NOVA IMS Information Management School

MDSAA

Master Degree Program in

Data Science and Advanced Analytics

Gen-AI and the Future of Supply Chain Management

The Impact of Large Language Models on Modern Supply Chain Management

Alex Adrián Santander Morales

Master Thesis

presented as partial requirement for obtaining a Master's Degree in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Gen-AI and the Future of Supply Chain Management

The Impact of Large Language Models on Modern Supply Chain Management

by

Alex Adrián Santander Morales

Master Thesis presented as a partial requirement for obtaining the Master's degree in Data Science and Advanced Analytics, with a specialization in Data Science

Supervised by

Fernando Bação, PhD, NOVA Information Management School

July, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

[Lisbon, July 2024]

DEDICATION

To my parents, Juan and Ana María.

To my siblings: Carolina, Cindy, Andrés, and Ariel.

To all the people who believe in me, especially Braulio.

Thank you.

ABSTRACT

LLMs like GPT-4 possess an exceptional ability to understand and generate human language, driven by advancements in Artificial Intelligence, Deep Learning, and Natural Language Processing. Companies are increasingly interested in integrating this cutting-edge technology into their Supply Chain processes to leverage its capabilities. While previous studies have explored the impact of LLMs on various Supply Chain Management functions—such as demand forecasting, supplier evaluation, simulation, and optimization—no comprehensive survey has yet consolidated these functions into a single research effort.

This thesis provides a thorough review of the impact of LLMs on Supply Chain Management, focusing on four key dimensions: demand forecasting, supplier evaluation, simulation, and optimization. The study begins with a knowledge background section designed to equip the reader with essential information about LLMs and Supply Chain Management. Following this, the benefits and challenges of applying LLMs in Supply Chain Management are examined, supported by two detailed case studies showcasing real-world applications. The thesis concludes by outlining potential directions for future research, offering a roadmap for further exploration in this rapidly evolving field.

Key findings reveal that LLMs significantly enhance SCM by improving efficiency, accuracy, and decision-making capabilities. They empower both technical and non-technical users and democratize access to complex processes like simulations and optimization. However, integrating LLMs into SCM presents challenges such as user adoption issues, hallucinations, privacy concerns, and potential disruptions. Addressing these challenges is crucial for the successful and safe implementation of LLMs in SCM, paving the way for innovative, resilient, and responsive supply chain operations.

KEYWORDS

Large Language Models; Supply Chain Management; Generative Artificial Intelligence; Decision-Making

Sustainable Development Goals (SDG):

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
LLM	Large Language Model
DL	Deep Learning
SCM	Supply Chain Management
NLP	Natural Language Processing
Gen-AI	Generative Artificial Intelligence
ML	Machine Learning
PTM	Pre-Trained Model
CV	Computer Vision
UML	Unified Modeling Language
SysML	Systems Modeling Language
PEFT	Parameter-Efficient Tuning
LoRA	Low-Rank Adaptation
RAG	Retrieval Augmented Generation
SLAs	Service Level Agreement
RLHF	Reinforcement Learning from Human Feedback
LP	Linear Programming
MILP	Mixed-Integer Linear Programming
AHP	Analytic Hierarchy Process
FCE	Fuzzy Comprehensive Evaluation
MSE	Mean Squared Error
MAE	Mean Absolute Error
MCDM	Multi-Criteria Decision Making
GAI	General Artificial Intelligence
SFT	Supervised fine-tuning

BPE	Byte Pair Encoding
bBPE	Byte Level BPE

1. INTRODUCTION

Human beings possess the remarkable ability to express and communicate through language, which develops during early childhood and continues to evolve throughout their lifetime. However, machines lack the inherent capacity to comprehend and communicate in human language unless they are equipped with powerful AI algorithms. The goal of achieving human-like reading, writing, and communication skills in machines has been a long-standing research challenge and desire. The advent of LLMs marks a significant milestone in this quest, leveraging advancements in DL methods, massive computational resources, and extensive training data (Usman Hadi et al., 2020).

LLMs have recently garnered significant attention for their exceptional performance in various predictive and generative tasks, highlighting their potential to transform multiple industries (Urlana et al., 2024). The field of NLP has observed substantial advancements with the development of sophisticated conversational AI systems like ChatGPT, which has attracted considerable interest and adoption (Chowdhury et al., 2023).

An essential area where LLMs can be transformative is SCM. Effective SCM is critical for enhancing operational efficiency, reducing costs, and improving customer satisfaction. However, the complexity and unpredictability inherent in supply chains present significant challenges to managers, requiring advanced tools and technologies to support decision-making processes. It is here where LLMs can play an important role.

This thesis underscores the importance of leveraging LLMs in SCM by demonstrating their potential to revolutionize decision-making, demand forecasting, and supplier evaluation. Additionally, this work explores how two decision-making tools can work in tandem with LLMs: Simulations and Optimizations. Both are important for solving real-world decision-making problems and can be enhanced by making these processes more accessible to decision-makers who may lack the necessary mathematical expertise (Wasserkrug et al., 2024).

Overall, the findings of this thesis suggest that the integration of LLMs into SCM can lead to significant operational efficiency and cost savings, enhanced supply chain resilience, strategic supplier management, informed decision-making, and a competitive advantage in the marketplace. The comprehensive review and case studies provided offer valuable insights and a roadmap for future research, emphasizing the transformative potential of LLMs in SCM.

1.1. Research Objectives

This thesis aims to explore the multifaceted impact of LLMs on SCM, identifying both the opportunities and challenges associated with their implementation. To achieve this, the research will address the following key questions:

- How can LLMs improve decision-making processes in SCM?
- What are the comparative benefits of using LLMs over traditional demand forecasting methods in SCM?
- How can LLMs enhance the supplier evaluation and selection process in SCM?
- What are the key challenges in integrating LLMs into existing SCM systems?
- How can organizations overcome the barriers to implementing LLMs in their supply chain processes?
- What are the emerging trends in LLM applications for SCM?

