Liminal

# Ensuring security in the era of generative AI

Foundational paper

liminal.ai

# Table of contents

Executive summary

# In the rapidly evolving landscape of artificial intelligence, generative AI is a transformative force, reshaping how businesses operate and interact with data.

The unparalleled pace of growth, as exemplified by ChatGPT's rise to 100 million users in a mere 60 days, underscores generative AI's potential and widespread appeal. A Harvard Business Review case study vividly illustrates this impact: when pitted against a control group without AI tools, a consulting team equipped with generative AI completed tasks 12.2% more efficiently, 25% faster, and with 40% higher quality. This efficiency gain, scaled across a large workforce, can be equivalent to significantly expanding the team size while enhancing output quality.

## Despite these advantages, integrating generative AI in regulated industries like healthcare, insurance, and manufacturing brings pressing data privacy and security challenges.

Numerous data points from Liminal's research highlight a growing trend of generative AI use in the workplace, often without formal approval or adequate training on safe and ethical usage:

□ 91% of respondents stated they are familiar with generative AI, and **64% reported using generative AI tools at work at least weekly.**

□ **8% of enterprise employees could be circumventing restrictions and using generative AI tools** despite company policies or other restrictions.

□ Data privacy and security are the number one factors preventing or deterring workers from using generative AI, with **47% of respondents citing concerns in this area.**

□ **63% of respondents reported that they would be comfortable sharing at least some personal or proprietary information with generative AI tools**, regardless of company policy.

A November 2023 *report from Salesforce*[1] highlights similar concerns:

□ **Over a quarter (28%) of workers are currently using generative AI at work,** and over half are without the formal approval of their employers.

□ **Nearly 7 in 10 global workers** have never completed or received training on how to use generative AI safely and ethically at work.

□ **Nearly half (47%) of workers** believe mastering generative AI would make them more sought after in the workplace.

□ **Over half (51%) of workers believe generative AI will increase job satisfaction**, and 44% believe they will earn more than those who haven't mastered the technology.

The haphazard approach to adopting generative AI poses significant risks for organizations in various areas, including regulatory compliance, intellectual property protection, and potential damage to reputation.

This foundational paper delves into these challenges, **proposing a layered approach to security encompassing policy, process, and technology**. It emphasizes the necessity of robust security measures in generative AI systems to safeguard against data breaches, ensure regulatory compliance, and maintain consumer trust.

[1] "More than Half of Generative AI Adopters Use Unapproved Tools at Work", Salesforce, November 15, 2023

# Introduction

## The Dawn of Generative AI: Opportunities and Imperatives

Generative AI, a subset of artificial intelligence characterized by its ability to create novel content, is rapidly reshaping the technological and business landscape. From crafting human-like text to generating predictive analytics, its applications are as diverse as they are impactful. However, this revolutionary technology also presents unprecedented security challenges. The capacity of generative AI to process and reproduce vast amounts of data raises significant concerns about privacy, data protection, and ethical usage.

**As generative AI becomes increasingly integral to business operations across various sectors, the imperative for robust security measures has never been more pronounced.**

This document explores the current landscape of generative AI, highlighting its potential applications and the accompanying necessity for comprehensive security strategies. Additionally, we explore the current generative AI technology stack and where technology can solve specific security gaps.

We focus on delineating a **multi-layered approach to security that addresses the unique challenges posed by generative AI**, ensuring its responsible and secure deployment across industries.

## Emerging Security Challenges

As the capabilities of generative AI continue to evolve, the technology is finding its way into a myriad of applications, from assisting doctors with writing insurance claims appeals to analyzing substantial financial datasets to provide deep insights. However, with this rapid expansion comes a spectrum of security challenges (chief among these being data privacy, security, and sovereignty).

## At their core, generative AI tools are powered by Large Language Models (LLMs).

These models operate by processing and analyzing vast datasets, enabling them to discern patterns among words and phrases, subsequently generating contextually appropriate responses. LLMs rely on shared data for refinement and optimization, meaning any data submitted via a prompt can be used to train the models going forward.

