new dimension to data processing, necessitating a comprehensive approach to security. The following sections detail the key security considerations to ensure the confidentiality, integrity, and availability of data and AI-driven processes within the platform.

2.1
Data Privacy & Compliance

The handling of data within Matillion Data Productivity Cloud, particularly in the context of AI and LLM interactions, adheres to stringent data privacy principles and comply with an evolving landscape of regulatory requirements.

- **Data handling policies:**
  Robust data handling policies are paramount. These policies should govern the entire lifecycle of data, from ingestion and processing to storage and disposal, with a specific focus on the transmission of data to and from LLM providers. Data minimization techniques should be employed to ensure that only the minimum necessary data is processed by LLMs.

- **User data isolation and encryption:**
  A fundamental security principle is the isolation of user data. Data should be strictly segregated, with strong access controls to prevent unauthorized access. Encryption, both in transit and at rest, is essential to protect data confidentiality. All data transmitted to LLM providers is encrypted using industry-standard protocols, and encryptions applied to any data stored within Matillion's systems. Matillion integrates to these services, and uses protocols like HTTPS & TLS to guarantee that communications are encrypted from end-to-end.

2.2
Secure AI-Powered Pipeline Authoring

The use of AI to assist in data pipeline authoring introduces new security considerations related to the integrity and trustworthiness of the generated pipelines.

- **Preventing unauthorized access and modifications:**
  Access to AI-powered pipeline authoring tools, such as Matillion Matillion's Maia, must be strictly controlled. Role-based access control (RBAC) should be used to ensure that only authorized users can create or modify pipelines using Matillion's Maia, according to your company policy.

- **Auditing and logging AI-generated pipeline changes:**
  All changes to data pipelines, including those generated by AI, must be thoroughly audited and logged. This provides a clear audit trail, enabling the detection of unauthorized modifications or malicious activity.

- **Mitigating potential biases in AI-generated code:**
  AI models can perpetuate or amplify existing biases in the data they are trained on, potentially leading to the generation of biased or unfair data pipelines. Matillion Data Productivity Cloud should employ techniques to detect and mitigate these biases, promoting fairness and transparency in data processing.

2.3 Secure LLM Integrations in Data Pipelines

The integration of LLMs into data pipelines for data transformation and enrichment requires careful consideration of the security implications of transmitting and processing data with external AI services.

- **Safe handling of sensitive data in LLM-powered transformations:**
  When using LLMs to transform sensitive data, it is crucial to implement appropriate safeguards. This may include techniques such as data masking, anonymization, or differential privacy to protect data confidentiality. Also, services like Amazon Bedrock or Snowflake Cortex allow to use built-in guardrails so that language model responses associated with harmful and non-ethical content - such as violent crimes, hate, sexual content, self-harm and more - are automatically filtered out, that communications are encrypted from end-to-end.

- **Rate limiting and API security:**
  To prevent abuse and ensure the availability of LLM services, rate limiting and API security measures should be enforced. This includes setting appropriate rate limits for API requests and implementing security best practices. If Cloud AI providers providing a free-tier are outside of your policy, it is best to ensure that rules are set to block access to these endpoints.

- **Detecting and mitigating hallucinations in AI responses:**
  LLMs are known to sometimes generate inaccurate or fabricated information, a phenomenon known as "hallucination." Matillion Data Productivity Cloud employs techniques to detect and mitigate these hallucinations, ensuring the accuracy and reliability of data processed by LLMs.

- **Ensuring secure access to external LLM APIs:**
  Access to external LLM APIs must be strictly controlled. This involves the implementation of strong authentication and authorization mechanisms, such as API keys, access tokens, and multi-factor authentication (when available), to prevent unauthorized access. Regular security audits of API access controls are essential. This governance process sits outside of Matillion, so customers are expected to have a continuous scrutiny on these accesses.

- **Model monitoring and version control:**
  LLMs are subject to ongoing development and refinement. It is therefore essential to implement robust model monitoring and version control mechanisms. This includes tracking model versions, monitoring model performance, and establishing procedures for addressing model updates and potential vulnerabilities.

- **Handling adversarial inputs and prompt injection risks:**
  LLMs are vulnerable to adversarial inputs and prompt injection attacks, which can manipulate model behaviour or lead to the disclosure of sensitive information. Matillion Data Productivity Cloud implements robust input validation and sanitization techniques to mitigate these risks.

Find out more or join a live demo

Visit Matillion.com/maia

matillion