SecureAIFlow

# Zero Trust at the AI Boundary:

## Security & Deployment Guide

Data flows, network boundaries, and compliance posture
for on-premises and SaaS deployment models

Author: BOUAOUDA ACHRAF

Version 1.0 · 2026
secureaiflow.com

© 2026 SecureAIFlow. All rights reserved .secureaiflow.com

## Abstract

This document describes the security architecture, data flows, and network boundaries of SecureAIFlow (SAF) across its two supported deployment models: on-premises and SaaS. It is intended for security architects, compliance officers, and procurement teams evaluating SAF for regulated enterprise environments.

The document covers: (1) the threat model and scope of detection, (2) the shared security responsibility model, (3) data flow and network boundary specifications for each deployment model, (4) on-premises integration points, (5) regional deployment for SaaS, (6) compliance posture for applicable regulatory frameworks, and (7) a security controls summary.

This document does not disclose proprietary detection methodology, model architecture, or internal implementation details. Those aspects are available under non-disclosure agreement upon request.

## Contents

© 2026 SecureAIFlow. All rights reserved .secureaiflow.com

# 1. Introduction and Scope

AI-assisted tools : embedded in IDEs, web browsers, and CI/CD pipelines , transmit user prompts to third-party Large Language Model (LLM) providers. These prompts routinely contain source code, configuration files, environment variables, and business data in which credentials, API keys, secrets, and personally identifiable information (PII) may be embedded. This transmission path exists outside the access controls, audit logging, and data classification policies that govern conventional workflows.

SecureAIFlow addresses this gap by intercepting prompts before transmission, identifying sensitive material, and replacing it with deterministic pseudonyms. The LLM provider receives the semantic content of the prompt without receiving the sensitive values. Responses are post-processed to restore original values within the customer's environment.

## 1.1 Intended Audience

This document is written for:

- Security architects and engineers evaluating SAF for enterprise deployment
- Compliance officers assessing regulatory alignment (GDPR, PCI DSS, SOC 2, ISO 27001, NIS2, EU AI Act)
- Procurement and legal teams conducting vendor due diligence
- CISOs reviewing security controls and the responsibility boundary

## 1.2 Scope of Detection

SAF detection covers two primary categories of sensitive material:

| Category | Examples | Primary Exposure Context |
|---|---|---|
| Credentials and secrets | API keys, passwords, access tokens, private keys, connection strings, service tokens | All contexts: IDE, browser, SAF UI, automated pipelines |
| PII (via integration) | Customer identifiers, financial references, emails, domain-specific sensitive fields defined by the customer | Non-IDE contexts primarily: direct browser interaction with AI interfaces, SAF web UI, support and analyst workflows |

**NOTE** PII detection is available through optional integration with customer microservices or databases. SAF does not perform PII detection by default without this integration. See Section 4.4 for integration details.

## 1.3 User Base

SAF is designed for use across all organizational roles that interact with AI systems, not exclusively software developers:

- **Developers and engineers:** IDE-integrated workflows, code review, configuration management

- **Data analysts:** Direct browser interaction with AI interfaces; may paste data extracts containing customer identifiers or financial fields
- **Support agents:** AI-assisted ticket resolution; may include customer PII or account credentials in prompts
- **Operations and DevOps teams:** Automated pipelines forwarding configuration or infrastructure data to LLMs

## 1.4 Interaction Contexts

Sensitive data exposure through AI prompts occurs across the following interaction contexts:

- **IDE-integrated AI assistants:** Tools receiving code files or selections as context (VS Code, JetBrains, Cursor, Windsurf, GitHub Copilot)
- **Browser-based AI interfaces:** Direct use of ChatGPT, Claude, Gemini, and equivalent by any organizational user
- **SAF web UI:** SAF's own interface for multi-LLM interaction, in which users may paste prompts containing sensitive data
- **API-driven pipelines:** Automated workflows forwarding application code, configuration, or data to LLMs for analysis or transformation

Traditional secret scanning tools operate at the repository commit layer and do not address any of these real-time transmission contexts. SAF is designed specifically for this gap.

