ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14

A Framework for Trustworthy Agentic Automation in Enterprise Environments

Luca Bianchi
Asst Professor

Abstract

Activity for its own sake is not a strategy. Enterprise strategy demands a response to genuine business needs: meeting customer service levels, keeping the supply chain functional, detecting and managing fraud, and satisfying regulatory requirements. The shape and scale of that response should reflect both established organizational goals and the enterprise's risk appetite.

Where these needs are pressing but the business is not yet fully equipped to address them, planning becomes essential. Several paths are available. An organization might choose the simplest route to meeting service levels. Alternatively, it may be more practical to invest early effort in reducing long-term workload — through prototyping or piloting, for example. In some cases, a single substantial investment can sustain performance at the desired level for an extended period with minimal further effort. In other cases, a response may be mandated by a central function rather than chosen locally.

Where enterprise strategy calls for such responses, agentic AI offers a technology capable of fulfilling them. Financial operations and fraud detection illustrate two areas where a best-practice Retrieval-Augmented Generation (RAG) approach is particularly well suited. Financial operations resemble an advisory service: interactions are largely one-off, with numerous edge cases. Fraud detection, by contrast, requires high-volume monitoring with zero tolerance for false negatives. Both domains carry an added demand for explainability and regulatory compliance.

Agentic AI can address these needs directly. At a narrow or operational level, agentic services can be deployed to replicate critical functions or guard against failure deep within the organization. However, its most strategically significant application lies in improving the availability and quality of information across the business ecosystem. A strategic RAG layer, delivering full-featured Information-as-a-Service, opens the door to invention and reinvention throughout the organization. Building this capability requires ensuring that information is accurate, complete, and current — and ensuring that the work of maintaining it is either automated or routed to wherever it can be handled most effectively within the ecosystem.

Keywords—Agentic AI Strategy, Enterprise Risk Appetite, AI-Driven Decision Systems, Strategic RAG Architecture, Information-as-a-Service, AI in Financial Operations, Fraud Detection Systems, Explainable AI Compliance, Enterprise AI Adoption, AI Governance Frameworks, Risk-Based AI Deployment, Intelligent Information Systems, Business Process Automation, High-Volume Monitoring Systems, AI-Driven Advisory Services, Data Quality Management, Autonomous AI Services, Enterprise Knowledge Systems, AI-Enabled Strategy Execution, Scalable AI Infrastructure.

I. Introduction

Enterprise organizations are investing significant resources to harness the potential of Artificial Intelligence (AI) across their operations. However, the overwhelming focus on the prospects of Autonomous AI—systems that operate independently of direct human involvement—can mask important challenges surrounding the governance of these technologies. An equally valid but often overlooked class of AI technologies that act on behalf of human decision-makers—Agentic AI—is gaining traction across enterprise use cases. Agentic AI is characterized by the use of Machine Learning or other AI technologies to provide intelligence-based support capabilities in decisions with a high level of human involvement. These systems free decision-makers from time-consuming repetitive tasks and processes,

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14

allowing them to focus their cognitive resources on the tasks that matter—including strategic decision-making, stakeholder engagement, simulation, and experiment prototyping. Fulfilling this key function requires extensive knowledge and insight into diverse, disparate, and unconsolidated data sources across the organization acting as the true proxy of the user and effectively making recommendations and decisions that align closely with the goals of the user.

Agentic AI technologies can be classified into three layers—Data, Model, and Interface—similar to the levels of an RAG architecture. However, enterprise stakeholders must also remain vigilant concerning the RAG systems that enable Agentic AI use cases to ensure that deployment is safely managed. A RAG architecture should address a range of concerns during the deployment process. These security and risk considerations, if neglected, can expose organizations to wide-ranging risks in these high-stakes use cases.

II. Theoretical Foundations of Agentic AI

Enterprise automation systems seek to fulfil business objectives without human intervention; conversely, agentic AI systems operate with human-in-the-loop control but deploy high levels of agency. Exploring the relevance of agentic agency and autonomy to enterprise automation reveals several characteristics of human decision-making: an agent generally knows the state of the world when making decisions; agents can distinguish between and prioritise different types of decisions; decisions are understood as taking action in the world, that is, as changing the current state of the world. While agentic AI systems may be governed by established policies that define acceptable operating conditions and boundaries, there are few limitations on their internal mechanics. AI systems typically supporting enterprise operations are model-driven, that is, the decisions they make are a product of a single inference made using a supervised machine learning model trained on labelled historical data. While these systems are demonstrably useful, they are not operated as autonomous systems, as they do not have the capability to work towards their own goals; agents supporting the enterprise thus possess agentic—rather than autonomous—agency. The difference has implications for strategy development and governance in the enterprise.

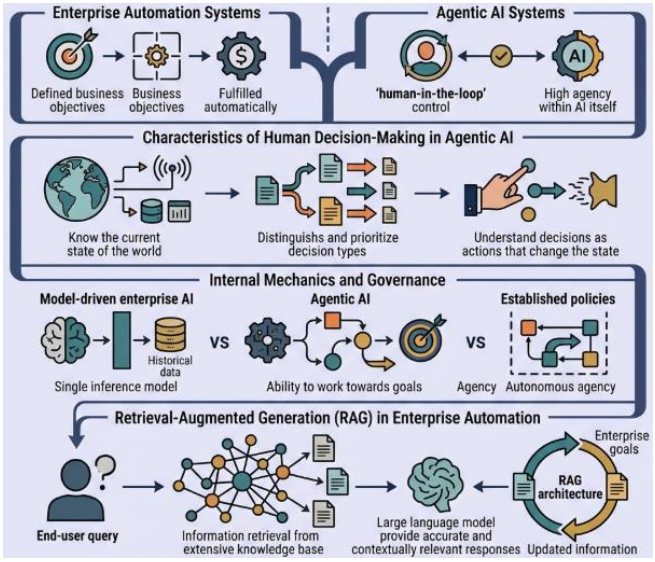


Fig 1: A Concise, Innovative, Research-Grade Title and Enterprise and Agentic AI Systems

To the author's knowledge, the application of these principles to enterprise automation has not previously been addressed. Retrieval-Augmented Generation (RAG) systems offer an alternative approach, combining large language models with information retrieval techniques to provide accurate, relevant, and temporally appropriate responses to end-user queries. The primary reason for the success of RAG systems is knowledge retrieval; response quality improves as the size of the knowledge base increases and the probability of the end user asking a question that has been previously seen by the model approaches 1. RAG systems are thus well-suited for enterprise-facing applications, where the knowledge base is frequently updated and aligned with the user's goals. While the application of RAG systems to enterprise automation has been addressed, most implementations do not leverage the architecture. A novel architectural framework supporting RAG-driven enterprise automation uses end-user queries as the main input to the RAG architecture.