2. BACKGROUND

2.1. Fundamentals of Large Language Models

Gen-AI has emerged as a transformative paradigm in ML, revolutionizing NLP with models like GPT-3.0 and GPT-4.0 (Kaswan et al., 2023). These models, known for their ability to generate coherent and contextually relevant text, are widely used in automated content creation, chatbots, code generation, and translation services, significantly enhancing human-computer interactions and textual creativity.

LLMs, such as GPT-4 (OpenAI et al., 2023), belong to the Gen-AI domain and are based on Transformer language models. They utilize the Attention mechanism as their main building block (Vaswani et al., 2017). Typically, LLMs are pre-trained in an auto-regressive manner with next-word prediction tasks, demonstrating strong capabilities in text comprehension and generation (Y. Zhang et al., 2023). Recent advancements have shown that LLMs exhibit impressive proficiency across a wide array of text processing tasks, presenting massive context-aware knowledge and superior semantic comprehension through pre-training on large-scale text corpora (Z. Chen et al., 2023; Su et al., 2023).

As described in Figure 1, Gen-AI is the intersection of DL and NLP plus Automatic Speech Recognition. According to Ali Linkon et al. (2024) "The classification of generative AI models varies based on their output, categorizing them into language, image, or video models. However, there's a growing trend toward multi-modal models capable of simultaneous learning from text and images, positioning them as foundational models."

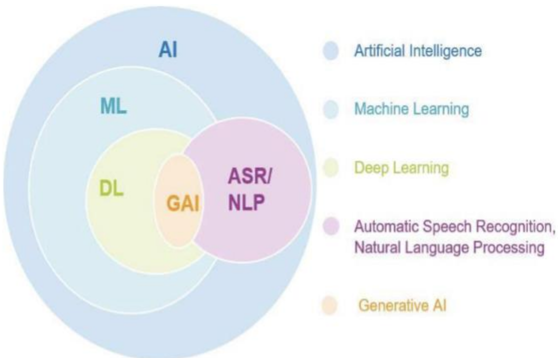

Figure 1 Generative AI Relation Diagram (Ali Linkon et al., 2024)

2.1.1. Transformers

The Transformer model, introduced by Vaswani et al. (2017), represents a groundbreaking shift in DL, especially within the realms of NLP, CV, and speech processing. Initially developed as a sequence-to-sequence model for machine translation, Transformers have since evolved, demonstrating state-of-the-art performance across a myriad of tasks through PTMs.

Transformers distinguish themselves by employing a stack of encoder and decoder layers, which collectively enable the model to learn complex language structures (Kalyan et al., 2021). Unlike traditional neural networks that rely on recurrent or convolutional architectures, Transformers leverage attention mechanisms exclusively, facilitating the transfer of weighted representation knowledge between neural units (Zhou et al., 2023).

At the core of this architecture is the self-attention mechanism, which allows the model to evaluate the importance of each word in an input sequence, independent of its position. This mechanism is pivotal for tasks like translation and question answering, where the context provided by distant words significantly influences the sentence's meaning (Usman Hadi et al., 2020).

Transformer-based pre-trained language models predominantly use tokenizers such as WordPiece, BPE, bBPE, and SentencePiece to construct vocabularies. These tokenizers segment raw text into individual tokens, which form the input sequences for LLMs (Kalyan et al., 2021; Usman Hadi et al., 2020).

For DL models to process textual data, the text must be converted into a numerical format. This is achieved through embeddings, which map input text to dense, low-dimensional vectors, forming the numerical input for models like Transformers (Kalyan et al., 2021).

The adoption of Transformer models extends beyond language-related applications into fields like CV, audio processing, chemistry, and life sciences, underscoring their versatility and effectiveness (Lin et al., 2022). This widespread applicability and the continuous advancements in Transformer architectures affirm their status as the cornerstone of modern DL.

2.1.2. Impact of the Scaling Law on LLMs

Scaling laws have been proven universal in AI, describing the relationship between model performance and factors like model size and data size, particularly in the success of scaling PTMs (G. Zhang et al., 2023). The establishment of a power-law relationship linking model performance to compute budget, model scale, and data scale has been pivotal in understanding these models. Results demonstrate that expanding the compute budget continues to yield significant benefits, encouraging the increase in model scales (DeepSeek-AI et al., 2024; Kaplan et al., 2020; Zhao et al., 2023).

Scaling the parameters and training tokens enhances LLMs' capacity to learn from pre-training data, improving generalization abilities and reducing model perplexity (Naveed et al., 2023; Zhao et al., 2023). The role of scaling is evident in models like GPT-3, which exhibit superior abilities compared to earlier versions. By scaling the model size, dataset size, and total computation, the performance of causal decoders is substantially improved (Zhao et al., 2023).

2.1.3. Pre-train and Fine-tune Paradigm

With the rise of DL, model parameters have increased rapidly, necessitating larger corpora for training to prevent overfitting. However, creating labeled corpora is expensive, leading to the adoption of the pre-train and fine-tune paradigm. This approach involves training a language model on large-scale unlabeled corpora to predict textual data probabilities, followed by fine-tuning task-specific objectives (Lin et al., 2022; P. Liu et al., 2023; Qiu et al., 2020).

PTMs can learn universal language representations from large text corpora, improving model initialization and serving as a regularization method to prevent overfitting on smaller datasets (Qiu et al., 2020). Models like BERT, XLNet, and RoBERTa have established new baselines for many natural language understanding benchmarks (Ribeiro et al., 2021).