# 2. Threat Model

## 2.1 Primary Threat

The primary threat addressed by SAF is the inadvertent transmission of sensitive material embedded in prompts to LLM providers operating outside the organization's security perimeter.

| Attribute | Description |
|---|---|
| Threat type | Data exfiltration through AI interaction channel |
| Actor | Unintentional (any AI user) : not necessarily malicious |
| Vector | LLM API request containing embedded credential or PII material |
| Target | API keys, passwords, private keys, access tokens, customer PII, financial data |
| Impact | Sensitive data exposure to third-party infrastructure; potential downstream breach or regulatory violation |
| Existing controls gap | Secret scanners operate at git commit layer only; DLP tools cover file/email, not AI prompt layer |

## 2.2 Sensitive Data Categories

Table 1 describes the credential and secret categories in scope for SAF detection.

**Table 1 : Credential Categories in Scope**

| Category | Examples |
|---|---|
| Cloud provider credentials | AWS access/secret keys, GCP API keys, Azure credentials, DigitalOcean tokens |
| AI and LLM API keys | OpenAI, Anthropic, Hugging Face, and equivalent provider keys |
| Source control tokens | GitHub PATs, GitHub OAuth tokens, GitLab PATs |
| Payment platform keys | Stripe live/test keys and equivalent |
| Communication platform keys | Twilio SIDs, Slack bot tokens, SendGrid keys, Mailgun keys |
| Infrastructure credentials | HashiCorp Vault tokens, Kubernetes service tokens, SSH/RSA private keys , etc |
| Database credentials | Connection URIs (PostgreSQL, MySQL, MongoDB, Redis), standalone passwords |
| Session and auth tokens | JWTs, internal API tokens, Basic Auth headers, application secrets |
| PII (via integration) | Customer identifiers, emails, financial references : configured per deployment |

# 3. Shared Security Responsibility Model

Security responsibilities are divided between SecureAIFlow and the customer. This division varies by deployment model. The tables below define the boundary for each model.

## 3.1 On-Premises Deployment

| SecureAIFlow Responsible For | Customer Responsible For |
|---|---|
| Detection algorithm accuracy and correctness | VM and  infrastructure provisioning and hardening |
| New version releases and detection rule updates | Rules configuration |
| Integration connector software (VS Code, browser, API endpoint) | Network configuration, firewall rules, and perimeter security |
| Documentation, security guidance, and release notes | Data residency |
| Responding to reported detection issues | Integration configuration (secret manager, database, internal API) |
| Support | |

**NOTE** SecureAIFlow provides software releases and detection updates. Patching of the operating system, virtual machine infrastructure, and third-party dependencies within the customer environment is the customer's responsibility.

## 3.2 SaaS Deployment

| SecureAIFlow Responsible For | Customer Responsible For |
|---|---|
| Detection algorithm accuracy and correctness | API key management and rotation |
| New version releases and detection rule updates | User identity and authentication |
| SaaS infrastructure availability, patching, and security | Acceptable use policy enforcement for all AI users |
| TLS encryption of all data in transit | Interpretation and use of KPI dashboard data |
| Zero data retention enforcement (technical controls) | Integration configuration for PII-aware detection |
| Regional data residency (customer region assignment) | |
| Anonymized KPI metrics and customer dashboard | |

**NOTE** In SaaS deployment, SecureAIFlow acts as a data processor under GDPR Article 28. A standard Data Processing Agreement is incorporated by reference into the Terms of Service and applies automatically for all customers.A copy is available at secureaiflow.com/legal-dpa.

