International Review of Business, Trade, and Economics
ISSN: 3069-230X

PRIME OPEN ACCESS

Volume 2, Issue 4

Research Article

Date of Submission: 02 Sep, 2026
Date of Acceptance: 30 Sep, 2026
Date of Publication: 08 Oct, 2026

Governing Enterprise Knowledge with On-Premise Retrieval-Augmented Generation: The AI42 Design Framework

Cristian Cosentino*, Natale Motta, Carlo Sarugeri, Tommaso Maggisano and Massimiliano Bacchetta

Moveo S.p.A. Largo Richini 6, 20122 Milan, Italy

***Corresponding Author:** Cristian Cosentino, Moveo S.p.A. Largo Richini 6, 20122 Milan, Italy.

Citation: Cosentino, G., Motta, N., Sarugeri, C., Maggisano, T., Bacchetta, M. (2026). Governing Enterprise Knowledge with On-Premise Retrieval-Augmented Generation: The AI42 Design Framework. *Int Rev Bus Trades and Econ, 2*(4), 01-16.

Abstract

Enterprise adoption of generative artificial intelligence is constrained less by access to capable language models than by the difficulty of connecting those models to fragmented, sensitive, and differently governed organizational knowledge. This paper presents Governed Enterprise Retrieval-Augmented Generation (GE-RAG), a design framework that treats retrieval-augmented generation as a data-governance control surface rather than only a question-answering technique. The framework is instantiated through AI42, an on-premise enterprise AI platform developed by Moveo. Following a design-science approach, the study derives governance requirements from the data-governance, information-retrieval, trustworthy-AI, and technology-adoption literature and traces them to observable architectural mechanisms in the industrial artifact. The analysis identifies six linked capabilities: governed knowledge onboarding, identity- and context-aware retrieval, bounded evidence construction, source-grounded generation, auditable interaction, and feedback-driven improvement. The on-premise boundary strengthens organizational control over storage, processing, model execution, and operational telemetry, but is explicitly treated as an enabling condition rather than proof of compliance or answer correctness. The paper contributes a governance-to-architecture mapping and a preregistration-ready evaluation protocol covering retrieval effectiveness, groundedness, access-control conformance, task-completion time, workload, usability, and user trust. The resulting methodology shows how enterprise RAG can make existing knowledge more accessible while preserving decision rights, provenance, and accountability. Empirical productivity claims are intentionally deferred until the proposed controlled user study is completed with real organizational data.

Keywords: Data Governance, Retrieval-Augmented Generation, Enterprise AI, On-Premise AI, Data Sovereignty, Knowledge Management, Auditability, Human-Centered Evaluation

Introduction

Organizations already possess much of the knowledge required to improve decisions and execute recurring work. It is distributed across policies, technical documentation, project repositories, contracts, reports, correspondence, and the experience of employees. The central difficulty is not simply storing these materials. It is ensuring that the right person can locate the right evidence, in a usable form, for a legitimate purpose, while the organization retains control over access, provenance, retention, and accountability.

Generative language models appear to offer a natural interface to this knowledge. A general-purpose model can interpret natural-language questions and produce fluent responses, yet fluency alone does not establish that an answer is based on current corporate sources, that the user was entitled to see those sources, or that the interaction can later be inspected. Research on hallucination and unfaithful generation shows why these distinctions matter [1,2]. Long context windows do not remove the problem: model performance can vary with the position of relevant evidence, making indiscriminate context expansion an unreliable substitute for retrieval and evidence selection [3].

The current relevance of AI is also visible in the breadth of applications moving from experimental models toward operational workflows. Recent studies combine multimodal and generative models for surgical-video documentation, use explainable generative methods to interpret complex user opinions, analyze political polarization at scale, and distribute intelligent monitoring across edge-cloud infrastructures [4,5,6,7,8]. This expansion increases organizational value, but it also amplifies the importance of data quality, provenance, explainability, privacy, infrastructure control, and human responsibility. The question for enterprises is therefore no longer whether AI can generate useful outputs, but whether those outputs can be produced within a governable and reviewable operating model.

Retrieval-augmented generation (RAG) combines a language model's parametric capabilities with an external, non-parametric knowledge source retrieved at query time [9]. In an enterprise setting, however, the useful unit is not merely retrieval plus generation. A deployable system must govern what enters the knowledge base, which identity and purpose constrain retrieval, how evidence is ranked and bounded, what provenance accompanies the answer, how use is recorded, and how corrections propagate. These concerns correspond closely to the data-governance literature, which frames governance in terms of decision rights, accountability, policies, processes, and controls over data assets [10,11].

This paper argues that enterprise RAG can become an operational mechanism for data governance when the governance rules are enforced inside the question-answering path. The argument is developed through AI42, an industrial platform created by Moveo to place generative AI around organizational knowledge and processes while retaining an on-premise execution option. The study focuses deliberately on the platform's RAG and conversational components. Other product functions are outside scope so that the relationship between data governance, retrieval, grounded generation, and everyday knowledge work can be examined precisely.

The study addresses three research questions:

- **RQ1**: Which data-governance responsibilities can be operationalized within an enterprise RAG and chat architecture?
- **RQ2**: How does an on-premise RAG artifact instantiate those responsibilities without disclosing product-sensitive implementation details?
- **RQ3**: Which empirical design can validly measure answer quality, governance conformance, user accep¬tance, and time reduction in everyday enterprise tasks?

The paper makes four contributions. First, it introduces the Governed Enterprise RAG (GE-RAG) framework, which links organizational decision rights to controls distributed across the RAG lifecycle. Second, it provides an artifact-grounded case analysis of AI42 at a level that is technically meaningful but does not reveal proprietary prompts, thresholds, model weights, or deployment recipes. Third, it distinguishes data locality from governance: on-premise deployment improves controllability, but organizational policies, identity management, evaluation, and human responsibility remain necessary. Fourth, it specifies a complete evaluation protocol that separates retrieval, generation, governance, and human outcomes rather than compressing them into a single accuracy score.

The remainder of the paper reviews the relevant literature, explains the research method, presents the GE-RAG framework and its instantiation in AI42, reports the analytical governance mapping, and defines the empirical study required to test productivity and user-centered claims.

Background and Related Work

Data Governance as Organizational Accountability

Data governance is broader than data administration. Khatri and Brown conceptualize it through decision domains and loci of accountability, distinguishing the determination of rights and responsibilities from the execution of day-to-day data management [10]. A structured review by Abraham, Schneider, and vom Brocke organizes the field around interdependent dimensions that include strategy, organization, roles, policies, processes, standards, measurement, technology, and people [11]. This literature makes clear that a technically centralized repository is not, by itself, a governed repository.

Contingency research further cautions against a universal operating model. Weber, Otto, and Oesterle show that governance arrangements depend on organizational characteristics and must align roles, decision areas, and responsibilities [12]. Reviews of governance activities similarly find that implementation spans policy definition, stewardship, quality management, compliance, communication, and monitoring [13]. In practice, the same document may therefore be available to different people under different purposes, may require an accountable owner, and may need to be superseded or removed without silently persisting in downstream indexes.

The current ISO/IEC 38505-1:2026 places governance of data within organizational governance and addresses the effective, efficient, and acceptable use of data created, collected, stored, secured, protected, or controlled by information systems [14]. This perspective is especially relevant to generative AI because a conversational layer increases the reach of existing information. It can reduce the effort required to use organizational knowledge, but it can also scale ambiguity, stale content, and inappropriate access if governance constraints do not travel with the data.

