Capturing Tacit Knowledge through Generative AI: A Context-Driven Feedback Architecture for Reliable Output

Dr. Darren W Pulsipher

Organizations seeking to leverage Generative AI (GenAI) for high-*stakes, unstructured data processing face a critical challenge: how to reliably capture and scale the tacit knowledge of Subject Matter Experts (SMEs) without costly model fine-tuning or risking sensitive data exposure.*

1. Introduction

Tacit knowledge—insight gained through experience that experts rely on to interpret complex, ambiguous data—is essential for accurate decision-making across domains like government, finance, legal, and healthcare cite:[nonaka1995knowledge; wiig1993knowledge]. However, this form of knowledge is difficult to capture and leverage in traditional systems, and even harder to preserve and utilize at scale.

In today's data-driven world, organizations face significant challenges in effectively processing and interpreting unstructured data. While traditional systems struggle with complex, ambiguous information, advances in generative AI technology offer promising new capabilities for data interaction and analysis. However, these AI systems often lack the critical context and domain expertise needed for reliable, production-ready results. The technology gap between raw AI capabilities and practical business needs has led to issues with accuracy, consistency, and trust in AI-generated outputs [1]. An emerging solution involves capturing expert knowledge and injecting it into AI workflows through dynamic context engineering and prompt injection.

2. The Challenge: AI Without Expert Context

GenAI models are designed to be general-purpose, trained on massive corpora. While this makes them powerful for language generation, it also leads to:

- Inconsistencies in domain-specific interpretation
- High variability in outputs for similar inputs
- Hallucinations and incorrect inferences [2]
- Lack of traceability in decisions [3]
- Barriers to using sensitive or classified data due to privacy risks

Fine-tuning models with domain-specific data is often cost-prohibitive, resource-intensive, and inflexible. Instead, we propose a **contextual feedback loop** where SME interactions form the basis of an evolving guidance system—what we call a **Context Atlas**.

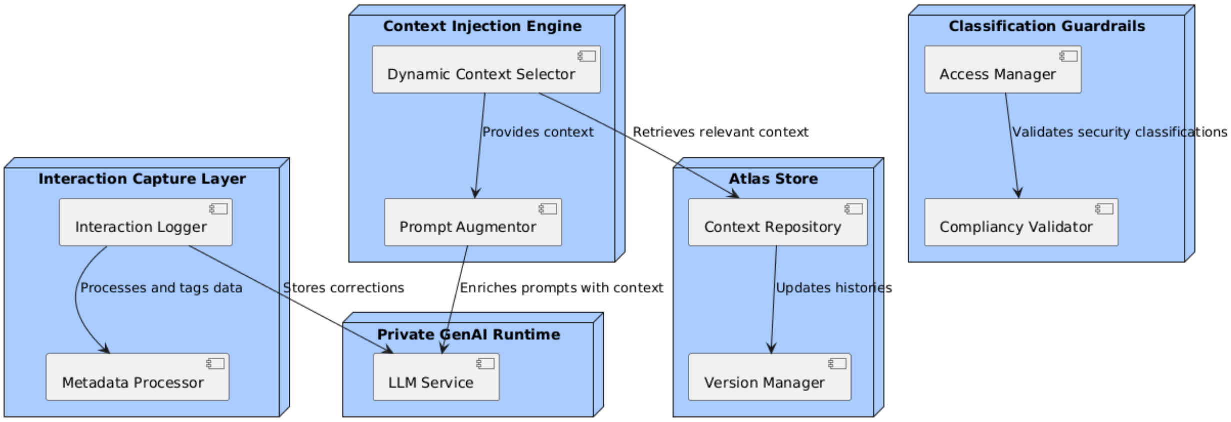
3. Architecture Overview: Capturing Tacit Knowledge via Context Mapping

The **High-Level Architecture Diagram** shows how the framework delivers secure, context-aware AI interactions by integrating document management, context injection, dynamic prompt augmentation, and expert feedback. This diagram provides a foundational reference for understanding how components interact, demonstrating how the system processes queries, enriches them with context from the **Context Atlas**, and manages them securely.

