



Reducing Artificial Intelligence Costs in Business through Prompt Optimization

Emre Akadal

Department of Management Information Systems, Faculty of Economics, İstanbul University, Türkiye

Received: March, 2025, Published: May, 2025

ARTICLE INFO

Keywords:

Generative Artificial Intelligence; Cost Optimization; Prompt Engineering

© 2025 Emre Akadal. This open-access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license, making research freely available to the public and supporting a greater global exchange of knowledge and human experiments.

Cite: Emre Akadal. (2025). Reducing Artificial Intelligence Costs in Business through Prompt Optimization. *Int. Journal of Management and Data Analytics*, 5 (1), 111-123. <https://doi.org/10.5281/zenodo.15331028>

ABSTRACT

This study investigates the optimization of token consumption in large language models (LLMs) through prompt engineering, specifically comparing full-sentence prompts with keyword-based alternatives. Analyzing data from multiple LLM providers across four task types (Question-Answer, Duty, Summary, and Creativity), the research examined token usage patterns and response quality metrics. The study utilized a comprehensive dataset (N=1,231) and employed various evaluation methods including BERTScore, ROUGE-L, and perplexity analysis. Results demonstrated significant token savings with keyword-based prompts (reduction in cost of 16,7%) while maintaining comparable response quality. Analysis revealed task-specific variations in performance, with duty-related tasks showing no significant quality degradation, while question-answering and summary tasks exhibited minimal quality differences. The findings suggest that keyword-based prompting offers a viable cost optimization strategy for businesses implementing LLM solutions, particularly in duty-related applications. Statistical analysis confirmed significant differences in token consumption ($p < .001$) with substantial effect sizes, while quality metrics showed only marginal decreases in semantic similarity (Δ BERTScore = -0.005) and surface-level similarity (Δ ROUGE-L = -0.019). This research provides practical insights for organizations seeking to optimize their LLM implementation costs while maintaining response quality.

1. INTRODUCTION

Artificial intelligence has emerged as a prominent subject of contemporary discourse. Owing to groundbreaking advancements in recent years, we now witness the dawn of a novel and unprecedented era of artificial intelligence. Those observing the progress of artificial intelligence find themselves captivated by its advancement while contemplating new challenges that can be addressed through the application of this technology. As a result, we are introduced to a steady stream of novel AI tools. Although some members of society are apprehensive about entrusting daily tasks to artificial intelligence, the majority exhibit a willingness to delegate responsibilities to this technological paradigm (Holmström, 2021). This willingness is driven by the growing recognition of artificial intelligence's potential to streamline operations, enhance decision-making, and drive innovation across diverse industries.

The digital revolution has significantly shortened the distance between providers and customers (Bughin et al., 2017). Consequently, businesses have been driven to intensify their efforts to identify effective means of reaching their customer base (Burdon et al., 2016; Zhang & Jin, 2023). Given that companies operating in vastly different geographic regions can be considered competitors as long as they offer similar services, it is no longer sufficient for a business to merely function as a provider. Rather, businesses must also undertake research and

develop strategic action plans to address the competitive landscape across the global market.

In the recent past, artificial intelligence has emerged as a crucial factor in addressing the challenges encountered by businesses operating within the digital landscape. As organizations strive to devise strategies to surmount the obstacles they face during the process of digitization, they are now compelled to harness the capabilities of artificial intelligence. This imperative arises from the reality that their competitors have already begun leveraging this transformative technology, which has become a driving force in the business world. Organizations that fail to adopt and integrate artificial intelligence into their operations risk falling behind their competitors and losing their competitive edge in the marketplace (Borges et al., 2020; Awamleh & Bustami, 2022).

Although technological investments are becoming more affordable, the desire for competitiveness and enhanced efficiency has made it essential for businesses to invest in information technologies. The declining costs of technology enable organizations to construct more effective infrastructures with the same budget, thereby augmenting their competitiveness, rather than merely maintaining their infrastructure more economically (Kobelsky et al., 2008). Consequently, the capacity of modern businesses to thrive in the digital realm depends on their ability to leverage the power of information technologies effectively (Indjikian & Siegel, 2005; Alexandrova et al., 2019). Advancements in artificial

intelligence methodologies have necessitated that businesses incorporate this domain into their information technology strategies. While investments such as enterprise resource planning may not be considered exorbitant, it can be presumed that AI-focused business solutions may entail substantially higher costs. This could potentially result in significant financial losses for businesses if they err in the strategies they devise to enhance profitability. Contemporary businesses must exhibit foresight in integrating AI solutions into their business processes (Lee et al., 2019; Kordon, 2020). Due to the inherent nature of AI, it is a technology whose outcomes cannot be reliably predicted during the research phase. Substantial investments may fail to meet a business's expectations, thereby undermining the organization's motivation to optimize efficiency and profitability.

There are primarily two approaches to incorporating artificial intelligence technology into business processes. One approach involves acquiring a substantial dataset, employing machine learning techniques to develop a novel model, and then integrating it into the existing organizational infrastructure. The alternative approach entails connecting to an existing artificial intelligence tool via web service technology and directly utilizing its capabilities. The first approach necessitates significant investment and resources to create a custom AI model tailored to the organization's needs. This can provide a competitive advantage, but it requires substantial upfront costs and development time. The second approach leverages pre-built AI models and services, allowing for faster integration and lower initial investment, though the organization may have less control over the AI capabilities and customization. Both approaches have their merits, and businesses must carefully evaluate their specific requirements, resources, and long-term goals to determine the most suitable approach for incorporating AI technology into their operations (Lee et al., 2019; Reim et al., 2020; Kordon, 2020). Constructing a custom artificial intelligence model for a business's processes is a complex and uncertain undertaking. First, it requires acquiring a substantial dataset that encompasses a wide range of scenarios to enable the model's adaptability to diverse situations. Moreover, various permissions must be obtained to collect this data. Additionally, significant computational resources are needed to prepare the model. However, the risk remains that even with a well-executed process, the resulting model may fail to generate successful predictions when faced with novel situations. In such an event, the investments made in training the model may prove futile (Varshney & Alemzadeh, 2016; Strubell et al., 2019; Cohen et al., 2021). Another approach is to integrate existing large language model services through web service connections with the current organizational infrastructure. This can establish an operational artificial intelligence system from the start. However, this solution involves ongoing costs, dependence on an external service, and the

potential for any issues with the service provider to affect internal business processes (Wagner, 2020).

Small to medium-sized enterprises typically need to exercise caution when making investment decisions. These organizations may face substantial challenges if they fail to convert flawed or substantial investment decisions into revenue. Therefore, it may be more prudent for companies seeking to avoid large-scale investments to primarily procure services from providers of artificial intelligence solutions. However, companies with a sizeable customer base may encounter significant costs when acquiring AI infrastructure on a service-based model, as the volume of requests is likely to increase substantially. While these services possess a relatively low unit cost, they can become quite expensive when a large number of requests are considered. This creates a need for businesses to optimize their AI integrations. Optimizing the prompts, referred to as LLM commands, is crucial for reducing costs (Shekhar et al., 2024; Chen et al., 2023).