© 2026 SecureAIFlow. All rights reserved .secureaiflow.com

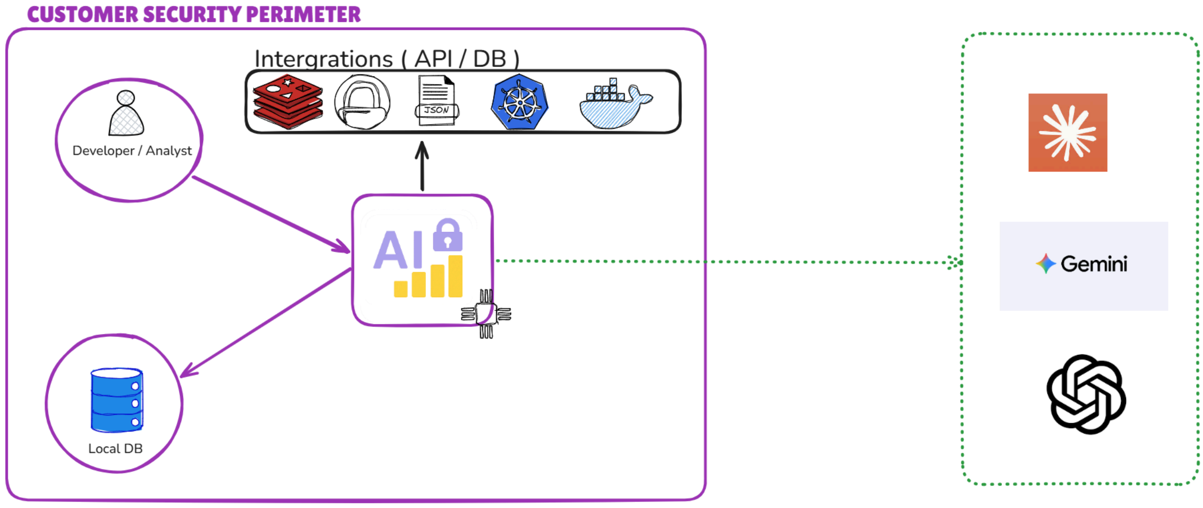
# 4. On-Premises Deployment

In on-premises deployment, the SAF detection engine runs exclusively within the customer's infrastructure. SecureAIFlow has no network access to the customer environment and no visibility into prompts, credentials, or audit logs. The customer assumes full infrastructure responsibility as defined in Section 3.1.

## 4.1 Architecture Overview

Figure 1 illustrates the on-premises deployment topology, including optional integration points.

**Figure 1 : On-Premises Deployment Topology**



## 4.2 Data Flow Specification

Table 2 specifies each data flow step, data content, and boundary status.

**Table 2 : On-Premises Data Flow**

| Step | From → To | Data Content | Boundary Status |
|---|---|---|---|
| 1 | Tooling → SAF Engine | Raw prompt (code, config, env vars, business data) | Within customer perimeter |
| 2 | SAF Engine ↔ Secret Manager | Vault lookup for managed credentials (optional) | Within customer perimeter |
| 3 | SAF Engine ↔ Customer DB / Microservice | PII schema or identifier lookup (optional) | Within customer perimeter |
| 4 | SAF Engine → Local Storage | Pseudonym map: token ↔ original value | Within customer perimeter |

© 2026 SecureAIFlow. All rights reserved .secureaiflow.com

| Step | From → To | Data Content | Boundary Status |
|---|---|---|---|
| 5 | SAF Engine → Audit Log (DB / file) | Detection event: timestamp, pseudonym ID, detection stage | Within customer perimeter |
| 6 | SAF Engine → LLM Provider | Redacted prompt: credentials and PII replaced by pseudonyms | Crosses boundary : no sensitive content |
| 7 | LLM Provider → SAF Engine | LLM response containing pseudonym references | Crosses boundary : no sensitive content |
| 8 | SAF Engine → Tooling | Restored response: pseudonyms replaced with original values | Within customer perimeter |

## 4.3 Network Boundary Definition

Table 3 defines what data does and does not cross the customer's network boundary.

