Informatica

White Paper

Informatica Intelligent Data Management Cloud™:

CLAIRE® Security and Privacy Overview

Where data & AI come to LIFE

Contents

Summary

This paper outlines Informatica's approach to security, privacy and ethical AI development for CLAIRE® within the Intelligent Data Management Cloud™ (IDMC) platform. Key points covered in the paper include:

1. **Data Usage and Protection**
   - CLAIRE uses customer metadata for training within the customer environment, however, customer metadata or data is not used for large language model (LLM) training across customer environments.
   - Strong encryption (AES 256-bit) is used for data at rest, with customer-managed key options available.
2. **AI Model Development**
   - CLAIRE implements both customer-specific and agnostic models.
   - Training uses anonymized metadata, public datasets and Informatica documentation.
   - Customers can opt out of metadata sharing for AI training.
3. **Data Residency and Access**
   - Multiple locales for model training ensure data does not cross geographical boundaries.
   - Strict access controls are in place, adhering to user permissions.
4. **Ethical AI Principles**
   - Focus on enhancing human productivity in data management.
   - Commitment to data security, transparency and responsible AI democratization.
   - Avoidance of AI applications that could cause harm or violate rights.
5. **Secure Development Lifecycle**
   - Multiple layers of security testing including automated scans and internal and third-party assessments.
   - Threat modeling and security architecture reviews are integral to the process.
6. **Compliance and Certifications**
   - IDMC holds various certifications including SOC Type II, ESN, UK Cyber Essential Plus, and FedRAMP.

This comprehensive approach demonstrates Informatica's commitment to secure, responsible and innovative AI-powered data management.

Introduction

CLAIRE® is a metadata-powered AI engine that is a key component of the Informatica Intelligent Data Management Cloud (IDMC) platform. It leverages artificial intelligence and machine learning to enhance and automate data management capabilities.

CLAIRE GPT is a generative AI (GenAI)-powered IDMC service that provides a natural language interface to help data analysts, data scientists and data engineers discover data for analytics, explore metadata to determine whether the asset is fit for use, find data insights and automate ELT pipelines to build data products.

This paper provides an overview of the security and privacy considerations of CLAIRE, detailing how Informatica approaches metadata and data, implements encryption, trains AI models and adheres to ethical as well as responsible AI principles.

We will explore the types of data used by CLAIRE, the separation between customer-specific and agnostic models, data residency practices and options for customers to control their data usage. Additionally, this paper outlines Informatica's secure development lifecycle, risk mitigation strategies and compliance certifications – all aimed at ensuring the responsible and secure use of AI within the IDMC platform.

By providing transparency into these practices, Informatica aims to build trust with our customers and demonstrate our commitment to data privacy, security and ethical AI development as we continue to innovate in intelligent data management.

Metadata and Data Used for CLAIRE Training

Metadata and Data Stored in Intelligent Data Management Cloud™

Informatica Intelligent Data Management Cloud (IDMC) stores different types of information:

Metadata

- **Operational and usage metadata**: Includes information extracted from service and activity logs, such as connection and schedule data. It also includes information about how cloud services are used, such as transactions conducted in the customer's data marketplace.
- **Technical metadata**: Includes data schemas, rules and data profile statistics, and design metadata that define integration tasks and processes, such as data sync, data replication, mappings and templates, task flows, process definitions and data lineage.
- **Business metadata**: Information related to customer data, including data classifications and glossaries designated by the customer, crowdsourced information such as data product ratings and comments, etc.

The collection of metadata by IDMC is necessary to provide cloud services and cannot be disabled.

Data

Depending on how IDMC features are configured and enabled, customers may need to store partial or full records of business data from applications and endpoints accessed through IDMC.

However, CLAIRE does not use business data stored in applications for either AI or large language model (LLM) training. For more information, see Assets Used to Train CLAIRE Models below.

Encryption for Data and Metadata at Rest

Any data or metadata persisted in the IDMC multi-tenant data repository is encrypted using the AES encryption algorithm, which uses a 256-bit key. The key is unique to the customer's tenant. By default, the key is rotated once a year, but it can be configured to be rotated every 90, 120 or 180 days.

Customer-managed Encryption Keys

IDMC supports customer-managed keys. While Informatica uses strong encryption practices, customers can utilize their own encryption keys to safeguard data. With this feature, a customer can hold their encryption keys and maintain the authority to encrypt and decrypt their data within the Informatica cloud environment. This gives customers confidence that their data remains confidential and protected, even from Informatica as the service provider.