### Why security is at the forefront

This behavior of leveraging user-inputted data for model optimization is particularly problematic when a generative AI model is provided with sensitive or protected data. For public models, this can mean that organizations no longer have control over how their proprietary and customer data is used or shared. In private or internally-deployed models, organizations can find themselves in possession of highly personal consumer or employee information, which, if not managed appropriately, can put them out of compliance with HIPAA or other data protection regulations.

# These data challenges give rise to three significant risk categories that organizations must address when dealing with generative AI.

## Regulatory Compliance Risk

This primarily revolves around the potential for regulated data to be either leaked outside an organization's boundaries or inappropriately stored and accessed internally. Personally identifiable information (PII), protected health information (PHI), payment card industry (PCI) information, and other compliance-defined data types exist in abundance in every organization, regardless of industry. The accidental sharing of such highly sensitive information can lead to severe legal and punitive consequences, resulting in violations of data protection regulations like HIPAA, CCPA, GDPR, and others. The unauthorized disclosure of PII, PHI, and PCI data triggers significant fines and penalties and breaches compliance requirements, eroding consumer trust and confidence.

## Sensitive Data + IP Risk

Another prevalent concern within organizations regarding generative AI centers on the threat of intellectual property (IP) or other proprietary data leaks. This apprehension is well-founded because IP assets typically constitute over 65% of the total value of Fortune 500 companies[2]. In 2023, Samsung encountered three instances in a single month where employees inadvertently shared trade secrets[3] with OpenAI's LLMs, disseminating precious and confidential company information beyond Samsung's control.

## Reputational Risk

Reputational risk in the context of generative AI focuses on the concern that models may ingest or output offensive, violent, discriminatory, or derogatory language. The possibility of association with hate speech, racism, sexism, ageism, or any other forms of socially unacceptable behavior poses a significant threat to a brand's integrity and public perception.

[2] "Decoding the Worth: The World of IP Valuation", TTConsultants, September 15, 2023
[3] "Samsung workers made a major error by using ChatGPT", TechRadar, April 04, 2023

## Addressing the Security Challenges

The vast majority of enterprises will not train their own models. Instead, they will leverage commercially available LLMs such as Microsoft Azure OpenAI's suite, Google's Gemini models, or any of the thousands of open-source models. Many will fine-tune these models to fit their requirements.

**In these situations, enterprises must employ a comprehensive multi-tiered security strategy for generative AI, including a blend of policy frameworks, procedural measures, and technological defenses.**

### Policy

Organizations must first develop comprehensive security policies addressing regulatory compliance and ethical considerations in AI deployment. These policies should be a foundation for guiding AI development and use, ensuring alignment with legal standards and ethical norms.

### Process

Regarding the process layer, it is critical to implement risk assessment and management strategies. Successful processes involve identifying potential security vulnerabilities in AI systems and establishing protocols for incident response and continuous monitoring. Additionally, companies must provide robust training to all employees regarding the usage of generative AI. These processes reinforce policy adherence and ensure security measures are effectively implemented and maintained.

### Technology

With effective policies and processes implemented, the focus shifts to deploying advanced security tools at the technology layer. This deployment includes leveraging state-of-the-art encryption methods, implementing data anonymization technologies, and utilizing AI-powered cybersecurity solutions. These technological measures are crucial in protecting AI systems from external threats and ensuring the integrity and confidentiality of the data they process.

# Liminal's placement in a multi-layered security approach

## Liminal is the technology security layer for organizations looking to deploy and use generative AI.

The Liminal Platform provides tools for setting and enforcing granular security controls and data governance policies, enabling organizations to securely utilize generative AI. With Liminal, organizations can manage user access, permissions, and data governance rules regardless of application or model. Liminal enables organizations to align technology with effective policy, ensuring compliance with regulatory standards and ethical guidelines.