**Table 3 : On-Premises Network Boundary**

| Data Element | Crosses Boundary | Rationale |
|---|---|---|
| Credential values (secrets) | Never | Pseudonymized before any outbound transmission |
| PII fields (when integration active) | Never | Pseudonymized before any outbound transmission |
| Raw prompt content | Never | Processed locally; and forwarded to Gen AI |
| Source code | Never | Contained within raw prompt; same boundary rule applies |
| Pseudonym map and keys | Never | Customer-controlled local storage only |
| Audit logs | Never | Written to customer-controlled storage; |
| SAF software telemetry | Never | No telemetry transmitted in on-premises configuration |
| Redacted prompt (pseudonyms only) | Yes : to LLM provider | No sensitive content; pseudonyms are opaque tokens |
| LLM response (pseudonyms) | Yes : from LLM provider | No sensitive content; restoration occurs locally |

## 4.4 Integration Points

On-premises deployment supports the following integration points. All integrations are optional and configured by the customer.

**Table 4 : On-Premises Integration Points**

| Integration | Supported Systems | Purpose |
|---|---|---|
| Secret manager | HashiCorp Vault, AWS Secrets Manager, Azure Key Vault | Deterministic detection of all managed credentials with 100% precision. Any value |

| Integration | Supported Systems | Purpose |
|---|---|---|
| | | registered in the vault is detected regardless of format or encoding. |
| Customer REST API | Customer-defined internal LLM endpoints | Routes redacted prompts to internal or self-hosted LLM endpoints instead of third-party providers. Enables fully air-gapped deployments. |
| Database / microservice | PostgreSQL, MySQL, internal microservices | Enables PII-aware detection by providing SAF with customer-defined sensitive field schemas or identifier patterns. No customer data is copied to SAF : lookup only. |
| Audit log persistence | PostgreSQL, MySQL, or file system | Persists SAF detection events to customer-controlled storage for compliance, forensic analysis, and SIEM integration. |

**NOTE** All integration traffic remains within the customer perimeter. No integration data is transmitted to SecureAIFlow infrastructure.

## 4.5 Security Controls : On-Premises

- **Pseudonymization:** Credential and PII values replaced with deterministic opaque tokens before any network transmission. Mapping maintained in customer-controlled storage.

- **Audit logging:** All detection events logged locally with timestamp, detection stage, and pseudonym reference. Logs contain no original sensitive values.

- **Secret manager integration:** When connected, achieves 100% precision on vault-registered credentials regardless of encoding or format.

- **PII integration:** Optional lookup-based integration with customer databases or microservices for domain-specific PII detection. No customer data stored by SAF.

- **Air-gapped operation:** When integrated with a customer REST API endpoint, all LLM traffic can be contained within the customer perimeter with zero internet egress.

- **Zero telemetry:** No usage, diagnostic, or operational data is transmitted to SecureAIFlow infra in on-premises configuration.

- **Latency:** End-to-end detection and pseudonymization completes within 500 milliseconds on a standard CPU virtual machine. No GPU required.

## 4.6 Compliance Posture : On-Premises

| Framework | Relevant Requirement | SAF On-Premises Control |
|---|---|---|
| GDPR Art. 32 | Technical measures for data security | Credentials and PII pseudonymized before leaving the data controller's environment. No personal data transmitted to SAF. |

© 2026 SecureAIFlow. All rights reserved .secureaiflow.com

| Framework | Relevant Requirement | SAF On-Premises Control |
|---|---|---|
| GDPR Art. 44 | Data transfers to third countries | All sensitive data processing occurs within the customer's jurisdiction. No credential or PII data leaves the EU boundary. |
| PCI DSS Req. 3 | Protection of cardholder data | Credentials never stored by SAF. Pseudonymization is in-memory; mapping is customer-controlled. |
| PCI DSS Req. 12 | Audit trail | Detection event logs generated locally under exclusive customer control. |
| SOC 2 (Confidentiality) | Protection of confidential information | SAF has zero logical access to customer sensitive material in on-premises deployment. |
| ISO 27001 A.8 | Information asset management | Credentials and PII classified and pseudonymized before any external interface. SAF is a software tool, not a data processor. |
| NIS2 Art. 21 | Cybersecurity risk measures | Prompt interception prevents exfiltration through AI supply chain. Full audit trail under customer control. |
| EU AI Act | Human oversight requirements | Interception layer maintains governance over sensitive data flow through AI interactions. |