The sections below explore key systems, including:

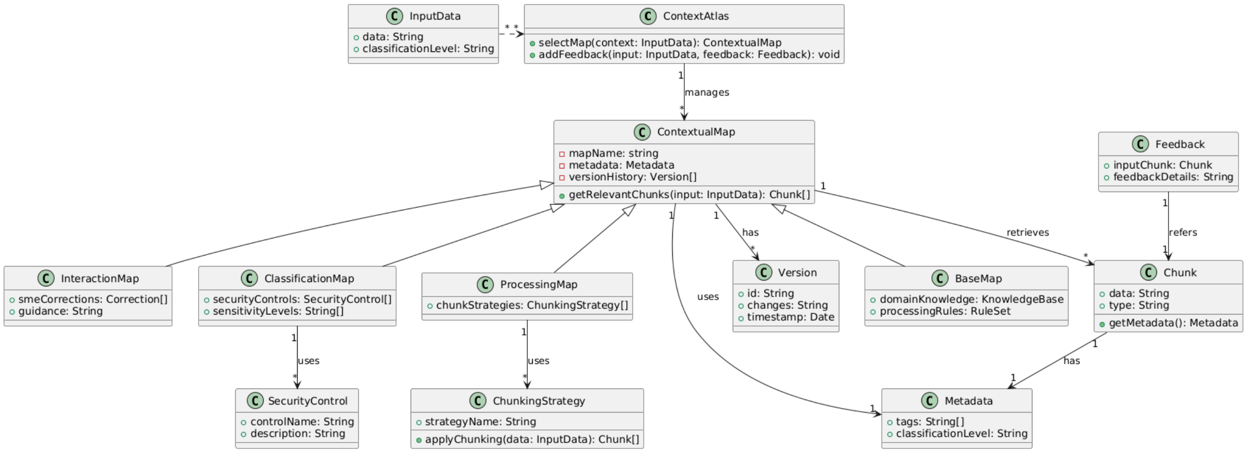
- **Document Management** for secure data ingestion and retrieval,
- **Query Processing and Context Augmentation** for real-time prompt enrichment using expert-validated contextual maps,
- **Security Layers** that ensure data privacy through classification guardrails and runtime protections,
- And the **Feedback Mechanism** that drives system improvement through SME (Subject Matter Expert) interaction.

The high-level architectural view clarifies how these components work together to deliver robust, scalable, and secure AI-powered solutions.



3.1. Key Concept: The Context Atlas

The **Context Atlas** is a structured knowledge layer that sits alongside the GenAI model. It serves as an intelligent mapping framework that continuously evolves through expert interaction.



The Atlas is constructed from multiple integrated components:

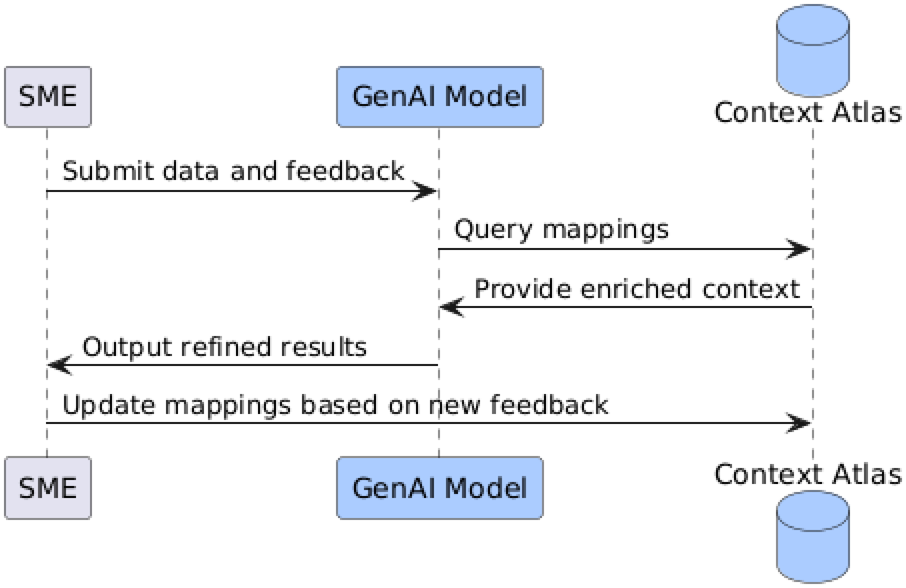
- Multiple contextual maps that evolve based on SME interaction level and data processing needs:
  - Base Map: Core domain knowledge and fundamental processing rules
  - Interaction Map: Captures evolving SME corrections and guidance
  - Classification Map: Security and data sensitivity controls
  - Processing Map: Chunk-specific handling strategies
- Data chunking strategies tied to each map type:
  - Semantic chunking for domain knowledge
  - Token-based chunking for interactions

  - Classification-aware chunking for sensitive data
  - Stream processing chunks for real-time data
- Dynamic map selection by LLM based on:
  - Input data analysis
  - SME prompt evaluation
  - Required processing context
  - Security classification
- Bidirectional associations between input patterns, contextual requirements, and target output behaviors
- Granular security classification metadata that enables context-aware data handling and ensures appropriate processing isolation based on sensitivity levels [4]
- Versioned interaction histories that enable auditing and rollback capabilities
- Metadata tags for domain-specific knowledge categorization and retrieval

This comprehensive atlas functions as the LLM's navigational system, dynamically injecting relevant expert guidance into each query and enabling SMEs to systematically correct, annotate, and evolve the model's behavior over time—all without requiring resource-intensive retraining cycles.