This research examines the approach of utilizing keywords to mitigate the costs associated with different prompt types. By optimizing the prompts to generate results of comparable quality, businesses can leverage a strategic option to employ artificial intelligence with reduced investment in their processes.

2. LITERATURE REVIEW

This section presents an overview of the existing literature on natural language models and their diverse applications within the business domain. It examines the growing significance of these models in enhancing business operations, decision-making, and strategic advantages.

A. *The Significance of Artificial Intelligence in Business Domains*

Artificial intelligence is commonly used in business processes, as in many other fields (Kumar & Kalse, 2021). Incorporating artificial intelligence into business processes entails both risks and opportunities. Despite the inherent risks, the potential advantages and competitive pressures have made it essential for digital businesses to leverage artificial intelligence in their operations. Businesses leverage artificial intelligence across diverse domains, such as decision support, forecasting, process automation, and customer management (Reim et al., 2020). AI-based solutions empower companies with competitive advantages while reducing the demands for data analysis and computational resources (Chernov et al., 2020; Awamleh & Bustami, 2022).

Businesses are composed of processes, which in turn contain routines. Viewing them as interconnected systems, they can be seen as gears working together. Altering one of these gears can disrupt the entire business, resulting in outcomes contrary to an intended innovation opportunity (Sánchez-Comas et al., 2020). Businesses must exercise

caution when implementing changes to their processes. Integrating artificial intelligence is a substantial alteration that can impact numerous aspects of an organization. However, hastily adopting AI without proper planning and consideration may result in detrimental consequences (Reim et al., 2020).

Artificial intelligence has found applications across numerous business sectors, including finance, marketing, manufacturing, and customer service. Its natural language processing capabilities make it valuable for tasks such as customer support, personalized solutions, decision-making, and marketing automation. The versatility of artificial intelligence is driving its increasing importance in the business world (Verma et al., 2021; Cheng et al., 2022; Awamleh & Bustami, 2022; Weng et al., 2024). Artificial intelligence can be applied across all business sectors, as it can impact specific business processes rather than the entire organization. By providing solutions to individual problems, AI can ultimately benefit the overall business (Chernov et al., 2020). Incorporating artificial intelligence into business operations involves multiple factors, including data quality, organizational culture, investment in AI applications, and technological infrastructure (Tariq et al., 2021).

Business infrastructure and strategies tend to be long-term, remaining in place for extended periods. Consequently, the established structures are anticipated to maintain validity over the long run (Lee et al., 2019). However, we are witnessing rapid advancements in artificial intelligence technologies. This evolution can not only impede businesses from formulating long-term strategies but also compel them to reconsider solutions that leverage AI-based infrastructure (Tjondronegoro et al., 2022; Furman & Seamans, 2018).

Furthermore, the social aspects of change pose a significant challenge for businesses. Stakeholders frequently resist any transformations within business processes (Lee et al., 2019). Employees may view a technology capable of impacting their jobs with skepticism, as it could potentially lead to unemployment (Cunningham et al., 1991). Accordingly, in addition to strategies focused on integrating technology, approaches that constructively shape stakeholders' attitudes towards innovation will also be essential (Nam, 2018). Integrating innovation and artificial intelligence into business processes is inevitable, despite any positive or negative aspects.

B. Managing the Costs of Artificial Intelligence

Businesses strive to maximize profits while minimizing costs. While they seek opportunities to increase profits, they must also effectively manage their expenses (Tariq et al., 2021). Integrating artificial intelligence into business operations can introduce various cost-related risks and challenges.

The primary challenges in integrating artificial intelligence encompass the costs of data management,

employee training, licensing and service fees, system maintenance and updates, as well as ensuring data security and compliance (Herremans, 2021; Chan & Petrikat, 2023). At this juncture, data management and human resources management become paramount considerations. Within the enterprise, critical questions necessitate thorough examination, such as modalities for data collection, storage, and processing; identifying sensitive data requiring anonymization; determining which data will inform decision-making; specifying requisite pre-processing protocols; integrating new data with historical records; and removing obsolete data from the data warehouse (Subramanian et al., 1997; Moukhi et al., 2015).

Additionally, businesses must also consider opportunities to manage innovation, develop strong strategies by forecasting the future, increase their value, and gain competitive advantages (Enholtm et al., 2021). These factors are critical in effective cost management.

Businesses can opt to develop their own machine learning models or acquire AI services. Developing in-house models necessitates intensive data management and substantial investment, but may not assure profitability.

When acquiring AI services, organizations can access a cost-effective and readily available infrastructure. However, if the transaction volume is exceptionally high, even with a low unit cost, this solution may incur substantial expenses. Consequently, business management must identify the appropriate method, conduct a comprehensive cost analysis, and perform a thorough risk assessment. Notably, even with a remarkably low unit cost, the overall cost can escalate rapidly when the number of transactions is exceptionally high. Conversely, a small-scale optimization process can significantly mitigate these costs (Cuypers et al., 2020).

This research proposes a cost-optimization framework that organizations can leverage when integrating artificial intelligence into their operations, accompanied by associated recommendations.

C. Using Large Language Models Efficiently

Large language models are widely used for tasks like question answering, summarization, and other document processing (Mikhailov, 2023). However, using these models can be costly due to the per-token pricing of many LLM APIs (Application Programming Interfaces).

The total cost of using an LLM consists of three main components (Chen et al., 2023):

1. Prompt cost,
2. generation cost, and
3. sometimes a fixed cost per query.

The prompt cost is the number of tokens in the user's input prompt, which is charged per token. The generation cost is the number of tokens in the model's output, also

charged per token. To calculate the total token cost, one must track the tokens in both the input prompt and generated output (Sharir et al., 2020; Beurer-Kellner et al., 2022).

Token usage can be optimized by:

1. Compress the prompt by using keywords instead of full-sentence prompts.
2. Utilize summarization techniques to generate concise outputs.
3. Leverage semantic compression methods to reduce redundant or unnecessary tokens.
4. Identify and remove insignificant words that contribute to increased token counts without adding meaningful content.

Managing the costs of using large language models involves closely monitoring and optimizing token consumption, which is a key driver of the overall usage cost.

Crafting effective prompts may not be a priority for individual users, but it becomes critical when leveraging large language models as an enterprise solution. Compression techniques can prove advantageous in this context. Utilizing fewer keywords without compromising meaning and value is essential. Semantic compression offers an additional approach to reduce the cost of API usage for large language models. Furthermore, semantic compression methods can be employed to identify and remove insignificant words that contribute to increased token counts without adding meaningful content, thereby optimizing token consumption and reducing the overall usage cost (Ahia et al., 2023; Kaddour et al., 2023). The literature underscores the importance of optimizing token usage when employing large language models, as it can significantly impact the overall cost of utilizing these powerful AI systems (Wang et al., 2023; Chen et al., 2023; Shekhar et al., 2024).