A. Agentic Agency and Autonomy in Enterprise Contexts

Agents with agentic agency possess the capacity to plan and instigate decisions that achieve their objectives. By analogy, an agent that acts autonomously can be defined as one that exhibits conduct such that, if it existed in the world, it would be considered responsible for some aspects of the

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14

choices that determined that conduct and, hence, for the state of the world resulting from the choices made.

A distinction is made between decision-making, goal-setting, and action. Decision-making aligns the agent with subsidiary motives or desires, reflecting the agent's more distal objectives. As indicated earlier, RAG naturally supports decision-making via hierarchy and threading. The outputs returned, aligned with different levels of detail, orientation, and abstraction, are stored as predecessors. Goal-setting defines which immediate focus is to be pursued via the appropriate foregrounding of subsidiary motives. The action cycle actually calls on only the action components of the four-part model. The big difference between the action component of agentic AI and that in the original focus change theory is that a temporal-dynamic internal model is used. Indeed, the decision-formation, goal-setting, and action components of agentic AI together fulfil the very definition of a self-constructed, self-consistent, internal simulation of an agent's environment as an extract of a full-blown predictive world model, specifically tuned to the situation the agent is currently in. The key difference with current autonomous systems is that the enterprise information systems producing the KRR outputs now fulfil their assigned roles within an actual enterprise.

Equation 1: RAG response generation

Step 1: User submits query.

$$q = \text{user query}$$

Step 2: Retrieve relevant knowledge.

$$K_q = R(q, D)$$

Step 3: Combine query with retrieved context.

$$C = q + K_q$$

Step 4: Generate answer.

$$y = G(C)$$

Final:

$$y = G\big(q + R(q, D)\big)$$

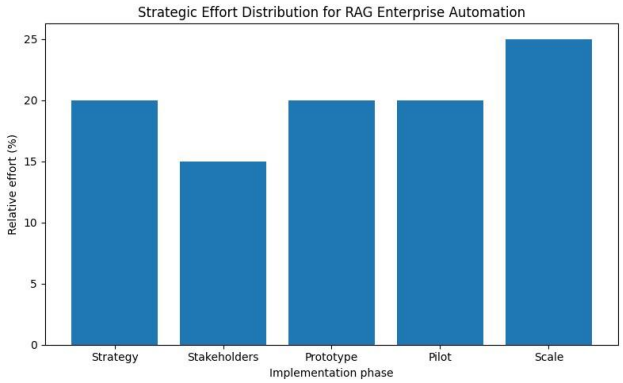
B. Retrieval-Augmented Generation: Principles and Benefits

The principles and architecture of Retrieval-Augmented Generation (RAG) are articulated to establish a comprehensive framework for architecting RAG-centric automation solutions, and to serve as a reference for specifying the requisite external data sets. RAG enablers are defined and grouped into three categories: knowledge retrieval, knowledge grounding, and model alignment.

Leverage of domain-specific knowledge enhances the veracity, reliability, and interpretability of Machine Learning (ML) models, including performance on classification and forecasting tasks. Such knowledge can also reduce response time thus speeding up decision-making cycles. Retrieval-Augmented Generation (RAG) is a concept that encompasses a methodology, a specific architecture, a set of principles, and supporting enabling components and technologies. RAG systems retrieve contextually-relevant knowledge, such as textual passages, from information sources external to the generative model at inference-time, and use that knowledge as additional input with the goal of improving the quality of the generated outputs. Categorizing RAG methods into three classes based on the source of knowledge referenced during generation is also informative: 1) segment-indexed approaches that leverage language model embeddings; 2) search-indexed techniques that determine shortest distances in dense vector spaces; and 3) search-indexed methods that use traditional Information Retrieval (IR) systems and pipelines.

The overall architecture of a RAG system can be divided into two components Data and Model layer. The generic data flow follows the order: 1) data generation; 2) data storage ; 3) data preparation (indexing); and 4) knowledge retrieval readying the model for the inference phase. The inference phase can be further decomposed into the three main stages: 1) context preparation; 2) model inference; and 3) model-specific post-processing with feedback capability. RAG systems are advantageous, since they make knowledge retrieval a first-class function, location-agnostic, and enable knowledge-store pipelines that transform and prepare knowledge repositories for efficient retrieval."

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14



III. METHODOLOGY

The study presents an architectural framework for enterprise automation using Retrieval-Augmented Generation principles and mechanisms. Supported by a Data Layer, overseeing content creation, collection storage, security, and veracity; and a Model Layer comprising agentic AI elements leveraging agentic Decision Models, the framework enables autonomous business decisions that strive to secure and ensure compliance with relevant governance requirements. Actions across the enterprise are thus supported, controlled, and integrated while Employing Well-Stacked Causal Policies that provide both decision-making and action management fidelity.

The architectural foundations of both layers of the proposed RAG-centric automation solution are establishment. Data sources, repositories, ontology, provenance, and quality assurance mechanisms are cataloged and RAG principles are recalled, together with the requirements for an agentic Decision Model capable of handling the demands imposed on action management. The resulting definitions serve as preparatory groundwork for agentic decisions and actions supported by RAG-based automation, gauging the consequent impact on model-layer responsibilities and control.

A. Architectural Framework for RAG-Driven Automation

An architectural framework for RAG (Retrieval-Augmented Generation)-driven automation is proposed by detailing the components of RAG-powered enterprise automation, specifying their interfaces, describing the data pipelines and flows among the components, and planning the evaluation.

The data layer enumerates the data sources of the enterprise automation architecture; describes the ontological representation of the information in the sources; specifies the processes that ensure the lineage and quality of the data; outlines governance practices; and identifies the retrieval indexes that support the grounding and alignment of the agentic AI components. The model layer focuses on the agentic AI components of the architecture: the constituent agents are identified, and their roles, prompts, policies, decision logs, interfaces, and orchestration are described. The overview highlights the RAG architecture; illustrates the interactions among the components; delineates the control flows; and defines the data stores at the core of the system. The approach and methods are general enough to apply to any enterprise-wide automation scenario enabled by RAG technology.

IV. OBJECTIVE OF THE STUDY

The objective is to provide an architectural framework that supports the development and deployment of enterprise automation systems leveraging retrieval-augmented generation. The telecommunications industry serves as the primary context, owing to the ubiquity, strategic importance, and operational scale of these organizations. The emerging area of research addresses many of the pertinent enterprise topics, such as strategic planning, technology roadmaps, stakeholder engagement, prototype development, pilot execution, and system scaling. The current study generates a comprehensive foundation that consolidates these activities; describes the specific cybersecurity requirements for enterprise-grade deployments; details the physical security and risk management safeguards that must be in place for operations; identifies the key evaluation metrics needed to demonstrate system readiness; and presents case studies for three real-world applications—financial operations, fraud detection, and supply chain optimization.