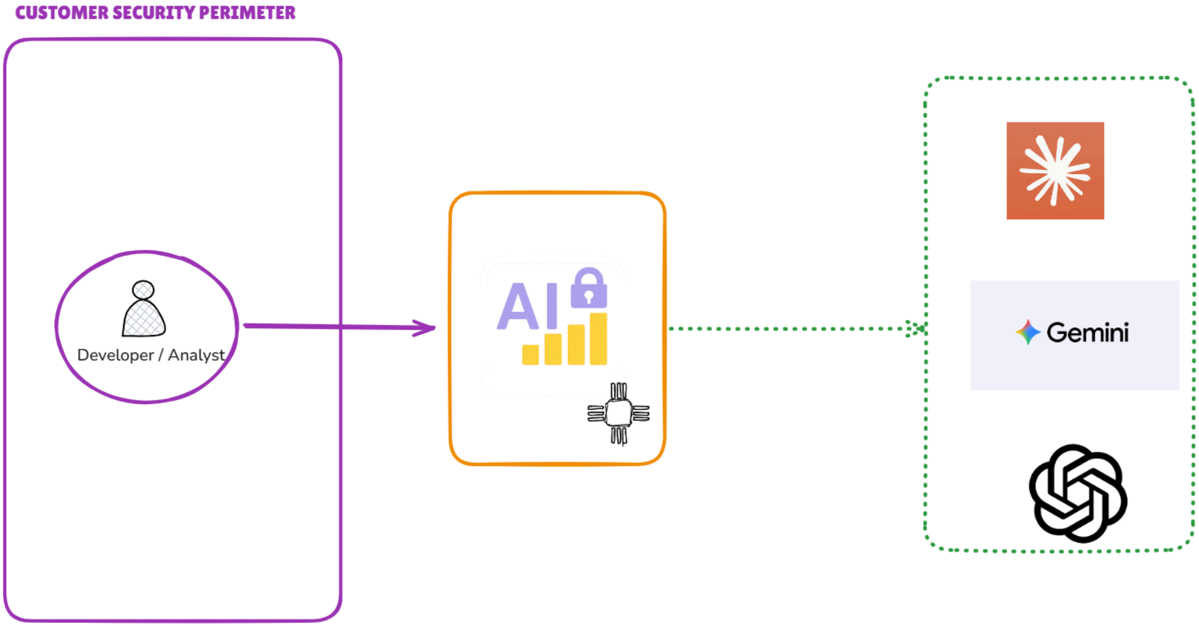
# 5. SaaS Deployment

In SaaS deployment, prompts are routed through SecureAIFlow-hosted infrastructure for detection and pseudonymization before being forwarded to the LLM provider. The security guarantee is zero data retention: SAF processes prompts in transit but does not persist any prompt content, credential values, PII, or source code fragments. SaaS infrastructure is provisioned within the customer's geographic region.

## 5.1 Architecture Overview

Figure 2 illustrates the SaaS deployment topology.

**Figure 2 : SaaS Deployment Topology**



## 5.2 Regional Deployment

SecureAIFlow SaaS infrastructure is deployed within the customer's geographic region. Region assignment is determined at account provisioning and is not changed without customer consent.

| Customer Region | SAF Infrastructure Region | Cloud Location | Cross-Region Transfer |
|---|---|---|---|
| European Union ( france ) | eu-west-3 | Paris, France | None |
| United States | us-central | United States Central | None |
| Other regions | On request | Determined at provisioning | None |

© 2026 SecureAIFlow. All rights reserved .secureaiflow.com

**NOTE** All prompt processing occurs within the assigned region. No prompt data, credential values, or PII is transferred between regions at any point in the detection pipeline. EU customers' data does not leave the European Union.