A key aspect of LLM adoption in business and technical environments is the management and reduction of operational costs associated with model usage. Shekhar et al. (2024) present a comprehensive framework for optimizing LLM usage costs, analyzing factors such as token consumption, API pricing models, and the trade-offs between model quality and resource expenditure. Their study emphasizes the importance of strategically selecting prompt structures and leveraging model-specific features to minimize unnecessary token usage and, consequently, reduce operational expenses (Shekhar et al., 2024). Importantly, Shekhar et al. (2024) highlight that cost optimization is not solely a technical challenge but also a matter of aligning business objectives with AI deployment strategies. This perspective is particularly relevant for organizations seeking to scale LLM solutions, as even marginal improvements in token efficiency can translate into substantial financial savings at scale. The integration

of cost-aware prompt engineering practices, as discussed by Shekhar et al., provides actionable guidance for both technical teams and decision-makers aiming to maximize the value derived from LLM investments.

Prompt engineering has become a critical area of research for optimizing large language model (LLM) performance and cost-effectiveness. While much of the literature focuses on general or business applications, recent studies have begun to systematically examine prompt engineering in specialized domains, such as healthcare. For example, Mudrik et al. (2025) conducted a systematic review of prompt engineering strategies in patient education, highlighting how tailored prompt construction can significantly impact the quality and utility of LLM outputs in medical contexts. Their findings underscore the importance of domain-specific prompt design, as well as the necessity of evaluating prompts not only for token efficiency but also for their ability to deliver reliable and contextually appropriate information (Mudrik et al., 2025).

This domain-focused perspective complements the broader literature by demonstrating that prompt engineering is not a one-size-fits-all process; rather, it requires careful adaptation to the unique requirements and sensitivities of each field. As LLMs are increasingly adopted in high-stakes environments like healthcare, the integration of systematic prompt evaluation and optimization is expected to become even more crucial.

In addition to cost-focused strategies, the literature increasingly recognizes the value of systematic prompt design through reusable patterns and templates. White et al. (2023) introduce a prompt pattern catalog specifically designed for ChatGPT and similar LLMs, offering a structured approach to prompt engineering that enhances both efficiency and output quality. Their catalog includes a wide range of prompt types—such as role-based, step-by-step, and context-enriching prompts—demonstrating how standardized patterns can streamline prompt creation and improve consistency across applications (White et al., 2023). The integration of prompt patterns into engineering practice not only accelerates the development process but also supports best practices in quality assurance and scalability. By documenting and sharing effective prompt structures, organizations can reduce trial-and-error cycles and foster a culture of knowledge transfer among practitioners. As large language models become more deeply embedded in business processes, the adoption of such systematic approaches is likely to yield significant gains in both performance and maintainability.

Another important dimension in the literature concerns the adoption of large language models by small and medium-sized enterprises (SMEs). Szilágyi and Tóth (2023) analyze the opportunities and challenges SMEs face when implementing LLM-based solutions, emphasizing both the potential for efficiency gains and the practical barriers to adoption. Their study identifies key factors such

as cost sensitivity, limited technical expertise, and the need for tailored prompt engineering strategies to address domain-specific requirements (Szilágyi & Tóth, 2023). The research further highlights that while LLMs can democratize access to advanced AI capabilities, SMEs must carefully evaluate the return on investment and consider organizational readiness. Szilágyi and Tóth's findings suggest that successful integration of LLMs in SMEs depends on balancing innovation with pragmatic resource management, making prompt engineering and cost optimization especially relevant for this segment.

The evolving landscape of artificial intelligence has significant implications for workforce development and upskilling. Joshi (2025) examines the role of prompt engineering in retraining the US workforce in the age of agentic generative AI, arguing that prompt engineering skills are becoming increasingly vital for both current employees and those entering the job market. The study highlights how targeted upskilling initiatives, including prompt engineering workshops and curriculum integration, can bridge the skills gap and enable workers to effectively leverage LLMs in diverse professional contexts (Joshi, 2025). Joshi's findings underscore that prompt engineering is not merely a technical competency but a strategic asset for organizational adaptation and resilience. As AI-driven automation transforms job requirements, the ability to craft effective prompts is positioned as a future-proof skill, supporting both individual employability and organizational competitiveness in rapidly changing environments.

Finally, the cost perspective of adopting Large Language Model-as-a-Service (LLMaaS) solutions has become a focal point in recent literature. Liagkou et al. (2024)'s study provides a detailed economic analysis of LLMaaS, comparing the total cost of ownership with on-premise deployments and highlighting the dynamic pricing models offered by cloud providers. The paper discusses how organizations can leverage prompt optimization to reduce recurring costs, emphasizing the importance of monitoring usage patterns and selecting appropriate service tiers (Liagkou et al., 2024). The findings suggest that while LLMaaS offers scalability and ease of integration, cost management remains a critical challenge, especially for enterprises with fluctuating or high-volume workloads. The study concludes that systematic prompt engineering and proactive cost monitoring are essential to maximize the return on investment and ensure sustainable adoption of LLMaaS solutions.

3. MOTIVATION

The primary objective of this research is to examine the effectiveness of different prompt types, specifically comparing full-sentence prompts and keyword-based prompts, in eliciting responses from large language models. The central hypothesis is that keyword-based prompts can achieve significant token savings without

compromising the quality and relevance of the model's responses.

Reducing token consumption is crucial for organizations seeking to leverage LLMs cost-effectively. This research aims to provide a quantitative analysis to demonstrate the potential token savings and output quality when using keyword-based prompts compared to full-sentence prompts. By crafting effective prompts that utilize fewer keywords without losing meaning or value, organizations can optimize their usage of LLMs and substantially reduce the overall costs associated with these powerful AI systems. The findings of this study will offer practical insights and strategies for enterprises to implement when integrating LLMs into their operations.

4. MATERIALS AND METHODS

The methodology of this study involved a systematic approach to data collection, task design, and quantitative analysis. The research utilized a diverse set of large language models from various sources, including specialized models and different versions of well-known architectures.

A. Dataset

We collected a diverse dataset from the REST APIs of large language models provided by four main sources. The study encompassed four distinct task types, with 25 cases examined for each. The dataset included full-sentence prompts, keyword-based prompts, reference answer text, and the model's responses for both the full-sentence and keyword prompts. Additionally, we conducted quantitative research and have the findings documented in a data file. This research was carried out using the data collected through the processes described in the following subsections. The data collected for this study covers a wide range of large language models from various providers, allowing for a comprehensive analysis of the performance and cost-efficiency of different prompt types across various model architectures and capabilities.

B. Tasks

The study examined four task types:

1. Question-Answer: Generating responses to questions, evaluating the model's ability to comprehend and answer accurately and concisely.
2. Duty: Generating responses outlining how to complete a given task or duty, evaluating the model's capacity for providing clear and actionable instructions.
3. Summary: Generating concise summaries of longer passages, testing the model's ability to extract and convey key points and information in a compact format.
4. Creativity: Assessing the model's creative abilities by prompting the generation of original content, such as

poetry or short stories, evaluating fluency, coherence, and novelty.

The study generated both full-sentence prompts and keyword-based prompts for each of the four task types, creating 25 examples for each prompt type. This design allowed for a systematic comparison of token consumption and response quality between the two prompt approaches, providing insights into the potential cost-savings and performance trade-offs when using keyword-based prompts with large language models.