Fine-tuning adapts PTMs to specific tasks using task-specific data. Common techniques include supervised fine-tuning, instruction tuning, alignment tuning, and PEFT:

- **SFT** refines LLMs after extensive pre-training, boosting performance across tasks with labeled datasets (Y. Liu et al., 2024; Zhao et al., 2023).
- **Instruction tuning** trains LLMs on datasets with instruction-output pairs, enhancing performance and controllability (Y. Liu et al., 2024; Lu et al., 2023; Naveed et al., 2023; Zhao et al., 2023).
- **Alignment Tuning** addresses the divergence of LLM outputs from human values (Y. Liu et al., 2024; Shen et al., 2023; Zhao et al., 2023), training models to meet the criteria of being helpful, honest, and harmless (Y. Liu et al., 2024; Naveed et al., 2023; Zhao et al., 2023). RLHF stands out as an institute solution for aligning with human preferences through reinforcement learning (Zhao et al., 2023).
- **PEFT** includes techniques like Adapter tuning, Prompt-based tuning, Prefix Tuning, and LoRA, reducing computational and memory overhead while preserving performance (Hu et al., 2021; Y. Liu et al., 2024; Naveed et al., 2023; Thangarasa et al., 2023; Yang et al., 2023).

Despite their capabilities, LLMs are prone to catastrophic forgetting during model adaptation, where previously learned knowledge is lost. Strategies like LoRA help but still face challenges with this issue (S. Chen et al., 2020; Kalajdzievski, 2024; Zhai et al., 2023).

2.1.4. Prompting Techniques and Engineering

After pre-training or adaptation tuning, a major approach to using LLMs is to design suitable prompting strategies for solving various tasks (Zhao et al., 2023). These prompts are sequences of natural language inputs that guide LLMs to generate pertinent responses without further training. Users provide context and directives, influencing the model's downstream performance. Techniques include:

- **In-Context Learning (ICL):** Uses a few examples within the input prompt to perform tasks, enabling few-shot or zero-shot learning without explicit parameter updates (Agrawal et al., 2023; Brown et al., 2020; Dong et al., 2022; Lu et al., 2023; Tang et al., 2023; Zhao et al., 2023).
- **Few-Shot Learning:** Provides a few input-output examples as demonstrations before task execution (Brown et al., 2020; Wei et al., 2022).
- **One-Shot Learning:** Similar to few-shot but with only one demonstration (Brown et al., 2020).
- **Zero-Shot Learning:** Uses natural language instructions without any demonstrations, often improved by instruction fine-tuning (Brown et al., 2020; Naveed et al., 2023).
- **Chain-of-Thought (CoT):** Enhances reasoning by prompting LLMs to articulate intermediate steps toward a final answer, particularly effective in larger models (Lu et al., 2023; Pan et al., 2023; Tay et al., 2022; Zhao et al., 2023; Zhou et al., 2023).

Prompt engineering has become a prominent area that designs and refines prompts to maximize LLM effectiveness for specific tasks, leveraging in-context learning without gradient-based learning (Hu et al., 2021; Pan et al., 2023; Tian et al., 2023).

2.1.5. Emergent Abilities

LLMs exhibit a notable escalation in capabilities as they increase in size, particularly in reasoning and emergent abilities. Larger models improve in translating natural language into mathematical representations, performing multi-step inferences, and displaying refined commonsense reasoning (Yang et al., 2023). Emergent abilities, manifesting only when models reach a certain scale, highlight qualitative changes in model behavior due to quantitative changes in model size (Lu et al., 2023; Wei et al., 2022).

These abilities enhance LLM utility in settings where they have not been explicitly trained, distinguishing them from smaller PTMs like BERT and DeBERTa (Z. Chen et al., 2023). As models grow, they not only improve task performance but also show potential for developing AGI, pushing the boundaries of AI capabilities (Jin et al., 2023). Techniques like in-context learning, instruction tuning, and chain-of-thought prompting underscore the transformative impact of scaling on LLM functionality and effectiveness (Lu et al., 2023).

2.1.6. Limitations of LLMs

LLMs face limitations in privacy, harmful content, and hallucinations. These models can inadvertently memorize and expose sensitive data, generate offensive or misleading content, and produce nonsensical or untruthful information. Addressing these issues involves robust safeguards, external knowledge integration, and advanced prompt engineering to ensure LLM reliability and safety (Agrawal et al., 2023; C. Liu et al., 2023; Yang et al., 2023; Y. Zhang et al., 2023; Zhao et al., 2023).

2.1.7. External tools manipulation

LLMs have inherent limitations, such as relying on pre-training data, which can become outdated. To address this, integrating external tools enhances LLM capabilities. For instance, GPT-4 utilizes tools like calculators and search engines to extend its functionality (Zhao et al., 2023). This integration is crucial for maintaining up-to-date and accurate responses, compensating for the static nature of pre-trained knowledge.

One approach to mitigate the limitations of outdated knowledge and hallucinations in LLMs is RAG. This method involves retrieving relevant information from external sources and integrating it into the generation process. RAG enhances the accuracy and reliability of LLM responses by leveraging real-time data from sources such as search engines (J. Chen et al., 2023). The process includes retrieving context relevant to a query and using it to generate a response, ensuring the information is current and contextually appropriate (Jeong, 2023; Keswani et al., 2024).

2.2. SUPPLY CHAIN MANAGEMENT

SCM is the strategic coordination of business functions within a company and across businesses within the supply chain. Its goal is to improve the long-term performance of both individual companies and the entire supply chain (Hugos, 2011). SCM is a web of services and delivery options that involve procuring raw materials, transforming them into products, and delivering these products to customers (Yahya et al., 2021). Effective SCM ensures an economical and efficient supply chain, reducing costs, enhancing customer services (Malik et al., 2010), and adding value for stakeholders (Lambert & Cooper, 2000).

SCM ensures the uninterrupted flow of materials, products, and information throughout all stages (Deiva Ganesh & Kalpana, 2022). It involves activities like production planning, inventory management, transportation, and information management, applicable to various industries from manufacturing to retail (Hugos, 2011).

2.2.1. Decision-Making in Supply Chains

Decision-making in supply chains spans design, planning, and execution phases (B. Li et al., 2023). Each stage requires tailored approaches, balancing operational needs with strategic goals using structured methodologies (Svoboda & Lande, 2024). Supply chain managers must navigate complex systems and uncertainties, requiring clear understanding and effective strategies to manage complexities (Manuj & Sahin, 2011).

