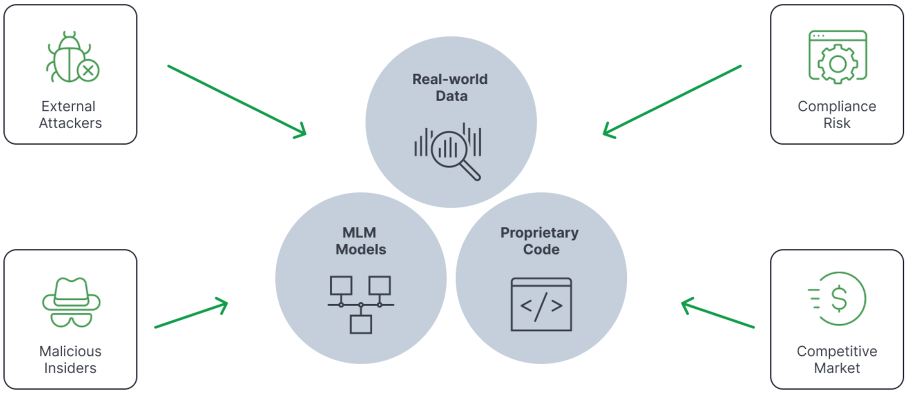
anjuna®

WHITE PAPER

Confidential Computing for Artificial Intelligence and Machine Learning Workloads

AI Adoption is Blocked by Data Privacy and Security

While enterprises aim to quickly adopt artificial intelligence (AI) to be competitive in today's rapidly-evolving landscape, the regulatory and reputational risk of a data breach remains high. According to Gartner, 41% of organizations have experienced an AI privacy breach or security incident, and over 50% are the result of a data compromise by an internal party (source). Data scientists and machine learning (ML) developers need the full power of models, code, and real-world data to deliver the best results, but traditional security controls impair or prevent productive work entirely. Frictionless security - which defends against modern threats without causing new blockers - is essential to enable an agile and successful AI strategy.



How does Anjuna Seaglass enable Confidential AI?

Anjuna Seaglass makes it easy to implement Confidential AI: AI protected using Confidential Computing. Unlike traditional security software, Confidential Computing uses secure enclaves to provide hardware-level isolation, confidentiality, and integrity protection for sensitive data and ML models: true zero trust computing. This creates a fortress to protect your most valuable intellectual property and pave the path to securely get data in and out. Confidential Computing immediately shuts out at least 77 high-threat attack techniques (out of a total of 277 defined by MITRE) and helps address 7 of the OWASP LLM Top 10.

Confidential AI reduces the risk of handling sensitive or regulated training data, stops malicious attacks, and prevents even privileged insiders from accessing training data, inputs, outputs, and models. This enables new use cases involving sensitive data and cross-organization collaboration in financial services, medical analytics, AI-as-a-service, and other regulated industries.

Anjuna Seaglass application-agnostic approach allows you to deploy all ML architectures (RNNs, CNNs, GANs, transformers/GPTs, etc.), and frameworks (PyTorch, Tensorflow, Keras, ONNX, etc.) on any of the leading public clouds (AWS, Azure and GCP), enabling you to focus on delivering value for your customers.

At-a-Glance

Anjuna Seaglass makes it easy to protect AI and ML workloads using Confidential Computing. Data and ML models are protected at all times from attackers and insiders alike, which enables new use cases and revenue through better privacy and security.

What You Can Achieve:

- **Get full-fidelity insights on your data** instead of limiting yourself with data masking, synthetic data, or other techniques that obfuscate PII but create new risks from biased distributions.
- **Earn customer trust by locking out insider access to their data,** while still being able to use state-of-the-art machine learning for better, data-driven decisions.
- **Enable secure collaboration** using confidential data clean rooms, maintaining tight access control to the underlying data and models.
- **Provide access to private training datasets and pre-trained ML models** to enable new monetization strategies without losing control of your intellectual property.
- **Simplify and ensure compliance** with security, privacy, and data sovereignty regulations.