In terms of process, Liminal offers a platform for centralized oversight across all engagements with generative AI, including **direct interactions with LLMs**, **third-party applications containing generative AI, and custom applications built with generative AI functionality**. The platform includes real-time alerting and observability (vital for ongoing monitoring and management of AI interactions), streamlining the enforcement of security policies, and facilitating quick responses to potential security incidents.

**Liminal provides a comprehensive suite of tools to protect sensitive data within all generative AI interactions.** The Liminal Engine, a core platform component, detects non-compliant data types like PII, PHI, PCI, and unique organizational intellectual property. It employs intelligent masking and redaction, protecting sensitive information (while retaining rich context to ensure the highest value output) before submission to any generative AI model.

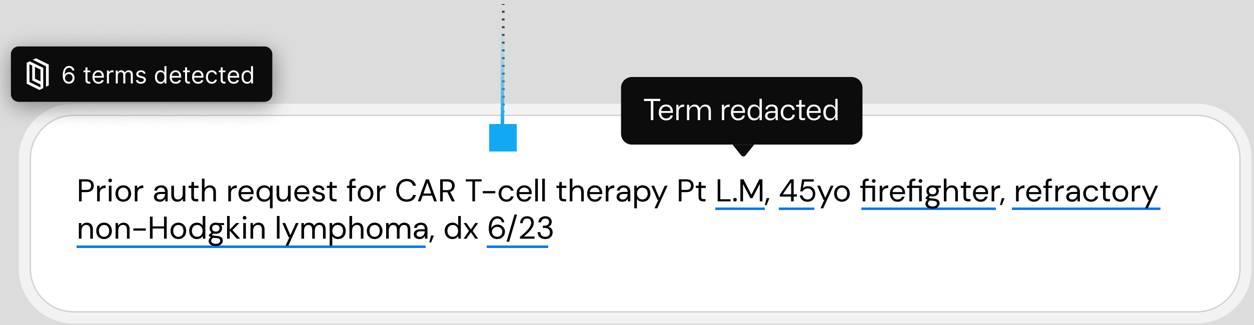
The model output returns to the Liminal Engine for re-seeding (rehydration) with sensitive data as appropriate (to ensure a consistent, contextually accurate user experience). This mapping of sensitive information to compliant alternatives is retained only for the duration of the exchange. This strategy ensures regulatory compliance with frameworks such as HIPAA.

# Enabling Horizontal Security for Generative AI with Liminal

There are three paths to engaging with generative AI. The first is direct interaction with a generative AI model, such as OpenAI's GPT-4 or Anthropic's Claude 2. The second path is engaging with applications containing generative AI functionality, such as Microsoft's Office 365 Copilots. The third is building custom applications that leverage generative AI to provide a chat experience or retrieval augmented generation[4] (RAG). **Liminal provides horizontal security across all three engagement paths.**

## Direct Interactions with Generative AI Models

Liminal is model-agnostic by design. The Liminal Spaces web application provides a secure chat experience with any number of models that a company has licensed. As a user engages with the chat interface, Liminal applies the policies determined by the administrators in-stream, highlighting non-compliant or sensitive data.



Prior auth request for CAR T-cell therapy Pt L.M, 45yo firefighter, refractory non-Hodgkin lymphoma, dx 6/23