3.2. Feedback Loop Mechanism

The system operates through an adaptive feedback loop that continually improves the AI's responses based on expert input. This loop ensures that the system maintains high accuracy while building a growing repository of verified contextual knowledge through multiple contextual maps.

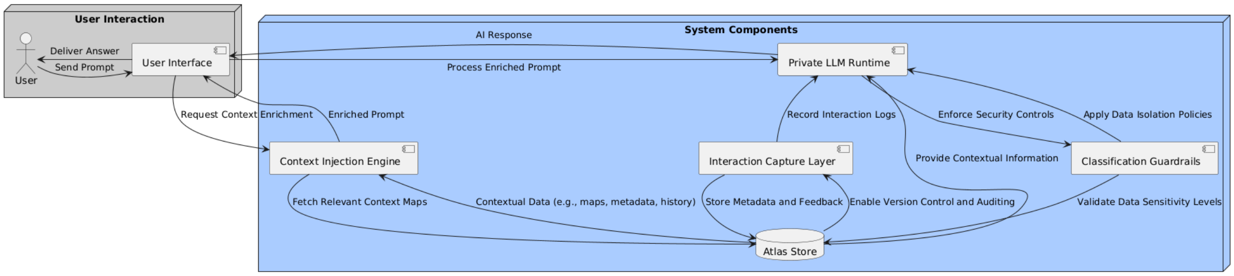


1. User submits prompt
2. LLM generates initial response while dynamically selecting appropriate contextual maps based on:
    - Data classification level
    - Domain expertise required
    - Input pattern recognition
    - Processing workflow stage
3. SME reviews and corrects using selected contextual maps
4. Feedback is stored in relevant maps within Context Atlas
5. Future queries dynamically select and use appropriate Atlas context maps

This creates a continuous learning cycle that preserves expert-driven control and accountability while enabling systematic improvement without model retraining [5].

4. Technical Components

The system comprises several key technical components that work together to enable secure, contextually-aware AI interactions. The following diagram illustrates how these components interact to process queries, inject context, and maintain security boundaries.


User Interaction
User
Deliver Answer
Send Prompt
User Interface
System Components
AI Response
Process Enriched Prompt
Private LLM Runtime
Request Context Enrichment
Enriched Prompt
Record Interaction Logs
Enforce Security Controls
Apply Data Isolation Policies
Context Injection Engine
Interaction Capture Layer
Provide Contextual Information
Classification Guardrails
Fetch Relevant Context Maps
Contextual Data (e.g., maps, metadata, history)
Store Metadata and Feedback
Enable Version Control and Auditing
Validate Data Sensitivity Levels
Atlas Store

4.1. Private GenAI Runtime

A containerized execution environment (e.g., on-prem or VPC-isolated) that provides complete data isolation and security through infrastructure-level controls. This component ensures all AI processing occurs within organization-defined trust boundaries, preventing sensitive data exposure while enabling high-performance inference. The runtime leverages hardware security features and encrypted memory spaces to maintain Zero Trust principles throughout the execution lifecycle [6].

4.2. Interaction Capture Layer

A comprehensive system for recording and processing all SME interactions with the AI system. This includes:

- Detailed logging of corrections and modifications made to AI outputs
- Capture of SME rationale and contextual guidance
- Metadata tagging for domain categorization
- Version control of interaction histories
- Performance metrics and quality indicators

4.3. Context Injection Engine

A sophisticated AI-powered system that dynamically enriches prompts with relevant expert context from the Atlas. Key capabilities include:

- Real-time context selection based on input analysis
- Dynamic prompt augmentation
- Context relevancy scoring
- Automatic map selection and blending
- Security classification awareness

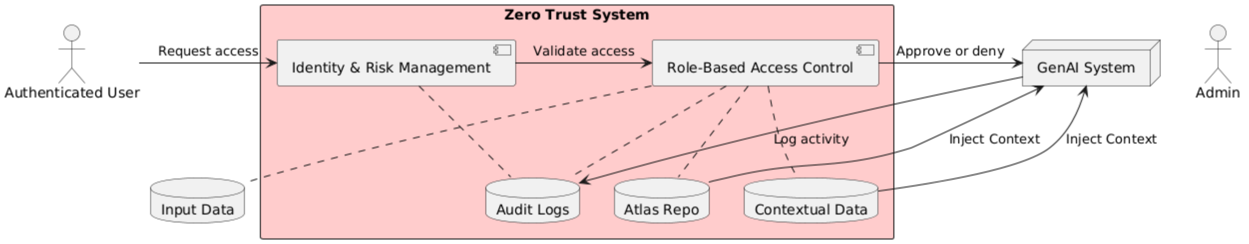
4.4. Atlas Store

A highly-structured, versioned, and queryable database optimized for storing and retrieving contextual information. The store includes:

- Hierarchical context maps
- Annotated interaction histories
- Example repositories
- Bidirectional relationship graphs
- Classification metadata
- Version control mechanisms

4.5. Classification Guardrails

A multi-layered security control system leveraging Zero Trust principles to ensure sensitive data remains properly isolated based on established classification policies [7]. The system assumes no inherent trust and requires continuous verification of every access attempt.