What Makes Anjuna Seaglass Unique?

FROM ZERO TO VALUE IN MINUTES

If you're just getting started, Anjuna Seaglass provides ready-to-run templates to protect common ML use cases: generative AI, data clean rooms, and inference serving. Be up and running with confidential computing in a few minutes, and build with agility on top of the template for your own use case.

USE YOUR PREFERRED SOFTWARE AND ML TOOLCHAIN

A general-purpose approach lets you choose the technology that meets your needs. You don't need to refactor your pipelines to use a specific model format or retrain your data scientists to use a new ML framework. Simply run your existing ML workloads with Anjuna Seaglass.

TRUST WITH POLICY-BASED VERIFICATION

Your customers need to be assured that their ML workloads are producing outputs that have not been tampered with. With a verifiable encryption and policy management capabilities, they can see proof of their output integrity. Confidential Computing ensures that your code is always verified for authenticity, which rejects maliciously-modified code before it can compromise your data.

LOCK OUT INSIDER THREATS AND CLOUD SERVICE PROVIDERS

Anjuna Seaglass creates fully-isolated hardware-based secure enclaves that cannot be infiltrated, locking out rogue insiders from your own company as well as the cloud service provider. Gain customer trust in your data privacy, easing SOC II compliance and your security team's workload.

Path To Success

Anjuna Seaglass is delivered as software that you run in your own infrastructure. Below are a few examples of common deployment patterns we see. For each, the path to success is the same:

Step 1: Run the Anjuna-provided template project - 15 minutes

Step 2: Use a custom model instead of the Anjuna-provided model - 2-6 hours depending on complexity

Step 3: Update clients to fetch encryption keys and encrypt data - 2 hours

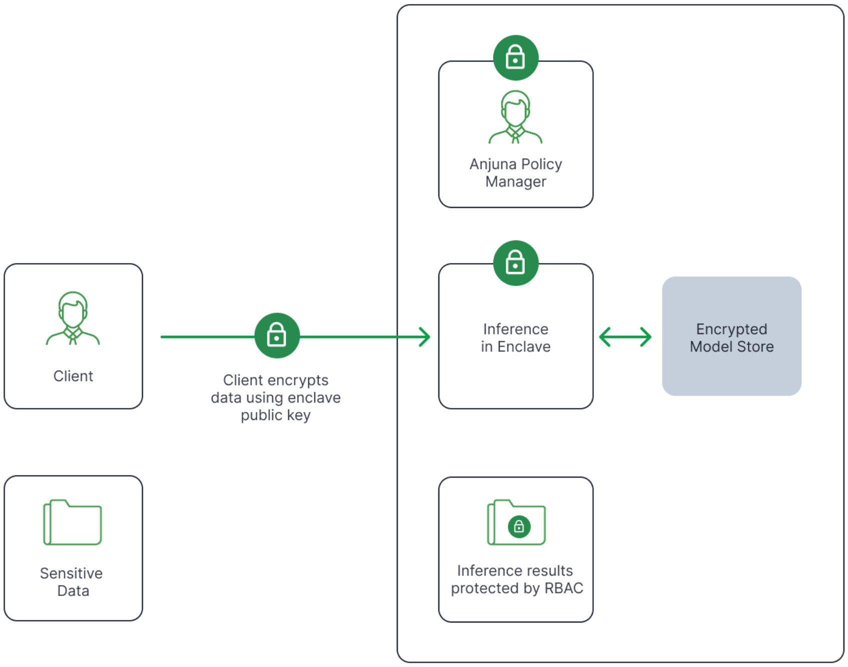
Step 4: Testing - varies based on organization

Step 5: Update CI/CD pipelines - varies based on the organization

Anjuna customers accelerate their deployment by an average of 6 months using Anjuna Confidential AI.

Reference Architecture: Secure Inference

In the Secure Inference pattern, end users can run complex ML workloads on their sensitive data without the risk of data breach. Confidential Computing protects the input data, output data, and the model itself from insider sysadmins, attackers, and cloud service provider admins.