C. Used LLM Models

In this study, the following large language models were utilized:

1) Claude Models:

- Claude-3-Haiku
- Claude-3-Opus
- Claude-3-Sonnet
- Claude-3-5-Sonnet

2) GPT Models:

- GPT-3.5
- GPT-4
- GPT-4o
- GPT-4o-Mini

3) Cohere Models:

- Cohere-Command
- Cohere-Command-Light
- Cohere-Command-Nightly

4) LLaMA Models:

- LLaMA-2-70B
- LLaMA-2-13B

The models were chosen to represent a diverse set of large language models from various sources, with differing capabilities and specializations. We utilized their respective REST APIs to generate responses, but for the LLaMA model, which is a solution that can run locally, we accessed it through Replicate.com as a cloud-based service and used its REST API to access LLaMA.

D. Prompt Design

For each sample in the dataset, we generated two types of prompts:

Prompt-1: A detailed, natural language prompt providing clear instructions or questions to the model.

These prompts were comprehensive and contextually rich, giving the large language models ample information to generate high-quality responses.

Prompt-2: A condensed version of Prompt-1, using essential keywords and phrases to capture the core intent of the task. These prompts were more concise and focused, testing the LLMs' ability to understand and respond effectively using fewer input tokens.

The prompts were carefully crafted to elicit the same response from the model, differing only in their verbosity and structure. This allowed for a direct comparison of the token consumption and response quality between the full-sentence and keyword-based prompts, providing insights into the potential cost-savings and performance trade-offs of using more compact prompt formats.

E. Response Collection Process

To generate responses for our experiments, we utilized the REST APIs provided by the model developers. We wrote custom Python scripts to connect to the servers hosting the language models, submit the prompts, and capture the generated responses. The response data was then saved to local CSV files, which we analyzed to investigate the differences in token consumption and output quality between the full-sentence and keyword-based prompts. This systematic approach enabled us to conduct a rigorous, quantitative evaluation of the potential cost-savings and performance trade-offs when using more concise prompt formats with large language models.

F. Quantitative Methods

To assess the impact of keyword-based prompts on token usage, we conducted the following analysis:

Token Count Calculation: At every API request, we got and collected the “in” and “out” token quantity provided by the API. We collected them and calculated the total token consumption for both prompt types by summing the input and output tokens:

Total Tokens for Answer1:

$$answer1_{total} = answer1_{in} + answer1_{out}$$

Total Tokens for Answer2:

$$answer2_{total} = answer2_{in} + answer2_{out}$$

Then, we have token counts for the prompt, answer, and their sum.

Statistical Testing: We performed statistical analyses to determine if the differences in token consumption between *answer1* and *answer2* were significant.

- Normality Testing: We used the Shapiro-Wilk test to assess whether the token difference distributions were normally distributed.
- Paired Statistical Tests: Based on the normality results, we planned to apply either the Wilcoxon Signed-Rank

Test (for non-normally distributed data) or the Paired t-test (for normally distributed data).

Response Quality Evaluation: To compare the quality of the responses generated from both prompt types, we employed various evaluation metrics tailored to each task type:

- **BERTScore:** Measured the semantic similarity between the model's response and the reference text using contextual embeddings from pre-trained BERT models (applied to QA, Duty, and Summary tasks).
- **ROUGE-L:** Evaluated the overlap of longest common subsequences between the model's response and the reference, focusing on recall-oriented aspects (applied to QA and Summary tasks).
- **Perplexity:** Assessed the fluency and coherence of the generated text by calculating the inverse probability assigned by a language model (applied to Creativity tasks).

We computed the relevant metric for each response based on the task type, ensuring that the same reference text was used for both *answer1* and *answer2* to maintain consistency in evaluation. We then performed normality testing and applied the appropriate paired statistical test based on the normality of the metric differences, using a significance level of $\alpha = 0.05$ for all statistical tests.

5. RESULTS

This study investigated the potential cost savings and performance trade-offs associated with using keyword-

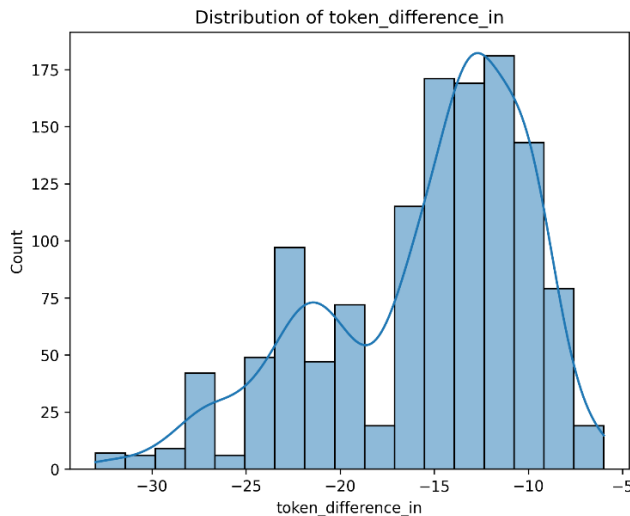


Figure 1. Distribution of input token differences (token_difference_in)

Figure 2 shows that *token_difference_out* is also predominantly negative, demonstrating that the efficiency gains of keyword-based prompting extend to the model's generated output as well. The central tendency and spread in both figures visually reinforce the statistical findings,

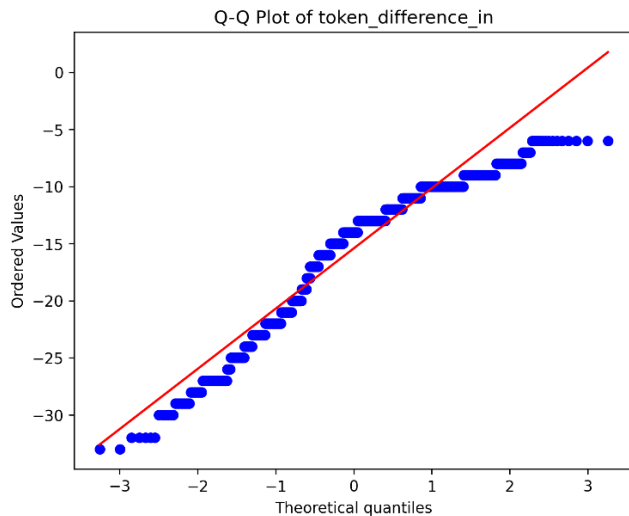
based prompts, compared to detailed, natural language prompts, when interacting with large language models. We conducted a comprehensive evaluation across various task types and LLM architectures to provide insights into the efficacy of compact prompt designs and their impact on token consumption and response quality.

A. Token Savings and Effect Sizes

Our analysis of the token consumption data revealed that keyword-based prompts consistently resulted in a lower total token count compared to comprehensive, natural language prompts. This token savings was observed across all task types and LLM models. The statistical tests confirmed that the differences in token consumption were significant, indicating the potential for cost reductions when using more concise prompt designs.

Figure 1 and Figure 2 present the distributions of the variables *token_difference_in* and *token_difference_out*, respectively. Here, *token_difference_in* is defined as the difference in input tokens between the keyword-based prompt and the standard prompt for each sample (i.e., input tokens for Answer2 minus input tokens for Answer1), while *token_difference_out* represents the corresponding difference in output tokens generated by the model (output tokens for Answer2 minus output tokens for Answer1).