Customer-specific vs. Customer-agnostic CLAIRE Models

IDMC implements various AI models with different scopes to deliver CLAIRE functionality:

- **Customer-specific models**: CLAIRE models that are tailored to meet the unique needs of individual customers. These models are trained using data and metadata specific to each customer, allowing Informatica to provide services customized to that particular customer.
- **Customer-agnostic models**: CLAIRE models that serve all customers, including CLAIRE GPT. These models are trained on Informatica product documentation, public data sets and aggregated anonymized metadata.

Assets Used to Train CLAIRE Models

CLAIRE customer-agnostic models are trained on metadata, which includes technical metadata, operational and usage metadata and customer business metadata.

This metadata is fully anonymized and aggregated before being used for training, ensuring that no sensitive information is exposed during the process.

Informatica never accesses or uses customer data stored in IDMC for training CLAIRE customer-agnostic models. This approach ensures that the customer's data remains private and secure.

Informatica does not access customer data endpoints (such as Salesforce, SAP or Snowflake) or other cloud applications, databases or data warehouses for CLAIRE or LLM training.

The only exception is the customer-driven fine-tuning of Master Data Management (MDM) match and merge tenant-specific models, which is explicitly performed by the customer. This data is not used to cross-train CLAIRE models and is only utilized by customer-specific CLAIRE models.

Informatica does not reuse data from prompt conversations that occur within CLAIRE GPT or any other IDMC service that uses CLAIRE GPT. Informatica does not use specific customer prompt commands for cross-training CLAIRE models, only for customer-specific models.

Finally, other sources of public data are used to train Informatica customer-agnostic large language models, such as:

- General world knowledge: wiki data, books, public articles
- Verticalized domain knowledge: industry terms, industry metrics, industry systems
- Informatica documentation and Knowledge Base articles

The following table shows which assets are used to train CLAIRE:

Asset Type	Opt Out	Cross Training	Org Training
Technical metadata	Yes	Yes	Yes
Operational and usage metadata	Yes	Yes	Yes
Business metadata	Yes	Yes	Yes
Customer data stored on IDMC	N/A	No	Yes (User explicit)
Customer data on applications	N/A	No	No
Prompt commands on CLAIRE AI	Yes	No	Yes (Feedback)
General world knowledge data	N/A	Yes	N/A
Verticalized domain knowledge data	N/A	Yes	N/A
Informatica Knowledge Base data	N/A	Yes	N/A

CLAIRE Model Fine-Tuning

Customers can refine selected CLAIRE models based on their own data. The data and the refined models are unique and are not shared across other tenants. They are stored within the context of the encrypted tenant and are available only to the specific customer.

Data Access

At Informatica, we recognize and understand the importance of the security and privacy of customer data. CLAIRE GPT provides data exploration capabilities so customers can ask questions regarding data using natural language.

These data exploration tasks fetch data in real-time, do not persist it in any repository, and only use it in the user session on CLAIRE GPT while it is open. The data is only viewable during the active user session. When the user logs back in, it will no longer be visible.

The CLAIRE GPT service will honor the platform RBAC model and permissions. For example, users will only be able to perform data exploration over connections for which they have permissions.