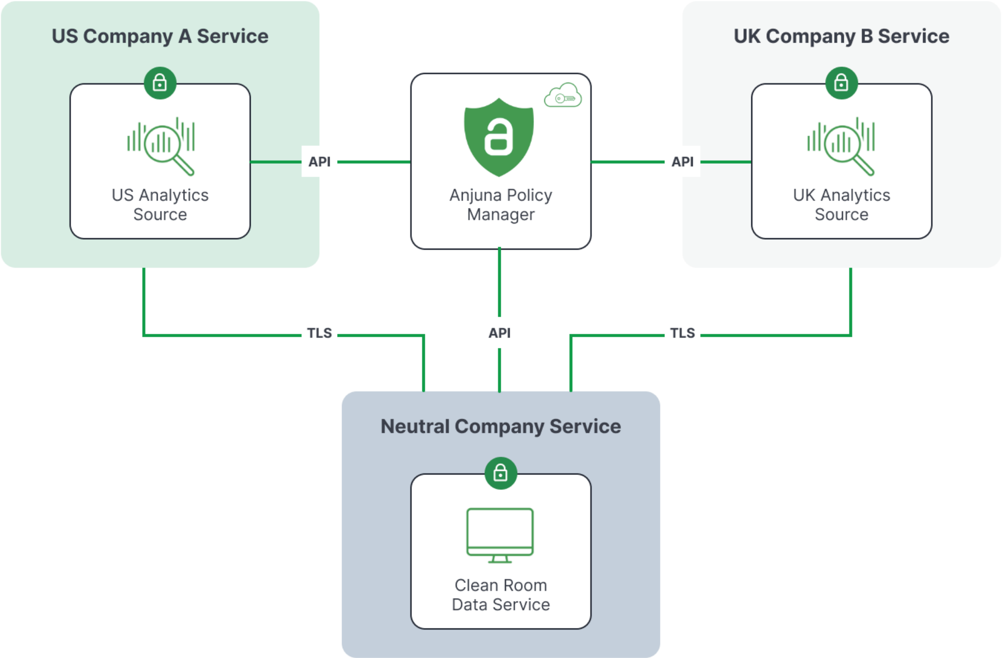
In this architecture, the end user first performs verification of the enclaves involved (also referred to as “remote attestation” in confidential computing jargon). Once it has verified their safety, the client can encrypt its data and upload it for inference. Only the enclave can decrypt the sensitive client data and perform its inference within the enclave. No sensitive information is ever in plaintext - even a sysadmin with root access to the inference service can’t access the data.

Customer examples:

- One government agency used Anjuna Seaglass to securely run ML workloads in the public cloud for the first time ever, enabling the use of cloud-scale computing power without compromising national security.
- A major bank was able to securely analyze real data, significantly improving prediction results compared to their previous models built on synthetic data.

Reference Architecture: Data Clean Room

In the Data Clean Room pattern, multiple parties can combine their data and learn from the combined dataset, without revealing PII or trade secrets.



In this architecture, multiple parties encrypt their data in a way that can only be decrypted within an enclave, the "data clean room". The resulting analysis outputs are aggregated results which do not reveal the underlying data.

