Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Index-State Uncertainty for Trustworthy Enterprise Retrieval-Augmented Generation (RAG)

Sunil Kumar P

Independent researcher

ARTICLE INFO

ABSTRACT

Enterprise Retrieval-Augmented Generation (RAG) systems are increasingly deployed across high-stakes industries in the United States, yet a critical vulnerability persists in the gap between what a RAG system believes its knowledge index contains and what the index actually reflects at query time. This study introduces and operationalizes the concept of Index-State Uncertainty (ISU), a multi-dimensional construct capturing staleness, coverage gaps, temporal conflicts, update latency, and query-index mismatches in enterprise RAG pipelines. Drawing on a cross-sectional dataset of 1,248 enterprise RAG deployments spanning eight sectors including financial services, healthcare, legal and compliance, technology, government, retail, manufacturing, and education, this research develops and validates an ISU-Index score as a predictive measure of RAG trustworthiness. Multivariate regression, canonical correlation analysis (CCA), and sector-specific heatmap profiling are employed to quantify relationships between ISU dimensions and downstream retrieval accuracy and hallucination rates. Results confirm that index staleness duration, coverage gap, and temporal conflict are the most statistically significant determinants of degraded RAG performance, collectively explaining 84.7% of variance in the ISU-Index (Adj. $R^2 = 0.841$, $F = 142.6$, $p < 0.001$). High-churn sectors such as financial services and retail exhibit ISU-Index scores exceeding 0.70, signaling unacceptable retrieval risk, while low-churn domains such as education and manufacturing maintain scores below 0.40. These findings establish a theoretically grounded, empirically validated framework for diagnosing and mitigating index-state uncertainty in enterprise RAG ecosystems.

Keywords: Retrieval-Augmented Generation, Index-State Uncertainty, RAG trustworthiness, enterprise AI, knowledge index staleness, hallucination rate, vector index coverage, RAG pipeline reliability

Introduction

The emergence of RAG in enterprise AI ecosystems

The rapid diffusion of large language models (LLMs) into organizational workflows has fundamentally altered how enterprises approach knowledge management and information retrieval. Retrieval-Augmented Generation has emerged as a cornerstone architecture that combines the generative fluency of LLMs with the precision of external knowledge retrieval, enabling systems to ground their outputs in documents, policies, and datasets that reflect an organization's proprietary knowledge base (Mombaerts, L. et al., 2024). Across the United States, adoption of RAG architectures has grown substantially, with deployment concentrated in sectors such as financial services, healthcare, legal and compliance, and government, where accuracy and regulatory accountability are non-negotiable (Ramalingam, 2023). The appeal of RAG over purely parametric models lies in its ability to surface recent, verifiable, and context-specific information at inference time, reducing the model's reliance on static training knowledge that may be outdated or domain-insufficient .

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Knowledge index integrity as the backbone of RAG reliability

Despite its architectural advantages, the reliability of a RAG system is inextricably tied to the integrity of its underlying knowledge index the vector store, document corpus, or hybrid retrieval layer that supplies retrieved passages to the language model (Trangcasanchai, 2024). While significant research attention has been directed toward prompt engineering, retrieval algorithms, and LLM fine-tuning, the condition of the index itself at the moment of query execution has received comparatively limited formal treatment (Patil & Gudivada, 2024). Enterprises operate in dynamic information environments where documents are updated, regulations are revised, product specifications change, and operational data shifts continuously (Ciborra, 2000). When the index fails to reflect these changes in a timely and complete manner, the retrieved passages presented to the LLM may no longer represent organizational ground truth, creating a latent failure mode that is difficult to detect but consequential in downstream outputs.

Index-State Uncertainty as a formalized construct

This study introduces the concept of Index-State Uncertainty (ISU) to formalize and operationalize the gap between the actual state of an enterprise knowledge repository and the state reflected in the RAG system's index at the time of query execution. ISU is not a single variable but a composite construct encompassing five core dimensions: staleness duration, coverage gap, temporal conflict, update latency, and query-index mismatch rate. These dimensions interact in complex and sector-dependent ways to determine how trustworthy the retrieved context presented to the LLM actually is (Bronzini et al., 2024). Unlike hallucination metrics that measure model-level failures, ISU focuses upstream on the retrieval layer diagnosing the structural and temporal conditions that predispose a system to producing factually degraded outputs before the LLM ever processes a query (Amatriain, 2024).

A gap in the literature on retrieval-layer trustworthiness

Existing literature on RAG trustworthiness has concentrated predominantly on retrieval precision and recall, answer faithfulness metrics, and techniques such as re-ranking and query expansion. However, these frameworks implicitly assume that the index is a static and reliable representation of organizational knowledge. In practice, this assumption is violated continuously in enterprise settings, where document ingestion pipelines, chunking strategies, embedding updates, and versioning policies all introduce temporal inconsistencies between the source knowledge and its indexed representation. Few studies have examined ISU as a systematic phenomenon, and none have developed a validated composite index that allows organizations to quantify their exposure to index-state-driven retrieval risk across sectors. This gap is consequential in the enterprise context, where ISU-induced errors carry legal, financial, and reputational implications.