Enterprise automation systems that leverage retrieval-augmented generation face security and risk management challenges that must be addressed for enterprise deployment. Security protects the information technology resources from malicious actors, while risk management seeks to ensure the availability of the systems for authorized use. Every possible risk to confidentiality, integrity, and availability, as well as misuse scenarios, must be identified for their potential severity and likelihood of occurrence. Risk mitigations are then implemented as controls based on the accepted level of risk.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14

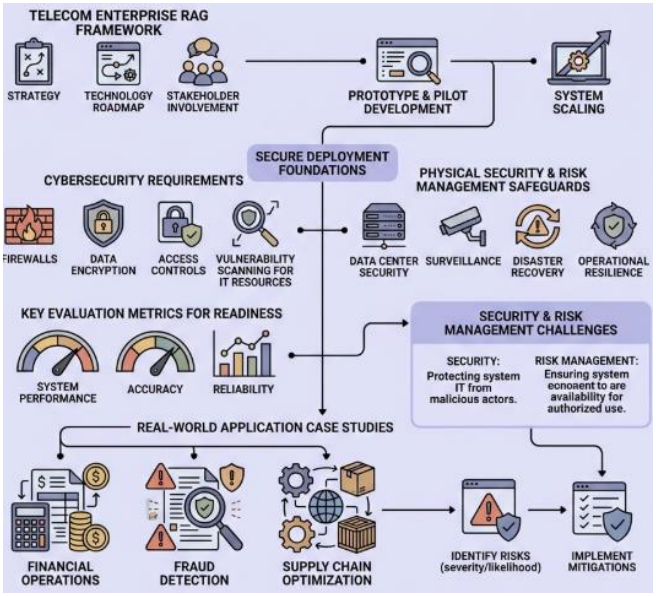


Fig 2: Architectural Framework And Security Protocols For Enterprise Rag Automation In Telecommunications

A. Defining the Study's Aims and Scope

Integration of agentic AI with retrieval-augmented generation for secure enterprise automation. Methodological framework comprising an agentic AI archetype supporting RAG-centric automation. An agentic AI archetype for enterprise automation. Main Aim: To define the concept of agentic artificial intelligence within an enterprise context, contributing to a detailed architectural framework for scheduling and supporting enterprise automation based on retrieval-augmented generation principals. Researched Questions: How can agentic agency and autonomy be defined for artificial intelligence systems deployed within enterprise environments? In what ways do the agentic AI frameworks facilitate planning, control, and risk management of automation in enterprise environments? Hypothesis: The design principles of agentic AI, integrated underpinning architectural description, and RAG operational data requirements define the basis for agentic AI enterprise automation models. Organisation: Relevance and definition of agentic agency and autonomy are presented, providing a foundation for clarifying enterprise-focused decision-making and control. The examination proceeds to detail a methodology for RAG-centric automation, articulating the elements, interactions, and data pipelines of the supporting RAG-driven architectural framework.

The study amalgamates the architectural and theoretical foundations of agentic AI and retrieval-augmented generation to define the concept of agentic artificial intelligence within an enterprise context. It aims to contribute to a detailed architectural framework for scheduling and supporting enterprise automation based on retrieval-augmented generation principles. The main question concerns how agentic agency and autonomy can be defined for artificial-intelligence systems deployed within enterprise environments. A further query asks in what ways the agentic-AI framework facilitates planning, control, and risk management of automation in enterprise contexts. The initial hypothesis states that the design principles of agentic AI, integrated with an underpinning architectural description and retrieval-augmented-generation operational data requirements, define the basis for agentic-AI enterprise-automation models. Relevance and definition of agentic agency and autonomy are therefore established, providing a foundation for clarifying enterprise-focused decision-making and control.

Equation 2: Precision for fraud detection

Step 1: True positives are correct fraud alerts.

$$TP$$

Step 2: False positives are normal cases wrongly flagged.

$$FP$$

Step 3: Total predicted fraud alerts:

$$TP + FP$$

Final:

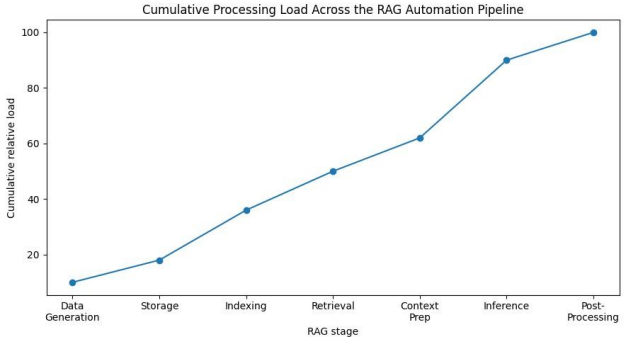
$$Precision = \frac{TP}{TP + FP}$$

V. RESEARCH SUMMARY

Together, these elements provide a comprehensive framework for RAG-centric automation and prepare the ground for expanding the application of agentic AI in enterprise settings. A theoretical model synthesizes the architectural and conceptual foundations; the required datasets for future implementations are identified; and a validation plan encompassing prototyping, piloting, and scaling trajectories is proposed.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14

The exploration demonstrates the viability of Deploying agentic AI capabilities within a RAG architecture for enterprise applications. Financial operations offer an initial testing ground, where fraud detection illustrates the potential for accurate, timely, and explainable outcomes that meet regulatory requirements. The model is then extended to supply chain optimization, with the aim of delivering holistic end-to-end visibility, reducing risk exposure, enhancing inventory accuracy and lowering lead times.



A. Comprehensive Framework for RAG-Centric Automation

The research synthesizes the architectural and theoretical foundations of agentic AI and Retrieval-Augmented Generation (RAG) for enterprise automation. The comprehensive framework delineates the required architecture and associated data for effective execution, thereby advancing scholarship on RAG-empowered solutions. Specifically, relevant theories are integrated, elementary concepts are clarified, and research questions are articulated and addressed. Together, the sections identify key gaps and contributions.

Agentic agency and autonomy are defined in enterprise contexts. Decisions, goals, and action cycles are analyzed, and the framework is compared with autonomous system standards. Governance implications are also discussed. The principles, architecture, data flows, and benefits of Retrieval-Augmented Generation are articulated alongside its knowledge retrieval, grounding, and alignment mechanisms.

VI. ARCHITECTURAL OVERVIEW OF RAG-BASED AUTOMATION

Agentic AI with RAG enables a variety of enterprise automation scenarios. An architectural overview distinguishes the data layer, establishing data management, sources, format, and defences; and the model layer, housing the RAG components—including agentic AI elements—along with their interfaces, inputs, outputs, and orchestration. The data layer covers data sources, supporting ontology, data lineage, quality controls, governance, and knowledge retrieval indexes.

Building on Retrieval-Augmented Generation and agentic AI foundations, the framework supports enterprise applications of RAG-driven automation. These are scenarios involving repetitive decision-making based on repeated sensory observations—such as monitoring, categorisation, classification, selection, and optimisation—where business risk and supporting controls are sufficiently bounded. Use cases in these categories, explored in detail elsewhere, encompass financial operations, fraud detection, supply chain risk mitigation and vulnerability discovery. Other scenarios may also involve agency, but involve more sophisticated RAG automations that leverage risk-aware learned models, such as customer engagement, sales forecasting, and large-scale image analysis.