Retrieval-Augmented Generation and Enterprise Knowledge Access

RAG was introduced as a way to combine generative models with explicit non-parametric memory, improving access to external knowledge and enabling provenance beyond what is stored in model parameters [9]. Dense Passage Retrieval demonstrated the effectiveness of learned dense representations for passage retrieval [15]. At the same time, lexical approaches remain strong baselines, especially under domain shift, and the BEIR benchmark shows that no retrieval family dominates every heterogeneous task [16]. Rank fusion methods offer a simple way to combine complementary rankings [17]. These findings support enterprise architectures that can combine sparse and dense signals, then rerank a bounded candidate set instead of depending on a single similarity function.

Evaluation must also preserve provenance. KILT evaluates both downstream task performance and the ability to identify supporting evidence [18]. RAGAS separates dimensions such as context relevance, faithfulness, and answer relevance, while ARES uses calibrated model-based judges supplemented by human annotations [19,20]. The frameworks are useful for iteration, but automated judgments cannot replace domain review where corporate decisions, contractual interpretations, safety, or compliance are involved.

Recent domain-specific work has moved toward audit-oriented RAG. Cosentino, Squillace, and Marozzo present an auditable architecture for financial document intelligence that preserves evidence and intermediate artifacts around analyst-facing outputs [21]. The present study takes a different unit of analysis. Rather than optimizing a financial document workflow, it studies how a reusable enterprise platform can translate data-governance responsibilities into controls across knowledge onboarding, role-aware retrieval, conversational use, monitoring, and organizational feedback.

Adjacent research further illustrates why enterprise AI governance must extend beyond answer accuracy. Multimodal surgical-documentation systems depend on reliable temporal evidence, clinical provenance, and human validation explainable opinion interpretation and large-scale political analysis raise questions of transparency, bias, and accountable use of socially sensitive data machine unlearning makes retention, deletion, and the continuing influence of training data explicit governance concerns; and edge-cloud monitoring shows that latency, data movement, and infrastructure placement are part of the system's operational design [4-6,8,22]. These works are domain-specific, whereas GE-RAG seeks a reusable control layer for enterprise knowledge interaction: source authorization, scoped retrieval, bounded evidence, provenance disclosure, auditable use, and corrective feedback.

Data Sovereignty, Privacy, and Trustworthy AI

On-premise deployment is frequently presented as synonymous with privacy. That equivalence is too strong. Local execution can keep model inference, indexes, documents, and interaction telemetry within infrastructure controlled by the organization. It can also reduce exposure to third-party service terms and external API dependencies. Yet locality does not determine whether collection is lawful, retention is proportionate, access is appropriate, logs are protected, or generated answers are correct.

The GDPR establishes principles and obligations for personal-data processing, including data minimization, purpose limitation, privacy by design and by default, and security of processing [23]. The NIST Privacy Framework similarly treats privacy risk as an enterprise risk-management concern across the data lifecycle [24]. For AI-specific governance, the NIST AI RMF and its Generative AI Profile emphasize governance, mapping, measurement, and management of risks throughout the system lifecycle [25,26]. The EU AI Act adds a risk-based legal framework whose applicability depends on the system's purpose and deployment context [27]. An enterprise platform should therefore provide mechanisms that help an organization implement its obligations; it cannot confer compliance independently of the use case and operating model.

Within the Italian legal order, this European framework is complemented by Law No. 132 of 23 September 2025, in force from 10 October 2025. The statute promotes correct, transparent, responsible, and human-centric AI; requires attention to the quality, reliability, security, appropriateness, and transparency of data and processes; protects human decision-making and oversight; and treats cybersecurity throughout the AI lifecycle as an essential precondition. It also designates AgID and ACN as national AI authorities, while preserving the competences of the data-protection and sectoral authorities [28]. The law expressly states that it does not create new obligations beyond Regulation (EU) 2024/1689, but it gives the European framework an operational national setting. For enterprise RAG, the implication is not that local deployment guarantees legal compliance; rather, traceable sources, scoped access, audit records, lifecycle security, and human escalation become concrete governance capabilities that organizations can map to their specific role, risk classification, and use context.

Locality also does not neutralize application-layer threats. Retrieved documents can contain malicious or misleading instructions, and indirect prompt-injection research demonstrates that external content may manipulate LLM-integrated applications when data and instructions are not adequately separated [29]. Governance must consequently include content onboarding, document trust, prompt isolation, output constraints, access checks, monitoring, and human escalation.

Human adoption and organizational value

The business value of enterprise AI cannot be inferred from model benchmarks alone. The Technology Acceptance Model relates adoption to perceived usefulness and perceived ease of use while UTAUT integrates performance expectancy, effort expectancy, social influence, and facilitating conditions [30,31]. These constructs are relevant because a RAG assistant changes how employees search, synthesize, and verify knowledge rather than merely adding a new user interface.

Usability and workload should be measured separately from answer quality. The System Usability Scale (SUS) provides a compact global usability measure and NASA-TLX captures perceived workload across multiple dimensions [32,33]. A system may produce accurate answers but increase verification effort, or reduce time while creating overreliance. Accordingly, credible enterprise evaluation requires behavioral measures, task outcomes, source-checking behavior, and qualitative evidence in addition to questionnaires.

Stream	Established contribution	Remaining enterprise gap addressed here
Data governance	Decision rights, accountability, roles, policy, quality, and lifecycle controls	Translation of governance responsibilities into the live retrieval and chat path
RAG and information retrieval	External memory, dense and sparse retrieval, fusion, reranking, provenance-aware benchmarks	Identity-aware and purpose-bounded retrieval over heterogeneous corporate knowledge
Trustworthy and privacy-aware AI	Risk management, privacy principles, documentation, and monitoring	Concrete control points inside an on-premise conversational knowledge artifact
Human-computer interaction	Usefulness, acceptance, usability, workload, and trust measures	Joint evaluation of answer quality, governance conformance, and time-on-task

Table 1: Summarizes the Gap Addressed by GE-RAG

Research Method

Design-Science Orientation

The research follows a design-science orientation because its central object is an information-system artifact built to address an organizational problem [34]. The process reflects the stages articulated by Peffers et al.: problem identification, definition of objectives, design and development, demonstration, evaluation planning, and communication [35]. The current paper reports the first five at different levels of maturity. AI42 provides the implemented industrial artifact and its demonstration context; the governance mapping provides an analytical evaluation; and the controlled organizational study is specified as a preregistration-ready protocol rather than reported with fabricated outcomes.

This boundary is deliberate. A product description can establish that a control exists in software, but it cannot show that users save time or that answers are more accurate in a particular organization. Conversely, a user questionnaire can indicate perceived usefulness but cannot prove access-control correctness or retrieval quality. The study therefore separates artifact evidence from future empirical evidence.

Evidence Base and Analysis Procedure

The artifact analysis used four evidence categories:

- The platform's service topology and deployment configuration;
- Source-level behavior in the RAG, chat, document, identity, feedback, and audit components;
- User-interface behavior for source disclosure and document-scoped interaction; and
- Moveo's public description of AI42's on-premise and data-sovereignty goals [36]. The review used a September 2026 development snapshot supplied by the product team. Customer content, personal prompts, secrets, and model weights were excluded.

The analysis proceeded in five steps. First, governance objectives were synthesized from the literature into six operational questions: who authorizes knowledge, who may retrieve it, which evidence supports an answer, where processing occurs, how use is reconstructed, and how problems trigger improvement. Second, the system was decomposed into control points along the data and interaction lifecycle. Third, each control point was traced to observable artifact behavior. Fourth, controls were classified as implemented, configuration-dependent, or evaluation-required. Fifth, the mapping was reviewed for negative cases, including stale documents, insufficient evidence, unauthorized access attempts, deletion, retrieval failure, and malicious content.

