INSTABASE

Security and Trust

Leveraging Third-Party Large Language Models with Instabase

Last Updated: August 21, 2023

Contents

Frequently Asked Questions

Which third-party large language model ("LLM") providers do you support?

We currently support OpenAI (US) and Azure OpenAI[1] (EU).

Will LLM providers use my data to retrain their models?

We have an agreement with OpenAI that prevents them from using your data to improve or train their models.

Will LLM providers retain my data?

Our agreement with OpenAI prevents them from retaining your data.

Does your use of LLMs comply with leading security & privacy practices?

Instabase, including the use of LLMs, is designed to be compliant with our applicable security and privacy frameworks (SOC2 Type II, GDPR, HIPAA, CCPA).

Where is my data processed?

LLM providers process data in the US (OpenAI) or EU (Azure OpenAI[1]).

Can I trust the AI to not hallucinate and introduce bad data into my systems?

Instabase reduces the risk of hallucination with a technique called grounding. Grounding allows us to leverage the reasoning power of LLMs, but using the knowledge base of our choice (e.g. customer documents and files) rather than the internal trained knowledge of an LLM, which can lead to inaccuracies.

To ensure quality results, we provide comprehensive data validations and business rules to act as guardrails and identify potential errors. We also include a native human review interface for customers to review and modify results as needed.

[1] Due to Azure OpenAI API rate limitations, access to the service will not be available through Instabase's account. Customers may use their own Azure accounts and customer-provided endpoints with the Instabase Platform. Customers are responsible for ensuring their Azure OpenAI service levels are adequate to support their desired throughput requirements.

Introduction

Generative AI, powered by large language models ("**LLMs**"), has unlocked unprecedented opportunities for enterprise transformation and productivity. These cutting-edge AI models possess a remarkable capacity for processing human language and are positioned to revolutionize the way organizations extract meaningful information from unstructured content.

At Instabase, we've built proprietary technology that enables LLMs to provide a day-one understanding of even the most complex unstructured data without any additional training. This intelligence from the get-go dramatically accelerates AI utility and adoption, unlocking numerous use cases that were once too costly and time-intensive or offered too low an ROI to pursue in the past.

In a recent survey by KPMG, 77% of executives rank generative AI as the emerging technology that will have the biggest impact on their business. However, the vast majority (92%) are concerned about the risks of implementation, ranking cybersecurity, privacy, and liability as the top focus areas[2].

Instabase is built from the ground up with security, privacy, and compliance in mind. We've designed processes and controls to comply with the requirements of highly regulated industries and global security frameworks. This white paper details Instabase's approach to leveraging third-party LLMs in a secure, compliant, and trustworthy manner, including how we interact with LLM providers and the controls we've developed to handle customer data.

Our approach is three-fold:

- **Data Privacy and Security**
  Secure and confidential handling of customer data

- **Model Accuracy and Trust**
  Enhancing reliability of model outputs

- **Security Frameworks and Compliance**
  Designed to comply with global standards

Last updated: 8/21/2023

Currently Supported Third-Party LLM Providers	
OpenAI	Azure OpenAI[1]

Currently Supported Deployment Options
Instabase Single Tenant Software-as-a-Service (SaaS)

[2] Generative AI: From buzz to business value, KPMG, 2023

1. Data Privacy & Security

We understand that data privacy and security are of paramount importance to our customers. With some of the most advanced and performant LLMs offered as SaaS-based services, control and confidentiality of customer data is central to making an informed decision in line with regulatory and security requirements.

As a platform for unlocking unstructured content, customers submit input files such as documents, images, and text that are processed in memory by Instabase. The resulting output files, or extracted data, is then stored or delivered to a downstream system (CRM, ERP, Core Systems, Data Warehouse, etc.).

At Instabase, we strive to ensure that our customers clearly understand how their data is processed. Our primary objective is to enable the safe and secure use of our cutting-edge technology, while maintaining transparency throughout the process.

Data Privacy

Instabase's privacy program is designed to keep customers in control of their data and comply with applicable regulations.

CONFIDENTIALITY

As a result of the controls we have implemented, our agreements with our third-party LLM providers ensure that they cannot retain customer data or use it to improve their models.

OWNERSHIP

Instabase, and our LLM providers, do not claim any ownership to the service inputs (data or prompts). In addition, all rights to the output of our services (responses) will be assigned to customers as long as they comply with our Terms of Use.

RETENTION

Customers are in full control over what data is retained by Instabase. We provide the choice of storing data in either Instabase-managed storage (with the option to apply user-defined retention policies) or in customer-managed storage.

ACCESS CONTROL

Instabase provides role-based access controls for our customers to govern data access within the Instabase platform. Access is restricted to geographies defined in our Terms of Use. By default, Instabase employees do not have access to customer installations.

If customers require support or assistance, they can facilitate access by (a) a screen sharing session or (b) provisioning an Instabase employee with access via the customer's identity provider.

Data Security

Instabase provides comprehensive security to protect customer data such as encryption, network controls, data governance, and auditing.

