WHITE PAPER

Generative AI Security

in business applications

Table of Contents

This Document Will Cover the Following:

- Generative AI in Business Applications: An exploration of how Generative AI is being integrated into various business domains and the benefits it brings.
- Security Concerns: Detailed discussion of the primary security challenges, including data privacy, authentication issues, and the risks of overreliance on AI technologies.
- Best Practices and Mitigation Strategies: Practical advice and recommendations on safeguarding your applications against the inherent vulnerabilities of Generative AI.
- Conclusion: Summation of the critical points discussed, with a strategic outlook on maintaining security while leveraging Generative AI for business innovation.

By the end of this white paper, business leaders, IT experts, and policymakers will be equipped with essential insights and tools to ensure that Generative AI is implemented securely and ethically within their organizations.

Introduction

Generative AI, characterized by its ability to create new content and solutions from existing data sets, stands at the forefront of the next wave of digital transformation. Its applications range from content creation and product design to predictive analytics and customer service enhancements. As these technologies evolve, they promise significant benefits for businesses seeking efficiency, innovation, and competitive advantage. However, this rapid advancement also introduces complex security challenges that must be addressed to safeguard business interests, customer trust, and compliance with regulatory standards.

The objective of this whitepaper is twofold: to delve into the specific security concerns that generative AI presents in business applications and to outline strategic measures that organizations can adopt to navigate these challenges effectively. By examining the landscape of generative AI in business, assessing the impact of security risks, and recommending best practices for risk mitigation, this document aims to equip business leaders, IT professionals, and policymakers with the knowledge needed to harness the power of generative AI securely and responsibly.

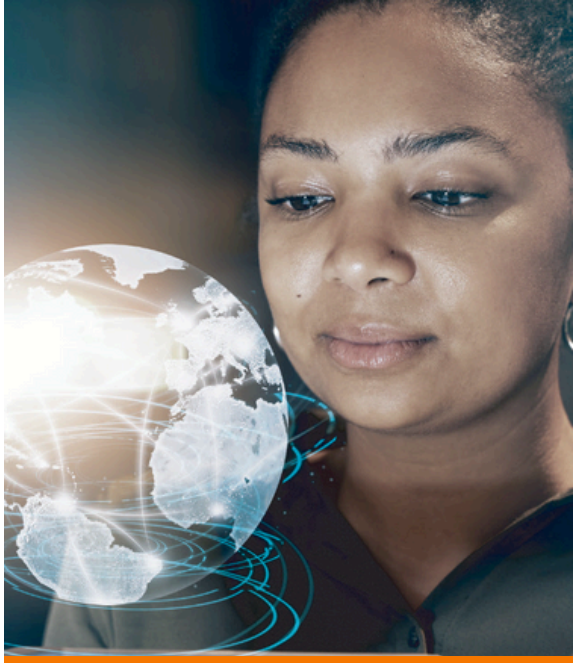
Generative AI in Business Applications

The deployment of generative AI within business applications marks a shift in business operations, offering enhancements in efficiency, innovation, and customer engagement. There are many potential applications in different areas of business that can bring value to companies.

Streamlining Operations and Decision-Making

Generative AI also plays a pivotal role in operational efficiency and strategic decision-making. In logistics and supply chain management, AI models predict and simulate complex scenarios, optimizing routes and inventory levels to minimize costs and improve service delivery. Financial institutions employ generative AI for risk assessment and fraud detection, leveraging its capacity to analyze datasets and identify patterns.

By strategically integrating generative AI, businesses can harness its potential to drive innovation, streamline operations, and foster more meaningful connections with customers. As this technology continues to evolve, its role in shaping the future of business will only grow, underscoring the importance of secure implementation.



Revolutionizing Customer Interactions

Generative AI transforms how companies engage with their customers, enabling the creation of personalized guidance when interacting. From customized promotion suggestions to guidance on building trust, AI algorithms analyze consumer data to deliver tailored experiences, significantly enhancing conversion rates and brand loyalty. AI-driven chatbots and virtual assistants also redefine customer service, offering 24/7 support and personalized interactions, thereby elevating the overall customer experience.