## 5.3 Data Flow Specification

Table 5 specifies each data flow step in SaaS deployment.

**Table 5 : SaaS Data Flow**

| Step | From → To | Data Content | Persistence at SAF |
|---|---|---|---|
| 1 | User tooling → SAF API | Raw prompt over HTTPS/TLS within assigned region | Not persisted : in-memory only |
| 2 | SAF API → Detection Engine | Prompt for analysis | Not persisted : processing only |
| 3 | Detection Engine → Memory | Credential/ pseudonym mapping | Not persisted : request-scoped memory; discarded at end of request |
| 4 | SAF API → LLM Provider | Redacted prompt: pseudonyms replacing sensitive values | Not persisted at SAF |
| 5 | LLM Provider → SAF API | LLM response with pseudonym references | Not persisted : processed in transit |
| 6 | SAF API → User tooling | Restored response with original values | Not persisted at SAF |
| 7 | Detection Engine → KPI Store | Request count, detection rate, latency (no content fields) | Persisted : anonymized metrics only |

## 5.4 Network Boundary Definition

Table 6 defines data boundary and retention status in SaaS deployment.

**Table 6 : SaaS Network Boundary**

| Data Element | Retained by SAF | Rationale |
|---|---|---|
| Credential values | Never | Pseudonymized in-memory; discarded after processing |
| Custom fields | Never | Pseudonymized in-memory; discarded after processing |
| Prompt content (raw) | Never | In-memory only; not written to storage |
| Source code fragments | Never | Contained in prompts; same in-memory-only rule |
| Pseudonym mapping | Never | Discarded at end of request; not persisted |
| LLM responses | Never | Processed in-memory; not logged or stored |

© 2026 SecureAIFlow. All rights reserved .secureaiflow.com

| Data Element | Retained by SAF | Rationale |
|---|---|---|
| Cross-region data transfer | Never | All processing within customer-assigned region |
| Anonymized KPI metrics | Yes | Request count, detection rate, latency : no content fields |

## 5.5 Zero Data Retention Controls

- **In-memory processing only:** All prompt content is held in request-scoped memory. No write operations to disk, database, or object storage are performed for prompt content, credentials, or PII.

- **Request isolation:** Each request is processed in an isolated memory context. Pseudonym mappings are not accessible after the request lifecycle ends.

- **No prompt logging:** Application logs do not record prompt content, credential candidates, PII values, or pseudonym mappings. Only operational metadata is logged.

- **KPI metrics are content-free:** Retained metrics contain no content fields: aggregate counts, rates, and durations only. No credential type, variable name, PII field, or value is included.

- **TLS in transit:** All communication between user tooling and SAF API, and between SAF API and LLM provider, is encrypted using TLS.

- **Regional containment:** All processing occurs within the customer's assigned region. No cross-region data transfer occurs at any step.

## 5.6 Compliance Posture : SaaS

| Framework | Relevant Requirement | SAF SaaS Control |
|---|---|---|
| GDPR Art. 28 | Data processor obligations | SAF acts as processor; DPA required for EU customers. Processing limited to detection only; no secondary use of data. |
| GDPR Art. 32 | Security of processing | TLS encryption in transit; in-memory-only processing; zero persistent storage of personal data. |
| GDPR Art. 5(e) | Storage limitation | Prompt content not retained beyond processing transaction. Retention period for sensitive content: zero. |
| GDPR Art. 44 | International transfers | EU customers are served exclusively from Paris (eu-west-3) or another selected EU region. No data leaves the European Union. |
| PCI DSS Req. 3 | Cardholder data protection | Credential values pseudonymized in-memory; not persisted. SAF retains no payment credential data. |
| SOC 2 (Confidentiality) | Protection of confidential information | Zero data retention is auditable: no prompt logs exist to be breached or subpoenaed. |

| Framework | Relevant Requirement | SAF SaaS Control |
|---|---|---|
| EU AI Act | Transparency and data governance | Regional containment and prompt interception support human oversight and governance obligations. |

# 6. Deployment Model Comparison

Table 7 provides a side-by-side comparison of both deployment models across key security, compliance, and operational dimensions.