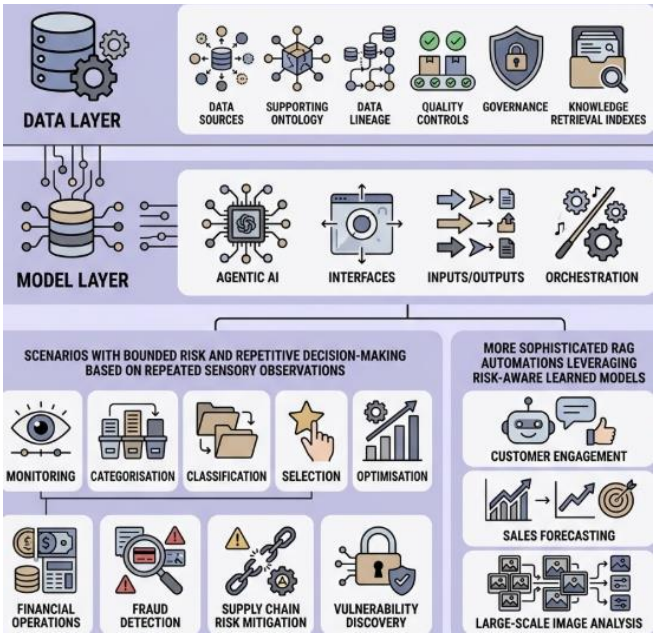


Fig 3: A Modular Architecture for Enterprise Agentic RAG: Formalizing Bounded and Risk-Aware Cognitive Automation

A. Data Layer and Knowledge Management

The data layer for RAG-centric automation encompasses the data sources supplying input data and knowledge to the model layer; an ontological model ensuring common semantics, vocabulary, and structure; data lineage capturing

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14

the origin, transformation, and integration of data; quality controls providing guidelines and monitoring for each data source; governance policies defining who is responsible for data input, access, quality, and retention; and indexes supporting fast, scalable retrieval. Data source management must address aspect analyses, exploitable features, and volume–velocity–variety trade-offs for both structured and unstructured sources. Structured data sources should provide clearly articulated and validated data pipelines that minimize data preparation costs. Computationally-expensive data sources, which cannot be cached or worked with in batch, should provide data in small, frequent batches to enable quick reactions. For unstructured data sources, aspect graphs, heuristic measures, data-lineage management, and shadow signal analysis can help design and manage measures of signal quality and data reliability.

Knowledge management ensures that the right information reaches the right actors, be they human or artificial. Effective decision making requires knowledge to be accessible, expressive, trustworthy, and timely. Knowledge-retrieval indexes help identify knowledge relevant to the current situation context, proposed action, or implied risk within the automation system. Natural-language processing-based information retrieval can simplify knowledge-base administration by allowing untrained actors to augment the repository without requiring its underlying structure to conform with conventional engines.

Four knowledge-management facilities are essential: defining an ontology that enables the use of a common vocabulary and semantics for the underlying information; managing knowledge sources and their proxies; managing the provenance of the available knowledge; and implementing quality controls that validate, track, and present quality indexes for each knowledge source being exploited.

Table 1: Components of RAG-Based Enterprise Automation Architecture

Area	Main Components	Purpose
Data Layer	sources, ontology, lineage, quality, governance, indexes	trustworthy retrieval
Model Layer	retrieval agent, risk agent, domain agent, orchestrator	agentic decision support
Security	RBAC, identity, audit logs, privacy, segmentation	safe deployment
Evaluation	accuracy, latency, throughput, hallucination, alignment	validation
Use Cases	fraud detection, finance, supply chain	enterprise automation

B. Model Layer: Agentic AI Components

The Model Layer implements the core agentic AI components within the RAG architecture. The goals, decisions, and actions of each agent are delineated, along with the accompanying prompts and policies. Interaction protocols and required interfaces are specified. The orchestration process—how agents communicate with and call each other—is presented in a decision tree format.

The essential agents for enterprise RAG applications include a Business Knowledge Retrieval Agent, Risk and Control Analyst Agent, Domain-Specific Agent (or set of agents), and an Orchestration Agent. Each agent's responsibilities are recorded in an agent log containing the description, goal, decision-making process, domain-specific knowledge, key interactions with other agents, roles, and prompts. These components form the foundation for additional enterprise-specific agentic AI components—such as a Financial Operations Agent, an Incident Classification Agent, and so on—that will be developed in the course of these case studies.

VII. Security and Risk Management in Agentic AI Systems

To safeguard against malicious exploitation and reduce unintentional misbehavior, agentic AI systems must include robust security and risk-management capabilities. Such countermeasures address access control, identity management, data privacy, compliance, auditability, and ongoing risk assessment. Security controls and policies for data, models, and underlying infrastructure systems must be clearly defined; the principle of least privilege and zero-trust access must be applied throughout.

Access control and identity management encompasses the characteristics and policies of the roles supporting interaction with the agentic AI components. Role-based access control

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14

(RBAC) dictates permissions configured against the personas—also known as roles—representing system users at any given time. Defined personas are supported by defined authentication methods, such as simple identification and password matching for internal users, with support for a variety of technologies for other classes of interaction with the environment or the model, including OIDC, JWTs, onetime or time-based passwords, PKI devices, and so on. Consideration is given to the use of federated or centralized identity systems. Data privacy, compliance, and auditability address aspects related to the use of sensitive, personally identifiable, financial, or otherwise regulated data in the operation of the agents as well as provisions necessary for the auditing of actions performed by the agents. A description of sensitive data types and handling practices that achieve alignment with GDPR, CCPA, and other relevant regulation should be included.

Equation 3: Recall for fraud detection

Step 1: True fraud detected:

$$TP$$

Step 2: Fraud missed by the system:

$$FN$$

Step 3: Total real fraud cases:

$$TP + FN$$

Final:

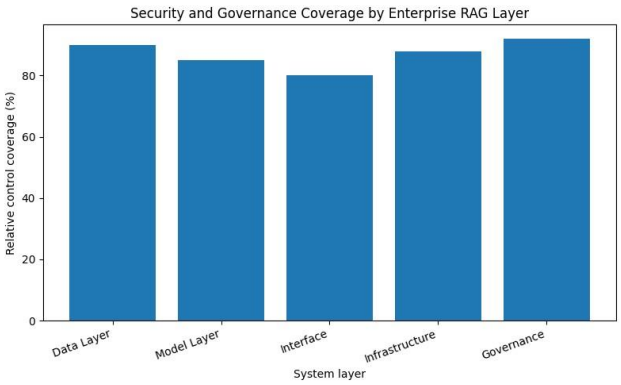
$$Recall = \frac{TP}{TP + FN}$$

A. Access Control and Identity Management

Access control and identity management together grant or restrict access to resources based on the identity of subjects (humans, systems, processes, etc.) and the appropriate permissions. Adopt a role-based access control approach, granting only those permissions required for performing specific roles or functions. Authenticated actions that exceed those permissions must be logged explicitly and investigated afterward. Strong authentication and authorization are essential to minimize the risk of unauthorized actions.

