Journal of Management and Economics

TYPE Original Research
PAGE NO. 56-61

Check for updates

# Enhancing Enterprise Knowledge Management through Chain-of-Thought Prompting in Large Language Models: A Framework for Secure and Intelligent Decision Support

OPEN ACCESS

SUBMITED 01 June2025
ACCEPTED 16 June2025
PUBLISHED 30 June 2025
VOLUME Vol.05 Issue 06 2025

**Dr. A. Sterling**
Department of Information Systems and Strategic Management

**J. K. Vance**
Department of Information Systems and Strategic Management

**Abstract:** The rapid exponential growth of big data has rendered traditional Knowledge Management (KM) systems insufficient for real-time decision-making. While organizations possess vast repositories of unstructured data, the ability to synthesize this information into actionable insights remains a critical bottleneck. This study explores the integration of Large Language Models (LLMs) into enterprise KM frameworks, specifically focusing on the efficacy of Chain-of-Thought (CoT) prompting and Active Prompting strategies to enhance reasoning fidelity. Drawing upon recent advancements in foundation models, including GPT-4 and LLaMA, we propose a novel architecture that couples generative AI with Role-Based Access Control (RBAC) protocols to ensure data security and governance. Our methodology involves a comparative analysis of zero-shot reasoning versus iterative CoT prompting across simulated business intelligence scenarios. The results indicate that CoT prompting significantly mitigates logical fallacies and improves the contextual relevance of outputs, effectively transforming LLMs from passive text generators into active reasoning engines. Furthermore, the integration of RBAC mechanisms addresses the critical challenge of information segregation in hierarchical organizations. This research suggests that synergizing advanced prompting engineering with robust security standards allows organizations to leverage their data assets more effectively, fostering a "learning organization" culture that supports competitive advantage.

**Keywords:** Large Language Models, Chain-of-Thought Prompting, Knowledge Management, Role-Based Access Control, Business Intelligence, Big Data Analytics, Generative AI.

## INTRODUCTION

The contemporary digital landscape is defined by an unprecedented accumulation of data. As organizations transition into the era of hyper-connectivity, the volume, velocity, and variety of information—collectively known as "Big Data"—have outpaced the human capacity for manual processing. Gandomi and Haider [10] note that while Big Data offers the potential for profound insight, it often manifests as noise rather than signal, creating a paradox where information abundance leads to decision paralysis. Consequently, the domain of Knowledge Management (KM) is undergoing a radical transformation. Historically, KM systems were designed as static repositories, intended to capture and store explicit knowledge. However, as Gagnon et al. [9] illustrate in the context of healthcare, effective KM requires a "learning organization" approach, where knowledge is not merely stored but actively synthesized and disseminated to facilitate adaptive decision-making.

Into this gap steps the paradigm of Generative Artificial Intelligence, specifically Large Language Models (LLMs). The release of models such as OpenAI's GPT-4 [5] and Meta's LLaMA [6] has demonstrated that neural networks can achieve performance comparable to human cognition on a wide array of professional benchmarks. These models are not limited to simple pattern matching; they exhibit emergent behaviors that resemble reasoning. However, the deployment of LLMs in enterprise environments is fraught with challenges. Chief among these are the issues of "hallucination"—the generation of plausible but factually incorrect assertions—and the lack of verifiable logic trails.

To address these cognitive deficits, recent research has pivoted toward prompt engineering. Techniques such as Chain-of-Thought (CoT) prompting, as explored by Wang et al. [1] and Kojima et al. [2], suggest that encouraging a model to verbalize its intermediate reasoning steps significantly improves performance on complex tasks. Simultaneously, Diao et al. [4] have proposed "Active Prompting" to further refine these capabilities by identifying uncertain questions and annotating them for the model.

Yet, cognitive capability is only half the equation for enterprise adoption. The other half is governance. In a corporate environment, not all knowledge is intended for all users. The principles of Role-Based Access Control (RBAC), a standard established by Ferraiolo et al. [7], must be rigorously applied to prevent AI systems from becoming vectors for data leakage. This creates a tension between the holistic synthesis capabilities of LLMs and the compartmentalization required by security protocols.

This paper proposes a unified framework that leverages CoT prompting to enhance the analytical utility of LLMs within KM systems, while simultaneously integrating RBAC constraints to ensure security. By examining the intersection of advanced prompting strategies, fallacy recognition [3], and business intelligence applications [8], we aim to provide a roadmap for organizations seeking to leverage AI for sustainable competitive advantage.

## METHODOLOGY

Our research methodology employs a multi-tiered approach to evaluate the integration of LLMs into enterprise Knowledge Management systems. We adopt a design science research strategy, constructing a theoretical prototype that simulates the interaction between a user, a secure data repository, and an LLM reasoning engine.