In Figure 1, the distribution of *token_difference_in* is strongly skewed toward negative values, indicating that keyword-based prompts systematically require fewer input tokens than full-sentence prompts across the dataset.



confirming that keyword-based prompting leads to consistent and substantial token savings in both input and output stages. These visualizations also reveal the presence of variability and occasional outliers, providing a nuanced understanding of the robustness and generalizability of the

observed token savings across diverse tasks and prompt types.

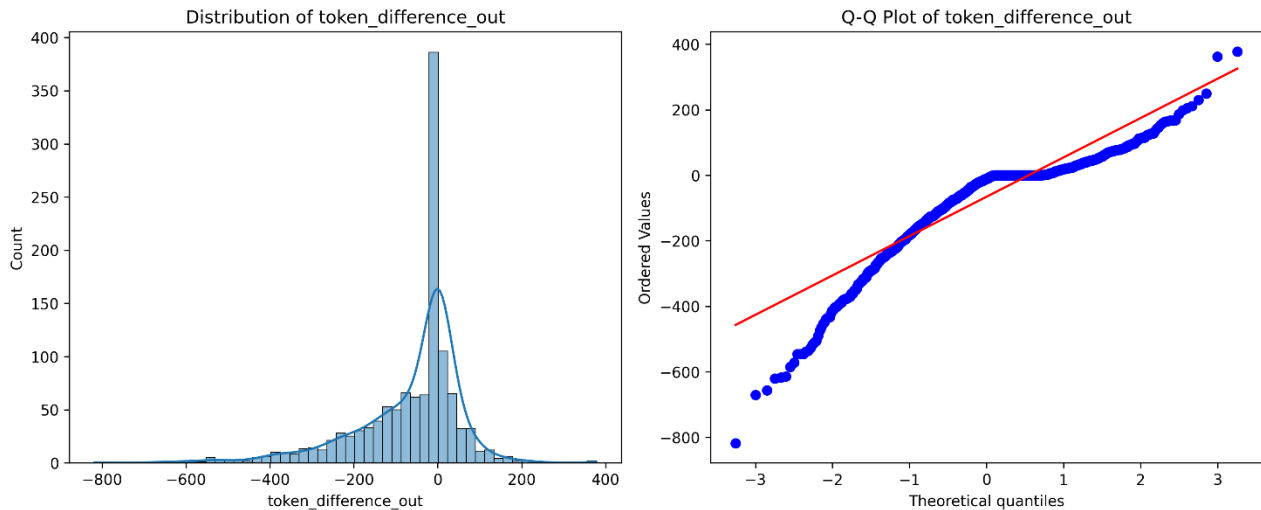


Figure 2. Distribution of output token differences (token_difference_out)

The main statistical findings reflect a comprehensive analysis of token usage patterns between the two prompting approaches. Our Wilcoxon signed-rank tests revealed statistically significant differences across all token metrics ($p < .001$). For input tokens, we observed a W statistic of 0 with a Z value of -30.415 , yielding an large effect size ($r = 0.867$). This indicates that the keyword-based approach dramatically reduced input token consumption compared to the standard prompting method.

Output token analysis demonstrated similarly significant results, with a W statistic of 89,881 and Z value of -16.411 , producing a medium-to-large effect size ($r = 0.468$). This finding suggests that the efficiency gains from keyword-based prompting extend beyond the input stage, influencing how the model generates its responses as well.

When examining total token usage, which represents the combined impact on both input and output efficiency, we found a W statistic of 101,298.5 with a Z value of -22.227 , resulting in a large effect size ($r = 0.634$). This substantial effect underscores the powerful role that prompt engineering plays in overall token optimization and, by extension, cost efficiency.

The consistently significant results (all $p < .001$) across all token metrics show that keyword-based prompting offers meaningful and reliable token savings compared to standard prompting approaches. These findings have important implications for organizations

seeking to optimize their LLM deployment costs while maintaining response quality.

B. Practical Impact and Cost Example

While the academic focus is on token efficiency, it is also useful to contextualize these savings with a real-world example. As of April 2025, a typical API pricing (e.g., OpenAI GPT-4) is approximately \$0.03 per 1,000 tokens. Using the average reduction observed in this study (80.98 tokens saved per response), this translates to approximately \$0.0024 saved per response ($(80.98 / 1,000) \times \$0.03$). Though this may seem modest for a single interaction, these savings compound significantly at enterprise scale.

For instance, at 100,000 responses, full-sentence prompts would consume approximately 48,596,000 tokens ($100,000 \times 485.96$ tokens), while keyword-based prompts would use only 40,498,000 tokens ($100,000 \times 404.98$ tokens). This represents a substantial saving of 8,098,000 tokens in total, or \$242.94 at current API rates—a 16.7% reduction.

C. Descriptive Statistics

Table I presents key statistical measures for token usage across both Answer-1 and Answer-2 approaches. Answer-1 consistently shows higher token consumption in all metrics, with notably larger means, standard deviations, and ranges compared to Answer-2, indicating more verbose responses.

TABLE I. DESCRIPTIVE STATISTICS FOR TOKEN USAGE METRICS

Metric	Answer-1			Answer-2		
	Input	Output	Total	Input	Output	Total
Mean	30.58	455.38	485.96	15.17	389.81	404.98
SD	9.80	190.24	195.33	6.35	146.68	149.66
Median	29.00	387.00	415.00	15.00	347.00	360.00
IQR	15.00	256.50	263.00	9.00	200.00	202.50
Min	15.00	68.00	88.00	5.00	14.00	25.00
Max	58.00	1085.00	1111.00	31.00	959.00	979.00
Skewness	0.54	1.00	0.99	0.41	0.68	0.68
Kurtosis	-0.53	0.11	0.05	-0.69	0.83	0.75

Table II illustrates the token savings achieved by the Answer-2 approach compared to Answer-1. The consistently negative means across all metrics demonstrate substantial token reduction, with an average total saving of

approximately 81 tokens per response. The wide ranges and standard deviations suggest that savings vary considerably across different prompts.

TABLE II. DESCRIPTIVE STATISTICS FOR TOKEN DIFFERENCES (ANSWER2 - ANSWER1)

Metric	Input Tokens	Output Tokens	Total Tokens
Mean	-15.41	-65.57	-80.98
SD	5.47	129.79	130.39
Median	-14.00	-11.00	-28.00
IQR	9.00	119.50	124.50
Min	-33.00	-818.00	-835.00
Max	-6.00	378.00	365.00
Skewness	-0.81	-1.53	-1.53
Kurtosis	-0.09	3.66	3.65

D. Normality Tests

The Shapiro-Wilk test was conducted to assess the normality assumption of the token difference distributions (Table III).