Identity and access management must be bidirectional—the agentic AI system must introduce itself and its context to external systems participating in an action, and that introduction must suffice for the receiving system to identify the agentic AI system as trusted or untrusted. Authentication can follow a conventional approach and utilize API token-based authentication for trusted internal services. The reverse direction—introducing the calling service to the agentic AI system and its context—necessitates a separate mechanism. The authorizing models perform this introduction based on the policy and prompt upon each invocation, validating that the configuration is currently authorized for the calling system. Model repository health checks defending against subliminal channels provide additional reassurance of the model's integrity.

The model orchestrator may itself be considered an agentic AI system. Requesting entities must be logged, including any superseded permissions. As another level of monitoring, deviation from established invocation patterns warrants internal alerting.



B. Data Privacy, Compliance, and Auditability

Incorporating agentic AI into enterprise applications presents significant privacy concerns due to the volume and potential sensitivity of the input/output data. The undertaking must adhere to the principles of data minimization, storage restriction, and purpose limitation, as specified by the GDPR and similar regulations. Consequently, all personally identifiable information must be promptly stripped from input data. Model outputs need to be classified and archived according to the least-privilege principle; only those individuals actively participating in the decision-making process should retain access to the data beyond the processing instance. User-generated data must be kept only for as long as required for exploratory or analysis use cases. Retention periods must align with consent terms defined in user agreements. Data entries that have exceeded their retention period should, at a minimum, be marked for deletion.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14

The classification and retention decisions must be recorded in a publicly available log that can easily be cross-referenced by automation agents using a centralised knowledge repository. Such a log can comprise a simple table, allowing the data handling decision to be efficiently checked. Input/output data classification, handling, and retention must also follow auditability principles to enable the tracking of data-related events. Each interaction must therefore record a unique audit ID; log entries maintained on a secure platform must capture the user identifier (if applicable), the audit ID, the prompted command, the output data, the decision-responsible component, and the retention period. Beyond ensuring compliance, publicly accessible audit logs can further guarantee adherence to the least-privilege principle: users interested in inspecting the previous decisions and the rationale behind them can do so without directly accessing sensitive data.

VIII. Secure Deployment Practices for Enterprise RAG Architectures

Deploying Retrieval-Augmented Generation architectures in enterprise environments requires stringent security measures to protect sensitive information and safeguard against misuse. The term RAG encompasses a full-stack architecture that integrates multiple subsystems operating in concert, meaning that a secure deployment approach must cover all parts of the deployment—from hardware and operating systems to networks and model management.

Infrastructure hardening and network segmentation must be complemented by controls, monitoring mechanisms, and isolation strategies for both trusted and untrusted environments. Supply chain and model risk management should comprise vendor vetting, data provenance and versioning, and an ongoing assessment of model-associated risks.

To ensure safety, security, and compliance with data protection regulations, the resulting enterprise RAG architecture must adhere to the five principles of identity management. Special attention should be paid to the access control and identity management components, specifying the roles and permissions required by different users and services and specifying all actions subject to authentication and authorization, in accordance with the principle of least privilege. The principles of data privacy and compliance further prescribe that any data stored or processed will be done so only for as long as necessary and with full consideration of the GDPR and CCPA regulations. Final principles include logging all access to and manipulation of sensitive data and the establishment of a traceable data handling process.

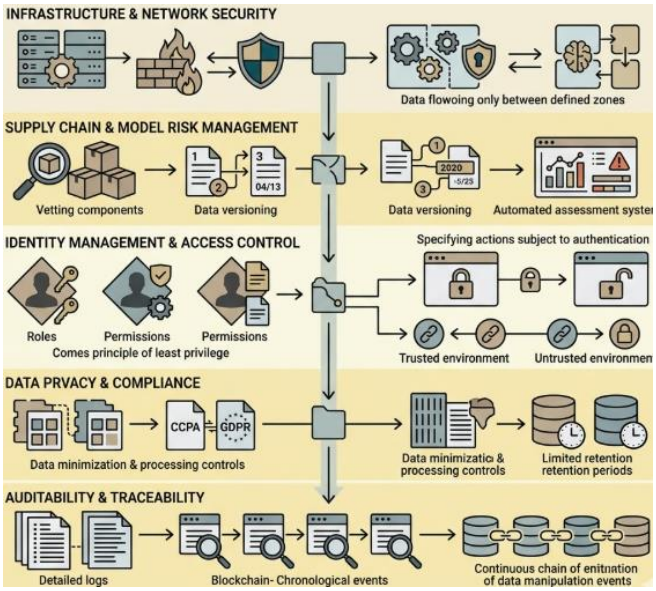


Fig 4: Securely Deploying Enterprise Rag Architectures: Multi-Layered Defenses

A. Infrastructure Hardening and Network Segmentation

Establishing design principles for the RAG ecosystem is an important part of a secure enterprise environment. Security controls should be deployed against the common vulnerabilities and exposures (CVEs) identified on external network penetration points of a RAG framework. The CSIRTs recommends that every component in the environment be monitored with a SIEM solution such as an open-source ELK stack. Network and host-based firewall controls, such as Suricata, should be deployed to build defense-in-depth capabilities. Incoming traffic should be segmented based on a ZTNA model to minimize potential compromises on exposed services. All network transport services need to be encrypted and hardened to reduce the exposure surface area. Interfaces and APIs for installation packages should be whitelisted with a vendor or package manager control.

Vulnerability scanners and patch management solutions should be configured to monitor the SaaS stack, database management system, and containerized runtime. An attack path should be built to identify existing CIA triad violations and security controls that can prevent attacks. Regular threat

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14

modeling based on Trike, STRIDE, Visual SLA, Attack Trees, or the DREAD ratings method should be conducted in a collaborative red-team exercise. External penetration testing visibly demonstrated an attack path and uncovered weaknesses in outdated packages used by the RAG framework. An attack surface reduction strategy is essential to protect against ransomware attacks that have caused several companies to go out of business. By applying network segmentation to prevent lateral movement during lateral movement, the complexity of an enterprise environment can be managed by providing a reasonably high amount of privileges to systems-on-privilege accounts.

B. Supply Chain and Model Risk Management

Proposed secure deployment practices for enterprise RAG architectures cover infrastructure hardening, network segmentation, and supply chain/model risk management.

Enterprise models, including their RAG constituent components, are often hosted by third parties or offer cloud services that others consume. Risk management is required for instantiating agents of all types in these contexts, including those supporting infrastructure-as-code and network-as-code. Vendor vetting should consider aspects such as operating model maturity with security operations practices in place, corporate reputation, management, external audits as applicable, product risk assessments, reliability assessments of public services, and sources of model data. Provenance and versioning should be established for any models that are not public domain or in wide acceptable use. Ongoing risk assessment, especially of proprietary models, is vital.