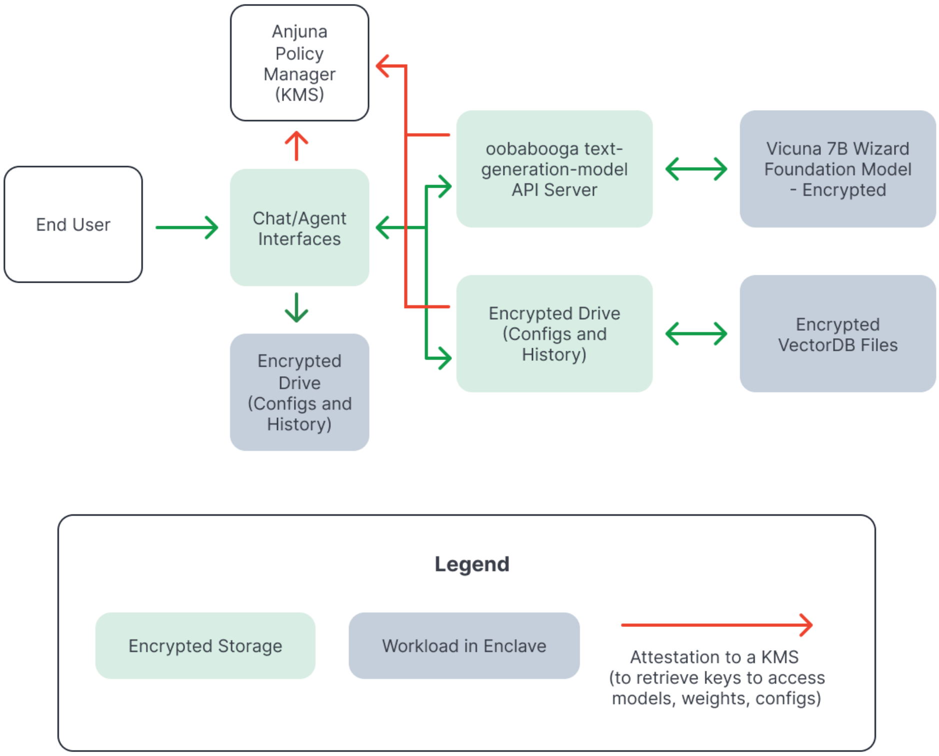
Customer examples:

- A bank was able to accelerate their data clean room development by months, enabling collaboration with other organizations without exposing any confidential data.
- A specialist SaaS AI company for healthcare research spent months struggling to build their own cloud-native multi-party clean room, which delayed revenue growth. Anjuna Seaglass enabled custom AI model execution using AMD SEV-SNP in just days and set the course for future GPU use.

Reference Architecture: Private LLM

In the Fine-Tuned Large Language Model (LLM) with Vector Embeddings pattern, an organization can utilize the power of cutting-edge AI combined with their internal confidential data and provide secure access to their employees.

Corporate LLM Protected with Confidential Computing in the Cloud


Anjuna Policy Manager (KMS)
End User
Chat/Agent Interfaces
Encrypted Drive (Configs and History)
oobabooga text-generation-model API Server
Encrypted Drive (Configs and History)
Vicuna 7B Wizard Foundation Model - Encrypted
Encrypted VectorDB Files
Legend
Encrypted Storage
Workload in Enclave
Attestation to a KMS (to retrieve keys to access models, weights, configs)

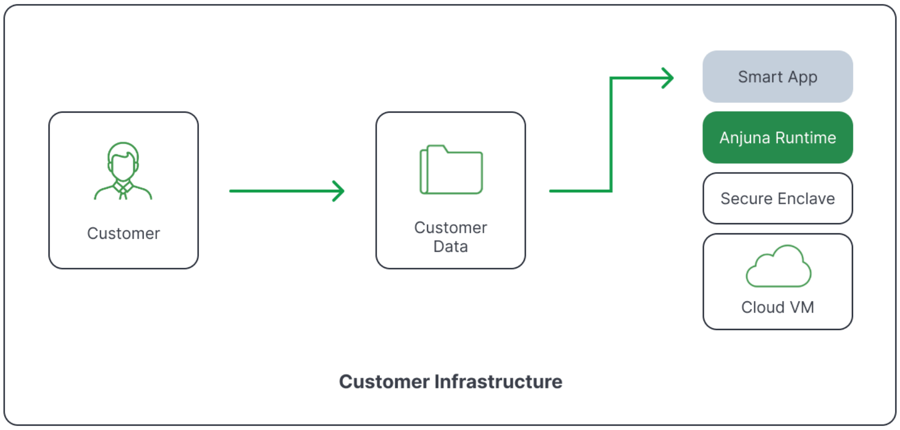
In this architecture, internal corporate data sources are added to an open-source LLM using vector embeddings, and the model is fine-tuned for a particular task. All data is encrypted using enclave keys, eliminating the risk of leaking powerful model weights and the associated proprietary information.