The paper intentionally omits product-sensitive details whose disclosure would add little scientific value: exact system prompts, model-selection rules, threshold values, proprietary workflow logic, customer-specific schemas, credentials,

and infrastructure sizing. Scientific inspectability is preserved through the governance model, control definitions, state boundaries, evaluation variables, and disclosure of limitations.

Unit of Analysis

The unit of analysis is an enterprise knowledge interaction: an authenticated user asks a question; the platform constructs a query under organizational scope; candidate evidence is retrieved and ranked; bounded context is presented to a locally controlled language-model service; an answer and source information are returned; and the interaction can produce feedback and audit evidence. This unit is sufficiently narrow for causal evaluation and sufficiently complete to expose governance dependencies.

The Governed Enterprise RAG Framework

GE-RAG models enterprise question answering as a governed set of responsibilities rather than a direct dialogue between a user and a language model. Figure 1 presents these responsibilities around the AI42 conversation: authorizing knowledge, resolving identity and purpose, retrieving bounded evidence, linking claims to sources, controlling the processing boundary, and learning through audit and feedback. Governance surrounds the interaction because provenance, privacy, quality, decision rights, and accountability affect every stage.

Principle 1: authorize knowledge before retrieval

The first governance decision occurs before a query is submitted. A document should enter a corporate knowledge space only through an accountable onboarding process. The process records an owner, origin, business purpose, sensitivity, applicable audience, lifecycle state, and—where available—validity dates. Extraction and chunking generate derived objects, but those objects inherit the governance metadata of the source.

This principle prevents a common failure mode in which a technically searchable index becomes a shadow repository. If ownership and classification are absent at onboarding, later access filters operate on incomplete semantics. GE-RAG therefore treats the vector index as a derived representation, not the authoritative source of governance truth.

Principle 2: Resolve Identity and Purpose Before Search

Authentication answers who the user is; authorization answers what the user may do. The retrieval request should be resolved against both user identity and organizational context, such as role, corpus, department, project, or explicitly selected documents. Filters must constrain candidate generation rather than merely hide sources in the interface after retrieval.

Purpose is more difficult to encode than role. A user may be entitled to a document but not to every use of it. GE-RAG therefore recommends a progressive model: role- and corpus-based enforcement as a minimum,

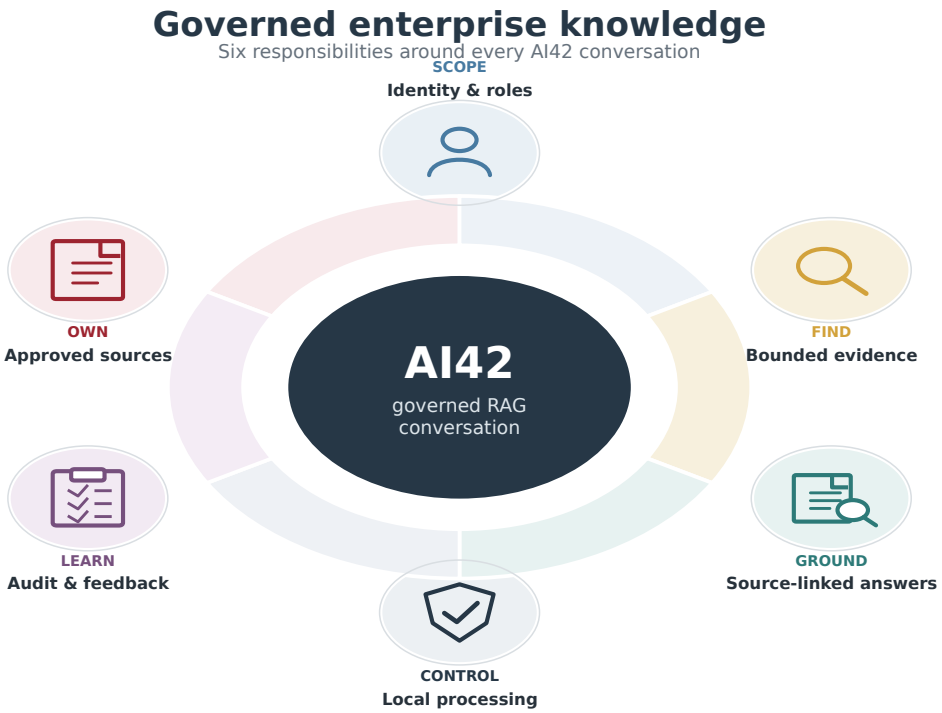


Figure 1: The Icon-Based GE-RAG Responsibility Model

Six Complementary Responsibilities Surround the AI42 Conversation and Connect it to Decision Rights, Provenance, Privacy, Quality, and Accountability

documented purpose and task categories for higher-risk workflows, and explicit approval or abstention when the request

crosses a defined boundary.

Principle 3: Build Bounded, Diverse, and Inspectable Evidence

Retrieval produces candidates, not facts. Enterprise corpora contain duplicate files, versions, contradictory policies, tables, scans, and documents written for different audiences. The evidence builder should combine complementary sparse and dense signals, rerank candidates, remove duplicates, preserve document diversity when the question spans sources, and fit the final evidence into a controlled context budget.

Bounding serves governance as well as efficiency. It limits unnecessary disclosure to the model, reduces the probability that irrelevant content dominates generation, and creates a finite evidence set that can be shown to a reviewer. When evidence is weak, unavailable, or outside authorized scope, abstention is preferable to an unsupported answer.

Principle 4: Generate Answers as Claims Linked to Sources

The generator should receive a clear separation between instructions, user intent, conversation state, and retrieved evidence. The output should identify the sources used and, for consequential tasks, support claim-level attribution. Provenance is not equivalent to correctness: a citation may be irrelevant, outdated, or misinterpreted. It nevertheless makes verification possible and creates an observable surface for evaluation.

Conversation memory should also be bounded. Prior turns can clarify references and sustain a task, but may introduce stale intent or cross-context leakage. GE-RAG treats memory as a governed input with explicit inclusion rules rather than an unlimited transcript appended to every prompt.

Principle 5: keep Deployment Boundaries Explicit

An on-premise system can place document storage, embeddings, indexes, model inference, and interaction logs within infrastructure controlled by the organization. This supports data-sovereignty objectives and enables internal decisions about capacity, model lifecycle, network exposure, and retention. The meaningful property is not the label on-premise but the documented data-flow boundary: which components process which data, which outbound connections exist, which administrators can access the system, and which backups or telemetry leave the boundary.

GE-RAG therefore requires a deployment manifest and data-flow inventory. Optional external services must be disabled by default for sensitive workflows or explicitly disclosed and governed. Local model execution does not eliminate the need for encryption, access management, patching, segmentation, backup protection, or incident response.

Principle 6: Make Use Observable and Correction Actionable

Governance requires evidence of operation. Each interaction should create structured events covering actor, action, resource, outcome, time, and relevant configuration or model version. Integrity mechanisms should make silent alteration detectable, while retention and access to audit data must themselves be governed. User feedback should be attached to a specific conversation turn and feed a triage process rather than directly modifying production behavior.

The resulting improvement loop connects users, stewards, product owners, and risk functions. Negative feedback may indicate a retrieval defect, a stale source, a generation problem, a confusing interface, or an inappropriate policy boundary. Treating every negative rating as a model problem would hide these distinct causes.

AI42 as an Industrial Instantiation

Architectural Overview

AI42 implements GE-RAG through modular services deployed on a private application network. Figure 2 gives a user-facing view of the resulting enterprise knowledge experience rather than a replicable system blueprint. It emphasizes the movement from governed documents to a source-bearing answer inside an organization-controlled boundary, without disclosing host topology, secrets, proprietary prompts, or customer-specific configuration.