Ending Supply Chain Security Risks and Management summarizes actions to manage supply chain security and service delivery risks and to protect the supporting company's assets against supply chain risks, with aspects that can contribute to a deployment risk management framework for other cloud-based services. The actions support robust supply chain and delivery partner security for organization-critical delivery, enabling enhanced security and resiliency for organization-bound supply chain delivery.

Table 2: Performance and Evaluation Metrics for Agentic AI Systems

Metric	Formula Use
Precision	fraud alert correctness
Recall	fraud capture rate
F1 Score	balance between precision and recall
Latency	response speed
Throughput	request-handling capacity
Risk Score	likelihood × impact

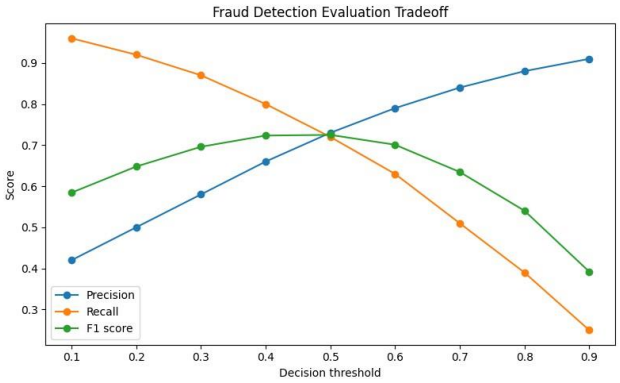
IX. EVALUATION METRICS AND VALIDATION METHODOLOGIES

Performance, robustness, and reliability are the core metrics for evaluating any enterprise model without agentic agency. For agentic AI with RAG, auxiliary safety, hallucination, and alignment metrics help assess the adequacy of prompt engineering and policy selection. Factual knowledge accuracy is also an important component of overall system performance connected to the knowledge retrieval layer. Detecting misaligned and incorrect responses reduces operational risk and enhances the controllability and explainability of the system.

Performance, robustness, and reliability metrics include accuracy, latency, throughput, fault tolerance, and resilience under load and failure conditions. Auxiliary metrics assess the end-to-end alignment of generated responses with the prescribed control goals, the factual accuracy of generated knowledge, and any inherent risks of initiating or responding to dialogues. Testing for these properties together provides sufficient coverage of the operation and response characteristics of agentic AI and RAG in enterprise applications.

Accuracy quantifies the ability of the deployed model to meet the requirements of the specific production use case over a given test set. It can be computed directly using ground truth indicators for all outputs generated during normal operation. Latency measures the time taken to produce a response once it is requested. Throughput expresses the number of concurrent interactions or requests the model chains can service while remaining within predefined limits for both latency and fault tolerance.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14



A. Performance, Robustness, and Reliability Metrics

A variety of metrics enable comprehensive performance evaluation. Accuracy is assessed by precision, recall, F1 score, and area under the receiver operating characteristic curve for unsupervised models, and by prediction error for Bayesian regression models. Latency measures per-query model response time, throughput is roll rate per time unit, and fault-tolerance and resilience tests evaluate system reactions to hardware faults.

Performance metrics do not encompass security, safety, or trustworthiness criteria directly. However, incident rates (inadequate alignment, unacceptable level of misinformation, and negative feedback from stakeholders) affect operational longevity and overall risk; they therefore warrant separate consideration. Performance risk facets include factual hallucination frequency, degree of alignment with specified guiding principles, and efficacy of current-risk-monitoring infrastructure.

Equation 4: F1 Score

Step 1: Combine precision and recall.

$$Precision + Recall$$

Step 2: Use harmonic mean.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Final:

$$F1 = \frac{2PR}{P + R}$$

B. Safety, Hallucination, and Alignment Benchmarks

Safety is crucial when deploying RAG systems in businesses or production settings. Not only must the model generate an appropriate response to an input prompt, but it must do so in a safe manner and with justifiable reasoning. The classical evaluation of safety, alignment, and hallucination for large language models is also relevant for agentic AI using RAG architecture.

There are many factors to consider when determining the safety of a model. For RAG systems, safety measures examine whether the model takes appropriate actions for queries that may lead to inappropriate actions, untrustworthy Hallucination, or unsafe behavior according to predefined constraints. Safety measures assess whether the model refrains from producing hateful, biased, fake, violent, abusive, or sexual content for inputs that may trigger such behavior. Hallucination measures verify whether the model's response is factually aligned with knowledge sources relevant to the query. Finally, alignment measures assess whether the model behaves in a risk-averse manner by speaking with a human before deciding to undertake an action with which the user may not be comfortable.

X. Case Studies: Enterprise Applications of Agentic AI with RAG

Enterprise Agentic AI Systems were proposed to harness the benefits and mitigate the dangers of RAG-driven automation. An extensive list of enterprise applications was compiled and subsequently re-evaluated to reflect a specific focus on agentic systems (i.e. systems that behave independently of continuous human supervision). The resulting list emphasises applications that share similar technologies, goals, and execution patterns, converge on the structure of agentic RAG systems, and have been deployed or piloted by multiple enterprises. The enterprise functions described thus far represent automation use cases rather than agentic applications: while the agentic dimensions are important for enterprise risk management, many of the applications are characterised by continuous human supervision of the model, dual-involvement system prompts, and in-person 'enabled agent' behaviour rather than decision-making and execution capacity afforded by policy switches. Two such applications were covered: A Banking Enterprise Detection Model and A Retail Enterprise Supply Chain Model.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14

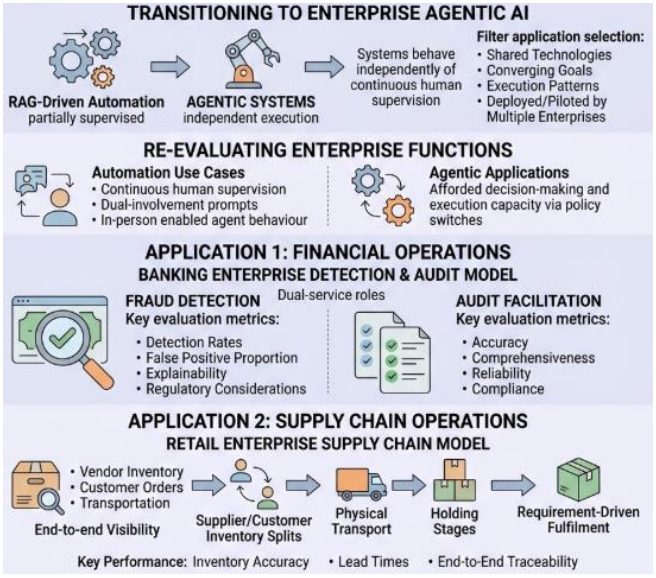


Fig 5: Deployed Enterprise Applications & Risk Management

Agentic AI applied to enterprise financial operations encompasses the dual-service roles of fraud detection and audit facilitation. Fraud detection models are evaluated based on detection rates, false positive proportions, explainability, and regulatory considerations, while audit facilitation systems are analysed with respect to accuracy, comprehensiveness, reliability, and compliance. The enterprise supply chain agent model addresses supplier and customer performance visibility by tracking vendor inventory, customer orders, and transportation; distinguishing supplier/customer inventory splits; incorporating physical transport, inspection, and holding stages; and supporting requirement-driven fulfilment. Key performance indicators include inventory accuracy, lead times, and end-to-end traceability. The applications illustrate the dual-service financial operation roles of fraud detection and audit facilitation.