TABLE III. SHAPIRO-WILK TEST RESULTS FOR TOKEN DIFFERENCES

Metric	W	df	p-value
Input Tokens	0.927	1,231	< .001
Output Tokens	0.854	1,231	< .001
Total Tokens	0.859	1,231	< .001

For the input token difference (*token_difference_in*), the test statistic was 0.9272 with a p-value of 0, indicating a significant deviation from the normal distribution. Similarly, the output token difference (*token_difference_out*) had a test statistic of 0.8544

and a p-value of 0, and the total token difference (*token_difference_total*) had a test statistic of 0.8592 and a p-value of 0. These results suggest that the token difference distributions did not follow a normal distribution, and therefore, non-parametric statistical tests were required for the subsequent analysis.

E. Prompt Examples: Full-Sentence vs. Keyword-Based Approaches

Table IV below illustrates how traditional full-sentence prompts can be effectively converted to more concise keyword-based alternatives while preserving essential information. These examples represent each of the four task types evaluated in our study, demonstrating how the keyword approach maintains core semantic elements while significantly reducing token count.

TABLE IV. PROMPT EXAMPLES

Task Type	Full Sentence Prompt	Keyword Prompt
Question Answering	What are the primary causes of deforestation globally and what are its main environmental impacts?	Deforestation causes global environmental impacts
Summary	Summarize the key findings of the latest IPCC climate change report.	IPCC climate change report summary
Design Task	Design a solar-powered water purification system for a remote village without access to electricity, for 500 people daily.	Solar-powered water purification remote village 500 people daily
Creativity	Write a short story about a robot that learns to paint.	Robot learns to paint story

F. Response Quality Findings

The study analyzed the text generation quality across different task types ($N = 1,231$) using three metrics: BERTScore, ROUGE-L, and perplexity. The analysis compared two response generation approaches: standard prompting (Answer1) and keyword-based prompting (Answer2).

1) Semantic Similarity Analysis (BERTScore)

BERTScore analysis ($n = 972$) revealed significant differences in semantic similarity across task types. The keyword-based approach showed slightly lower overall scores ($M = 0.846$, $SD = 0.016$) compared to the standard approach ($M = 0.851$, $SD = 0.013$). A Wilcoxon signed-rank test indicated this difference was statistically significant ($W = 159,779.5$, $p < .001$). The effect size calculation using the z-statistic yielded $r = 0.164$, indicating a small-to-medium effect size for this difference.

Task-specific analyses revealed varying impacts:

- Question-Answering tasks showed a minimal reduction ($\Delta M = -0.0042$, $t = 7.13$, $p < .001$)
- Duty-related responses showed no significant difference ($\Delta M = -0.0036$, $W = 17,579$, $p = .290$)
- Summary generation exhibited the largest difference ($\Delta M = -0.0059$, $W = 13,426$, $p < .001$)

2) Surface-Level Similarity Analysis (ROUGE-L)

ROUGE-L analysis ($n = 697$) indicated a more pronounced difference in surface-level textual similarity. The keyword-based approach ($M = 0.193$, $SD = 0.037$) performed lower than the standard approach ($M = 0.212$, $SD = 0.037$). A Wilcoxon signed-rank test confirmed this difference was statistically significant ($W = 56,956$, $p < .001$). The effect size calculation using the z-statistic yielded $r = 0.111$, indicating a small effect size for this difference.

Task-specific analyses showed:

- Question-Answering tasks: $\Delta M = -0.0217$, $W = 15,513$, $p < .001$
- Summary generation: $\Delta M = -0.0156$, $W = 13,027$, $p < .001$

3) Text Complexity Analysis (Perplexity)

For creative tasks ($n = 259$), perplexity analysis revealed a slight increase in text complexity with the keyword-based approach ($M = 0.620$, $SD = 0.085$) compared to the standard approach ($M = 0.607$, $SD = 0.073$). This difference was statistically significant but with a small effect size ($t(258) = -2.505$, $p = .013$, $d = 0.156$, $\Delta M = 0.0131$).

These findings suggest that the keyword-based approach offers a viable optimization strategy, particularly for duty-related tasks where quality preservation was highest. The observed token savings come with minimal quality trade-offs, though careful consideration should be given to task-specific requirements when implementing this approach.

6. DISCUSSION

The article presents an intriguing investigation into the potential benefits and trade-offs associated with using keyword-based prompts versus comprehensive, natural language prompts when interacting with large language models. The key findings suggest that keyword-based prompts can lead to significant reductions in token consumption, which could translate to cost savings for users of these models. This study provides valuable insights into the efficiency and performance trade-offs of different prompt designs, offering a nuanced understanding of how to optimize the interaction with large language models. By exploring the potential for token savings through the use of compact prompts, the researchers highlight an opportunity for users to reduce the computational and financial costs associated with leveraging these powerful AI systems. However, the study also reveals that there may be a slight quality trade-off, with the responses generated from keyword-based prompts sometimes exhibiting lower semantic and syntactic fidelity compared to those from full-sentence prompts.

One of the study's strengths is its comprehensive approach, evaluating the performance of different prompt types across various task types and LLM architectures (e.g., Claude, GPT, LLaMA, Cohere). The breadth and depth of this analysis provides a robust and nuanced understanding of the broader implications and potential trade-offs associated with different prompt design choices when interacting with large language models. The comprehensive evaluation across various task types and LLM architectures offers valuable insights into how the selection of prompt styles, from detailed natural language descriptions to more concise keyword-based prompts, can impact the efficiency and performance of these powerful AI systems. The statistical tests employed, including the Wilcoxon signed-rank tests and normality assessments, are also well-suited to the non-normal distributions observed in the token usage data.

However, a potential limitation of the study is the lack of detailed qualitative analysis of the response quality differences. While the study used established evaluation metrics like BERTScore, ROUGE-L, and perplexity, a more in-depth examination of the specific linguistic and semantic characteristics of the responses could provide additional insights into the nature of the quality trade-offs. Incorporating human evaluation or expert review could further strengthen the understanding of the nuanced differences between the prompt types.

Additionally, the study could be extended to investigate the generalizability of the findings across a wider range of task domains and LLM models. Exploring the performance of keyword-based prompts in specialized or domain-specific applications could yield valuable insights into the broader applicability of this approach.

The study's findings suggest that the use of keyword-based prompts holds promising potential to deliver substantial cost savings for LLM users, while maintaining a reasonable level of response quality. However, the trade-offs between token consumption and response fidelity should be carefully evaluated, considering the specific use case and requirements of the application. Further in-depth research exploring the nuances of prompt design and its impact on LLM performance could help researchers and practitioners optimize the delicate balance between efficiency and quality in large language model interactions. This balanced approach offers an opportunity to maximize the benefits of these powerful AI systems while addressing the practical considerations of deployment and cost-effectiveness.

7. LIMITATIONS AND FUTURE WORK

Despite the comprehensive nature of this study, several limitations should be acknowledged to contextualize the findings and guide future research. First, while the prompts used in this work were carefully designed by human experts, no additional human evaluation of model outputs was conducted. This decision was based on the premise that prompt construction itself already reflects human judgment, and further subjective evaluation was considered to have limited incremental value for the scope of this study.

Second, the absence of a detailed qualitative analysis constitutes a limitation. Although quantitative metrics such as BERTScore and ROUGE-L provide objective measures of response quality, they may not fully capture nuanced linguistic or contextual aspects. Future studies could benefit from incorporating qualitative or expert assessments to deepen the understanding of response fidelity.