Behind this experience, the interaction layer contains the authenticated conversational interface and backend orchestration. The knowledge layer contains document processing, dense and sparse representation services, and a vector database. The model layer contains locally served language models and reranking capabilities. Identity and role management, configuration, conversation persistence, feedback, and audit services provide the assurance foundation. These modular boundaries permit independent evolution while keeping the user-facing process coherent.

Governed knowledge Onboarding

The platform accepts heterogeneous business documents through authenticated endpoints. The ingestion path validates supported formats and size limits, stores user- or corpus-scoped metadata, extracts content, and creates derived representations for retrieval. The system preserves document identity and metadata across chunks, enabling filters for document, area, corpus, and language-related use cases.

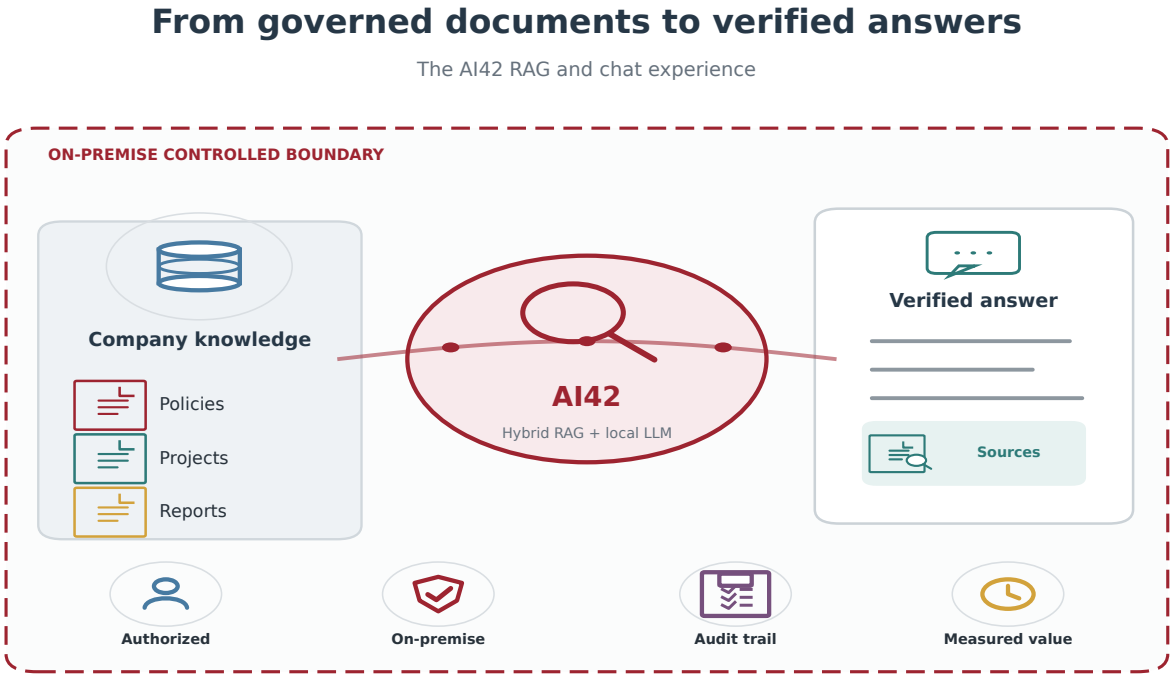


Figure 2: AI42 Enterprise knowledge Experience

Governed documents, scoped retrieval, local generation, source disclosure, identity, audit evidence, and measured human impact are represented as distinct user-visible or assurance concerns.

From a governance perspective, this is a useful foundation but not the end state. File validation and ownership establish technical scope; enterprise deployment should additionally integrate formal classification, account¬able stewardship, retention categories, approval state, and effective dates. The distinction is captured in the analytical mapping: technical metadata propagation is implemented, whereas enterprise policy completeness depends on deployment configuration and organizational process.

Hybrid and Hierarchical Retrieval

AI42 uses complementary dense and sparse query representations and combines their rankings. This choice aligns with evidence that lexical and semantic retrieval exhibit different strengths across domains [16]. Candidate passages are connected to parent document segments, and a reranking stage can refine the evidence set before generation. The context-selection logic also supports source diversity for questions that explicitly compare or combine documents.

This pipeline is important for governance because it produces a controlled evidence boundary. It can apply scope filters during retrieval, deduplicate candidates, restrict the number of sources, and reduce context when model limits are reached. These mechanisms improve inspectability, but their effectiveness must be tested on each organization's vocabulary, document structure, languages, and update patterns.

Conversational Generation and Source Disclosure

The chat service resolves output language independently from source-language filtering, allowing users to query documents across languages without automatically excluding source material written in another language. It assembles an evidence context, invokes a locally controlled model service, streams the answer, and returns source details that can include document labels, page or segment information, snippets, and retrieval scores when available.

Conversation histories are associated with authenticated owners and corpus context. Ownership checks prevent

a conversation identifier from silently loading another user's history. The service also supports recovery from context limits, resource exhaustion, and interrupted generation so that operational failures can be represented explicitly instead of appearing as valid answers.

Identity, Role, and Corpus Boundaries

Identity management is separated from application logic through an enterprise identity service. Application roles and corpus configuration determine which sections and knowledge spaces are available. Administrative overrides are explicit rather than implicit. In GE-RAG terms, identity is resolved before the search path and the selected organizational scope travels into retrieval.

Role-based access control is necessary but not sufficient. High-risk deployments should test negative authorization

cases, review default roles, minimize administrative overrides, and bind document classification to identity attributes or policy rules. Authorization must also be tested after corpus reconfiguration and document movement, not only at first ingestion.

On-Premise Processing Boundary

The supplied deployment topology supports local document storage, vector search, embeddings, reranking, language-model inference, identity, and application persistence. Components communicate over a private service network. This architecture allows an organization to operate the RAG and chat path without sending *ordinary* prompts and corporate evidence to an external generative-model API.

The qualification *ordinary* is important. A complete production statement requires inspection of every optional connector, telemetry path, update mechanism, backup destination, and administrator workflow. AI42's modularity makes such inspection possible, but the final boundary is a deployment property, not a universal product guarantee.

Feedback and Audit Evidence

Users can attach positive or negative feedback and comments to a specific conversation turn. The platform also records audit events and includes an append-only integrity chain. The chain can use keyed integrity verification when a secret is configured; otherwise it provides a weaker hash-linked tamper indicator. Production claims should therefore report the active configuration, key-management controls, verification cadence, unsigned legacy events, and retention policy.

These mechanisms create the basis for accountable operation: a reviewer can relate an outcome to a user, action, resource, and time, while feedback identifies turns requiring investigation. They do not automatically establish why a model produced a particular token. Their purpose is operational traceability, not a claim of mechanistic interpretability.

Analytical Results: Governance-to-Architecture Mapping

The principal analytical result is the control mapping in Table 2. It answers RQ1 and RQ2 by showing which governance responsibilities can be embedded in an enterprise RAG path and how AI42 currently represents them. Status labels deliberately distinguish existence from validated effectiveness.