Localization and Residency

Residency Observation

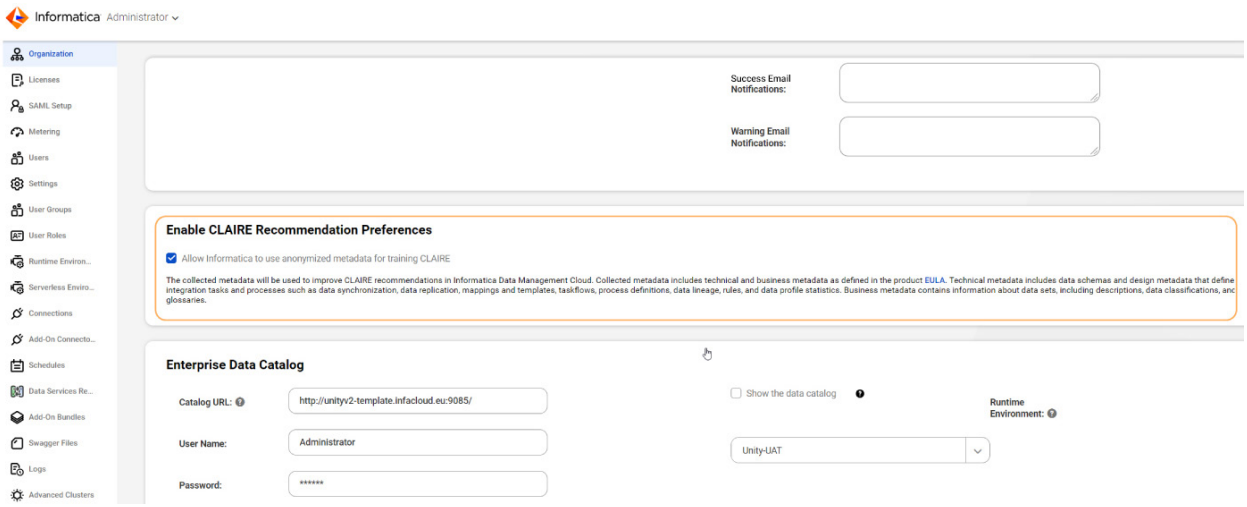
Complying with data residency regulations is a top priority. To ensure compliance, Informatica orchestrates multiple locales for LLM model training, safeguarding that metadata and data do not traverse beyond the delineated boundaries.

To adhere to regulatory frameworks, Informatica deploys distinct instances of CLAIRE models across designated geographies, encompassing the Americas, European Union, United Kingdom, APAC and beyond. This approach underscores our commitment to protecting data integrity and complying with regional compliance standards.

Opt-in, Opt-out and Disposition of Data

Opting Out of Metadata Sharing

Customers can opt out of providing metadata for training CLAIRE at any time through the Administrator counsel. This allows customers complete control over the use of their metadata at any point in time.



Note: Opting out may result in loss of non-critical CLAIRE functionality.

Disposition of Data

Informatica's policy is to retain processed customer data and customer-specific metadata for at least thirty (30) days after termination or expiration of a customer's subscription to the cloud service and to delete processed customer data and de-identify or delete customer-specific metadata within sixty (60) days of termination or expiration of customer's subscription to the cloud service.

Informatica will promptly comply to the extent practicable with written requests to destroy processed customer data within shorter time periods than those indicated above and provide written certification of destruction of processed customer data upon the customer's written request.

For more details on Informatica data retention and data destruction policies, please visit **Informatica Security Addendum**.

Development and Design Principles

Ethical and Responsible AI

Many Informatica products and services feature technology that enables our users to process information with an increasing degree of autonomy. At Informatica, we understand the profound impact of artificial intelligence (AI) that makes this automation possible, and we guide our AI development with an ethical, responsible and comprehensive set of principles.

These principles are designed to ensure that the AI technologies we create and deploy are developed and used in a way that respects human rights, contributes to societal benefits, upholds privacy and security, prioritizes transparency and explainability and strives for inclusivity and diversity.

We aim to democratize AI, providing tools that are accessible to all users regardless of technical expertise. Our commitment also extends to designing AI for deployment in ways that will not harm or undermine our values.

CLAIRE Principles

Informatica's main principles when designing and developing CLAIRE copilot and CLAIRE GPT applications and features are as follows:

- **Focus on Enhancing Human Productivity in Data Management**: We aim to develop AI technologies for data management making it easy for data teams and business users to manage their data effectively. By narrowing our focus, we aspire to deliver impactful solutions while tailoring our technologies to the unique needs and challenges within this area.
- **Ensure Data Security and Accountability**: We pledge to create AI technology that prioritizes data privacy and security and balances them against the functionality of our product features. AI development oversight includes third-party audits, robust feedback mechanisms and a dedicated oversight team. We will maintain documentary evidence of how our AI is trained. This will ensure transparency in our processes and trust in our operations.
- **Provide Transparency and Explainability**: We aim to create AI models that are effective and understandable. We leverage advanced explainability frameworks and tools to provide insights into how our models make decisions, allowing users to understand how our AI application reaches conclusions.
- **Design Delightful User Experiences**: We aim to harness AI to augment human productivity by crafting thoughtfully designed user experiences that delight end-users.
- **Democratize AI Responsibly**: We are committed to making AI accessible to a broad range of users while maintaining a strong focus on ethical and privacy considerations. We balance openness with robust control mechanisms designed to prevent misuse of technology and protection of data privacy.