Security Concerns for Generative AI in Business Applications

Data Privacy and Protection

The advent of generative AI in business applications brings to light significant concerns regarding data privacy and protection. As these systems rely on vast datasets to train and operate, they inherently risk exposing sensitive information, be it customer data, proprietary business insights, or confidential operational details.

Challenges

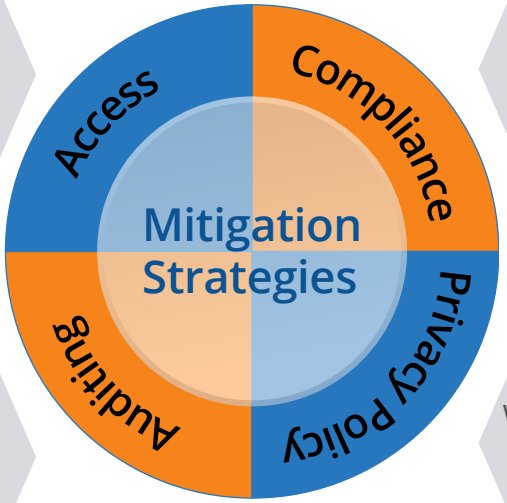
The primary challenge lies in ensuring that AI models access data securely and are not susceptible to breaches that could lead to unauthorized access or misuse. Additionally, there is concern for overreach when AI is accessing company data.

Access Control:

Companies must ensure that AI models only access data needed for their functions.

Auditing:

Regular audits should be established to monitor data usage, access, and potential vulnerabilities within AI systems.



Compliance with Data Protection Laws:

Companies must understand data protection laws such as GDPR in Europe and CCPA in California and implement specific safeguards for personal data.

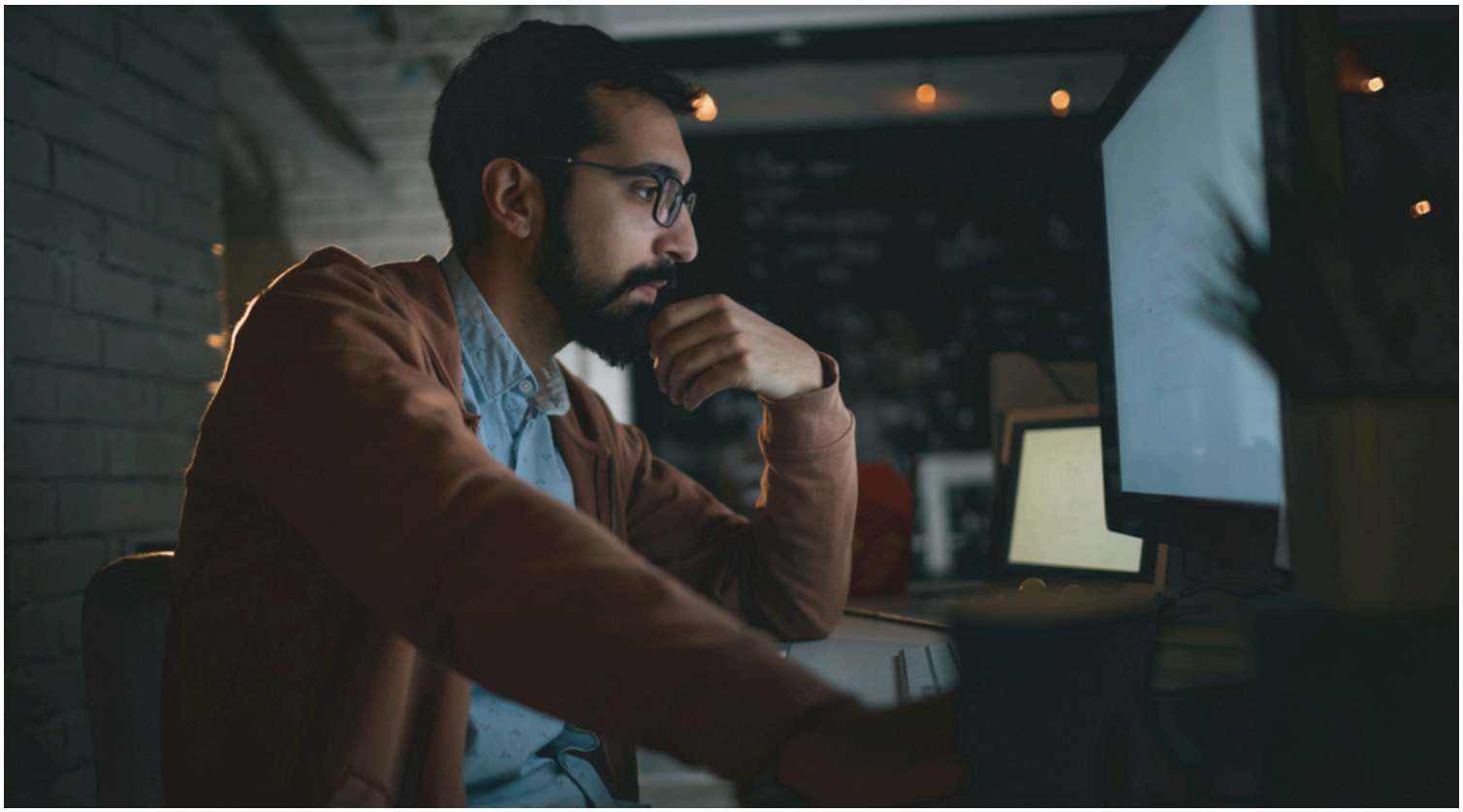
Understand your LLM's Privacy Policy:

When selecting a model for your AI use case, be sure to fully understand how they will use the data provided to the LLM for training and testing.

Businesses must prioritize these strategies to protect against data breaches, maintain customer trust, and ensure regulatory compliance. As generative AI continues to evolve, so must the approaches to securing the data that powers it.

Profound Logic

Using a holistic framework, like Profound AI, will help prevent data privacy issues by implementing best practices when interacting with the LLM. Specifically, Profound AI provides the following features to protect customers building AI functionality for their business applications.



Data Access Controls:

In the Profound AI IDE, customers select the specific tables and columns that will be accessible by the LLM. The framework also validates all SQL statements before they execute to ensure the LLM has not requested data outside of the defined tables and columns.

Logging:

Profound AI logs interactions with the LLM to allow for future auditing and troubleshooting. These logs are configurable to allow customers to fine tune the level of logging required. The logs are also consumable by enterprise monitoring software for automated notification of issues.

Data Access Exit Point:

An exit point is available for data access. This exit point allows customers to implement custom node.js code to implement additional validation before data passes to the LLM.

Authentication and Authorization

Generative AI is a powerful tool for enabling businesses to expand their capabilities. As businesses explore generative AI's capabilities, it is paramount to ensure that users are legitimate employees or customers and that they are authorized to perform the requested actions.

Challenges

As with any technology used to access business data and proprietary information, the challenge to identify users and authorize their activities is a crucial one. Improper implementation of authentication and authorization can leave your business susceptible to data exposure and system breaches.

Mitigation Strategies

- For internal business applications, your AI solution should integrate with the authentication used by that system. For more stand-alone implementations, companies should implement widely accepted methods such as OAuth 2.0 and JWT to secure access to your solution.

- Role-Based Authorization: After identifying a user, safeguards must be in place to prevent unwanted access to data or execution of logic.

Securing access to your AI agents is a top priority to safely integrate generative AI with your existing business applications and workflows. Implementing your own authentication and authorization systems without deep understanding of generative AI and security can leave your business exposed to risk of unexpected issues, both intentionally and accidentally.

Profound AI provides industry standard JWT authentication with more authentication methods to come. By integrating agents built with Profound AI into your existing business application screens, all use of the agents will be by authenticated users that are authorized to the screens where agents are deployed. This combined with proper access restrictions within the agents will ensure that all use of AI is by verified users with proper authority.

Prompt Injection

Generative AI gives users freedom to make requests that meet their specific needs. This freedom makes AI agents immensely powerful tools to enable users to be more productive and find unique solutions to business problems. This freedom can also be misused and kept under reasonable control in a business environment.