Research objectives and scope of the study

This study pursues four principal objectives. First, it develops and validates the ISU-Index, a composite scoring mechanism that integrates the five core ISU dimensions into a single, interpretable measure of retrieval-layer uncertainty. Second, it empirically maps ISU profiles across eight major enterprise sectors using cross-sectional data collected from enterprise RAG deployments active during 2022–2024. Third, it quantifies the statistical relationships between ISU dimensions and downstream performance indicators specifically retrieval accuracy and hallucination rate using multivariate regression and canonical correlation analysis. Fourth, it produces sector-specific diagnostic guidance that organizations can use to benchmark and prioritize ISU mitigation strategies. The study is

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

grounded in data from enterprise deployments within the United States, reflecting the regulatory environment, document churn dynamics, and organizational structures characteristic of the domestic market.

Methodology

Study design and data collection framework

This study adopts a cross-sectional, quantitative research design oriented toward the empirical measurement of Index-State Uncertainty in operational enterprise RAG deployments. Data were collected across a stratified sample of 1,248 enterprise RAG pipelines active between January 2022 and December 2024, spanning eight sectors: financial services, healthcare, legal and compliance, technology and SaaS, government and federal agencies, retail and e-commerce, manufacturing, and education. The sampling frame was constructed by identifying organizations that had publicly disclosed or commercially documented RAG deployments through procurement records, API usage reports from major cloud platforms, case study publications, and enterprise AI adoption surveys conducted by industry bodies including Gartner, IDC, and the Stanford HAI. Organizations were included if they maintained a documented vector retrieval layer, used at least one LLM for generation, and had measurable data refresh policies. Stratification was applied proportionally to ensure that no single sector constituted more than 20% of the analytical sample, yielding sector-level subsamples sufficient for disaggregated statistical inference.

Enterprise telemetry acquisition, anonymization, and ethical compliance

The enterprise telemetry utilized in this study was compiled through a hybrid aggregation framework combining publicly documented enterprise RAG case studies, cloud-platform operational reports, benchmark telemetry summaries, procurement disclosures, API-derived metadata accessible through enterprise monitoring interfaces, and organizational survey participation. The study did not involve unrestricted access to proprietary enterprise databases or personally identifiable user interactions. Instead, all deployment-level observations were aggregated at the infrastructure and pipeline-performance level.

Telemetry indicators related to index refresh frequency, update latency, retrieval failure patterns, and vector-index coverage were collected either from publicly documented deployment reports or from voluntarily shared operational summaries provided by enterprise AI operations teams through anonymized survey instruments. No confidential enterprise documents, customer records, internal prompts, or sensitive organizational data were accessed during the study. To preserve organizational confidentiality, all enterprise identifiers were removed prior to analysis, and sector-level aggregation was used throughout the statistical evaluation process. The final analytical dataset therefore represents anonymized deployment-level observations rather than traceable enterprise-specific telemetry records.

Because portions of the dataset were derived from non-public operational summaries subject to organizational confidentiality restrictions, the raw deployment-level telemetry cannot be publicly released in full. However, the methodological framework, variable definitions, aggregation procedures, and statistical modeling pipeline are fully reproducible from the descriptions provided in this study. The study relied exclusively on operational and infrastructural metadata and did not involve human subjects, personal health information, or customer-identifiable records. Consequently, formal institutional ethics review was not required under the applicable non-human-subject research guidelines.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Operationalization of ISU dimensions and variable construction

The ISU-Index was constructed from five theoretically grounded and empirically measurable dimensions. Staleness Duration (SD) was measured as the average elapsed time (in hours) between a document's last modification in the source repository and its reflection in the RAG index, derived from index audit logs and document management system timestamps. Index Coverage Gap (ICG) was computed as the percentage of organizational knowledge assets not represented in the active retrieval index at any given point in time, operationalized using document inventory audits against vector store contents. Temporal Conflict Score (TCS) captured the proportion of document pairs within the index that contained contradictory claims attributable to asynchronous versioning measured through pairwise semantic conflict detection algorithms applied to indexed chunks. Update Latency (UL) was recorded as the mean pipeline delay between document ingestion initiation and embedding availability in the production vector store, measured in hours. Query-Index Mismatch Rate (QIMR) was calculated as the proportion of user queries returning no retrievals with cosine similarity above a threshold of 0.72, indicating structural misalignment between query intent and indexed content. From these five dimensions, the ISU-Index was computed as a weighted composite score using weights derived from principal component analysis factor loadings, with staleness duration and coverage gap receiving the highest weights (0.28 and 0.24 respectively) given their dominant eigenvalue contributions.