Risk Mitigation

Informatica acknowledges the potential risks and ethical issues associated with AI applications. Accordingly, Informatica will not develop or deploy AI:

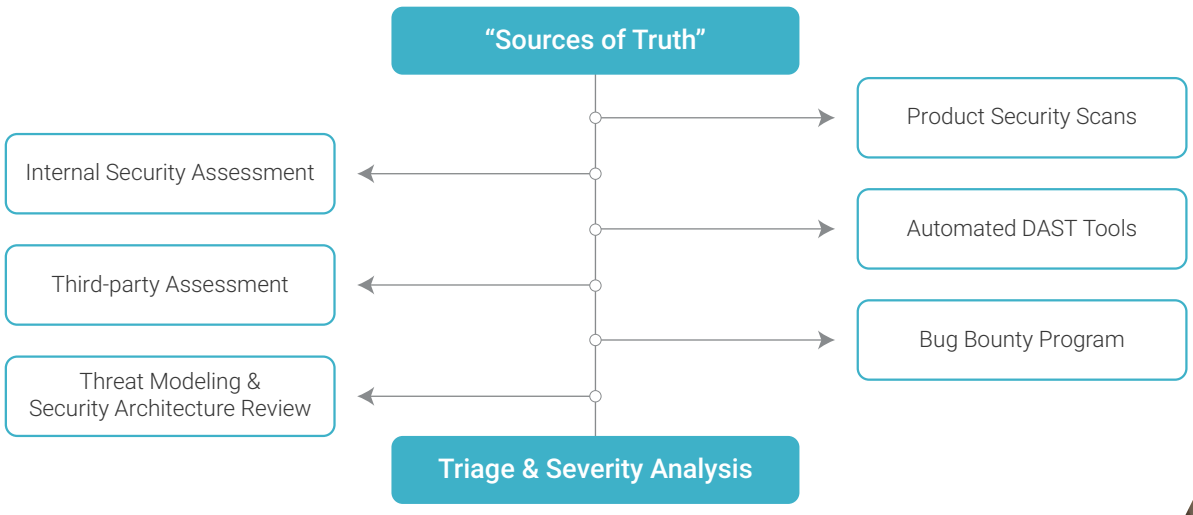
- In ways that are designed to cause or create an undue risk of harm to individuals or society.
- For purposes that are prohibited in the major jurisdictions in which we do business or otherwise constitute a clear threat to the safety, livelihoods and rights of people.

Informatica is committed to using AI to make the world a better place, not to harm or disadvantage any individual or group. We believe in using AI responsibly, and we are dedicated to following these principles as we develop and deploy AI technologies. We understand that the field of AI is rapidly evolving, and thus, we will reassess and update these principles to keep pace with technological advancements and emerging ethical considerations. We firmly believe that by adhering to these principles, we can drive progress while ensuring the responsible use of AI.

Secure Development Lifecycle

Informatica follows a thoughtfully crafted Secure Development Lifecycle to preemptively pinpoint and resolve security concerns in our products. This, in turn, cultivates a vigilant, security-conscious mindset among our product engineering teams and significantly reduces the likelihood of security breaches.

Informatica's approach to CLAIRE AI security spans several key stages, each designed to integrate security measures seamlessly within the development process, ensuring that security considerations are embedded from the outset and throughout the lifecycle. Vulnerability discovery is conducted through multiple layers of identification sources, or what we internally refer to as "sources of truth".



Vulnerability Identification — "Sources of Truth"

Product Security Scans

Informatica products undergo automated scans driven by continuous integration and continuous delivery (CI/CD) to identify vulnerabilities at various levels, from third-party components to application source code. This foundational aspect of our shift left security strategy enables early detection and remediation during the software development lifecycle (SDLC).

Internal Security Assessment

Informatica's team of cybersecurity engineers collaborate closely with development teams on each epic and story created during development sprints, ensuring continuous application security assurance. Informatica follows OWASP guidelines.

Automated DAST Tools

Integrated with CI/CD pipelines, dynamic application security testing (DAST) tools automate repetitive tasks in vulnerability discovery and analysis, providing rapid feedback to developers.