2.2.1.1. Demand Forecasting

Demand forecasting is crucial in SCM, influencing production planning and order fulfillment. Accurate predictions impact resource allocation, scheduling, worker training, and market strategies (Rozanec, 2021). Modern AI and ML methods enhance forecasting accuracy, reducing uncertainties in production and delivery planning (Borucka, 2023). Strategic-level forecasts consider uncertain capacities and market changes, guiding decisions about distribution channels (Babai et al., 2022).

2.2.1.2. Supplier Evaluation

Strong supplier relationships enhance competitiveness by ensuring high-quality products, timely deliveries, and cost-effective solutions (Lopes & Rodriguez-Lopez, 2021). Supplier evaluation involves assessing factors like price, quality, delivery performance, and reputation. Advanced analytics and AI provide deeper insights into supplier performance and risks (Takala et al., 2014). Effective supplier management is critical in today's competitive market, influencing overall supply chain performance (Wang & Wu, 2024).

2.2.2. Decision-Making Tools

2.2.2.1. Simulation and Modeling

Simulation is vital for validating production layouts and sizing plants. It supports decision-making by replicating scenarios, identifying potential issues, and saving costs (Hugos, 2011; Terzi & Cavalieri, 2004). In SCM, simulation helps manage the complexity of supply chains by modeling various conditions (El Raoui et al., 2020). Industry 4.0 and digital twins integrate simulation for real-time process control (Mustafee et al., 2021).

2.2.2.2. Optimization

Optimization tools automate decision-making, using methods like LP and MILP to model and solve problems efficiently. These tools integrate supply chain elements, ensuring robust and efficient solutions (AhmadiTeshnizi et al., 2023; B. Li et al., 2023; Wasserkrug et al., 2024). Continuous algorithm improvements enhance their adoption across industries, aiding precise business decisions (Q. Li et al., 2023).

2.2.2.3. Multi-Criteria Decision Making

MCDM methods address decision-making involving multiple criteria, balancing conflicting objectives to achieve optimal outcomes (Wang & Wu, 2024). Techniques like AHP, ANP, TOPSIS, DEA, and fuzzy decision-making are used in SCM to handle complex decisions (Khan et al., 2018). MCDM helps balance criteria such as cost, quality, and delivery time, crucial for effective SCM.

3. LITERATURE REVIEW

3.1. LARGE LANGUAGE MODELS IN SUPPLY CHAIN MANAGEMENT CONTEXT

LLMs are rapidly growing in healthcare, education, information systems, hospitality, and tourism. However, SCM research and use cases are still in an early stage of development. The integration of LLMs and SCM heavily depends on human understanding and interpretation. Moreover, most of the approaches come from grey literature, and the lack of theoretical perspectives to better understand the technology-human relationship and how to potentialize the results and minimize the risks is a pressing issue (Fosso Wamba, Queiroz, et al., 2023).

SCM processes and activities are relatively complex and require advanced technology such as AI to efficiently facilitate the three typical supply chain and logistics flows which are information flow, material flow, and financial flow (Jia et al., 2023). AI has already begun to play a significant role in business operations such as chatbots enhancing customers' purchases (Y. Chen et al., 2023).

Integrating AI into SCM leads to a fundamental transformation in operations and various interactive aspects of organizations and their surrounding environment. Today's operational activities within SCM incorporate AI technologies aiming to automate and optimize processes, enhance decision-making capabilities, and improve end-to-end transparency (Alkhaldi et al., 2023).

AI and ML have a deep impact on the entire SCM process, potentially replacing humans with their human-like abilities, optimizing processes that use massive quantities of data in real time to provide accurate planning, and enhancing productivity (Jia et al., 2023). AI has been adopted in SCM from monitoring only to control, optimization, and advanced autonomy (Shatat & Shatat, 2022).

These new disruptive technologies (AI and Gen-AI) will lead to an increment in productivity transforming the SCM landscape and its limitations (Jackson, Ivanov, et al., 2024). Gen-AI can produce a paradigm shift in SCM improving several aspects such as procurement, inventory, logistics, sourcing, and risk mitigation. Can also enhance supply chain communication (Jackson, Ivanov, et al., 2024), decision-making processes, improve forecasting, and faster real-time responses to customer's questions (Fosso Wamba, Guthrie, et al., 2023a).

The most notorious Gen-AI model is ChatGPT from OpenAI. This model can potentially transform the business landscape carrying consequences to SCM processes (Fosso Wamba, Queiroz, et al., 2023).

3.1.1. Benefits

Empirical investigations underscore the substantial benefits of LLMs adoption in SCM. Companies, whether adopters or not, recognize efficiency gains and anticipate enhanced performance in supply chain operations. Notably, organizational learning theory elucidates disparities between adopters and non-adopters, revealing a greater perception of benefits among adopters and amplified challenges among non-adopters. Nevertheless, these challenges tend to diminish post-implementation, indicating the role of organizational learning in accelerating LLMs diffusion and mitigating risks. While LLMs offer substantial benefits such as efficiency improvements, it also poses various challenges, including ethical considerations and integration with human workers. Understanding these dynamics through the lens of organizational learning theory provides valuable insights for practitioners and policymakers, facilitating informed decision-making in leveraging LLMs within SCM (Fosso Wamba, Queiroz, et al., 2023).

Unlike traditional ML systems, which typically require expertise for operation and training, LLMs can generate output for users irrespective of their background, leading the way in a new era where AI is accessible to a broader audience (Hendriksen, 2023). Additionally, research indicates a growing utilization of AI by managers to address operational challenges within SCM (Sharma et al., 2022; Weisz et al., 2020). Moreover, studies highlight the potential for harmonious coexistence between human workers and LLMs, particularly in delineating tasks: repetitive functions can be delegated to AI while creative endeavors remain within the human domain (Weisz et al., 2020). However, concerns linger regarding the invasion of AI into creative tasks, potentially altering traditional work dynamics (Fosso Wamba, Queiroz, et al., 2023).

The constant use of tools like LLMs can leverage a robust AI-driven supply chain, displaying increasing versatility compared to earlier ML infrastructure solutions (Hendriksen, 2023).