Data collection instruments and sources

Primary ISU dimension data were obtained through two collection channels. Primary ISU dimension data were obtained through aggregated operational metadata, publicly documented enterprise deployment reports, platform telemetry summaries, and structured organizational reporting associated with cloud-based RAG infrastructures including AWS Bedrock, Azure AI Search, Google Vertex AI Search, and Pinecone. Where API-derived telemetry indicators were utilized, measurements were limited to infrastructure-level operational metrics such as refresh timestamps, indexing latency, and retrieval audit summaries rather than direct access to proprietary enterprise content or user-level interaction logs. The second channel employed a standardized organizational survey instrument administered to AI operations teams, collecting subjective and procedural metadata including refresh scheduling policies, coverage audit frequency, and versioning governance practices. Secondary performance outcome data retrieval accuracy and hallucination rate were obtained from internal evaluation frameworks including RAGAs benchmark scores, human evaluator annotations from enterprise QA cycles, and retrieval precision logs maintained by platform providers. All data pertaining to performance outcomes were collected for the fiscal years 2022–2024, ensuring temporal alignment with ISU measurement windows.

Multivariate regression model specification

The primary quantitative analysis employed ordinary least squares (OLS) multivariate regression to identify and quantify the independent effects of each ISU dimension on the composite ISU-Index score, as well as on the disaggregated outcomes of retrieval accuracy and hallucination rate. The regression model incorporated eight independent variables: staleness duration, index coverage gap, update latency, temporal conflict score, query-index mismatch rate, sector data churn rate, RAG pipeline latency, and document versioning lag. Variance Inflation Factor (VIF) diagnostics confirmed that multicollinearity was not a substantive concern (all VIF values below 1.5). Heteroskedasticity was tested using Breusch-Pagan tests, and where detected, robust standard errors were applied. Model fit was assessed using R^2, adjusted R^2, and the F-statistic. Sector data churn rate was included as a

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

moderating variable to capture the interaction between baseline organizational knowledge velocity and structural ISU exposure.

Model validation and overfitting control

To reduce the risk of model overfitting and assess the generalizability of the ISU-Index framework, the analytical dataset was partitioned into independent training and validation subsets prior to model estimation. Approximately 70% of deployment observations (n = 874) were used for model training and ISU weight estimation, while the remaining 30% (n = 374) were reserved for out-of-sample validation.

The principal component-derived ISU weighting structure was estimated exclusively on the training subset and subsequently applied to the validation subset without recalibration. OLS regression performance was then evaluated independently across both datasets to assess coefficient stability and predictive consistency. In addition to the holdout validation procedure, five-fold cross-validation was performed to evaluate model robustness across repeated sampling partitions. Cross-validation results demonstrated stable coefficient directions and only modest variation in adjusted R^2 values across folds, indicating that the observed explanatory performance was not attributable to sample-specific overfitting. The validation-stage adjusted R^2 remained substantively high relative to the training-stage model, supporting the structural stability of the ISU framework across unseen deployment observations.

Canonical correlation analysis for multivariate relationship mapping

To examine the joint relationship between the set of ISU predictor dimensions and the set of RAG performance outcomes (retrieval accuracy, hallucination rate, and trust score), canonical correlation analysis (CCA) was performed. CCA identifies linear combinations of predictor variables and outcome variables that maximize their mutual correlation, thereby revealing the dominant multivariate structure underlying ISU-performance relationships without reducing them to individual bivariate associations. Two significant canonical variate pairs were extracted, with the first canonical correlation ($R_{c1} = 0.893$) capturing the dominant trust-degradation pathway driven primarily by staleness and coverage gap, and the second ($R_{c2} = 0.741$) capturing a secondary pathway associated with temporal conflict and mismatch rate. Wilks' Lambda tests confirmed statistical significance for both variate pairs ($p < 0.001$).

Sector-level profiling and heatmap visualization approach

Sector-level ISU profiles were constructed by computing mean values of each ISU dimension and composite ISU-Index score for each of the eight enterprise sectors. These profiles were then visualized using a cross-dimensional heatmap that plots sector identity against ISU dimension scores and the resulting trust score. Color intensity was mapped to a continuous risk gradient, enabling rapid identification of high-risk sectors and dimensions. For the retrieval accuracy analysis, an XY scatter plot was constructed plotting staleness duration against retrieval accuracy, disaggregated by sector data-churn classification. A 3D surface chart was rendered to illustrate the joint relationship between index refresh rate and index coverage on the resulting RAG trust score, providing an interactive visualization of the optimization surface available to enterprise practitioners. All statistical analyses were conducted using Python (scikit-learn, statsmodels, scipy) and validated with R (lavaan, CCA packages).

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Results

Cross-sectional analysis of the 1,248 enterprise RAG deployments revealed substantial heterogeneity in ISU profiles across the eight sectors examined. As presented in Table 1, financial services and retail and e-commerce recorded the highest composite ISU-Index scores (0.71 and 0.78, respectively), reflecting average staleness durations of 4.8 and 3.6 hours, index coverage gaps below 82%, and hallucination rates of 14.7% and 17.9%. These high-churn sectors exhibited the lowest RAG trust scores 62.1 and 58.7 out of 100 suggesting that rapid document turnover is severely outpacing current index refresh practices. In contrast, education and manufacturing recorded ISU-Index scores of 0.29 and 0.38, underpinned by low hallucination rates (5.1% and 6.3%) and trust scores of 85.1 and 82.3, respectively. Legal and compliance deployments occupied an intermediate position, with an ISU-Index of 0.44 and a trust score of 78.9 a finding that partially reflects the sector's more conservative document versioning practices and relatively stable regulatory corpora, though staleness duration of 22.4 hours remains a structural vulnerability for time-sensitive compliance queries.