Every Instabase Platform SaaS customer receives their own unique customer installation, which comprises three deployments (DEV, UAT, PROD). Each deployment is single tenant with no shared infrastructure components between environments within an installation or across customer installations.

We ensure industry best practices for handling data at rest and in transit. Any data stored on our file system or transmitted across networks is encrypted. This includes content processed by Instabase and any data transferred to model providers or our customers' infrastructure and systems.

ENCRYPTION

Instabase secures and encrypts customer data in transit and at rest. All client-server communication is encrypted over HTTPS and TLS 1.2 and above, while all data at rest is encrypted with AES-256.

PLATFORM SECURITY

Every SaaS deployment is single tenant and hosted in Instabase's account. There are no shared infrastructure components between deployments, meaning each deployment is completely isolated. Additionally, access to Instabase can be controlled through IP ACLs and AWS PrivateLink, and network ingress policies remain under customer control.

AUTHENTICATION

We support and encourage the use of SAML-based single sign-on and multi-factor authentication to help prevent unauthorized account access.

MONITORING

Instabase deploys monitoring across its computing resources with alert notifications set to the Security Incident Response Team (SIRT) for triage and response. This includes a SIEM system for near real-time monitoring, alerting, and investigation.

AVAILABILITY & CONTINUITY

The Instabase platform provides multi-region failover with committed uptime and is accessible by users in the regions defined here. Any data stored by Instabase is replicated in an alternate AWS region to ensure business continuity.

Data Residency

We offer various choices for our customers to comply with data residency requirements, including the ability to mount external cloud storage and choose the geographical regions where the Instabase and LLM services reside.

The systems used for data processing and storage are the Instabase Platform, third-party LLM providers, and file storage.

Instabase Platform

The platform operates on Amazon Web Services and is designed to process data in real time without persisting customer data in local storage.

Large Language Model Providers

Third-party LLM providers, such as OpenAI, are used for model inference and runtime processing only.