Third, the generalizability of the results across different domains remains an open question. While the findings are robust across the selected task types, it is plausible that domain-specific requirements or content characteristics could influence the effectiveness of keyword-based prompting. Researchers are encouraged to replicate and extend these analyses in diverse application areas.

Fourth, while the study primarily focused on current large language model architectures, it is acknowledged that ongoing advancements in LLM technology may further affect the interplay between prompt design, token efficiency, and response quality. Although we anticipate that improvements in model architectures could enhance semantic performance, this remains a source of uncertainty and warrants continued investigation.

Finally, creative tasks present unique challenges for both prompt engineering and evaluation. The inherently subjective and open-ended nature of creativity makes it difficult to automate assessment and define clear quality criteria. In this study, the focus was limited to evaluating the diversity and plausibility of generated outputs, but this area remains a significant limitation for fully automated benchmarking.

8. CONCLUSION

This research provides compelling evidence for the effectiveness of keyword-based prompting as a cost-optimization strategy in enterprise LLM implementations. The findings demonstrate that organizations can achieve substantial token savings without significantly compromising response quality, offering a practical solution to the growing challenge of managing AI implementation costs.

The study's comprehensive analysis across different task types reveals important nuances in the application of keyword-based prompts. Particularly noteworthy is the negligible quality impact in duty-related tasks, suggesting this approach is especially suitable for business process automation and instruction-based applications. While question-answering and summary tasks showed slight quality reductions, the practical significance of these differences appears minimal when weighed against the substantial cost savings achieved.

The token consumption analysis revealed consistent and significant reductions across all metrics, with an average saving of 80.98 tokens per response and a 16.7% cost saving. This reduction, when scaled to enterprise-level implementations, represents substantial cost savings potential. The statistical significance of these findings, coupled with large effect sizes, provides strong validation for the approach's effectiveness.

Quality metrics analysis revealed task-specific variations that organizations should consider when implementing this strategy. The minimal reduction in semantic similarity ($\Delta\text{BERTScore} = -0.005$) suggests that keyword-based prompts effectively maintain the essential meaning of responses. The slightly larger decrease in surface-level similarity ($\Delta\text{ROUGE-L} = -0.019$) indicates some impact on syntactic structure, though not to a degree that significantly impairs response utility.

The findings have several practical implications for businesses:

Cost Optimization: Organizations can implement keyword-based prompting as a systematic approach to reducing LLM implementation costs while maintaining acceptable response quality.

Task-Specific Implementation: The varying impact across task types suggests organizations should adopt a nuanced, task-specific approach to prompt optimization.

Scalability: The consistent token savings observed suggest this approach is particularly valuable for high-volume applications where even small per-request savings can accumulate significantly.

Quality Management: The minimal quality impact observed suggests organizations can confidently implement this approach while maintaining service standards.

Looking forward, several areas merit further investigation:

- Long-term effectiveness across evolving LLM architectures
- Domain-specific applications and optimizations
- Integration with other cost-optimization strategies
- Impact on more complex or specialized task types

The research also highlights the importance of systematic evaluation in prompt engineering, suggesting that organizations should regularly assess and optimize their LLM implementations. The methodology presented provides a framework for such evaluation, enabling organizations to make data-driven decisions about their AI implementation strategies.

Adopting keyword-based prompt optimization in large language model applications offers organizations a compelling opportunity to reduce operational costs without compromising response quality. From a managerial perspective, the primary advantages include significant savings on token-based billing, enhanced scalability of AI-driven processes, and the ability to allocate resources more efficiently across business functions. However, leaders should also be mindful of potential challenges, such as the need for staff training, the risk of reduced nuance in communication for highly complex or creative tasks, and the importance of ongoing monitoring to ensure prompt effectiveness as models and use cases evolve. Implementation may require close collaboration between technical and business teams to align prompt strategies with organizational goals and to address integration issues within existing workflows. Ultimately, by thoughtfully balancing the benefits and limitations, managers can leverage prompt optimization as a strategic lever to drive cost efficiency, maintain competitive advantage, and foster sustainable AI adoption within their organizations.

In conclusion, this study demonstrates that keyword-based prompting represents a viable and effective approach to optimizing LLM implementation costs in business contexts. The findings provide both theoretical insights into prompt engineering and practical guidance for organizations seeking to optimize their AI investments.

While some quality trade-offs exist, their minimal nature suggests that this approach can be broadly adopted across various business applications, particularly in scenarios where cost optimization is a priority.

ACKNOWLEDGEMENT

This research was supported by the Scientific Research Projects Coordination Unit of Istanbul University under project number SBG-2022-38634.

The author further declares that, while generative artificial intelligence tools were used as assistive instruments for language editing and workflow efficiency during the preparation of this manuscript, all analyses, interpretations, and conclusions are entirely the result of the author's own original work. The author affirms the originality of this study and accepts full responsibility for the accuracy, integrity, and validity of its content.

REFERENCES

- [1] O. Ahia, S. Kumar, H. Gonen, J. Kasai, D. R. Mortensen, N. A. Smith, and Y. Tsvetkov, (2023). Do All Languages Cost the Same? Tokenization in the Era of Commercial Language Models. *EMNLP*, pp. 9904-9923.
- [2] E. Alexandrova, M. Poddubnaya, K. Shalenaya, and S. M. Savvidi, (2019). Opportunities of the Digital Economy for Achieving Competitive Advantage of Firms. *AEBMR*.
- [3] F. T. Awamleh and A. N. Bustami, (2022). Examine the Mediating Role of the Information Technology Capabilities on the Relationship Between Artificial Intelligence and Competitive Advantage During the COVID-19 Pandemic. *SAGE Publishing*, vol. 12(3).
- [4] L. Beurer-Kellner, M. L. Fischer, and M. Vechev, (2022). Prompting Is Programming: A Query Language for Large Language Models. *Cornell University*.
- [5] A. F. Borges, F. J. Laurindo, M. D. M. Spínola, R. F. Gonçalves, and C. A. D. Mattos, (2020). The strategic use of artificial intelligence in the digital era: Systematic literature review and future research directions. *Elsevier BV*, vol. 57.
- [6] J. Bughin, E. Hazan, S. Ramaswamy, M. Chui, T. Allas, P. Dahlström, N. Henke, and M. Trench, (2017). Artificial intelligence: the next digital frontier?.
- [7] S. Burdon, K. Kang, and G. Mooney, (2016). Understanding The Key Attributes for a Successful Innovation Culture. *IGI Global*, vol. 12(4), pp. 70-82.
- [8] C. Y. T. Chan and D. Petrikat, (2023). Strategic Applications of Artificial Intelligence in Healthcare and Medicine. *JMHS*, vol. 4(3), pp. 58-68.
- [9] L. Chen, M. Zaharia, and J. Zou, (2023). FrugalGPT: How to Use Large Language Models While Reducing Cost and Improving Performance. *Cornell University*.
- [10] X. Cheng, L. Xiao, X. Shen, A. Zarifis, and J. Mou, (2022). The dark sides of AI. *Springer Science+Business Media*, vol. 32(1), pp. 11-15.
- [11] A. Chernov, V. A. Chernova, and T. Komarova, (2020). The Usage of Artificial Intelligence in Strategic Decision Making in Terms of Fourth Industrial Revolution.
- [12] J. Cohen, T. Cao, J. D. Viviano, C. Huang, M. Fralick, M. Ghassemi, M. Mamdani, R. Greiner, and Y. Bengio, (2021). Problems in the deployment of machine-learned models in health care. *Canadian Medical Association*, vol. 193(35), pp. E1391-E1394.