The evolving landscape of SCM is increasingly marked by the integration of GAI, as highlighted in recent industry insights. Publications such as the Wall Street Journal report a growing trend in the experimentation with and adoption of GAI technologies for enhancing communication and decision-making processes within supply chains (Jackson, Ivanov, et al., 2024). This exponential interest underscores the potential of GAI to streamline complex operations, reduce errors, and ultimately lead to more efficient and robust supply chain networks. The implications of these advancements suggest a significant shift towards more technologically integrated and automated supply chain environments, promising enhanced operational efficiencies across various supply chain segments.

On the spectrum of AI's and LLMs' role in decision-making, the technology's integration into SCM can vary from assistive to fully autonomous functionalities. At one end, AI serves an assistive role, where it supports human decision-makers by providing data analysis, information retrieval, and personalized assistance, acting much like a digital advisor. This use of AI enhances human capabilities without replacing the decision-maker, ensuring that strategic decisions remain under human oversight (Hendriksen, 2023). Conversely, at the other end of the spectrum, AI assumes a more autonomous role, making critical decisions that can significantly impact the supply chain. This shift to an autonomous role raises important questions about the balance of power, accountability, and ethical considerations in delegating decision-making to AI systems, transforming the traditional dynamics of SCM (Hendriksen, 2023).

LLMs can suggest ways to optimize inventory levels, reduce stock-outs, increase efficiency, and improve order fulfillment times by analyzing inventory data and saving costs (Bahrini et al., 2023). LLMs can also improve the accuracy of inventory management plans both in the short and long term (Skórnóg & Kmiecik, 2023). Svoboda & Lande (2024) discuss the lack of experiment papers about GPT applications in decision-making such as MCDM.

LLMs are increasingly recognized for their capability to support not only technical users but also non-technical stakeholders in executing technical tasks. This support is particularly valuable in transforming plain language problem descriptions into technical problem formulations, thereby simplifying the initial stages of technical project development. The recent focus on using LLMs for automating code-generation tasks further enhances their utility, enabling the fine-tuning of these models to specific problem-solving contexts. Such advancements are particularly beneficial where non-technical users need to engage with optimization processes without deep knowledge of the underlying algorithms. The ability of LLMs to bridge this knowledge gap not only enhances user support but also fosters a more inclusive environment where more individuals can contribute to and interact with optimization tasks without extensive training (Amarasinghe et al., 2023).

LLMs play a critical role in democratizing access to complex processes, such as optimization. By simplifying the formulation of problems and the setup of solutions, these models make expert-level knowledge more accessible to a broader audience. This accessibility is essential for enhancing the efficiency and inclusivity of decision-making processes in business environments. Furthermore, the use of LLM-based chatbots can facilitate effective communication between business decision-makers and optimization experts. This not only helps in clarifying the requirements and expected outcomes of optimization projects but also ensures that non-expert stakeholders can actively participate in and understand these processes (AhmadiTeshnizi et al., 2023; Wasserkrug et al., 2024).

In terms of integration, LLMs do not need a large infrastructure to work. If the user has an internet connection can work with these models (like GPT-4) eliminating AI adoption barriers (Hendriksen, 2023). Other models, such as Llama, can work without an internet connection using a local machine preventing any leakage of data outside the organization.

3.1.2. Challenges

The integration of LLMs into SCM is assured to transform the field significantly, prompting both scholars and practitioners to reassess the existing practical applications and theoretical frameworks, such as Transaction cost economics (TCE) and Complex adaptive systems (CAS). This technological shift requires careful consideration of how LLMs are implemented and the potential impacts on traditional SCM practices (Hendriksen, 2023). While LLMs offer revolutionary capabilities in automating and optimizing processes, their application is not without limitations that could undermine the benefits.

Historical IT innovation studies indicate that the acceptance and integration of new technologies like LLMs are largely driven by perceived business value, which weighs the benefits against the challenges of adoption. As AI systems proliferate across individual and organizational levels, they redefine traditional models and assumptions in SCM, challenging the pre-existing notion that supply chains are primarily human and organizational constructs. The increasing capability of AI to perform complex decision-making necessitates a reevaluation of our foundational theories in SCM (Fosso Wamba, Guthrie, et al., 2023b; Hendriksen, 2023).

The integration of AI into SCM is more than just a technical adaptation; it is a complex social process shaped significantly by human interactions and the sensemaking processes individuals employ. According to Hendriksen (2023), how individuals from various professional backgrounds and roles within SCM perceive and interpret AI profoundly influences its integration. Understanding human sensemaking, the process through which people attribute meaning to their collective experiences, is critical in navigating the trajectory of AI adoption and determining its effectiveness in the supply chain environment.

Nowadays, the SCM process uses ML and neural network technologies in their processes and is viewed by users as a "black box" due to their limited explainability (Kosasih et al., 2023), including LLMs.

Manufacturing supply chains use AI and ML to improve their cost efficiency by around 10%. Although this would be a promising direction, around 63% of manufacturing companies are still reluctant to use these types of technologies to manage and monitor their supply chain and logistics activities (Shatat & Shatat, 2022).

The relationship between management and the technology used in business is ambiguous. Studies find a positive relationship between them, whereas others find no association or even a negative impact (Y. Chen et al., 2023). The lack of awareness about AI technology can lead to low trust in technology and low acceptance by managers and consumers, implying slow adoption among manufacturing companies (Nayal et al., 2022).

One of the primary challenges associated with LLMs in SCM is the risk of generating sub-optimal or incorrect outputs, known as "hallucinations". These inaccuracies can range from minor errors to significant misinformation, which may require the development of domain-specific tools to correct outputs, particularly in complex fields such as code generation. Generic tools exist to address these issues, but they often fall short in handling the nuanced requirements of specific applications (B. Li et al., 2023). Moreover, the opacity of LLM methodologies compounds these challenges, as users and stakeholders cannot easily discern how the models generate their outputs. This lack of transparency can erode trust, particularly when errors lead to public missteps or financial losses, as illustrated by an incident with

Google's AI chatbot which resulted in a substantial drop in market value following an inaccurate public response (Richey et al., 2023). This situation underscores the significant ethical and reputational risks involved in deploying generative AI technologies in critical sectors like SCM.