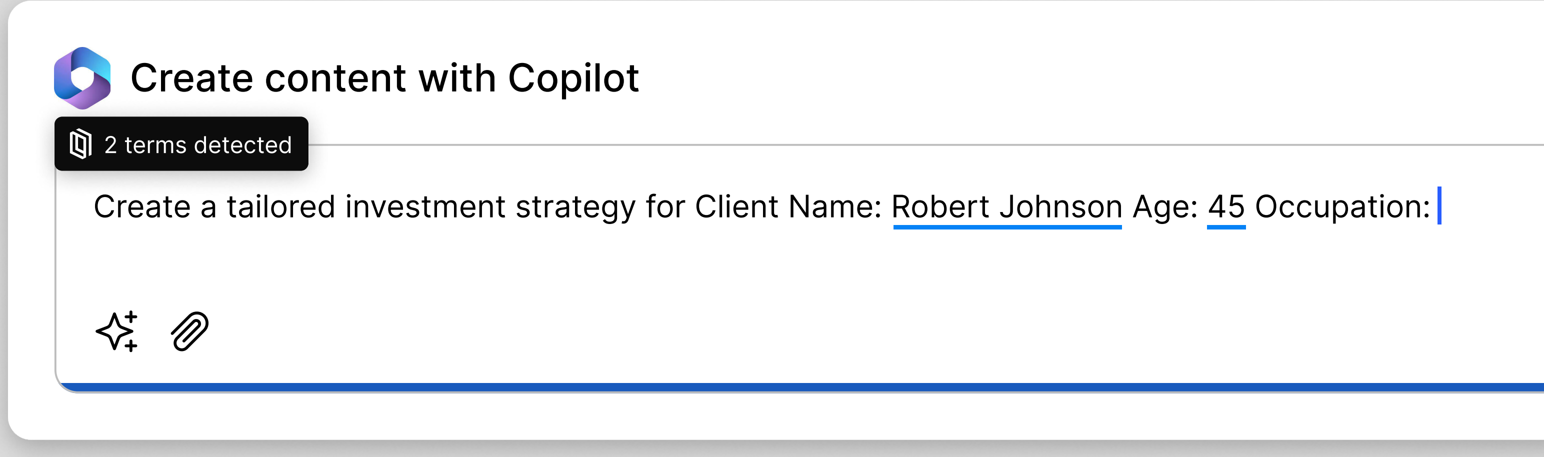
Upon submission, the Liminal Engine acts as an intermediary layer and appropriately cleanses the provided prompt (while maintaining context via intelligent masking) before its submission to the selected LLM. This behavior ensures that all direct interactions between users and AI models are secure and compliant with established policies (essential for maintaining the integrity of AI outputs and protecting sensitive information).

[4] "What is retrieval-augmented generation?", IBM, August, 22, 2023

## Using Generative AI-Enabled Third-Party Applications

Liminal also protects generative AI interactions within third-party client applications such as Microsoft's 365 Copilots, Slack's AI, and others. Via a lightweight desktop daemon or browser plugin, Liminal Streams monitors a desktop user's interactions with UI controls (such as text boxes) known to contain generative AI functionality.

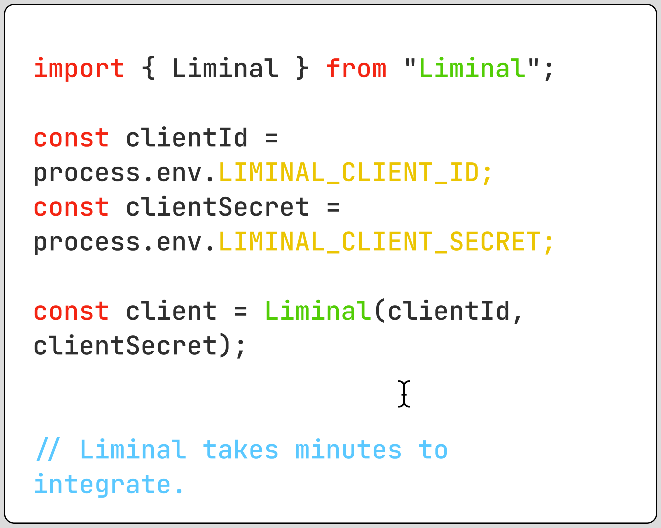
Upon detecting such an interaction, the input prompt is communicated to the Liminal Engine, where it is replaced with a cleansed version - without disrupting the intended user experience.