Customer examples:

- A Fortune 500 company added context to a public large language model for internal business insights, with full privacy and security provided by Anjuna, even from insider threats - Anjuna Seaglass saved them an estimated 6-8 months of engineering time.

Reference Architecture: Secure Self-Hosted

In the Secure Self-Hosted pattern, an organization can sell self-hosted (sometimes called "on-prem") access to their ML models to privacy-sensitive customers. Customer data never leaves their own environment, and the ML model owner's intellectual property is protected.



In this architecture, a proprietary ML model can be deployed into a customer environment, which preserves data privacy for its users. Normally, this risks attacks like the OWASP Top 10 threat "Model Theft", but the Anjuna Confidential Runtime directly prevents these attacks.

Customer examples:

- A leading customer support ISV was able to launch a self-hosted version of their product, unlocking millions of dollars in revenue from privacy-conscious EU customers

Comparison to Alternatives

Synthetic data uses algorithms to simulate real-world data with similar statistical properties. However, the results can be biased or completely incorrect, especially for long-tail distributions, since the quality depends entirely on how similar the synthetic dataset is to real data. With confidential computing, you can use real data for accurate results without putting your data at risk.

Differential privacy and K-anonymization are approaches which inject noise into individual data points so that the overall properties of the dataset are the same. Similar to synthetic data, the implementation risks biased results; unlike synthetic datasets, attackers can repeatedly query anonymized datasets to leak more information. Typically these approaches only operate on tabular data, which does not work for emerging generative AI use cases.

Encryption and tokenization are similar approaches which transform data into a garbled form, protecting it from outsiders who do not have access to the decryption key or the de-tokenization engine, respectively. But there are two major downsides to these approaches. First, if an application needs to process plaintext data, then an attacker can perform a memory-dumping operation to gain access to the plaintext data or keys needed to decrypt it. In confidential computing, hardware isolation prevents this attack. Second, encryption and tokenization can add data model complexity and processing overhead, which reduces time to value and increases operating costs. Confidential computing does not require modifying your data models and provides memory encryption through dedicated hardware.

Fully-homomorphic encryption is the closest theoretical comparison to confidential computing. Data is encrypted and then processed only in encrypted form. However, fully-homomorphic encryption today is 1,000-1,000,000 times slower than unprotected computation; a batch job that normally takes one hour would take over a month to complete. In comparison, Anjuna has observed overheads of 0-30%, or an extra twenty minutes for the batch job.

Zero-knowledge proofs have recently become more practical with the introduction of techniques like zk-SNARKs. However, zero-knowledge proofs can only prove statements about data known to the client, which has limited enterprise use cases. Confidential computing allows arbitrary computation over data, just like standard applications.

In conclusion, **confidential computing** offers several advantages over traditional approaches to securing AI and ML. Using Anjuna Confidential AI, you can process real data for accurate results with minimal overhead, without risking your data. You also get protection for your code integrity and protect your valuable model, which is not possible using any of the alternatives above.

Conclusion

Anjuna's Confidential AI capabilities helps you adopt cutting-edge AI and ML techniques safely, minimizing the risk from attackers and insider threats alike. Instead of toiling over noisy scanner outputs and patch lists, take a proactive approach to security and privacy with Anjuna Seaglass, and enable the highest-level of protection for your valuable code.

About Anjuna

Anjuna created the first Universal Confidential Computing Platform to run applications in any cloud with complete data security and privacy. Anjuna Seaglass isolates workloads in a hardware-assisted environment that intrinsically secures data in every state to create a zero trust environment for code and data. Anjuna Seaglass empowers enterprises to directly control application-level trust policies, ensuring that only trusted code can access sensitive data. Anjuna works with enterprises around the globe in industries such as financial services, government, and blockchain.

Get started with Anjuna Seaglass