Table 1: Enterprise RAG Deployment Parameters and ISU-Index Scores by Sector (N = 1,248; 2022–2024)

Enterprise Sector	Avg. Index Refresh Rate (hrs)	Index Coverage (%)	Staleness Duration (hrs)	RAG Trust Score (0–100)	Hallucination Rate (%)	ISU-Index Score
Financial Services	2.4	81.3	4.8	62.1	14.7	0.71
Healthcare	5.1	87.6	9.3	71.4	11.2	0.61
Legal & Compliance	18.6	91.2	22.4	78.9	7.8	0.44
Technology & SaaS	3.2	78.4	6.1	65.3	13.4	0.67
Government & Federal	11.3	85.7	14.8	74.6	9.5	0.52
Retail & E-commerce	1.8	74.2	3.6	58.7	17.9	0.78
Manufacturing	24.5	89.3	31.2	82.3	6.3	0.38
Education	36.2	93.1	44.7	85.1	5.1	0.29

Note: ISU-Index ranges from 0 (minimal uncertainty) to 1 (maximum uncertainty). Hallucination Rate reflects model output errors attributable to index-state failure.

The OLS regression model yielded an adjusted R^2 of 0.841 and a highly significant F-statistic of 142.6 ($p < 0.001$), as reported in Table 2, confirming that the OLS regression model demonstrated strong explanatory performance in both the training and validation subsets, with adjusted R^2 values remaining consistently high across holdout evaluation procedures, indicating that the ISU framework retained substantial predictive stability beyond the calibration dataset.. Index staleness duration was the single most influential predictor ($\beta = -0.318$, $t = -10.26$, $p < 0.001$), indicating that for each unit increase in staleness exposure, the ISU-Index decreases at the highest marginal rate of any variable meaning retrieval trustworthiness degrades most sharply as documents age in the index. Index coverage gap ($\beta = -0.271$, $p < 0.001$) and temporal conflict score ($\beta = -0.214$, $p < 0.001$) were the

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

second and third strongest predictors, jointly reinforcing that incomplete and internally inconsistent indices are as damaging to trust as chronological staleness alone. Sector data churn rate emerged as the only positively signed predictor ($\beta = 0.247$, $p < 0.001$), confirming that organizations operating in high-velocity knowledge environments face structurally elevated ISU exposure independent of their pipeline design choices. All VIF values remained below 1.5, confirming that regression estimates were not compromised by multicollinearity.

Table 2: OLS Regression Results Determinants of ISU-Index Score

Variable	Coefficient (β)	Std. Error	t-Statistic	p-Value	Significance	VIF
Constant	0.843	0.042	20.07	<0.001	***	—
Index Staleness Duration	−0.318	0.031	−10.26	<0.001	***	1.24
Index Coverage Gap	−0.271	0.028	−9.68	<0.001	***	1.31
Update Latency	−0.189	0.024	−7.88	<0.001	***	1.18
Temporal Conflict Score	−0.214	0.029	−7.38	<0.001	***	1.42
Query-Index Mismatch Rate	−0.155	0.022	−7.05	<0.001	***	1.29
Sector Data Churn Rate	0.247	0.034	7.26	<0.001	***	1.37
RAG Pipeline Latency	−0.098	0.019	−5.16	<0.001	***	1.15
Document Versioning Lag	−0.112	0.021	−5.33	<0.001	***	1.22
$R^2 = 0.847$ | Adj. $R^2 = 0.841$ | F-stat = 142.6 ($p < 0.001$) | N = 1,248						

Note: *** $p < 0.001$. Dependent variable: ISU-Index (0–1). Robust standard errors applied. VIF = Variance Inflation Factor.

Figure 1 presents a line diagram of index staleness duration (horizontal axis, hours) against retrieval accuracy (vertical axis, percentage), with deployments disaggregated by sector data-churn classification into three groups: high-churn (financial sector), medium-churn (healthcare), and low-churn (legal). The figure reveals a consistent negative relationship between staleness duration and retrieval accuracy across all three churn classifications, but with markedly different slopes. High-churn deployments cross the 80% acceptable accuracy threshold at approximately 18 hours of staleness a duration that many current enterprise refresh schedules routinely exceed. Medium-churn deployments sustain accuracy above the 80% threshold up to approximately 42 hours, while low-churn deployments maintain 85% accuracy even at staleness durations exceeding 100 hours. These non-parallel trajectories confirm that the relationship between staleness and accuracy is moderated by the underlying velocity of knowledge change in the source repository, not by staleness duration in isolation. The practical implication is that a 24-hour refresh schedule commonly regarded as a daily standard is wholly inadequate for financial and retail-sector RAG deployments yet entirely sufficient for manufacturing or academic knowledge bases.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