A. Financial Operations and Fraud Detection

Many enterprises collect vast customer transaction datasets. Fraud detection is often considered a high-level enterprise objective, but legal frameworks mandate various lower-level regulatory operations. For financial institutions, it is crucial to monitor customer transaction data in near real-time for suspicious activities. Providing trustworthy alerts is equally important. High false positive rates burden investigative teams, discredit fraud detection engines, and may even induce customer flight. Thus, auditors often require enterprise fraud detection processes to produce explanations.

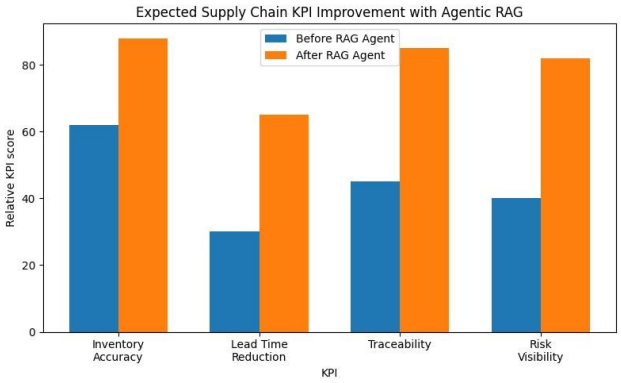
Agentic AI capability can potentially reduce false positives and improve explainability. Detection quality is assessed using actual credit card transaction datasets from a large financial institution. The analysis is conducted as follows: past criminal behavior is monitored with public data by an external vendor. When detected, the vendor provides a set of alerts, and at the end of the contract, the external company's crime identification is verified. The patterns significant for prediction are identified and used as features. Detection is conducted using supervised learning. The performance is verified by the business unit. The main questions examined in such a context are: what is the precision of the detection? Is explainability established? Are regulatory aspects considered?

B. Supply Chain Optimization and Visibility

Enterprise agents augmenting RAG capabilities may support supply chain management by accurately predicting product inventories at various locations and stages and thereby enhancing e2e visibility. Adapting a recent model for supply chain inventory control, an agent analyzes current inventories, forecasted demand, supply lead times, in-transit stocks, and inventory costs, then suggests restocking decisions that minimize total inventory cost. Evaluating whether suggested restocking actions are implemented accordingly provides the change control and operational check support also emphasized for other agent RAG combinations. Integration of the model's recommended inventory control with distribution centre-to-customer delivery could further increase e2e visibility and responsiveness.

A separate financial success metric quantifying bad debt write-offs as a proportion of total sales or as a proportion of receivables assessed updates the model. Detecting a drop in prediction performance with explainable AI—growing first-stage detection time, high-risk detections allocated low detection probability, or high-risk detections missed—triggers intensified business-user supervisory scrutiny, with a final validated assessment steering further action after a performance review. Integrating explainability into the actual AI probability underpinning regulatory disclosures could add even more robustness while aligning with the agency factors proposed for supervision.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14



XI. RESULTS

Strategic planning and stakeholder engagement provide direction and support for the implementation, followed by prototyping, piloting, and scaling.

Prototyping is a low-investment means of exploring the framework's potential. A pilot is a strategy-testing phase where the implementation doubles as a usability test; participants provide feedback but are mostly simply observing. After piloting one or more elements in production, the system can be scaled across the organization. While some components may scale easily, others might require customizations or even replacement for different contexts.

Automation of enterprise functions using Retrieval-Augmented Generation architecture and technology goes far beyond traditional conversational agents or chatbots, enabling wide-ranging application of agentic artificial intelligence, where RAG is an essential subset of the aggregate system architecture. Consequently, there is no shortage of potential enterprise use cases for RAG-enabled agentic AI – financial transactions, product development, customer interactions, supply chain operations, HR duties, industrial monitoring, and security incident response, to name a few. Within an RAG architecture, financial operations with fraud detection and supply chain management with visibility expose specific requirements that must be met for successful and effective use.

A. Strategic Planning and Stakeholder Engagement

Enterprise automation requires a well-defined strategy for planning, stakeholder engagement, and governance. Stakeholders must be identified, and their governance roles, responsibilities, interests, and expectations understood. A roadmap for the deployment, operation, and ownership of automation must be established, identifying required skill sets, training, and knowledge transfer. Continuity risk and change management must be addressed to ensure the solution continues to provide business value over time.

Key deliverables include a high-level enterprise automation strategy that aligns with top-down objectives and business drivers; position papers that build stakeholder buy-in; prototypes that demonstrate proof of concept and help in the identification of future support and risk management requirements; pilot implementations that address end-to-end solution management and support; and the delivery of skills and knowledge to promote a culture of intelligent business automation.

Fig 6: Strategic Effort Distribution

B. Prototyping, Piloting, and Scaling

Enterprise automation architectures based on Retrieval-Augmented Generation (RAG) can be complex and require the integration of diverse systems and data sources from organisational silos. Initial implementations therefore rely on external prototyping, piloting, and hardware-in-the-loop testing. External parties help develop minimum viable products and build foundational capabilities. Pilots then support stakeholders in validating short-term business cases, ensuring coverage of operational needs, and gaining experience, before scaling implementations for in-house use.

Pilot projects for RAG-driven agentic AI have focused on finance and fraud detection. In finance, the goal is to deliver a highly accurate chatbot capable of providing information on capital ratios, income statements, balance sheets, and cash flow statements without the risk of hallucination. Search engines expose relevant Glassbox documents containing the natural language provisions of the regulatory capital rules. The pilot eventually aims to confirm that a candidate detector

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14

achieves a balanced outcome in terms of the detection rate, false-positive rate, explainability, and considerations for the external auditor. During production use, attention to the financial regulatory obligations ensures that regulatory requirements are met, therefore supporting auditability.

A second pilot investigates agents that optimize supply chain operations. The central question is whether an end-to-end supply chain system connecting suppliers, manufacturers, transporters, distribution centres, retailers, and customers can be built as a minimum viable product. An ideal scenario is producing up-to-date inventory levels at all locations and the end customer, enabling automated stock replenishment by the fulfilment centre, automatic vendor orders for items below min stock levels, a clear view of the total lead time for items ordered by the retailer, and complete traceability of all items through the supply chain.