Governance responsibility	AI42 control surface	Evidence produced	Current analytical status	Required assurance
Knowledge ownership and scope	Authenticated upload, document identity, corpus and area metadata	Source identity and derived-chunk metadata	Implemented foundation	Add steward, classification, validity, retention, and approval fields
Authorized discovery	Identity, role and corpus resolution before retrieval; metadata filters	User and scope associated with request	Implemented, configuration-dependent	Negative access tests and policy review per deployment
Retrieval quality	Dense and sparse retrieval, fusion, parent reconstruction, reranking, deduplication	Ranked evidence set with metadata	Implemented	Curated enterprise benchmark and ablation study
Context minimization	Document limits, bounded evidence windows, context reduction and recovery	Finite context supplied to generator	Implemented	Verify that sensitive or irrelevant evidence is excluded
Provenance	Source details, labels, snippets, pages or segments where available	Answer-level evidence disclosure	Implemented foundation	Measure citation completeness and claim-source correctness
Controlled processing	Locally deployable model, index, document, and persistence services	Inspectable service and data-flow boundary	Implemented architecture	Verify outbound paths, backups, telemetry, and administrator access
Interaction accountability	Conversation ownership, structured audit events, integrity chain	Actor-action-resource-outcome records	Implemented; integrity strength configuration-dependent	Activate keyed verification, protect keys, schedule verification, govern retention
Corrective learning	Turn-level feedback with	Feedback linked to a specific output	Implemented collection	Define triage, stewardship

Lifecycle enforcement	Document deletion and scoped cleanup paths	Removal event and updated stores	Partially implemented	Test deletion propagation across derived indexes, caches, backups, and audit exceptions

Table: 2

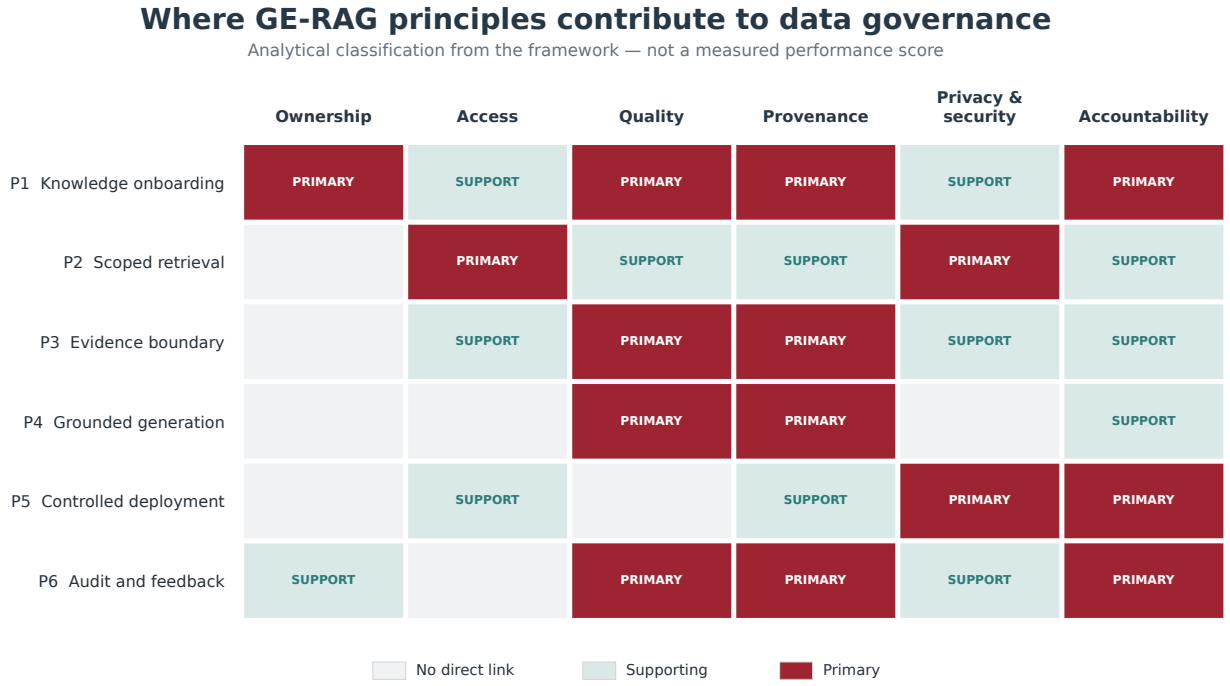


Figure 3: Analytical Relationship Matrix Between the Six GE-RAG Principles and Six Data-Governance Dimensions

Dark cells identify a primary relationship, light colored cells a supporting relationship, and gray cells no direct relationship.

Business value	Chat-based access to governed corporate knowledge	Task traces and user feedback	Evaluation required	Run controlled study; do not infer time savings from architecture alone

Figure 3 provides a second analytical view of the mapping. Each cell indicates whether a GE-RAG principle has a primary, supporting, or no direct relationship with six governance dimensions. The classification summarizes the control logic developed in this paper; it is not a measured performance score and should not be interpreted as evidence of compliance.

The mapping reveals three findings. First, governance can be made operational at the same points where the system transforms data: onboarding, retrieval, context construction, generation, delivery, and feedback. This reduces the gap between policy documents and runtime behavior. Second, several important claims are configuration-sensitive. Role models, corpus definitions, HMAC activation, connector settings, reten¬tion, and backup design can strengthen or weaken the same underlying artifact. Third, quality and value remain empirical questions. The architecture creates measurable evidence but does not predetermine the measurements.

The analysis also identifies a conceptual advantage of enterprise RAG over unbounded general-purpose chat. Because knowledge is retrieved from a governable external memory, source updates and access decisions can affect future answers without retraining the foundation model. This does not guarantee immediate consistency: indexes, caches, derived artifacts, and conversation histories still require lifecycle controls. The advantage is therefore governability, not automatic governance.

Evaluation Protocol

The evaluation is designed as a multi-layer study so that a failure can be attributed to retrieval, generation, access control, usability, or organizational process. Figure 4 presents the four evidence lenses that must be assessed together.

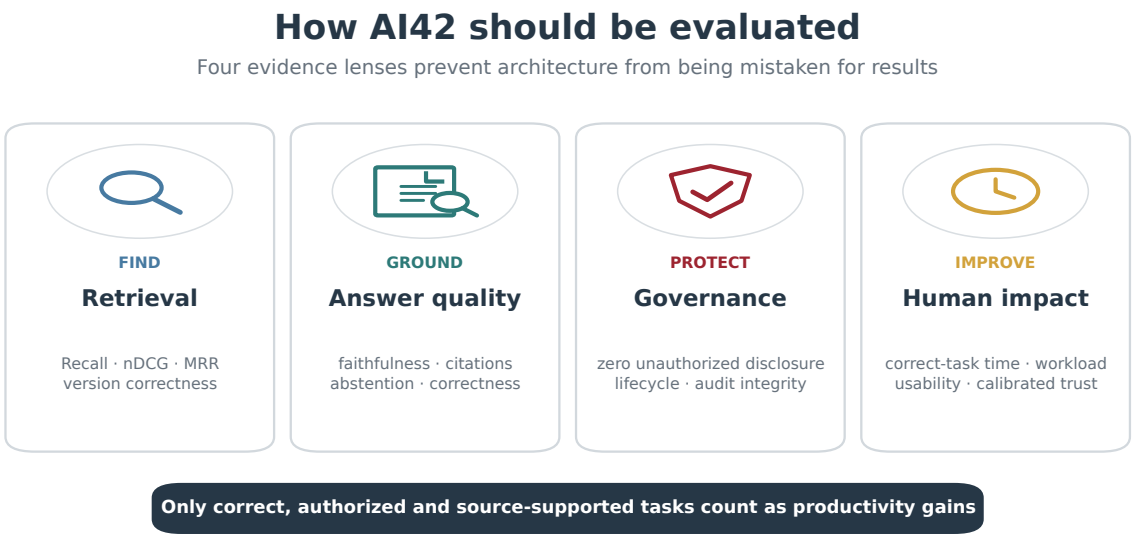


Figure 4: Four Evidence Lenses for Evaluating AI42: Retrieval Effectiveness, Grounded Answer Quality, Governance Conformance, and Human Impact

Productivity gains count only for correct, authorized, and source-supported tasks.