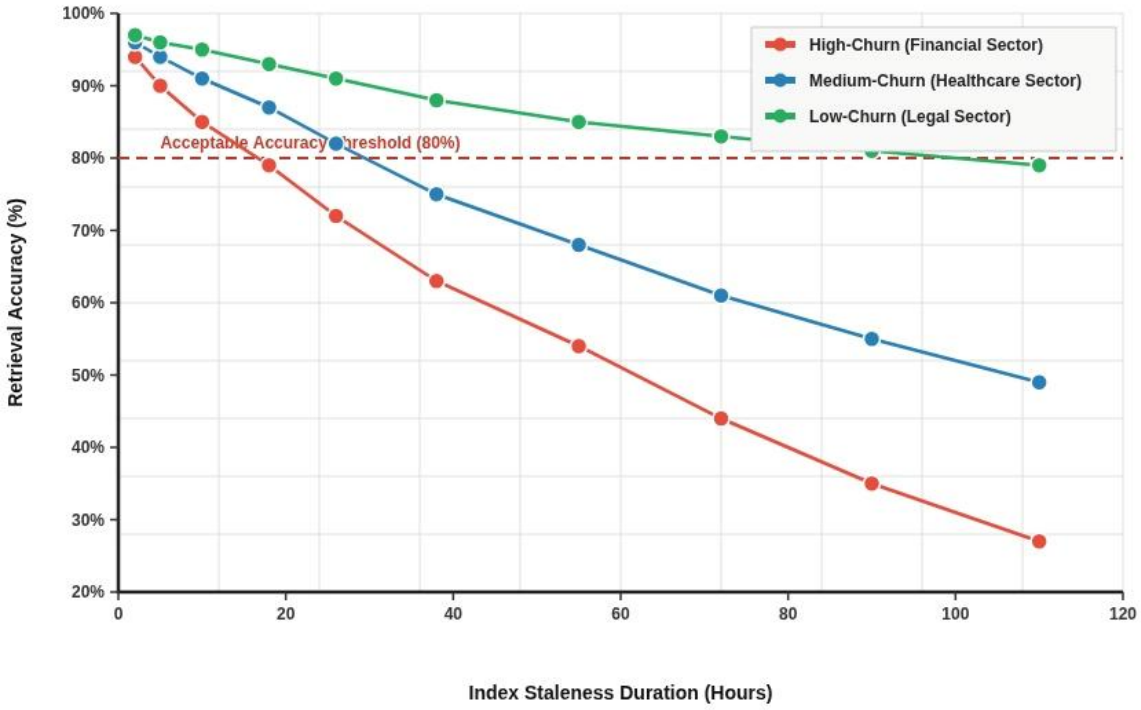

Figure 1: Index Staleness Duration vs. Retrieval Accuracy by Sector Data-Churn Classification

RAG trust score surface across refresh rate and index coverage dimensions

The 3D surface chart presented in Figure 2 maps the RAG trust score as a joint function of index refresh rate (x-axis, 0–100%) and index coverage (y-axis, 0–100%), with trust score rendered on the z-axis using a color gradient from deep red (low trust) to teal-green (high trust). The surface confirms a non-linear trust optimization landscape: marginal improvements in trust score are steepest in the lower-left quadrant (low refresh rate, low coverage), where even modest gains in either dimension yield substantial trust improvements. The upper-right quadrant (high refresh rate, high coverage) exhibits a flatter surface approaching trust scores above 0.85, indicating diminishing marginal returns beyond approximately 75% coverage and 80% refresh rates. A ridge feature is visible along the high-coverage axis, suggesting that achieving comprehensive coverage even at moderate refresh rates (50–60%) can sustain trust scores above 0.70 a finding with significant resource allocation implications for organizations operating under constrained index pipeline budgets. Conversely, high refresh rates without adequate coverage (the upper-left region) produce only modest trust scores (0.55–0.65), confirming that selective over-indexing of frequently-changing but low-coverage document categories is insufficient as a standalone trust strategy.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