Prompt injection is a technique where unintended or malicious instructions are embedded within the input given to an AI model, particularly one that processes natural language. This method aims to exploit the model's design to perform actions or generate outputs that developers did not intend and that might be harmful or unauthorized.

Challenges

Mitigation Strategies

Prompt Control:
By controlling access to the prompt, businesses can control the capabilities of the AI agent and implement guardrails to prevent exploitation.

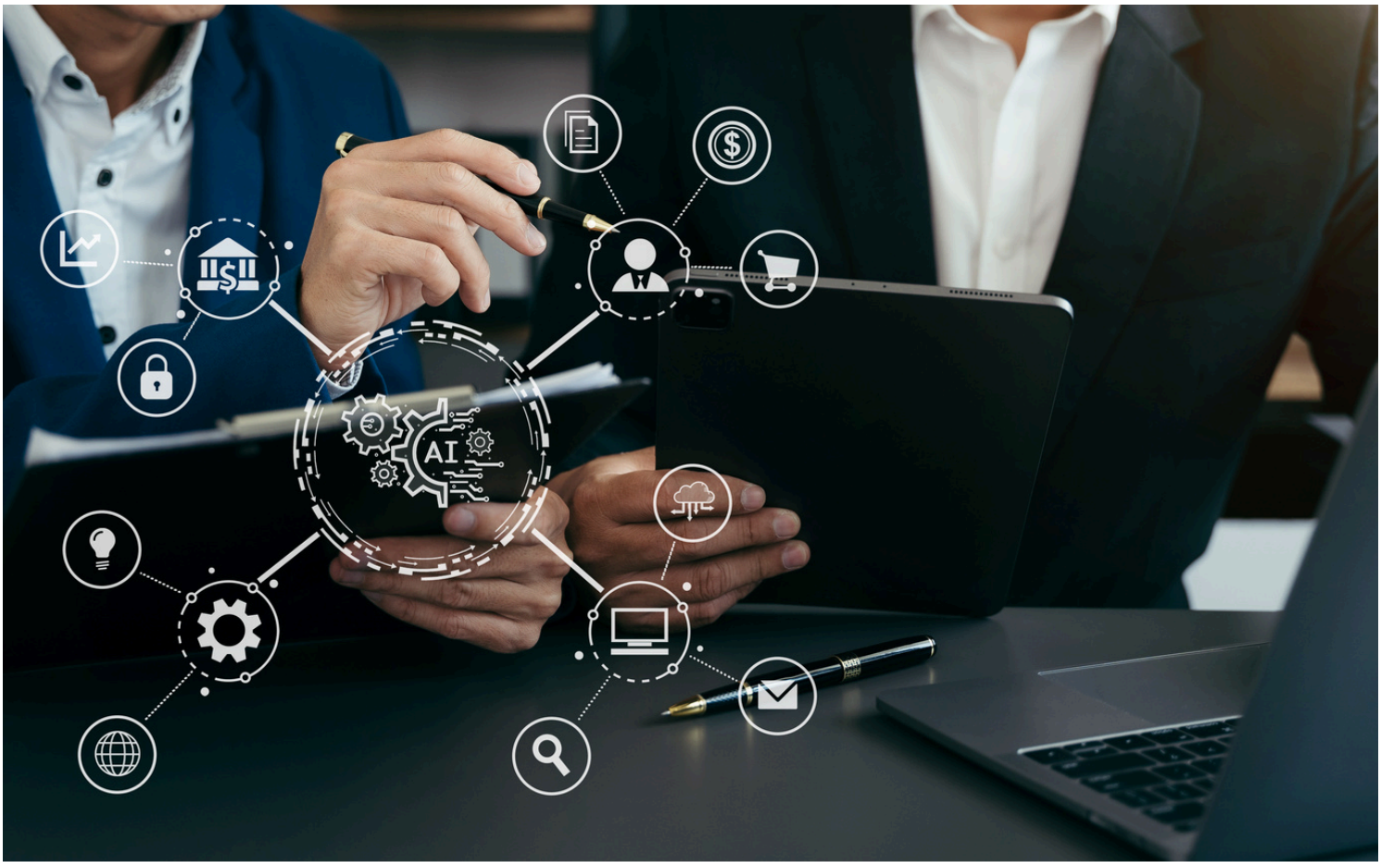
Input Sanitation:
Input from the end user should be examined to detect any possible injections to circumvent access controls or execute unintended functions.