Study Objectives and Hypotheses

The study evaluates four dimensions:

- **Effectiveness:** users complete enterprise knowledge tasks correctly and with sufficient evidence.
- **Efficiency:** users require less active time and fewer search actions than under the existing workflow.
- **Governance Conformance:** the system respects scope boundaries, supports provenance, records required events, and propagates lifecycle actions.
- **Human Experience:** users perceive the system as useful and usable without inappropriate trust or reduced verification.

The primary hypotheses are:

- **H1**, AI42 reduces median task-completion time relative to the existing search-and-read workflow;
- **H2**, AI42 improves task success and evidence completeness;
- **H3**, the full hybrid and reranked configuration improves retrieval effectiveness over sparse-only and dense-only alternatives;
- **H4**, unauthorized evidence exposure remains zero across the predefined negative-access test suite; and
- **H5**, users report higher perceived usefulness without a reduction in calibrated verification behavior.

Benchmark Construction

Each participating organization should create a versioned benchmark from documents users are legitimately entitled to access. Domain experts define questions before evaluating outputs and record: the task intent, acceptable answer, supporting source passages, valid alternative formulations, applicable role, prohibited sources, and an abstention expectation when the corpus is insufficient. Questions should include direct lookup, cross-document synthesis, temporal/version conflict, terminology variation, multilingual query, table-derived evidence, and unanswerable cases.

To avoid a benchmark that merely reflects one retrieval technique, experts should locate evidence manually and review candidate questions without seeing system rankings. Inter-annotator agreement should be reported for relevance and answerability. Disagreements should be adjudicated and retained as a difficulty signal rather than silently removed.

Offline RAG Evaluation

The retrieval experiment compares at least four configurations under the same corpus and queries: sparse-only retrieval, dense-only retrieval, hybrid fusion, and hybrid fusion plus reranking and hierarchical parent reconstruction. Retrieval is measured using Recall@k, precision@k, mean reciprocal rank, and nDCG@k. Source diversity and version correctness are added for multi-document and temporal questions.

Generation is evaluated with the same retrieved evidence budget across systems. Measures include answer correctness, answer relevance, completeness, faithfulness to the supplied context, citation coverage, citation precision, correct abstention, and unsupported-claim rate. RAGAS or ARES can accelerate iteration, but a blinded human review remains the primary test for a representative, risk-stratified [19,20]. Judges receive the question, authorized evidence, and

anonymized answer order; they do not receive the system label.

The study should report component ablations rather than only the best configuration. Removing corpus filters tests governance dependence; removing sparse retrieval tests exact-term sensitivity; removing reranking tests ordering; removing source-diversification tests cross-document questions; and replacing RAG with direct generation establishes whether the model is answering from organizational evidence or general prior knowledge.

Governance and Security Conformance Tests

The governance test suite is deterministic. It includes permitted and prohibited role-corpus combinations, direct document selection, conversation ownership, document deletion, corpus reassignment, stale-version replacement, audit-event completeness, integrity verification, service-unavailability handling, and controlled prompt-injection documents. The expected outcome for an unauthorized test is not a lower relevance score but zero disclosed content.

Deletion is tested end-to-end. After a source is removed or superseded, the evaluator queries the vector index, application metadata, preview caches, conversational recall, backups subject to retention policy, and user-visible source lists. Audit records may legitimately remain when retention is required, but their content and purpose must be documented. Prompt-injection tests use synthetic content and no production secrets.

Controlled user Study

The user study uses a randomized, counterbalanced, within-subject design. Participants complete matched tasks under two conditions: the organization's current workflow and AI42-assisted work. A Latin-square allocation controls order and learning effects. Participants are stratified by role and prior familiarity with the document domain. The final sample size is determined through an a priori power analysis using the smallest operationally meaningful paired time reduction; reporting an arbitrary fixed sample would be less defensible.

Each participant completes tasks drawn from six representative categories:

- locating a policy or procedure and citing the controlling passage;
- comparing two versions or documents;
- synthesizing a short evidence-based briefing;
- identifying that the available corpus cannot support an answer;
- answering a multilingual question from documents in another supported language; and
- resolving a realistic operational question under a role-specific scope.

Task start, answer submission, source opening, query reformulation, and abandonment are timestamped. The primary endpoint is active task-completion time for correct responses. Secondary endpoints are task success, evidence completeness, number of search actions, number of opened documents, verification time, and correction effort. Time saved is calculated only for valid paired tasks:

$$\Delta T_i = T_{i,\text{baseline}} - T_{i,\text{AI42}}, \qquad R_i = \frac{T_{i,\text{baseline}} - T_{i,\text{AI42}}}{T_{i,\text{baseline}}}.$$

The study reports medians and interquartile ranges alongside means and confidence intervals because task times are often skewed. Paired tests are selected after checking distributional assumptions; effect sizes and confidence intervals accompany p-values. Task correctness is analyzed with paired binary or ordinal methods as appropriate. Multiple secondary comparisons are corrected and clearly labeled exploratory.

User Evaluation and Qualitative Analysis

After each condition, participants complete NASA-TLX and short trust-calibration items. SUS and perceived-usefulness items are administered after the AI42 condition [30,32,33]. Trust questions avoid asking only whether the system is trusted; they test whether participants know when to verify, can identify source limitations, and reduce confidence when evidence is absent or conflicting.

Semi-structured interviews examine where the system saved effort, where it shifted effort into verification, how source disclosure affected decisions, and which organizational boundaries were unclear. Responses are coded using a documented thematic-analysis procedure with independent coding of a subset and resolution of disagreements [37]. Quotes are anonymized and published only with appropriate consent.

Reporting Template

Question	Primary measures	Baseline or comparator	Required reporting
Does retrieval find authorized evidence?	Recall@5/10, nDCG@10, MRR, prohibited-source exposure	Sparse, dense, hybrid, full pipeline	Per task type, role, language, and corpus
Are answers grounded and useful?	Correctness, faithfulness, citation precision/coverage, abstention	Direct LLM and RAG ablations	Human review plus automated metrics with judge model disclosed
Does AI42 reduce effort?	Correct-task time, search actions, documents opened, correction time	Current organizational workflow	Paired differences, effect size, confidence interval, exclusions
Is the system usable?	SUS, NASA-TLX, perceived usefulness	Within-subject condition comparison where applicable	Distribution, missingness, role differences
Does governance operate as designed?	Access violations, deletion propagation, audit completeness, integrity verification	Deterministic expected outcomes	Pass/fail per test and configuration manifest
Is trust calibrated?	Confidence-error alignment, source verification behavior, qualitative themes	Evidence-present vs. evidence-absent tasks	Overreliance and appropriate abstention rates

Table 3: Defines the Minimum Result Set

It should be completed with measured values before any empirical productivity claim is added to the abstract or conclusion.