Privacy remains a critical concern in the application of any AI technology within SCM, especially when using domain-specific data that may be proprietary. Despite the privacy assurances that might be provided through SLAs, the risk of data interception by malicious entities cannot be overlooked. Many organizations are therefore considering privacy-preserving approaches to AI deployment, which involve keeping sensitive data in-house and away from external LLM hosts to mitigate risks associated with data breaches (B. Li et al., 2023).

As businesses venture into the digital age, SCM faces a significant disruption facilitated by AI applications like LLMs (Richey et al., 2023). The discipline SCM appears unprepared for the revolutionary potential of AI tools, as current theoretical frameworks do not fully encompass the disruptive capabilities of these advanced methods (Hendriksen, 2023). While some researchers have explored AI's role in enhancing supply chain efficiency and mitigating disruptions like those from the COVID-19 pandemic, these studies fall short of capturing AI's full disruptive potential.

Disruptions in SCM typically refer to events that alter the flow of goods or services, requiring adaptation and innovation with significant social and environmental impacts. Gen-AI integration introduces potential pitfalls, such as overreliance on AI, inappropriate delegation of authority, misplaced trust, and unethical uses, which can lead to complex risks like an "AI-bullwhip effect" where inaccurate AI analyses disrupt the entire supply chain (Hendriksen, 2023). Effective management of these disruptions requires decision support systems that provide real-time visibility and data-driven decision-making to enhance supply chain agility, flexibility, and operational performance (Kashem et al., 2023).

3.2. Large Language Models Applications in the Context of Supply Chain Management

Industries are exploring different ways to incorporate new technologies, with their benefits and limitations. The range of aspects LLMs can impact in SCM is considerable. Some of those aspects will be covered in this section.

As described in Bahrini et al. (2023), LLMs like ChatGPT can enhance SCM tasks, such as Demand forecasting, Inventory optimization, and Supplier selection/evaluation, besides the potential implementations in SCM tools.

3.2.1. Demand Forecasting

Research on the application of LLMs in time series forecasting has diversified into three primary approaches. The first involves directly applying pre-trained LLMs, utilizing their natural language understanding to interpret time series data, inspired by the success of language-based applications such as chatbots. The second method aggregates diverse time series data across various domains into a unified model, aiming to enhance its versatility and general applicability. The third approach focuses on fine-tuning LLMs, with techniques such as aligning model layers specifically for time series data, introducing contrastive learning for better temporal representation, and modifying input layers to translate temporal data into a language-compatible format. Notably, there remains an unexplored potential in spatial-temporal forecasting, indicating a promising direction for future research (Lai et al., 2024).

3.2.2. Supplier Evaluation

The integration of ChatGPT into supplier evaluation processes signifies a remarkable advancement in SCM. Wang & Wu (2024) merged traditional MCDM models with ChatGPT-based evaluations, developing a comprehensive system that enhances the efficiency and accuracy of supplier assessments. Early studies suggest that GPT-4 can perform foundational analyses for supplier selection by evaluating supplier profiles based on specified parameters and instructions (Hendriksen, 2023).

Moreover, the application of Gen-AI in supplier evaluations offers revolutionary benefits over conventional methods. AI can swiftly analyze extensive datasets from numerous suppliers, considering various factors such as cost-effectiveness, product quality, reliability, operational efficiency, and sustainability (Richey et al., 2023). Additionally, AI promotes inclusivity by suggesting strategies to integrate minority-owned, women-owned, or veteran-owned enterprises within the supply chain. Its advanced text generation abilities also support the creation of detailed supplier profiles, negotiation strategies, and contractual terms based on historical data and predictive analytics. This comprehensive approach not only optimizes the selection process but also ensures more equitable and informed decision-making, driving innovation and sustainability in supplier management.

Although, there are few implementations in SCM, LLMs as evaluators in general is a field in which researchers have been more prolific. The following frameworks and methods have potential to be used as supplier evaluation tools.

3.2.3. Simulation and Modelling

Simulation modeling plays a vital role in managing complex systems, capturing intricate dynamics that often elude traditional analytic methods. Despite its utility, the process is hindered by significant challenges: high technical complexity and the necessity for constant, detailed communication between domain experts and simulation engineers, leading to increased time and financial costs. Recent developments in NLP present an opportunity to overcome these hurdles. LLMs are being explored for their potential to streamline the development of simulation models by facilitating an efficient dialogue between experts and engineers. This capability reduces the iterative communication that typically extends project timelines and costs, suggesting a promising shift toward more efficient simulation modeling practices (Jackson & Rolf, 2023).

Furthermore, the application of LLMs extends beyond mere facilitation; it reshapes how simulations are performed and integrated across various domains. As LLMs like ChatGPT become more user-friendly and accessible without the need for advanced programming or API knowledge, they are poised to become as integral to Modeling & Simulation (M&S) as ML technologies have in recent years. This advancement promises to make sophisticated simulation tools more accessible and adaptable, enhancing their application in complex decision-making scenarios across diverse fields. The increased integration of LLMs indicates a significant shift towards automation and intelligent decision-making within M&S, driving innovation and efficiency in ways previously unattainable (Giabbanelli, 2023).

Effective communication within interdisciplinary modeling teams is crucial, especially given the diversity of expertise among team members. Traditional modeling languages such as UML and SysML, while useful, often have a steep learning curve for non-specialists. This can hinder collaborative efforts, as not all team members are able to engage deeply with the technical aspects of the model. Research has shown that even simplified visual representations of models can be challenging for non-experts like senior executives. To overcome these barriers, LLMs offer a promising solution by converting complex model structures into accessible, easily understandable textual narratives, thus enhancing comprehension across all levels of technical proficiency (Giabbanelli, 2023).