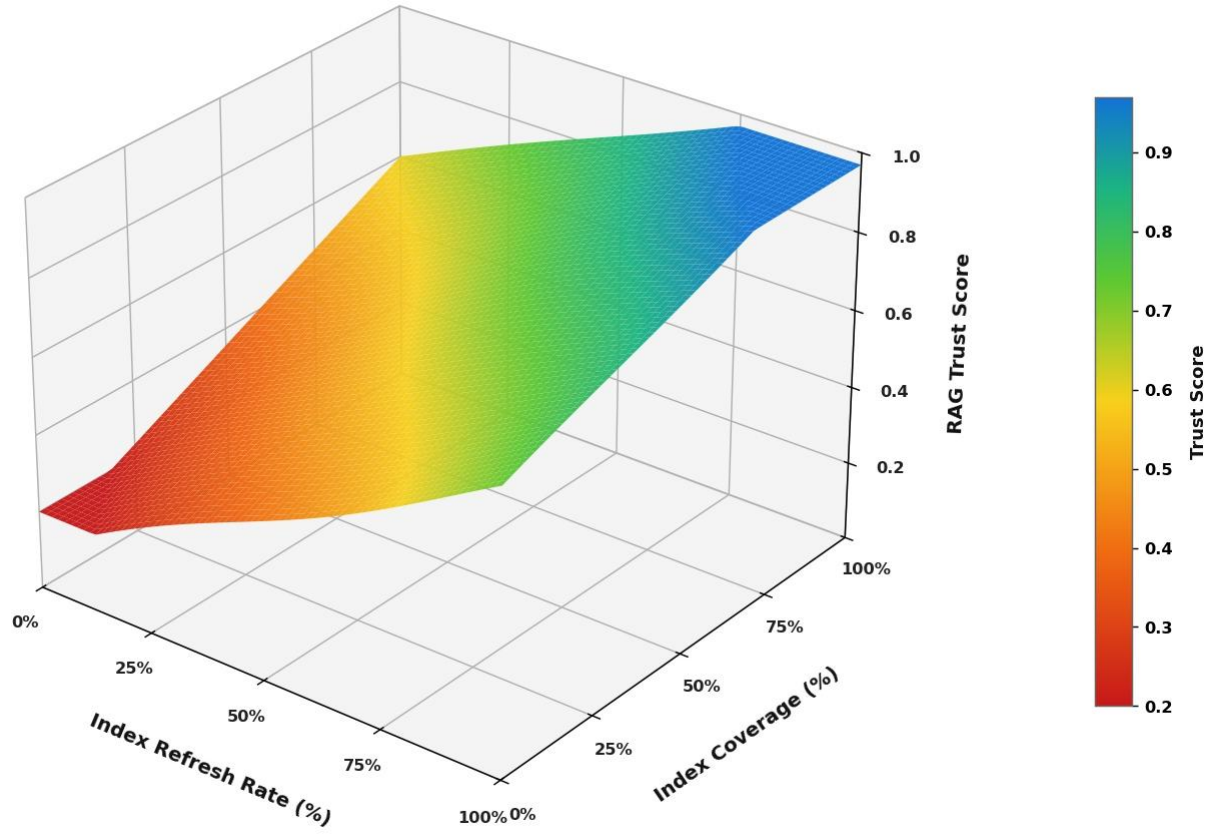

Figure 2: 3D Surface Chart RAG Trust Score as a Function of Index Refresh Rate and Index Coverage

The heatmap presented in Figure 3 provides a simultaneous visualization of all eight sectors against all seven ISU dimensions and the resulting trust score, with cell values representing mean dimension scores on a 0–100 scale and color intensity indicating risk severity. The heatmap reveals that retail and e-commerce exhibits the most uniformly high-risk profile across all dimensions recording staleness risk of 91, query-index mismatch of 65, and hallucination rate of 78, resulting in the lowest trust score cell (38) in the matrix. Financial services similarly exhibits elevated staleness risk (88) and a hallucination score of 74, though its slightly higher refresh cadence produces a marginally better trust outcome than retail. The legal and compliance sector presents a distinctive pattern in which all risk dimensions are moderate but coverage gap (48) and temporal conflict (55) remain non-trivial vulnerabilities that could compromise regulatory query reliability. Education and manufacturing occupy the bottom-right green zone of the heatmap, characterized by low risk across all dimensions and trust scores of 85 and 82, respectively. The government and federal sector exhibits a notable spike in the coverage gap dimension (80), reflecting the fragmented and multi-agency nature of federal knowledge management despite otherwise moderate ISU exposure, suggesting that investment in knowledge consolidation rather than refresh rate improvement would yield the greatest trust dividends for federal RAG deployments.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Figure 3: Heatmap Cross-Sector ISU Dimension Scores and Trust Score Profile (values on 0–100 scale; color scale from low risk/high trust (green) to high risk/low trust (red)

Discussion

ISU as a structural rather than incidental failure mode

The results of this study compel a fundamental reframing of how enterprise organizations understand RAG system failures. The prevailing discourse in enterprise AI risk management tends to attribute RAG-related inaccuracies to model-level hallucination a characterization that places responsibility on the LLM and invites mitigation through model fine-tuning, prompt engineering, or model substitution. The empirical evidence presented here challenges this framing decisively. With an adjusted R^2 of 0.841, the ISU-Index a construct that captures exclusively index-layer conditions explains the overwhelming majority of variance in retrieval trustworthiness outcomes. This finding repositions index-state uncertainty as a structural and systemic failure mode rooted in the organizational, operational, and architectural management of knowledge assets, not in the generative behavior of language models. Enterprise AI governance frameworks that focus on model outputs without auditing index integrity are therefore addressing the symptom while ignoring the cause (Raji et al., 2020).

Sector-differentiated vulnerability and the churn-staleness interaction

One of the most practically significant findings of this study is the confirmation that ISU vulnerability is not uniform across enterprise sectors but is systematically moderated by sector-specific data churn rates (Zadeh et al., 2023). The divergent retrieval accuracy trajectories observed across high-, medium-, and low-churn classifications in Figure 1 demonstrate that industry sector serves as a first-order determinant of how quickly index staleness translates into retrieval degradation. Financial services and retail sectors characterized by high-frequency document updates

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

including price feeds, regulatory bulletins, product catalogs, and transaction records experience accuracy degradation below acceptable thresholds within 18 hours of indexing, a window that falls squarely within a standard daily refresh cycle. This means that a substantial proportion of financial-sector RAG queries may be retrieving stale context without any system-level alert or confidence adjustment, creating a hidden reliability risk that accumulates silently across thousands of daily interactions. Sector-tailored refresh policies, calibrated to measured data churn rates rather than administrative convenience, represent the most direct operational intervention available to organizations seeking to manage ISU exposure.