### 2.1 Theoretical Framework and Model Selection

We selected two primary foundation models for this comparative analysis: GPT-4 [5] and LLaMA [6]. GPT-4 represents the state-of-the-art in proprietary, high-parameter models, offering superior zero-shot reasoning capabilities. LLaMA, conversely, represents the efficient, open-source alternative, allowing for local deployment which is often preferred in high-security enterprise environments. The framework is grounded in the concept of the "Learning Organization" [9], assessing the models not just on factual retrieval, but on their ability to facilitate organizational learning through synthesis.

### 2.2 Prompting Architectures

To test the reasoning fidelity of these models, we established three distinct prompting conditions:

1. Standard Zero-Shot Prompting: The model is presented with a complex business query (e.g., "Analyze the decline in Q3 regional sales") without any instructions regarding the reasoning process. This serves as the control group, testing the baseline capabilities described by Kojima et al. [2].

2. Iterative Chain-of-Thought (CoT) Prompting: Following the protocols outlined by Wang et al. [1], the model is instructed to "Think step-by-step" and decompose the problem into constituent variables (e.g.,

market trends, supply chain data, competitor activity) before aggregating a final answer.

3. Active Prompting with Fallacy Recognition: This advanced condition incorporates the fallacy recognition logic proposed by Alhindi et al. [3] and the active prompting mechanisms of Diao et al. [4]. Here, the model is explicitly instructed to scan the input data for logical inconsistencies (e.g., correlation vs. causation errors) before generating insights.

### 2.3 Security Integration Layer

A critical component of our methodology is the simulation of an RBAC governance layer. Drawing on the NIST standard defined by Ferraiolo et al. [7], we constructed a dataset tagged with security clearance levels (Level 1: Public, Level 2: Internal, Level 3: Confidential). The prompt engineering pipeline was designed to include a pre-processing step where the user's role ID is validated against the data tags. The LLM is then provided only with the data context accessible to that specific role. We evaluate the system’s ability to generate coherent answers when partial information is redacted due to security constraints.

### 2.4 Business Intelligence Scenarios

To ensure ecological validity, the test queries were derived from real-world Business Intelligence (BI) use cases as described by Patel [8]. These scenarios focus on competitive advantage analysis, requiring the models to synthesize structured sales data with unstructured qualitative feedback (e.g., customer reviews, employee emails). This tests the "Big Data" synthesis capabilities highlighted by Gandomi and Haider [10].

## RESULTS

The analysis of the interactions across the three prompting conditions revealed significant disparities in the quality, logic, and safety of the generated outputs.

### 3.1 Impact of Chain-of-Thought on Reasoning Fidelity

Under the Standard Zero-Shot condition, both GPT-4 and LLaMA exhibited a tendency toward "associative leaping." When asked to analyze sales declines, the models frequently jumped to the most statistically probable conclusion (e.g., "seasonal variance") without adequately weighing the specific context provided in the prompt.

However, the introduction of Iterative CoT Prompting produced a marked improvement. As suggested by Wang et al. [1], the instruction to generate intermediate reasoning steps forced the models to parse the input data more granularly. In the CoT condition, the models successfully identified subtle causal links, such as a supply chain disruption mentioned in a footnote of the provided text, which the zero-shot attempt ignored. The "step-by-step" mechanism appears to function as a self-correction loop, reducing the rate of hallucination by grounding the final output in the preceding logical steps.

### 3.2 Fallacy Recognition and Active Prompting

The application of Multitask Instruction-based Prompting for Fallacy Recognition [3] yielded the most robust results for decision support. When presented with data containing logical traps—such as a false dilemma or a post hoc fallacy—models using standard prompting often propagated the error. In contrast, the Active Prompting condition [4] empowered the models to flag these inconsistencies. For instance, when presented with a dataset suggesting a correlation between a new marketing campaign and a drop in sales, the CoT-enabled model explicitly stated, "While the timing coincides, there is insufficient evidence to establish causation without reviewing external market factors." This demonstrates a shift from mere text generation to critical analysis.

### 3.3 Model Efficiency and Deployment Considerations

In comparing the models, GPT-4 [5] consistently outperformed LLaMA [6] in complex reasoning tasks requiring broad world knowledge. However, LLaMA demonstrated surprising efficiency when the context was strictly limited to the provided enterprise documents. This suggests that for internal KM systems, efficiently fine-tuned open-source models (like LLaMA) utilizing CoT prompting may offer a more cost-effective and secure alternative to calling external APIs, aligning with the efficiency goals of modern BI infrastructures [8].

### 3.4 RBAC Compliance

The integration of the RBAC layer [7] proved effective but introduced semantic challenges. When high-value context was redacted due to low user clearance, the models occasionally struggled to bridge the logical gaps, resulting in generic or disjointed answers. This highlights a critical trade-off: strict security protocols can inadvertently degrade the quality of AI-generated insights if the model is not "aware" of the missing context.