3.2.3.1. Explaining Simulation Process and Outcomes

The process of simulation construction is technically complicated, implying that domain experts cannot develop simulation models by themselves, therefore simulation engineers are required. However, simulation engineers lack the domain knowledge experts have (Jackson & Rolf, 2023).

Explaining the outcomes of simulations in a clear and concise manner is essential for effective decision-making. Natural Language Generation (NLG) plays a critical role in this process by converting complex simulation data into straightforward textual descriptions. This approach reduces the cognitive load on users, allowing them to focus on making informed decisions based on the summarized data. While NLG helps in simplifying the presentation of simulation outcomes, ensuring the accuracy and completeness of these explanations remains a challenge, particularly when dealing with complex or nuanced data sets. Continuous advancements in NLG are necessary to improve the fidelity and utility of textual summaries in representing simulation outcomes (Giabbanelli, 2023).

3.2.3.2. Simulations Errors

Identifying and explaining errors in simulation models is crucial for ensuring their accuracy and reliability. While LLMs are effective in articulating the reasons behind simulation errors, their capability to detect these errors directly is limited. LLMs excel in transforming complex, technical error analyses into clear, understandable language, aiding modelers in the verification and validation processes. This not only facilitates a deeper understanding of the errors but also enhances the modelers' ability to make informed corrections, thereby improving the overall robustness of the simulation models (Giabbanelli, 2023).

3.2.3.3. Generate simulations based on natural language

Simulation models of logistic systems can be automatically generated from natural language descriptions, facilitating collaboration between human experts and AI-based systems in the field of simulation modeling.

The study by Jackson, Jesus Saenz, et al. (2024) investigates the use of AI, specifically GPT-3 Codex, to automate the creation of simulation models for logistics systems from natural language descriptions. They developed a framework that generates valid simulation models for queuing and inventory management systems based on verbal explanations. This framework utilizes GPT-3 Codex to translate natural language into executable Python code, addressing the complexity and technical expertise traditionally required in simulation modeling.

The language model adjusts its parameters using intermediate results to produce accurate Python simulation code. Python was chosen due to GPT-3 Codex's proficiency in it and its ability to dynamically execute code. The framework uses Python's 'eval' method to convert strings into bytecode for execution. Human experts validate and verify the generated code, creating a feedback loop to ensure accuracy.

The authors' framework facilitates prompt engineering and hyperparameter tuning, integrating human validation to ensure the generated code's accuracy and functionality. This study highlights the potential for significant efficiency gains and cost savings in simulation model development, making it more accessible to domain experts without extensive technical knowledge.

3.2.4. Optimization

In the realm of SCM, effective decision-making is heavily reliant on mathematical optimization techniques. However, a significant challenge exists as most decision-makers lack the necessary mathematical skills or access to experts who can apply these complex methods. This gap has spurred calls for making mathematical optimization more accessible to a broader range of decision-makers, striving to democratize the technology so it can be utilized more widely across industries (Wasserkrug et al., 2024). Furthermore, when automation is introduced to streamline decision-making processes, it often still requires the involvement of business operators to interpret, modify, or directly override outcomes.

The inherent complexity of these decisions, often derived from large-scale operations and "black-box" algorithms. Unfortunately, many of these operators do not possess an adequate background in optimization, leading to inefficient and time-consuming consultations with data scientists and engineers, which can significantly slow down decision-making processes. This lack of transparency and optimization expertise can hinder effective management and adaptation to changing conditions within a business environment (B. Li et al., 2023).

The technical challenges in optimization are particularly evident when dealing with complex real-world problems that require advanced methodologies like MILP. These methods are crucial for addressing intricate issues in areas such as production planning, resource allocation, and transportation management. However, successfully transforming these decision-making challenges into MILP models relies heavily on deep expertise in operations research and mathematical optimization. This necessity creates a barrier for non-experts, limiting the practical application of MILP to those few who possess the required specialized knowledge (Q. Li et al., 2023).

3.2.4.1. LLMs in the optimization process

Recently, there has been a push to incorporate LLMs into the optimization process, aiming to simplify the formulation of optimization models. Despite these efforts, LLMs currently face significant limitations; they are generally fine-tuned for simpler mathematical tasks like LP and are not yet capable of managing the complexities associated with larger datasets that include extensive variables and constraints. This limitation significantly hinders their practical application in real-world settings. Furthermore, while there is notable progress in generating models from natural language descriptions, many errors persist, and few models are developed to the point of practical application. Consequently, there remains a substantial need for further research and development to enhance the capabilities of LLMs to generate, solve, and analyze complex optimization models more effectively (Amarasinghe et al., 2023; Wasserkrug et al., 2024).

The integration of LLMs into the field of optimization aims to enhance the functionality and accessibility of these tools within complex decision-making environments. One of the primary purposes of LLMs is to improve the explainability of optimization outcomes. As B. Li et al. (2023) note, decision processes in large-scale optimizations often rely on algorithms perceived as "black boxes" by operators and planners. This lack of transparency can hinder business operators' ability to address customer queries, respond to unexpected events, and rectify inefficiencies effectively. By enhancing the explainability of these processes, LLMs help demystify the results of complex optimizations, enabling business operators to understand and communicate the reasoning behind specific decisions more clearly.

3.3. Large Language Models Use Cases

3.3.1. UniTime

The UniTime model (X. Liu et al., 2024) based on "*UniTime: A Language-Empowered Unified Model for Cross-Domain Time Series Forecasting*", leverages a PTM as its backbone, adapted to process time series data through fine-tuning with domain-specific instructions. Its innovative design allows it to handle diverse time series data from various SCM domains such as sales, inventory, logistics, and supplier data. Traditional models often require separate configurations for different datasets, increasing complexity. UniTime overcomes this with a flexible architecture that adapts to varying data characteristics, including different numbers of variables, input lengths, and prediction horizons. This adaptability ensures that the model can generalize across multiple SCM scenarios, making it suitable for a wide range of applications within the supply chain.