Equation 5: Enterprise risk score

Step 1: Risk depends on likelihood.

$$L$$

Step 2: Risk also depends on impact.

$$I$$

Final:

$$Risk = L \times I$$

XII. CONCLUSION

Agentic AI has the potential to automate enterprise decision-making and execution across a range of business functions. However, it is imperative that related systems be thoroughly evaluated, controlled, and governed—along the same lines as any autonomous system that can impact human lives. The avant-garde RAG architecture is particularly suited for the delivery of agentic capabilities. However, the significant security and risk management aspects that affect the deployment of RAG offer a useful lens for safe and controlled operation. Several security principles, incident detection and response operations, and system hardening measures constitute the protective fabric of any enterprise system. The measures discussed can therefore be used to guide secure deployment of enterprise RAG architectures.

Enterprise-wide deployment of RAG-based automation requires thorough evaluation, strategic planning, stakeholder buy-in, piloting, and scaling across business functions. RAG is particularly suited for experimentation in financial operations, such as portfolio monitoring, analysis, and model risk management; in fraud detection and investigation in areas with elevated levels of financial crime; and in the optimization of a business's overall supply chain. In all other business areas too, RAG is rapidly evolving into a standard automation platform that is cheap, fast, sophisticated, and readily understandable by the average user—attributes that shorten enterprise-approved build-test-use cycles to micro moments. Several factors shape the successful adoption of RAG for enterprise operations, but the most important is that the stakes have never been higher.

REFERENCES

1. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623).
2. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N. S., Chen, A., Creel, K., Davis, J. Q., Demszky, D., … Liang, P. (2021). On the opportunities and risks of foundation models. arXiv.
3. Mattaparthi, R. (2023). Deep Learning-Driven Combustion Anomaly Detection in Diesel Powertrains: A Multi-Sensor Fusion Approach for Real-Time ECM Adaptation. International Journal of Intelligent Systems and Applications in Engineering, 11, 1084.
4. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86–92.
5. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), Article 115.
6. Suresh, H., & Guttag, J. V. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 1–9).
7. Kolla, S. H., & Loganathan, R. (2023). Cloud-Native Deep Learning Architectures For Secure Generative AI Deployment In Enterprise Workflow Platforms. Journal of International Crisis and Risk Communication Research, 603-618.

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14

8. Hickman, E., & Petrin, M. (2021). Trustworthy AI and corporate governance: The EU's ethics guidelines for trustworthy artificial intelligence from a company law perspective. European Business Organization Law Review, 22, 593–625.
9. Harrison, T. M., & Luna-Reyes, L. F. (2022). Cultivating trustworthy artificial intelligence in digital government. Social Science Computer Review, 40(2), 494–511.
10. Pessach, D., & Shmueli, E. (2022). A review on fairness in machine learning. ACM Computing Surveys, 55(3), Article 51.
11. Mangala, N. (2022). Implementing Databricks Unity Catalog For Centralized Data Governance In Multi-Business-Unitenterprises. Journal of International Crisis and Risk Communication Research, 101-122.
12. Wang, X., Zhang, Y., & Zhu, R. (2022). A brief review on algorithmic fairness. Management System Engineering, 1, Article 7.
13. Stix, C. (2022). Artificial intelligence by any other name: A brief history of the conceptualization of "trustworthy artificial intelligence." Discover Artificial Intelligence, 2, Article 26.
14. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P. F., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35.
15. Bai, Y., Jones, A., Ndousse, K., Askell, A., Chen, A., DasSarma, N., Drain, D., Ganguli, D., Henighan, T., Joseph, N., Kadavath, S., Kernion, J., Conerly, T., El-Showk, S., Elhage, N., Hatfield-Dodds, Z., Hernandez, D., Hume, T., Johnston, S., … Kaplan, J. (2022). Constitutional AI: Harmlessness from AI feedback. arXiv.
16. Mangalampalli, B. M. (2023). AI-Driven Anomaly Detection in Healthcare Claims Data: A Business Intelligence Perspective. Journal of Rare Cardiovascular Diseases.
17. Weidinger, L., Mellor, J., Rauker, T., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, A., Balle, B., Kasirzadeh, A., Kenton, Z., Hawkins, W., Stepleton, T., Birhane, A., Williams, A., & Gabriel, I. (2022). Taxonomy of risks posed by language models. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 214–229).
18. Syed, S. (2023). Shaping The Future Of Large-Scale Vehicle Manufacturing: Planet 2050 Initiatives And The Role Of Predictive Analytics. Nanotechnology Perceptions, 19(3), 103-116.
19. Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K. R., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. In International Conference on Learning Representations.
20. Schick, T., Dwivedi-Yu, J., Dessi, R., Raileanu, R., Lomeli, M., Hambro, E., Zettlemoyer, L., Cancedda, N., & Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools. Advances in Neural Information Processing Systems, 36.
21. Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J.-R. (2023). A survey on large language model based autonomous agents. arXiv.
22. Yandamuri, U. S. (2023). An Intelligent Analytics Framework Combining Big Data and Machine Learning for Business Forecasting. International Journal Of Finance, 36(6), 682-706.
23. Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology.
24. Shinn, N., Cassano, F., Labash, B., Gopinath, A., Narasimhan, K., & Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. Advances in Neural Information Processing Systems, 36.
25. Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., Awadallah, A. H., White, R. W., Burger, D., & Wang, C. (2023). AutoGen: Enabling next-gen LLM applications via multi-agent conversation. arXiv.
26. Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy artificial intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. Information Fusion, 99, Article 101896.
27. Reddy, V. A. R. (2022). Designing Fault-Tolerant Data Ingestion Pipelines for High-Volume Healthcare Transactions. Frontiers in Health Informatics, 11, 861-889.
28. Laux, J., Wachter, S., & Mittelstadt, B. (2023). Trustworthy artificial intelligence and the European

ISSN: 3050-9696
E-ISSN: 3050-970X
EUROPEAN ADVANCED JOURNAL FOR SCIENCE & ENGINEERING
Volume: 02 Issue: 01
RECEIVED: JANUARY 06 REVISED: JANUARY 25 ACCEPTED: FEBRUARY 02 PUBLISHED: FEBRUARY 14

Union AI Act: On the conflation of trustworthiness and acceptability of risk. Regulation & Governance, 18(1), 3–32.

29. Simion, M., & Kelp, C. (2023). Trustworthy artificial intelligence. Asian Journal of Philosophy, 2, Article 8.
30. Sharma, S. (2023). Trustworthy artificial intelligence: Design of AI governance framework. Strategic Analysis, 47(5), 443–464.
31. Yandamuri, U. S. (2022). Big Data Pipelines for Cross-Domain Decision Support: A Cloud-Centric Approach. International Journal of Scientific Research and Modern Technology, 1(12), 227-237.
32. Kleizen, B., Van Dooren, W., Verhoest, K., & Tan, E. (2023). Do citizens trust trustworthy artificial intelligence? Experimental evidence on the limits of ethical AI measures in government. Government Information Quarterly, 40, Article 101834.
33. Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. Patterns, 4(9), Article 100804.