## DISCUSSION

The findings of this study suggest that the convergence of Large Language Models and rigorous prompt engineering represents a paradigm shift in Knowledge Management. We are moving away from the era of "search and retrieve" toward an era of "synthesize and reason." This section expands on the implications of these findings, specifically focusing on the cognitive architecture of the enterprise, the nuances of security in probabilistic systems, and the strategic value of automated insight.

### 4.1 The Cognitive Architecture of the Modern Enterprise

The implementation of Chain-of-Thought (CoT) prompting [1] does more than improve model accuracy; it fundamentally alters the cognitive architecture of an organization. Traditional KM systems, as described in early literature, relied on the user to provide the reasoning. The system provided the what (data), and the human provided the why (analysis). Our results indicate that CoT-enabled LLMs can begin to shoulder the burden of the why.

By breaking down complex problems into intermediate steps, the model mimics the cognitive processes of a human analyst. This aligns with the "Zero-Shot Reasoners" concept [2], where the model utilizes latent knowledge to construct logical bridges between disparate data points. For a "Learning Organization" [9], this is transformative. It implies that the knowledge base is no longer a passive library but an active participant in problem-solving.

For example, in the context of nursing management discussed by Gagnon et al. [9], a CoT-enabled system could not only retrieve a protocol for patient care but also reason through the specific comorbidities of a patient to suggest modifications to that protocol. This moves the system from informational support to decision support. The ability of the model to show its work—its chain of thought—is crucial here. It allows human operators to audit the AI's logic, fostering trust and enabling a "human-in-the-loop" verification process that is essential for high-stakes environments.

### 4.2 The Friction Between Generative AI and RBAC

One of the most complex aspects of deploying LLMs in the enterprise is the integration of rigid security frameworks like Role-Based Access Control (RBAC) [7] with the probabilistic, fluid nature of Generative AI.

In traditional software, access control is binary: a user either has access to a database row, or they do not. LLMs, however, operate on semantic vectors. If a model is trained on confidential data, that data exists diffusely within its parameters. Even with Retrieval-Augmented Generation (RAG) where data is injected at inference time, there is a risk of "context leakage" or "inference attacks" where a user might prompt the model to reveal information they are not authorized to see through deduction.

Our study utilized a pre-processing filter to strip data based on RBAC tags before it reached the model. While effective for preventing direct leakage, this approach creates the "logical gap" observed in our results. If an Executive (Level 3) asks "Why did sales drop?" the model sees the confidential merger failure and answers correctly. If a Manager (Level 2) asks the same question, and the merger data is redacted, the model sees a hole in the data.

To address this, future enterprise architectures must implement "Context-Aware Denial." Instead of simply failing or hallucinating when data is missing, the model should be trained (via Active Prompting [4] and Fallacy Recognition [3]) to recognize the information gap and state: "Based on the data available to your access level, the primary factors are X and Y. There may be additional strategic factors not currently visible." This preserves the integrity of the RBAC model [7] while maintaining the user's trust in the system's honesty. This aligns with the transparency required in the GPT-4 Technical Report [5], emphasizing the need for models to know what they do not know.

### 4.3 Strategic Business Intelligence and Competitive Advantage

Patel [8] argues that leveraging Business Intelligence (BI) is central to maintaining competitive advantage. The integration of LLMs amplifies this advantage by reducing the "time-to-insight." In traditional BI, the latency between data collection (Big Data) [10] and strategic action is often caused by the time required for analysts to query dashboards and interpret trends.

LLMs acting as "reasoning engines" compress this timeline. By automating the preliminary analysis through CoT prompting, organizations can monitor market signals in near real-time. For instance, utilizing LLaMA [6] models fine-tuned on internal communication channels could allow a company to detect a shift in customer sentiment weeks before it reflects in the quarterly sales figures.

Furthermore, the use of Active Prompting [4] allows the system to be proactive. Rather than waiting for a user to ask a question, an advanced KM system could run continuous background reasoning processes, flagging anomalies or opportunities. If the model detects a discrepancy between inventory levels and projected demand (a potential fallacy in planning), it could alert the relevant stakeholder. This shifts BI from a descriptive tool (what happened?) to a prescriptive tool (what should we do?), directly supporting the competitive maneuvering discussed by Patel [8].

### 4.4 Handling Hallucinations and Logical Fallacies

The propensity for LLMs to hallucinate remains the primary barrier to widespread adoption. Alhindi et al. [3] highlight the importance of fallacy recognition. In our expansion of this concept, we argue that CoT prompting serves as a prophylactic against hallucination. By forcing the model to externalize its logic, we increase the "inference cost" for the model to lie. To fabricate a conclusion, the model would have to fabricate a plausible chain of reasoning leading to it, which is statistically harder than simply predicting a false token.

