






# Secure Data, Preserve Privacy, Boost Accuracy



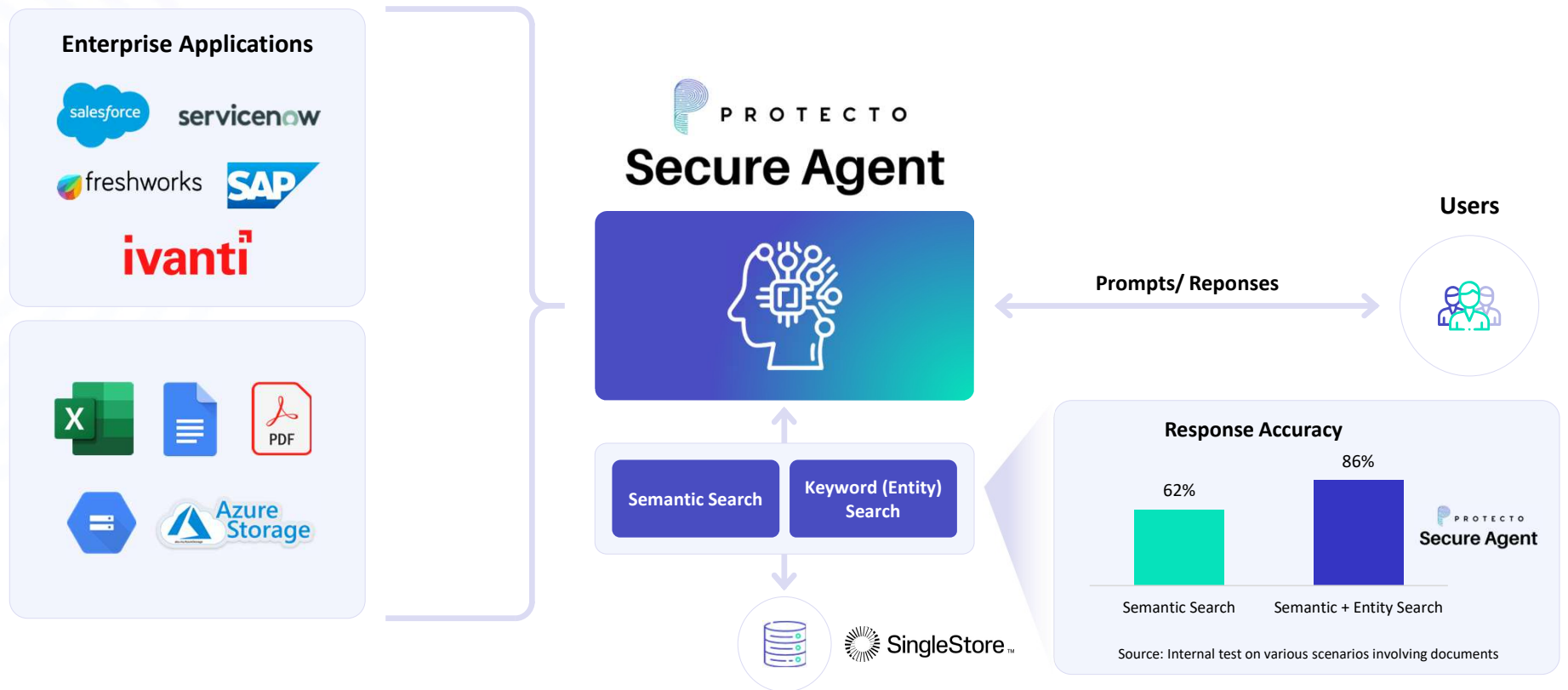
# The Hidden Risks of Sharing Data with AI



 Risk	 Description	 Impact
Privacy Violations	Exposure of sensitive data (PII, PHI).	Loss of customer trust, lawsuits, penalties
Data Leaks	AI models may memorize and expose confidential data.	Sensitive information becomes publicly accessible or misused.
Compliance Issues	Breaches of GDPR, HIPAA, or other regulations.	Hefty fines, legal challenges, and operational disruptions.
IP Leaks	Proprietary data or strategies could be exposed.	Loss of competitive edge and potential exploitation by competitors.

## Additionally, Protecto Improves Superior Accuracy

Leveraging Entity Recognition Capabilities & SingleStore's Hybrid Search



# Enhancing Response Accuracy with Entity Recognition



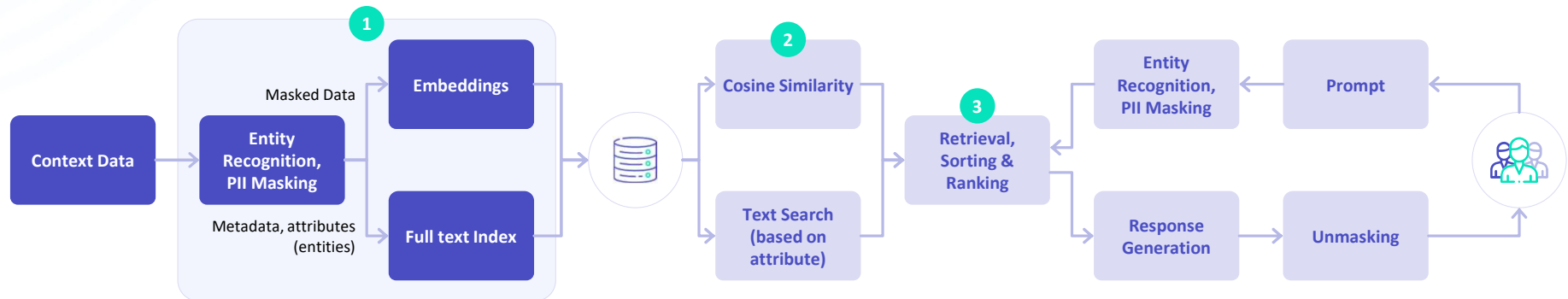
## Our Approach

- 1 Entity Recognition for Masking & Indexing During Preprocessing**
  - Identify attributes such as PII/PHI (e.g., name, location, medical record number) in chunks and documents.
  - Mask PII/PHI attributes as required for on use cases
  - Indexing - Store all the identified attributes (both masked and unmasked) as part of pre-processing.
- 2 Retrieval**
  - **Cosine Similarity Search**
    - Perform cosine similarity-based retrieval for semantic relevance.
  - **Text Search with Attributes**
    - Conduct text searches on identified PII attributes in chunks and documents.
- 3 Sorting & Ranking**
  - Rank chunks based on **Cosine similarity score and Presence of attribute keywords**



## Results: Superior Retrieval Accuracy

- Consistently outperforms cosine similarity-only Retrieval-Augmented Generation (RAG).
- Ensures contextually relevant and attribute-specific document retrieval



# Example: Hybrid Search Delivers Superior Relevant Retrieval