Third-Party Security Assessment

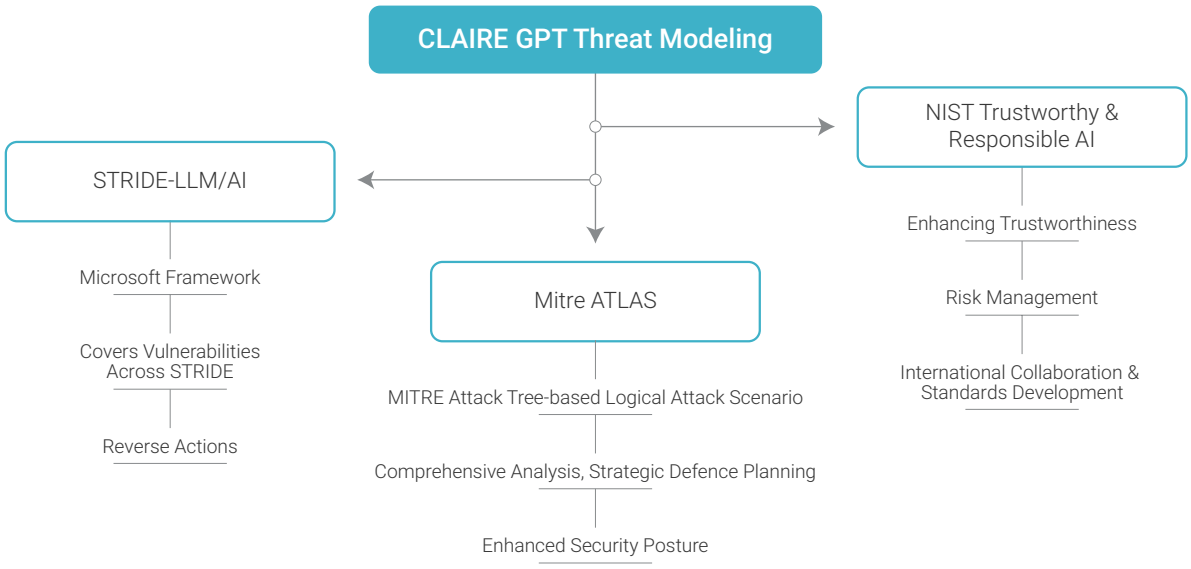
Informatica engages reputable third-party cybersecurity assessment providers to ensure continuous, industry-standard assessments of our products.

Bug Bounty Program

Through a private vulnerability rewards program, Informatica invites professional hackers to test our external attach surfaces for vulnerabilities and recognizes those hackers for their efforts.

Threat Modeling and Security Architecture Review

Informatica employs treat modeling (TM) and security architecture review (SAR) to systematically identify and assess potential threats, vulnerabilities and risks during the design phase of systems or applications.



Threat Modeling Framework

Triage and Severity Analysis

Informatica utilizes industry-standard practices (CVSS, Mitre Attack Framework, OWASP, EPSS, NIST) for classifying, risk-rating and describing cyberattacks and intrusions. Our product security team rigorously assesses each vulnerability identified to understand its potential exploitability, consequences, severity and impact.

Informatica IDMC Certifications and Compliance

Certifications and Compliance

The security of customer data is a critical objective of the IDMC platform. Informatica has established a risk-based information security program to protect Informatica and our customers' data security and privacy.

Informatica has voluntarily undertaken and/or is required by contractual obligation to perform in accordance with the standards listed below. These standards are measured by internal security teams and champions, third parties, and external assessment partners such as AICPA-accredited external audit firms.

Among others, Informatica is SOC Type II, ESN, UK Cyber Essential Plus and FedRAMP certified.

For a complete list of certifications, assessments and standards for IDMC, please visit **Informatica Trust Center**.

About Us

Informatica (NYSE: INFA), a leader in enterprise AI-powered cloud data management, brings data and AI to life by empowering businesses to realize the transformative power of their most critical assets. We have created a new category of software, the Informatica Intelligent Data Management Cloud™ (IDMC), powered by AI and an end-to-end data management platform that connects, manages and unifies data across virtually any multi-cloud, hybrid system, democratizing data and enabling enterprises to modernize their business strategies. Customers in approximately 100 countries and more than 80 of the Fortune 100 rely on Informatica to drive data-led digital transformation. **Informatica. Where data and AI come to life.™**

Worldwide Headquarters
2100 Seaport Blvd.
Redwood City, CA 94063, USA
Phone: 650.385.5000
Fax: 650.385.5500
Toll-free in the US: 1.800.653.3871

informatica.com
linkedin.com/company/informatica
x.com/Informatica

CONTACT US

IN09-4999-0824