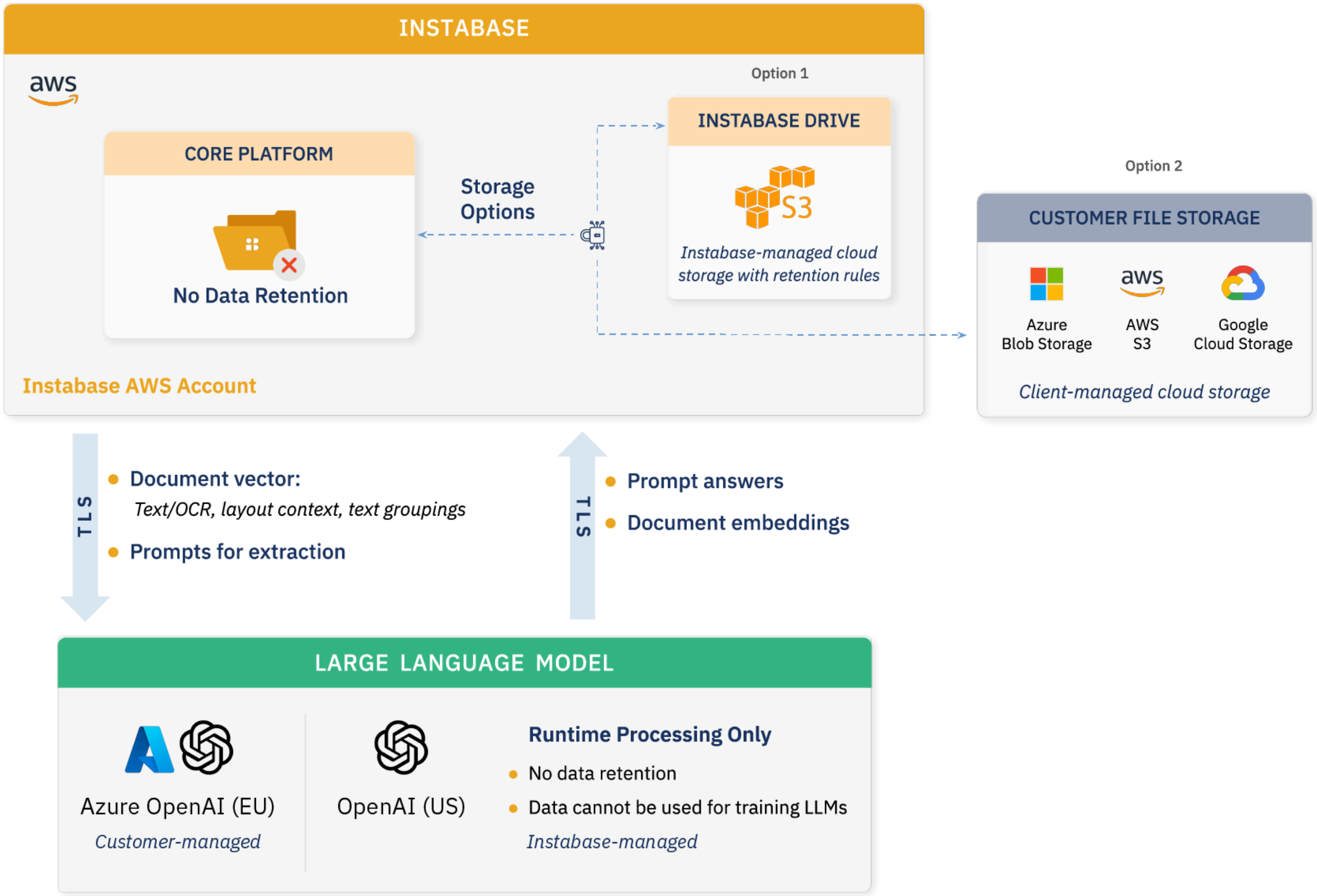
File Storage

Storage systems are used to manage customer data such as input and output files. Options include using the native Instabase Drive (AWS S3) or mounting external cloud storage such as Amazon S3, Azure Blog Storage, or Google Cloud Storage.

Please refer to the table below to understand the residency options and data handling.

	Residency	Retention
Instabase Platform	• Any AWS geographical region that complies with our Terms of Use	Processing only
LLM Provider	• OpenAI (US) • Azure OpenAI[1] (EU)	Processing only
File Storage (customer-managed)	• Customer-managed AWS S3, Azure Blob Storage, and Google Cloud Storage • Customer controls deployment region	Data retention controlled by the customer
File Storage (Instabase-managed)	• Instabase Drive (AWS S3) deployed in same region as the Instabase Platform	Ability to apply custom retention rules

Architecture Diagram



For more information regarding our enterprise-grade security and privacy practices, please visit our Trust Center.

2. Model Accuracy and Trust

Unlike ordinary software, large language models are massive neural networks typically trained on vast amounts of both publicly available data such as from the internet and data licensed from third-party providers. This makes them capable of performing incredibly complex tasks with command across a broad range of topics, but this can also present risks of bias and information reliability.

During the training process, LLMs compress this significant quantity of data into mathematical representations, which results in a loss of detail and information fidelity. Additionally, due to the public nature of the training data, it may contain bias and incomplete or conflicting information. The result is the capacity for LLMs to "hallucinate" or generate inaccurate or misleading information.

Instabase takes a multifaceted approach to address these concerns from knowledge grounding to data validation and review. Incorporating techniques to reduce risk of hallucination and verify model outputs is crucial to ensuring reliable data and responsible use.

Knowledge Grounding

Large language models have demonstrated an impressive ability to understand language and perform reasoning, but relying on their compressed knowledge base or memory of training data introduces opportunities for hallucination.

To mitigate this, Instabase leverages the reasoning capabilities of LLMs, but applies the model to use a customer-specific knowledge base in a process called grounding. This means the LLM output is based on the facts of the connected data source rather than any artificial output. We apply proprietary techniques to turn content into a custom knowledge base, grounding LLMs to generate results based only on our customers' source data.

Data Validation and Refinement

With any AI solution, it's prudent to validate results before using them to support live business workflows and decisions. Automated methods of data verification help scale this process by pinpointing potential errors for further investigation. A common approach is instituting a minimum required confidence score–a measure of probability that a model prediction is correct. However, relying on a confidence score alone is not sufficient to achieve high rates of accuracy and automation and in the world of generative AI, some LLMs don't output such probabilities and may not be proficient in providing estimates of correctness.

Instabase automatically applies a range of checks to verify data accuracy, including logical validations, business rules, external lookups, and cross-document comparisons. These rules provide a deterministic filter to ensure the outputs of LLMs and other AI models meet expectations. The platform also cleans, corrects, and normalizes data for use before sharing results with downstream systems.

Human Review

If our data validation system determines the model output to be incorrect or potentially incorrect or the process is of significant importance that a human must always verify results, such as in maker-checker processes, then Instabase provides a user-friendly experience for customers to review and modify values as needed. The interface allows for fast understanding of errors, tracing and comparing results to source content (provenance tracking), and the ability to re-validate any modified data before submission for additional precaution.

With fine-grained data validation tools, Instabase can tune the system's error detection sensitivity to meet customer-specific needs, maximizing the rate of automation at the desired level of accuracy.

3. Security Frameworks and Compliance

The world's largest organizations trust Instabase to process sensitive, business-critical data. To help customers meet their compliance obligations, we've designed, developed, and maintained our security program to comply with global security and privacy standards. This includes the administrative, organizational, technical, and physical controls designed to safeguard the confidentiality, integrity, and availability of customer data.

Instabase holds certifications and attestations for SOC 2 Type II and HIPAA and is designed to comply with GDPR and CCPA. Our certifications are renewed annually by independent bodies.

✓ CCPA ✓ GDPR ✓ HIPAA ✓ SOC 2

Instabase performs regular security and vulnerability testing, including third-party penetration tests, to ensure controls are effective against industry standards. As security threats change, Instabase continues to update its program and strategy to protect our customers' data. For additional information and documentation on our security program, please visit our Trust Center.