### Table 7 : Deployment Model Comparison

| Dimension | On-Premises | SaaS |
|---|---|---|
| Sensitive data retained by SAF | Never : not transmitted to SAF | Never : in-memory only |
| Prompt content retained by SAF | Never : not transmitted to SAF | Never : in-memory only |
| Data residency | 100% within customer perimeter | SAF-hosted, customer-assigned region |
| EU data boundary (GDPR Art. 44) | Fully controlled by customer | Enforced : EU customers on Paris region only |
| SAF access to sensitive data | Zero : no network path | Zero : in-memory processing only |
| Pseudonymization key ownership | Customer | Customer (per-session; not persisted at SAF) |
| Audit log ownership | Customer : local storage | Customer dashboard (KPI metrics only) |
| GDPR data processor role | Not applicable | SAF signs DPA as processor |
| Secret manager integration | Supported (Vault, AWS SM, Azure KV) | — |
| PII detection (via integration) | Supported | — |
| Air-gapped / internal LLM routing | Supported via customer API integration | Not applicable |
| Infrastructure managed by | Customer IT team | SecureAIFlow |
| SAF responsible for patching | Application layer only | Full infrastructure |
| Deployment effort | Medium : VM provisioning required | Low : API key configuration only |
| Latency | < 500ms on standard CPU VM | < 500ms + network round trip |
| Recommended for | Regulated industries; strict data residency; internal LLM routing | Teams requiring rapid deployment with strong regional privacy guarantees |

© 2026 SecureAIFlow. All rights reserved .secureaiflow.com

# 7. Security Controls Summary

Table 8 provides a concise security controls reference for procurement review and vendor questionnaire completion.

**Table 8 : Security Controls Summary**

| Control | On-Premises | SaaS | Notes |
|---|---|---|---|
| Credentials pseudonymized before transmission | Yes | Yes | Applies to all contexts and user types |
| PII pseudonymized before transmission | Yes (via integration) | — | Requires DB/microservice integration |
| Sensitive data persisted by SAF | No | No | Never written to any SAF storage |
| Prompt content persisted by SAF | No | No | On-prem: not transmitted. SaaS: in-memory. |
| Encryption in transit (TLS) | Customer-managed | TLS 1.2+ | SAF enforces TLS for all outbound API calls |
| Regional data residency | Customer perimeter | Customer-assigned region | EU: Paris. US: US Central. |
| Cross-region data transfer | Never | Never | Enforced by regional infrastructure assignment |
| SAF access to customer data | None | None (processing only) | No persistent access in either model |
| Audit log available to customer | Yes : full event log | Yes : KPI metrics | On-prem: full. SaaS: anonymized. |
| Secret manager integration | Yes | — | Vault, AWS SM, Azure KV |
| Air-gapped / internal LLM routing | Yes | No | Via customer API integration |
| Zero data retention guarantee | N/A : data not transmitted | Yes : technical controls | See Section 5.5 |
| GDPR DPA available | Not required | Yes : required for EU | Contact for DPA |
| SAF infrastructure patching | Customer responsibility | SAF responsibility | See Section 3 |
| CPU-only inference : no GPU required | Yes | Yes | < 500ms end-to-end |

## Notices

This document is provided for informational purposes only. It represents SecureAIFlow's current product architecture and practices as of the publication date, which are subject to change without notice.

This document does not create any warranties, representations, contractual commitments, or assurances from SecureAIFlow. The responsibilities and liabilities of SecureAIFlow to its customers are controlled by the applicable service agreement.

Customers are responsible for making their own independent assessment of this document and for determining its applicability to their regulatory obligations. This document does not constitute legal or compliance advice.

© 2026 SecureAIFlow. All rights reserved. · secureaiflow.com