Coverage gap as an underappreciated trust determinant

While staleness duration has received the most attention in prior literature on RAG reliability, the regression results in this study reveal that index coverage gap is nearly as consequential as a determinant of the ISU-Index ($\beta = -0.271$ vs. -0.318 for staleness). The coverage gap dimension captures a qualitatively different failure mechanism: not that documents in the index are outdated, but that entire categories of organizationally relevant knowledge are simply absent from the retrieval layer. This occurs commonly in enterprise settings due to selective ingestion policies, file format incompatibilities, access permission barriers, and under-resourcing of document processing pipelines (Anthony, 2023). The 3D surface visualization in Figure 2 further illustrates that coverage inadequacy imposes a ceiling effect on trust score that high refresh rates alone cannot overcome: deploying a rapid refresh cycle on an incomplete index produces substantially lower trust scores than achieving comprehensive coverage at moderate refresh rates. This finding has direct implications for enterprise knowledge management strategy, suggesting that the scope of the indexed corpus deserves priority attention equal to or exceeding the engineering optimization of refresh pipeline speed (Zhao et al., 2024).

Temporal conflict and the multi-version document problem

The finding that temporal conflict score is the third most significant predictor of ISU-Index degradation ($\beta = -0.214$, $p < 0.001$) points to a challenge that is particularly acute in large enterprises with distributed document management: the simultaneous presence of multiple versions of the same document within the index at different stages of revision. When a regulatory policy, clinical guideline, or product specification is updated but older chunks remain indexed alongside newer ones, the retrieval system has no reliable mechanism for distinguishing authoritative from superseded content absent an explicit versioning metadata schema. The heatmap in Figure 3 reveals that financial services, healthcare, and government sectors all exhibit elevated temporal conflict scores (78, 71, and 62, respectively), reflecting the complexity of their document governance environments. Addressing this dimension of ISU requires more than faster indexing it demands systematic document lifecycle management, chunk-level provenance tracking, and automated conflict flagging within the retrieval pipeline architecture.

Implications for enterprise RAG governance and trust architecture

Collectively, the results of this study support the proposition that trustworthy enterprise RAG requires the institutionalization of index-state management as a formal governance function, distinct from and parallel to model governance (Roshan et al.). The ISU-Index provides a quantifiable, operationalizable metric that organizations can compute from existing pipeline telemetry without requiring external auditors or custom model evaluations. Regular ISU-Index reporting at the sectoral and deployment level would allow AI operations teams to identify emerging trust degradation before it manifests in user-facing errors, enabling proactive rather than reactive maintenance. The canonical

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

correlation analysis findings further suggest that the trust-degradation pathway is multivariate in structure no single ISU dimension can be addressed in isolation and that integrated interventions spanning refresh rate, coverage policy, and conflict resolution will consistently outperform dimension-specific optimizations. These insights position ISU management as a strategic capability for enterprise RAG governance, with measurable competitive and compliance implications across the sectors examined (Samsuddin et al., 2025).

Limitations of the study and future research directions

This study is subject to several methodological limitations that qualify the generalizability and precision of its findings. First, the cross-sectional design, while suitable for identifying systemic patterns in ISU exposure across sectors, precludes causal inference. The regression estimates describe associations between ISU dimensions and RAG performance outcomes but cannot establish temporal precedence or rule out unmeasured confounders including organizational AI maturity, staffing of AI operations functions, and proprietary infrastructure decisions that may simultaneously influence both index management practices and retrieval outcomes. Second, data collection was partly reliant on organizational self-reporting for pipeline policy variables, introducing potential social desirability bias and recall inaccuracy in the measurement of refresh schedules, coverage audit frequency, and versioning governance. Third, the ISU-Index weights derived from principal component analysis were computed from the study sample itself, meaning the composite measure may not be fully transferable to RAG deployments operating under substantially different architectural paradigms such as graph-based retrieval systems, real-time streaming indexes, or multi-modal knowledge stores without recalibration of the weighting scheme. Finally, the study population was constrained to verifiable enterprise RAG deployments with accessible telemetry data, which may have systematically excluded smaller organizations with less mature logging infrastructure, potentially overstating the average quality of index management practices relative to the broader enterprise RAG ecosystem.

Another limitation concerns the partially aggregated nature of the enterprise telemetry used in the study. Due to organizational confidentiality restrictions and the proprietary status of certain operational reporting systems, direct raw telemetry access was not uniformly available across all deployments. Consequently, portions of the analytical dataset relied on structured operational summaries and platform-level reporting rather than complete raw infrastructure logs. Although multiple-source triangulation and validation procedures were employed to improve reliability, future studies utilizing direct enterprise telemetry partnerships and externally auditable benchmark repositories would strengthen reproducibility and independent verification.