Create a tailored investment strategy for Client Name: Robert Johnson Age: 45 Occupation:

## Building Generative AI-Powered Custom Applications

Liminal extends its protective capabilities to custom-built applications by providing Liminal Elements, a multi-programming language software development kit (SDK). **With four lines of code,** the SDK offers a Liminal client object that accepts input prompts, transmits those prompts to the Liminal tenant for application of policies to identified non-compliant and sensitive data, forwards the cleansed prompt to the selected model, then returns the rehydrated output responses to the caller.



```
import { Liminal } from "Liminal";

const clientId =
process.env.LIMINAL_CLIENT_ID;
const clientSecret =
process.env.LIMINAL_CLIENT_SECRET;

const client = Liminal(clientId,
clientSecret);


// Liminal takes minutes to
integrate.
```

**Liminal Elements** allows developers to ensure that their code-level interactions with generative AI contain the exact detection and cleansing capabilities of Liminal Spaces and Liminal Desktop.

## Retrieval-Augmented Generation (RAG) and Development Frameworks

Liminal Elements seamlessly integrates with Retrieval-Augmented Generation systems, combining information retrieval with generative models to produce more accurate and context-aware outputs. Liminal Elements ensures the data retrieval process adheres to strict security and privacy standards, protecting sensitive information.

Various generative AI application development frameworks, such as Griptape and Langchain, are essential for bringing generative AI models into production applications. Liminal ensures that data fed into and generated by these frameworks is monitored and controlled according to the predefined policies, regardless of the underlying framework used.

## Administration Features in the Liminal Console

### Real-Time Monitoring, Management, and Logging

The platform's real-time alerting, observability features, and auditable logs offer a proactive approach to security management. By providing immediate insights into potential security incidents, Liminal enables organizations to respond swiftly to threats, minimizing the risk of data breaches and unauthorized access.

### Policy Enforcement and Governance

Liminal offers robust tools for setting and enforcing security policies at a granular level. This capability is crucial for maintaining compliance with diverse regulatory standards and ensuring ethical AI practices. Organizations can tailor their security controls to meet specific needs, including specifying which models users can access and data types they can share. Security administrators can set policies and permissions at an organization, team, or individual user level, and the Liminal Platform includes IdP integration capabilities for streamlined setup.

### Deployment

Liminal is exclusively deployed into a single-tenant infrastructure to isolate each organization’s data and resources, eliminating the risks associated with shared environments. This approach significantly reduces the potential for data breaches and unauthorized access, offering a robust security layer appropriate for organizations with stringent compliance and privacy requirements. Liminal may also be deployed privately, with the same infrastructure and application services existing within an organization’s owned Virtual Private Cloud (VPC).

On average, a Liminal tenant takes less than 15 minutes to provision and under 1 hour to configure.

## Embracing the Future with Generative AI

Generative AI is not merely a fleeting trend; rather it represents a fundamental shift in the technological paradigm. Its rapid adoption and the significant efficiency gains it offers are reshaping industries and reimagining how work gets done.

The journey toward fully harnessing the potential of generative AI is not without its challenges, particularly concerning data security and privacy. As highlighted in our exploration, if not adopted and deployed thoughtfully, generative AI can pose significant risks in areas like regulatory compliance, sensitive data protection, and organizational reputation.

Addressing these challenges requires a robust, multi-layered security approach that integrates policy, process, and technology. The role of solutions like Liminal is indispensable in this context, offering comprehensive tools for policy enforcement, data governance, and transparency. By ensuring that generative AI interactions are secure and compliant, organizations can mitigate risks associated with this technology.

Generative AI represents a transformative chapter in the story of technological progress. With the right security measures in place, organizations can confidently step into this new era, harnessing the power of generative AI to drive greater productivity, improve operational efficiency, and push the boundaries of innovation.

Liminal

Liminal empowers regulated enterprises to securely deploy and use generative AI. With Liminal, organizations have complete control over data privacy, security, and sovereignty – across any generative AI model, in every application you use, and in every application you build. Learn more about horizontal security and Liminal by visiting **liminal.ai**.