Features include:

- Data classification enforcement with dynamic policy evaluation and verification at every access point
- Granular access control management using least-privilege principles and just-in-time access
- Security boundary maintenance through micro-segmentation and continuous trust verification
- Comprehensive audit logging of all access attempts, authorizations, and data movements
- Real-time compliance validation with automated policy enforcement and violation detection

5. Benefits of the Approach

The integration of a Context Atlas with generative AI systems offers several compelling advantages for organizations seeking to leverage AI capabilities while maintaining control, security, and accuracy.

5.1. Accuracy & Consistency

The approach delivers significant improvements in output quality and reliability through systematic context enrichment. Context mappings enhance response reliability by providing the AI system with domain-specific knowledge and expert guidance. As SME interactions continuously reinforce proper interpretation and handling of domain concepts, the system's outputs become increasingly aligned with real-world expert expectations and requirements [8].

5.2. Scalability

This architecture offers a highly efficient path to scale AI capabilities across an organization. As a cost-effective alternative to traditional model fine-tuning, it allows organizations to improve AI performance without expensive retraining cycles. SME corrections automatically scale across similar queries through the context mapping system, enabling broad impact from individual expert interactions. The Atlas continuously grows and evolves without incurring retraining overhead, providing sustainable long-term scalability [1].

5.3. Privacy & Security

The Context Atlas approach maintains robust security while enabling AI innovation. By keeping sensitive data within trusted boundaries and inheriting existing data classification schemes, it ensures AI systems operate securely within organizational constraints. The architecture is fully compatible with Zero Trust security principles and FedRAMP compliance requirements, making it suitable for high-security environments cite:[intel2023zerotrust; fedramp2023ai].

6. Case Example: Application in the Public Sector

A prototype implementation of this approach with the U.S. Census Bureau illustrated its impact. By capturing SME corrections into a reusable context atlas:

- LLM hallucinations were reduced by over 40%
- SME-to-SME variation in GenAI outputs dropped significantly
- Institutional knowledge was transformed into a structured resource usable by other teams [9] at scale.

Generative AI offers new ways to interact with unstructured data, but its success depends on integrating expert reasoning into the AI process. Traditional fine-tuning of models can be costly and take weeks or months of compute time, depending on model size and data requirements. This paper presents a method for capturing and applying SME knowledge without fine-tuning models, instead using an interaction-driven context mapping strategy. This process not only reduces hallucinations but improves reproducibility, transparency, and trust in AI-assisted workflows [1].

References

[1] R. Bommasani and others, “On the Opportunities and Risks of Foundation Models,” *arXiv preprint arXiv:2108.07258*, 2021, doi: 10.48550/arXiv.2108.07258.

[2] H. Palangi and others, “Contextual Augmentation in Generative AI: Reducing Hallucinations through Structured Feedback,” *Nature Machine Intelligence*, vol. 5, no. 4, pp. 321–335, 2023, doi: 10.1038/s42256-023-00664-2.

[3] OpenAI, “Reinforcement Learning with Human Feedback (RLHF): Guiding Generative Models.” AI Research Blog, 2023. doi: 10.48550/arXiv.2203.02155.

[4] F. R. A. M. P. PMO, “Security Assessment Framework for AI in Government Use Cases.” 2023. Available: https://fedramp.gov

[5] T. Brown and others, “Language Models Are Few-Shot Learners,” *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020, doi: 10.48550/arXiv.2005.14165.

[6] I. Corporation, “Zero Trust Architectures for AI in Multi-Cloud Environments.” Intel Whitepaper, 2023. Available: https://www.intel.com

[7] G. D. P. R. (GDPR), “Privacy and Security Requirements for Artificial Intelligence Systems.” 2018. Available: https://gdpr-info.eu

[8] A. Singh and M. Chen, “Iterative Feedback Loops for Domain-Specific Generative AI,” in *Proceedings of the ACM International Conference on Knowledge Engineering and Ontologies*, 2024.

[9] K. M. Wiig, *Knowledge Management Foundations: Thinking About Thinking – How People and Organizations Create, Represent, and Use Knowledge*. Schema Press, 1993.