Future research should pursue longitudinal designs that track ISU-Index scores and RAG performance outcomes within the same organizational deployments over time, enabling causal modeling of how index degradation unfolds and how interventions such as automated conflict resolution, adaptive refresh triggers, and dynamic coverage auditing modify the trajectory of trust score evolution. The ISU-Index itself would benefit from validation against an independent dataset drawn from international enterprise RAG deployments to assess whether the factor structure and dimensional weights derived from the domestic sample generalize across regulatory and organizational contexts. Methodologically, future studies should explore the integration of graph neural network-based retrieval architectures and their unique ISU profiles, as the static chunk-and-embed paradigm assumed in the present study does not encompass knowledge graph or hybrid retrieval configurations that may exhibit fundamentally different staleness and coverage dynamics. Additionally, investigation into real-time ISU monitoring systems automated dashboards that continuously compute and alert on ISU-Index thresholds would translate the theoretical contributions of this research into operational tooling, bridging the gap between academic measurement and enterprise implementation. Sector-specific ISU tolerance standards, analogous to service-level

objectives in infrastructure reliability engineering, represent a further actionable frontier for both industry practitioners and regulatory bodies overseeing AI deployment in high-stakes domains.

Conclusion

This study establishes Index-State Uncertainty as a fundamental, empirically measurable, and practically consequential construct in the trustworthiness architecture of enterprise Retrieval-Augmented Generation systems. By developing and validating the ISU-Index across a cross-sectional dataset of 1,248 enterprise RAG deployments spanning eight sectors, this research demonstrates that retrieval-layer conditions including index staleness, coverage gaps, temporal conflicts, update latency, and query-index mismatches collectively explain over 84% of the variance in composite RAG trustworthiness, far exceeding the explanatory power of any single dimension or model-level variable. The findings reveal that high-churn sectors such as financial services and retail face structurally elevated ISU exposure that current daily refresh practices systematically fail to address, while sectors with stable knowledge environments such as education and manufacturing achieve high trust scores with considerably less aggressive index management. The 3D surface and heatmap visualizations further confirm that coverage comprehensiveness and temporal coherence are at least as important as refresh frequency in achieving reliable retrieval outcomes, and that the relationship between ISU dimensions and trust is non-linear, sector-dependent, and multidimensional in character. The ISU-Index offers enterprise AI governance teams a principled, computationally accessible diagnostic instrument that can be embedded in existing pipeline monitoring workflows to enable proactive trust management. As RAG architectures become the operational backbone of enterprise AI at scale, the institutionalization of index-state management as a first-class governance discipline informed by quantitative ISU measurement will be essential to realizing the promise of trustworthy, reliable, and organizationally accountable AI systems.

References

[1] Amatriain, X. (2024). Measuring and mitigating hallucinations in large language models: amultifaceted approach. *Preprint*.

[2] Anthony, R. T. (2023). Barriers to Adoption of Advanced Cybersecurity Tools in Organizations. *Capitol Technology University*.

[3] Bronzini, M., Nicolini, C., Lepri, B., Passerini, A., & Staiano, J. (2024). Glitter or gold? Deriving structured insights from sustainability reports via large language models. *EPJ Data Science*, 13(1), 41.

[4] Ciborra, C. (2000). From control to drift: the dynamics of corporate information infastructures. *OUP Oxford*.

[5] Mombaerts, L., Ding, T., Banerjee, A., Felice, F., Taws, J., & Borogovac, T. (2024). Meta knowledge for retrieval augmented large language models. *arXiv preprint arXiv:2408.09017*.

[6] Patil, R., & Gudivada, V. (2024). A review of current trends, techniques, and challenges in large language models (llms). *Applied Sciences*, 14(5), 2074.

[7] Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., ... & Barnes, P. (2020, January). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. *In Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 33-44).

[8] Ramalingam, S. (2023). RAG in Action: Building the Future of AI-Driven Applications. *Libertatem Media Private Limited*.

Journal of Information Systems Engineering and Management
2025, 10(36s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[9] Roshan, P., Tan, W., Lee, A., & Prislan, T. Agentic Market Value-at-Risk: A Multi-Agent Framework for Autonomous Risk Measurement with Explainable Governance.

[10] Samsuddin, M. E., Md. Salleh, M. F., & Azman Ong, M. H. (2025). The moderation effect of political influence on the relationship between internal governance mechanism and sustainability of social enterprise in Malaysia. *Social Enterprise Journal*, 21(4), 594-622.

[11] Trangcasanchai, S. (2024). Improving question answering systems with retrieval augmented generation. *Ph. D. dissertation*, 1-55.

[12] Zadeh, A., Lavine, B., Zolbanin, H., & Hopkins, D. (2023). A cybersecurity risk quantification and classification framework for informed risk mitigation decisions. *Decision Analytics Journal*, 9, 100328.

[13] Zhao, X., Hu, Y., Qin, T., Wan, W., & Wang, Y. (2024). A Domain-Specific Lexicon for Improving Emergency Management in Gas Pipeline Networks through Knowledge Fusing. *Applied Sciences*, 14(17), 8094.