Access Controls:
Access to data and application functions should be deliberate and controlled. Agents should have the minimum amount of authority to perform their intended functions.

To mitigate the risks of prompt injection, developers and users of generative AI systems must implement robust validation and sanitization of inputs, employ strict access controls, and maintain awareness of the evolving nature of such exploits.

The Profound AI framework assists businesses with these mitigations through security functionalities at the product level.

IT Controlled Prompts	Agents are built by IT for specific use cases. End users cannot view or alter the prompts used to define agents.
Observable Input	Input from the user does not affect the prompt. Input is also logged for examination should an incident occur, and exit points are available for custom validations.
Built-in Access Control	Data access is read-only and limited to tables and columns IT permits. Any data manipulation, program calls, or system interaction is done through low-code routines built by IT for that agent.



Overreliance

Generative AI increases productivity and creativity of employees when used effectively. As users integrate AI into their everyday workflow, it is natural to become more reliant on the powerful features that AI provides.

Challenges

Overreliance on AI within business applications can pose significant security concerns, as it may lead to vulnerabilities and risks that could compromise business operations, data integrity, and stakeholder trust. When businesses become too dependent on AI solutions without fully understanding or managing their limitations and potential failure points, they expose themselves to various risks.

Mitigation Strategies

Human Oversight:
While AI can provide insights that are amazing, generative AI is not perfect. AI output should not be trusted blindly. Output should be fact checked periodically or when questionable.

Documentation of Functionality:
AI systems require frequent updates and maintenance both in tooling and models. As AI agents are built, they may perform certain tasks automatically as part of their design. It is important to document and understand these functions should they need to be performed manually during an outage of the AI.

Businesses should implement balanced strategies that combine AI capabilities with human expertise, ensuring that AI systems are supportive tools rather than infallible solutions. It is crucial to maintain robust oversight mechanisms, conduct regular audits and updates, and foster a culture of continuous learning and adaptation to address the security challenges associated with overreliance on AI.

The Profound AI framework enables customers to implement proper strategies to combat overreliance on AI.

New models can be implemented with simple configuration and agents moved to the new models individually when least disruptive.

Low-code routines to enable automation and advanced functions are clearly denoted in the agent configuration.

Conversations are logged for review by others for proper oversight.

Best Practices

As businesses increase their use of generative AI, some general best practices can go a long way in keeping usage safe and secure. Setting a proper foundation early will reduce the risk of security incidents and data leakage going forward.

Continuous Monitoring

AI usage should be monitored for anomalies and misuse. There are several aspects to consider when building a monitoring strategy.

- **Token Usage:**
  To properly understand your spending with a commercial AI model, it is important to understand the concept of tokens and to examine your AI agents' activity. You may find certain agents using excessive amounts of tokens and refining the agent configuration could reduce token usage and save considerable money.

- **Bias and Erroneous Results:**

  As your users encounter unexpected or incorrect results, have a system in place to report them. This will allow your IT staff to review the logs for that conversation and determine the reason for the result. Corrections can be made for the future by refining the agent or educating the user on better strategies for interacting with AI.

- **Random Review:**

  Periodically select random agent interactions from the logs and review them. Did the output properly answer the query? Did the result take too long to return? Did the result reveal unexpected or sensitive data? The answers to these questions will help with refining agent definitions and improving AI results over time

Staying Up to Date

The world of generative AI moves fast. New models, model updates, and tool updates are released frequently. New models bring faster speeds, more capabilities, and lower costs. Model and tool updates mitigate security issues, patch bugs, and introduce enhancements. If you are not staying up to date, you are not only missing improvements, but also putting your business at risk.