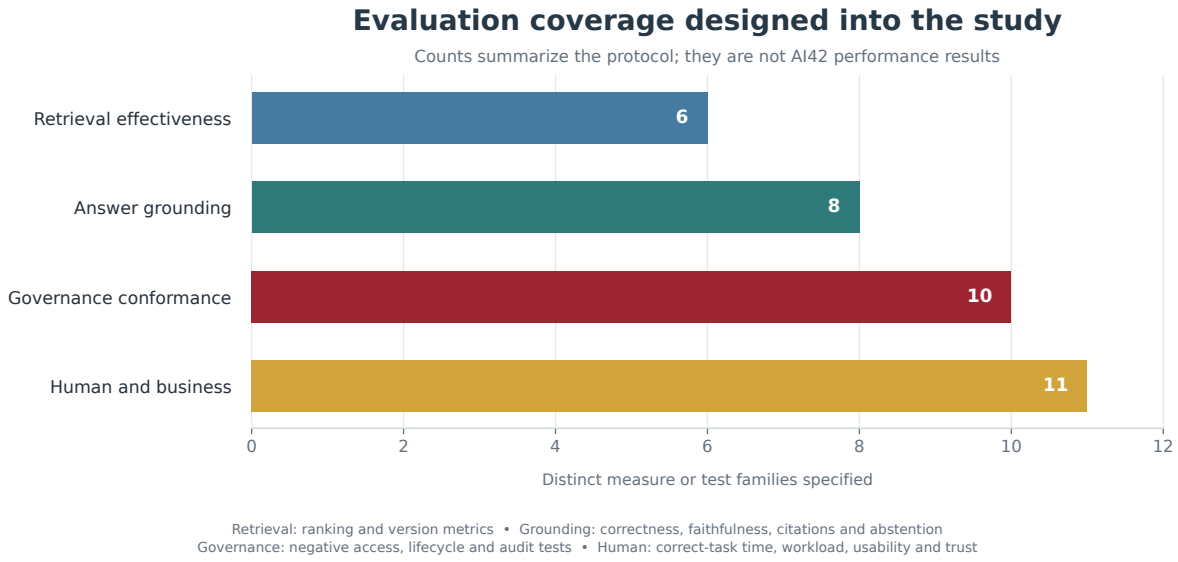


Figure 5: Coverage of the Proposed AI42 Evaluation Protocol

Bars count distinct measure or test families specified for retrieval, answer grounding, governance conformance, and human or business outcomes; they do not report system performance.

The protocol specifies more than a single accuracy measure. Figure 5 counts the distinct metric or test families defined for each evaluation layer. These values describe the breadth of the proposed study design, not observed AI42 performance. The emphasis on governance and human outcomes is deliberate: technical retrieval gains are insufficient if an answer crosses an access boundary, lacks usable evidence, or merely transfers effort from search to correction.

Managerial Implications

The framework changes the sequence of enterprise AI adoption. Instead of selecting a model first and connecting data later, an organization begins by identifying high-value knowledge tasks and the decision rights around their evidence.

It then creates a governed knowledge space, defines measurable failure conditions, and chooses models that operate within those boundaries. Model replacement becomes an infrastructure decision inside a stable governance envelope rather than a redesign of the entire operating model.

AI42 can support this transition because its RAG and chat components expose the points at which governance decisions become executable. Role and corpus scope can reduce indiscriminate access; source disclosure can make answers reviewable; local processing can strengthen control over data flows; and feedback and audit evidence can connect everyday use to improvement. The business case should nevertheless be built task by task. A measured reduction in one document-intensive workflow does not imply the same benefit for every function.

Managers should treat adoption as a joint program involving data owners, domain experts, security, legal or privacy functions, IT operations, and end users. Data stewards determine which sources are authoritative and current. Security and privacy teams validate the processing boundary and negative-access cases. Domain experts design the benchmark. Users reveal whether the assistant removes work or merely moves it. Product and operations teams monitor failures and close feedback loops.

The most defensible value narrative is therefore not that AI42 makes an organization omniscient. It is that the platform can shorten the path from authorized evidence to a reviewable answer while leaving the organization in control of what knowledge is available, to whom, under which operating conditions, and with what record of use.

Discussion

From Data Access to Governance by Interaction

Traditional governance often becomes visible to employees through repositories, approval processes, and policy documents. A conversational interface creates a new point of contact: users express intent in natural language, and the system must convert that intent into an authorized, evidence-bounded operation. This makes the chat interaction a governance event. The response is not only content; it is the outcome of identity resolution, source selection, model execution, and disclosure decisions.

This perspective helps explain why a RAG assistant can improve governance rather than merely consume governed data. Query logs reveal recurring knowledge gaps, failed retrieval reveals inconsistent terminology or poor metadata, conflicting evidence reveals version-management problems, and feedback reveals where users cannot translate policy into action. Used responsibly and with proportionate privacy controls, these signals can guide stewardship and knowledge-base maintenance.

The Role and Limits of On-Premise Deployment

On-premise deployment is valuable when an organization needs strong control over data location, network paths, model versions, availability, or integration with internal identity and monitoring. It may also reduce dependency on external APIs and make costs more predictable at stable utilization. These are architectural options, not universal outcomes: local infrastructure introduces capacity planning, patching, model operations, energy, hardware lifecycle, and specialist-skill requirements.

The paper therefore avoids a cloud-versus-on-premise dichotomy. A governed deployment could be fully local, private cloud, sovereign cloud, or hybrid if the organization can document the data flows and enforce equivalent controls. AI42's present case is on-premise because this boundary supports the product's design goals and the types of corporate knowledge considered. The scientific contribution is the explicit relationship between boundary control and governance responsibilities.

Generalizability

GE-RAG is intended to generalize at the control level, not at the exact implementation level. Every organization requires a way to authorize sources, scope retrieval, bound evidence, disclose provenance, monitor use, and correct errors. The technologies and policies used to implement those functions may differ. A small firm may use simple role groups and curated folders; a regulated enterprise may require attribute-based policies, legal holds, formal records management, and independent model validation.

The contingency view of data governance suggests that this variation is expected [12]. Evaluation should consequently publish the organizational context, task types, corpus characteristics, user roles, and active configuration. Without these details, even a statistically strong result is difficult to transfer.

Limitations and Threats to Validity

This study is authored by employees of the organization that develops AI42. That proximity provides access to the artifact but creates an evident risk of favorable interpretation. The paper mitigates the risk by separating observable mechanisms from empirical claims, declaring the commercial relationship, defining negative cases, and proposing blinded human evaluation. Independent replication or third-party assessment remains necessary.

The artifact analysis uses a development snapshot rather than a certified production deployment. Configuration can

materially change governance properties: a role model can be overly broad, optional outbound services can alter the processing boundary, an integrity key can be absent, or retention can be inconsistent. Results must therefore be accompanied by a configuration manifest and cannot be generalized from source capability to every installation.

The proposed user study may be affected by novelty, learning, and task selection. Counterbalancing reduces order effects, while matched tasks and preregistered exclusions reduce analytic flexibility. Real work remains more variable than a controlled study, so a follow-up longitudinal deployment should measure sustained use, reopened sources, corrections, escalations, and changes in knowledge-maintenance workload.

Automated RAG metrics introduce judge-model bias and may reward stylistic similarity rather than operational correctness. Human review is therefore primary for a risk-stratified sample. Human experts can also disagree or share organizational assumptions; adjudication rules and agreement measures should be reported.

Finally, the framework focuses on textual enterprise knowledge and conversational use. Structured decision automation, autonomous tool execution, multimodal evidence, and high-risk sector-specific systems require additional controls and are outside the scope of this paper.

Conclusion

Enterprise generative AI creates value only when it can use organizational knowledge without dissolving the boundaries that make that knowledge trustworthy and governable. This paper introduced GE-RAG, a framework that connects data-governance responsibilities to the technical stages of retrieval-augmented generation: knowledge onboarding, identity-aware search, bounded evidence construction, source-grounded generation, controlled deployment, auditable use, and corrective feedback.

The AI42 case shows how an on-premise enterprise platform can instantiate these controls through modular document, retrieval, model, identity, feedback, and audit services. The analysis does not treat architecture as proof of compliance, correctness, or productivity. Instead, it identifies what the artifact makes controllable and measurable, where assurance remains configuration-dependent, and which claims require empirical evidence.