However, this is not foolproof. As noted in the GPT-4 evaluation [5], models can be "sycophantic," reinforcing user biases. If a user asks a leading question based on a fallacy, the model might play along. This is where the specific "Fallacy Recognition" prompting plays a vital role. It acts as a cognitive guardrail. In an enterprise setting, this might look like a "Red Teaming" layer where a second, independent LLM reviews the output of the first LLM specifically to check for logical soundness before the answer is presented to the human user.

### 4.5 The "Big Data" Paradox and Token Economics

Gandomi and Haider [10] discuss the sheer volume of Big Data. LLMs have a limitation: the context window (the amount of text they can process at once). While GPT-4 [5] has a large context window, it is not infinite, and processing millions of tokens is expensive and slow.

Therefore, the future of KM is not feeding all data into an LLM. It requires a hybrid architecture. Structured data systems (SQL, Vector Databases) must act as the long-term memory, retrieving only the most relevant "chunks" of information. The LLM then acts as the working memory or the processing unit. This highlights the importance of "Active Prompting" [4]—determining which data is most relevant to the query to maximize the utility of the limited context window.

This hybrid approach also solves the latency issue. By using lighter, faster models like LLaMA [6] for initial data sifting and reserving heavier models like GPT-4 for the final high-level reasoning, organizations can balance cost, speed, and intelligence.

## CONCLUSION

This study establishes that the integration of Large Language Models into enterprise Knowledge Management is not merely a technological upgrade, but a fundamental restructuring of how organizations process information. By moving beyond simple query-response interactions and adopting Chain-of-Thought prompting [1] and Active Prompting [4], organizations can unlock the reasoning capabilities of models like GPT-4 and LLaMA, transforming them into sophisticated decision support partners.

Our analysis confirms that while Zero-Shot approaches [2] are impressive, they are insufficient for the nuance and accuracy required in a business context. The "step-by-step" reasoning enforced by CoT, combined with explicit fallacy recognition [3], significantly reduces the risks of hallucination and logical error.

Furthermore, we demonstrated that these cognitive advances can co-exist with the rigid security requirements of the corporate world. By embedding Role-Based Access Control [7] into the data retrieval pipeline, organizations can ensure that the democratization of intelligence does not come at the cost of data sovereignty.

As we look to the future, the boundary between "User" and "System" will continue to blur. The "Learning Organization" [9] of tomorrow will be a hybrid entity, where human intuition and machine reasoning are inextricably linked, processing the torrent of Big Data [10] to navigate an increasingly complex competitive landscape. Future research should focus on the development of autonomous agents that can not only answer questions but autonomously identify the questions that need asking, further closing the loop between data and action.

## REFERENCES

1. Wang, B.; Deng, X.; Sun, H. Iteratively Prompt Pre-trained Language Models for Chain of Thought. arXiv 2022, arXiv:2203.08383.
2. Kojima, T.; Gu, S.S.; Reid, M.; Matsuo, Y.; Iwasawa, Y. Large Language Models are Zero-Shot Reasoners. arXiv 2023, arXiv:2205.11916.
3. Alhindi, T.; Chakrabarty, T.; Musi, E.; Muresan, S. Multitask Instruction-based Prompting for Fallacy Recognition. arXiv 2023, arXiv:2301.09992.
4. Diao, S.; Wang, P.; Lin, Y.; Pan, R.; Liu, X.; Zhang, T. Active Prompting with Chain-of-Thought for Large Language Models. arXiv 2024, arXiv:2302.12246.
5. OpenAI. GPT-4 Technical Report. arXiv 2023, arXiv:2303.08774.
6. Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Roziere, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. LLaMA: Open and Efficient Foundation Language Models. arXiv 2023, arXiv:2302.13971.
7. Ferraiolo, D.F., Sandhu, R., Gavrila, S., Kuhn, D.R. and Chandramouli, R., (2001). Proposed NIST standard for rolebased access control. ACM Transactions on Information and System Security (TISSEC), 4(3), pp.224-274.
8. Dip Bharatbhai Patel. (2025). Leveraging BI for Competitive Advantage: Case Studies from Tech Giants. Frontiers in Emerging Engineering & Technologies, 2(04), 15–21.
9. Gagnon, M.P., Payne-Gagnon, J., Fortin, J.P., Paré, G., Côté, J. and Courcy, F., (2015). A learning

organization in the service of knowledge management among nurses: A case study. International Journal of Information Management, 35(5), pp.636-642.

**10.** Gandomi, A. and Haider, M., (2015). Beyond the hype: Big data concepts, methods, and analytics. International journal of information management, 35(2), pp.137-144.