It is important to establish a regular cadence for installing, testing, and deploying updates for all aspects of your AI solution. During this update cycle, it is also a good idea to educate your users on new capabilities as well.

Minimize Authority

Whether building your own AI solutions or utilizing tools like Profound AI, always start with a zero-access state. Add access to data elements, documents, and functions one at a time until your desired output is achieved. This minimizes the opportunity for data leakage, AI overreach, or unintended results that could cost your business time and money.

Once the desired output is achieved, thoroughly test your agent with requests of varying complexity. This will uncover any additional access needs that may have been missed during development.

User Training and Awareness

Educating your user base on proper usage of AI is just as important as your implementation of AI functionality. Simply creating agents and putting them out to users is a recipe for disaster.

Create an AI Acceptable Use Policy:

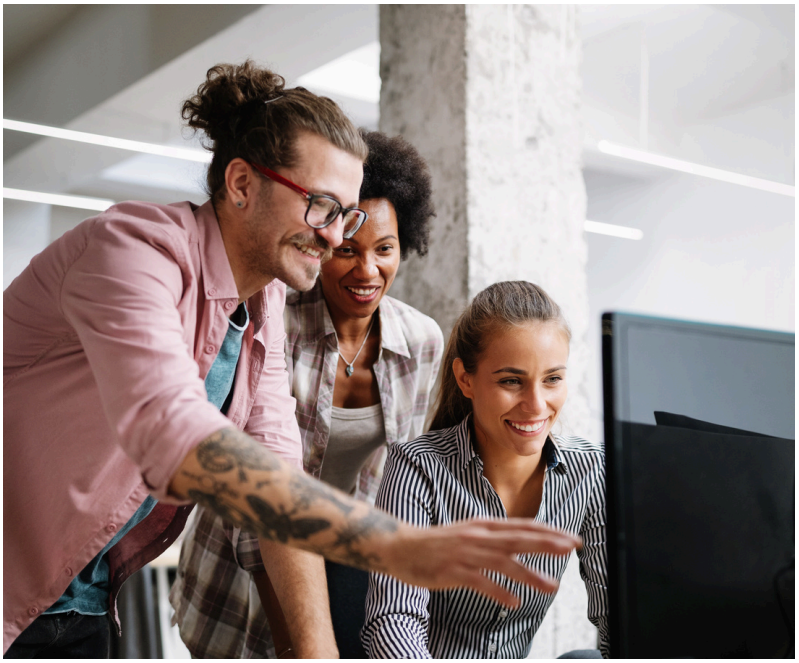
An AUP provides guidance to both IT and end users on responsible and ethical use of AI. This document will set the standards to which everyone will be held. This is the first step in any generative AI journey.

Offer Training on AI Usage:

Educate your users on how to use AI effectively in a business environment. Include guidance on how to ask questions of an agent to get desired results. Provide training, both internal and third party, to increase your users' skillsets to include understanding of generative AI. Effective use of AI is a skill that must be learned.

Compliance Education:

Outline the legal frameworks and standards relevant to AI applications in business, such as GDPR in Europe or HIPAA in the healthcare sector in the United States. Educate your IT staff and users on how to stay compliant with these regulations.



Partnering for Skills

When you do not have the skillset internally to properly implement complex solutions, such as generative AI, reach out to a partner that has the expertise to guide your business through building a roadmap, educating your staff, as well as planning, building, and delivering projects.

Profound Logic has experts and tools to ensure success at any stage of your AI journey.

Conclusion

As generative AI continues to revolutionize business applications, it brings forth both unprecedented opportunities and security challenges. This whitepaper has explored the landscape of generative AI in business, focusing on the importance of securing AI-driven processes and data. By adhering to best practices, implementing a holistic framework, being security focused, and fostering a culture of continuous vigilance and improvement, businesses can navigate the complexities of generative AI. Embracing these strategies will not only protect against potential threats but also unlock the full potential of generative AI to drive innovation, efficiency, and competitive advantage.

Ready to learn more about how Profound AI and impact your business? Schedule a demo and learn firsthand how AI can transform your business applications for the better!

Legal Disclaimer