The next step is the controlled evaluation defined in this paper. By jointly measuring retrieval quality, groundedness, access-control conformance, task-completion time, workload, usability, and trust calibration, organizations can decide whether AI42 improves a specific workflow on evidence rather than enthusiasm. In this form, RAG becomes more than a mechanism for answering questions: it becomes an operational interface between people, organizational knowledge, and the governance rules that protect both.

Acknowledgments

AI42 was conceived, designed, and developed by Moveo. The authors acknowledge the contribution of the wider Moveo engineering and product teams to the platform and to the operational knowledge that informed the design framework presented in this paper.

Author Contributions

Conceptualization, C.C. and N.M.; methodology, C.C., N.M., C.S., T.M., and M.B.; software and system design, C.C., C.S., T.M., and M.B.; business and organizational analysis, N.M. and C.S.; writing—original draft, C.C.; writing—review and editing, all authors. All authors have read and agreed to the submitted version of the manuscript. This statement must be confirmed by all authors before submission.

Conflict of Interest

All authors are employees of Moveo S.p.A., the organization that develops AI42. The paper therefore presents an industrial artifact in which the authors' employer has a commercial interest. No claim of independent product validation is made; the proposed empirical protocol is intended to support a subsequent independently reviewable evaluation.

Data Availability

No personal data or confidential customer documents are included in this design-science study. The source code and deployment configuration analyzed in the artifact case study are proprietary. The evaluation instruments, task templates, and reporting structure are included in the accompanying research package so that the study design can be inspected without disclosing product-sensitive implementation details.

References

1. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. ACM computing surveys, 55(12), 1-38.
2. Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020, July). On faithfulness and factuality in abstractive summarization. In Proceedings of the 58th annual meeting of the association for computational linguistics (pp. 1906-1919).
3. Liu, N. F., Lin, K., Hewitt, J., Paranjape, A., Bevilacqua, M., Petroni, F., & Liang, P. (2024). Lost in the middle: How

language models use long contexts. Transactions of the association for computational linguistics, 12, 157-173.
4. Cristian Cosentino, Hugo Georgenthum, Fabrizio Marozzo, and Pietro Liò. (2026)b Enhancing surgical documentation through multimodal visual-temporal transformers and generative ai. In Proceedings of the Fifth Machine Learning for Health Symposium, volume 297 of Proceedings of Machine Learning Research, pages 350–368. PMLR.
5. Cosentino, C., Marozzo, F., Felicetti, F., Lamonaca, F., & Liò, P. (2026, July). Metrologically Grounded Graph-Augmented Multimodal AI for Surgical Video Documentation in Smart Healthcare Environments. In 2026 IEEE International Workshop on Metrology for Living Environment (MetroLivEnv) (pp. 301-304). IEEE.
6. Cosentino, C., Gunduz Cure, M., Marozzo, F., & Ozturk Birim, S. (2026). Interpreting User Opinions: A Multidimensional Approach Leveraging Explainable AI and Generative Models. Machine Learning, 115(6), 140.
7. Belcastro, L., Cosentino, C., & Marozzo, F. (2026). Understanding Political Polarization With Generative AI: A Large-Scale Analysis of Voter Preferences on Social Media. IEEE Access, 14, 47389-47405.
8. Belcastro, L., Cosentino, C., Marozzo, F., Presta, A., & Trunfio, P. (2025, March). Empowering efficient drone monitoring with low-latency edge-cloud continuum platforms. In 2025 33rd Euromicro International Conference on Parallel, Distributed, and Network-Based Processing (PDP) (pp. 333-340). IEEE.
9. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. Advances in neural information processing systems, 33, 9459-9474.
10. Khatri, V., & Brown, C. V. (2010). Designing data governance. Communications of the ACM, 53(1), 148-152.
11. Abraham, R., Schneider, J., & Vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. International journal of information management, 49, 424-438.
12. Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all---a contingency approach to data governance. Journal of Data and Information Quality (JDIQ), 1(1), 1-27.
13. Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: an analysis of the literature. Journal of decision systems, 25(sup1), 64-75.
14. International Organization for Standardization. (2026). ISO/IEC 38505-1:2026 Information Technology—Governance of Data—Part 1: Application of ISO/IEC 38500 to the Governance of Data.
15. Karpukhin, V., Oguz, B., Min, S., Lewis, P., Wu, L., Edunov, S., ... & Yih, W. T. (2020, November). Dense passage retrieval for open-domain question answering. In Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP) (pp. 6769-6781).
16. Thakur, N., Reimers, N., Rücklé, A., Srivastava, A., & Gurevych, I. (2021). Beir: A heterogenous benchmark for zero-shot evaluation of information retrieval models. arXiv preprint arXiv:2104.08663.
17. Cormack, G. V., Clarke, C. L., & Buettcher, S. (2009, July). Reciprocal rank fusion outperforms condorcet and individual rank learning methods. In Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval (pp. 758-759).
18. Petroni, F., Piktus, A., Fan, A., Lewis, P., Yazdani, M., De Cao, N., ... & Riedel, S. (2021, June). KILT: a benchmark for knowledge intensive language tasks. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 2523-2544).
19. Es, S., James, J., Anke, L. E., & Schockaert, S. (2024, March). Ragas: Automated evaluation of retrieval augmented generation. In Proceedings of the 18th conference of the european chapter of the association for computational linguistics: system demonstrations (pp. 150-158).
20. Saad-Falcon, J., Khattab, O., Potts, C., & Zaharia, M. (2024, June). Ares: An automated evaluation framework for retrieval-augmented generation systems. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers) (pp. 338-354).
21. Cosentino, C., Squillace, S., & Marozzo, F. (2026). An Auditable LLM-RAG Architecture for Financial Document Intelligence and Decision Support. Future Internet, 18(6), 284.
22. Cosentino, C., Gatto, S., Liò, P., & Marozzo, F. (2026). Machine Unlearning: A Perspective, Taxonomy, and Benchmark Evaluation. Future Internet, 18(3), 174.
23. EU, G. (2016). General data protection regulation. Official Journal of the European Union.
24. Boeckl, K. R., & Lefkovitz, N. B. (2020). NIST privacy framework: A tool for improving privacy through enterprise risk management, version 1.0.
25. Tabassi, E. (2023). Artificial intelligence risk management framework (AI RMF 1.0).
26. AI, N. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile. NIST Trustworthy and Responsible AI Gaithersburg, MD, USA.
27. Lognoul, M. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act–AI Act). Revue du Droit des Technologies de l'information, (3-4), 145-189.
28. Italian Republic. Law no. 132 of 23 september 2025: Provisions and delegations to the government con¬cerning artificial intelligence, 2025.
29. Greshake, K., Abdelnabi, S., Mishra, S., Endres, C., Holz, T., & Fritz, M. (2023, November). Not what you've signed up for: Compromising real-world llm-integrated applications with indirect prompt injection. In Proceedings of the 16th ACM workshop on artificial intelligence and security (pp. 79-90).
30. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly, 319-340.
31. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view1. MIS quarterly, 27(3), 425-478.

32. John Brooke. (1996). SUS: A "quick and dirty" usability scale. In Patrick W. Jordan, Bruce Thomas, Ian L. McClelland, and Bernard Weerdmeester, editors, Usability Evaluation in Industry, pages 189–194. Taylor and Francis, London, UK.
33. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In Advances in psychology (Vol. 52, pp. 139-183). North-holland.
34. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research1. MIS quarterly, 28(1), 75-106.
35. Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of management information systems, 24(3), 45-77.
36. Moveo. (2026). Governare il dato per governare l'ai: dalla sovranità dell'informazione al valore per l'impresa.
37. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative research in psychology, 3(2), 77-101.