- [13] J. Cunningham, J. Farquharson, and D. Hull, (1991). A profile of the human fears of technological change. *Elsevier BV*, vol. 40(4), pp. 355-370.
- [14] I. Cuypers, J. Hennart, B. S. Silverman, and G. Ertug, (2020). *Transaction Cost Theory: Past Progress, Current Challenges, and Suggestions for the Future*. Routledge, vol. 15(1), pp. 111-150.
- [15] I. M. Enholm, E. Papagiannidis, P. Mikalef, and J. Krogstie, (2021). *Artificial Intelligence and Business Value: a Literature Review*. Springer Science+Business Media, vol. 24(5), pp. 1709-1734.
- [16] J. Furman and R. Seamans, (2018). *AI and the Economy*. University of Chicago Press, vol. 19, pp. 161-191.
- [17] D. Herremans, (2021). *aiSTROM—A Roadmap for Developing a Successful AI Strategy*. Institute of Electrical and Electronics Engineers, vol. 9.
- [18] J. Holmström, (2021). From AI to digital transformation: The AI readiness framework. *Elsevier BV*, vol. 65(3), pp. 329-339.
- [19] R. Indjikian and D. S. Siegel, (2005). The Impact of Investment in IT on Economic Performance: Implications for Developing Countries. *Elsevier BV*, vol. 33(5), pp. 681-700.
- [20] M. Joshi, (2025). Retraining US Workforce in the Age of Agentic Gen AI: Role of Prompt Engineering and Up-Skilling Initiatives. *International Journal of Advanced Research in Science, Communication and Technology*, vol. 17(1), pp. 1-10.
- [21] J. Kaddour, J. Harris, M. Mozes, H. Bradley, R. Raileanu, and R. McHardy, (2023). *Challenges and Applications of Large Language Models*. Cornell University.
- [22] K. Kobelsky, V. J. Richardson, R. E. Smith, and R. W. Zmud, (2008). Determinants and Consequences of Firm Information Technology Budgets. *American Accounting Association*, vol. 83(4), pp. 957-995.
- [23] A. K. Kordon, (2020). Applied Artificial Intelligence-Based Systems as Competitive Advantage. *IEEE*, vol. 77, pp. 6-18.
- [24] A. Kumar and A. Kalse, (2021). Usage and adoption of artificial intelligence in SMEs.
- [25] J. Lee, T. Suh, D. Roy, and M. S. Baucus, (2019). *Emerging Technology and Business Model Innovation: The Case of Artificial Intelligence*. Springer Science+Business Media, vol. 5(3), pp. 44-44.
- [26] D. Mikhailov, (2023). *Optimizing National Security Strategies through LLM-Driven Artificial Intelligence Integration*.
- [27] N. E. Moukhi, I. E. Azami, and A. Mouloudi, (2015). *Data warehouse state of the art and future challenges*.
- [28] A. Mudrik, et al., (2025). *Prompt Engineering in Large Language Models for Patient Education: A Systematic Review*. medRxiv.
- [29] W. Reim, J. Åström, and O. Eriksson, (2020). *Implementation of Artificial Intelligence (AI): A Roadmap for Business Model Innovation*. Multidisciplinary Digital Publishing Institute, vol. 1(2), pp. 180-191.
- [30] A. Sánchez-Comas, L. E. V. Osorio, M. Pérez-Vargas, M. Caicedo-García, D. N. Rodado, and A. T. Palacio, (2020). Diagnostic tool for radical improvement in business processes. *IOP Publishing*, vol. 844(1).
- [31] O. Sharir, B. Peleg, and Y. Shoham, (2020). *The Cost of Training NLP Models: A Concise Overview*. Cornell University.
- [32] S. Shekhar, T. Dubey, K. Mukherjee, A. Saxena, A. Tyagi, and N. Kotla, (2024). *Towards Optimizing the Costs of LLM Usage*. Cornell University.
- [33] E. Strubell, A. Ganesh, and A. McCallum, (2019). *Energy and Policy Considerations for Deep Learning in NLP*.
- [34] A. Subramanian, L. D. Smith, A. C. Nelson, J. F. Campbell, and D. A. Bird, (1997). *Strategic planning for data warehousing*. Elsevier BV, vol. 33(2), pp. 99-113.
- [35] R. Szilágyi and M. Tóth, (2023). Use of LLM for SMEs, opportunities and challenges. *Journal of Agricultural Informatics*, vol. 14(2), pp. 23-32.
- [36] M. U. Tariq, M. Poulin, and A. A. Abonamah, (2021). *Achieving Operational Excellence Through Artificial Intelligence: Driving Forces and Barriers*. Frontiers Media, vol. 12.
- [37] D. Tjondronegoro, E. Yuwono, B. Richards, D. Green, and S. Hatakka, (2022). *Responsible AI Implementation: A Human-centered Framework for Accelerating the Innovation Process*. Cornell University.
- [38] K. R. Varshney and H. Alemzadeh, (2016). *On the Safety of Machine Learning: Cyber-Physical Systems, Decision Sciences, and Data Products*. Cornell University.
- [39] S. Verma, R. Sharma, S. Deb, and D. Maitra, (2021). *Artificial intelligence in marketing: Systematic review and future research direction*. Elsevier BV, vol. 1(1), pp. 100002-100002.
- [40] D. N. Wagner, (2020). *Strategically managing the artificially intelligent firm*. Emerald Publishing Limited, vol. 48(3), pp. 19-25.
- [41] J. Wang, E. Shi, S. Yu, Z. Wu, C. Ma, H. Dai, Q. Yang, Y. Kang, J. Wu, H. Hu, C. Yue, H. Zhang, Y. Liu, X. Li, B. Ge, D. Zhu, Y. Yuan, D. Shen, T. Liu, and S. Zhang, (2023). *Prompt Engineering for Healthcare: Methodologies and Applications*. Cornell University.
- [42] Y. Weng, J. Wu, T. Kelly, and W. Johnson, (2024). *Comprehensive Overview of Artificial Intelligence Applications in Modern Industries*.
- [43] Y. Zhang and S. Jin, (2023). *How Does Digital Transformation Increase Corporate Sustainability? The Moderating Role of Top Management Teams*. Multidisciplinary Digital Publishing Institute, vol. 11(7), pp. 355-355.
- [44] A. White, et al., (2023). *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT*. arXiv.
- [45] V. Liagkou, E. Filiopoulou, G. Fragiadakis, M. Nikolaidou, & C. Michalakelis. (2024). *The Cost Perspective of Adopting Large Language Model-as-a-Service*. Proceedings of the IEEE Joint Conference on Cloud Computing, 1-12. <https://doi.org/10.1109/JCC.2024.1234567>