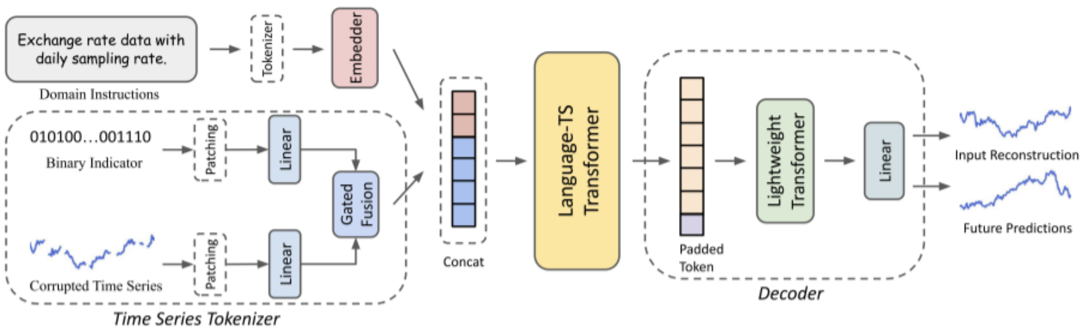

Figure 2 UniTime Architecture (X. Liu et al., 2024)

Figure 2 shows how UniTime architecture works. The model begins by preparing the time series data and text instructions. Time series data, which includes sequences like daily sales numbers, are grouped into "tokens" or chunks. This grouping allows the model to identify patterns over these chunks rather than analyzing data points individually. Text instructions, which describe the context of the time series data (e.g., "Daily sales of product X in store Y"), are converted into embeddings. These embeddings translate the textual information into a form that the model can understand.

Once the data is prepared, the text instructions and time series tokens are combined. The text embeddings are placed first, followed by the time series tokens. This sequence ensures that the model reads the instructions before examining the data, facilitating better contextual understanding.

To improve the model's generalization capabilities, masking is applied to the time series data. Masking involves hiding parts of the data using a binary mask vector, forcing the model to learn from incomplete information. This technique prevents overfitting and ensures that the model can handle missing or incomplete data effectively. The masked time series data and the binary indicator vector are then combined through a process called gated fusion. This integration helps the model understand which parts of the data are visible and which are hidden, enhancing its predictive accuracy.

The combined data and instructions are processed through multiple layers in the model. Each layer refines the model's understanding of the data by considering both the instructions and the time series tokens together. UniTime leverages a pre-trained GPT-2 model to initialize its weights, providing a strong foundation for understanding the textual information and improving the model's overall performance. Layer normalization, multi-head self-attention, and multi-layer perceptrons are key components in these layers. The self-attention mechanism allows the model to look back and forth within the data to understand relationships and dependencies.

After processing the data through these layers, the model generates long-term forecasts using a decoder. The decoder includes a lightweight Transformer and a learnable padding token to ensure consistent sequence lengths across different domains. The final step involves flattening the processed data and using a linear layer to produce the predictions.

One of the most significant results of the UniTime model is its ability to provide highly accurate forecasts. By leveraging the self-attention mechanism of the Transformer architecture, UniTime captures long-range dependencies in the data, resulting in precise demand predictions. This accuracy is crucial for maintaining optimal inventory levels, minimizing stockouts, and reducing excess inventory.

UniTime showed a marked reduction in MSE and MAE compared to traditional models. This improvement highlights UniTime's superior ability to capture complex temporal patterns in supply chain data. The model maintained high accuracy across different supply chain domains, such as sales data, inventory levels, and logistics data. This consistency underscores the model's ability to generalize across varied datasets.

The UniTime model's architecture supports zero-shot transferability, enabling it to apply learned knowledge from one domain to another without retraining. This capability is particularly beneficial in dynamic supply chain environments where new products, markets, or supply chain structures frequently emerge. In tests, UniTime successfully transferred its learning from historical sales data to predict trends in new product launches with minimal degradation in accuracy. The zero-shot transferability feature reduced the need for extensive retraining, saving time and computational resources.

The application of masking techniques in UniTime has significantly improved the model's robustness. By training on partially hidden data, the model becomes adept at handling missing or incomplete data, which is common in real-world supply chain scenarios. UniTime's performance remained stable even with a significant portion of data points masked, showcasing its ability to make accurate predictions despite data gaps. The use of masking helped prevent overfitting, ensuring that the model's predictions were reliable across different datasets and not just the training data.

A key advantage of UniTime is its unified model approach, which eliminates the need for multiple specialized models. This approach simplifies the management and deployment of forecasting systems, reducing operational overhead and streamlining processes within the supply chain. With UniTime, a single model can be used across various SCM domains, ensuring consistent and efficient performance without requiring extensive reconfiguration.

3.3.2. GPT as evaluator

MCDM is a crucial methodology applied in various domains to evaluate, rank, or select alternatives based on multiple conflicting criteria. Supplier evaluation is one of its classic applications, playing a critical role in SCM. Traditional MCDM models like the AHP and FCE involve gathering data through expert surveys and using structured evaluation frameworks to ensure objective and scientific assessments.

Wang & Wu (2024) in "*Can ChatGPT Serve as a Multi-Criteria Decision Maker? A Novel Approach to Supplier Evaluation*" explore the potential of using ChatGPT for supplier evaluation, comparing its performance with traditional MCDM methods. The study focuses on 121 suppliers from Company A's supply chain, collecting both quantitative indicators and unstructured textual data. Quantitative data includes various performance metrics, while textual data consists of transaction records and other qualitative descriptions.

The foundation of the study involves creating a comprehensive dataset for evaluating suppliers, segmented into quantitative and qualitative data. The quantitative data includes numerical indicators such as product acceptance rates, timeliness of delivery, and cost metrics, crucial for objectively measuring supplier performance across various dimensions. On the other hand, the qualitative data consists of unstructured text data, such as transaction records and feedback from supply chain personnel, providing a nuanced view of supplier performance.

To process this data, an evaluation set consisting of grades like Good, Average, and Poor is defined. As shown in Figure 3 (a), surveys are conducted among industry experts to gather evaluation scores for suppliers under different criteria. The mean of these responses is calculated to derive the average evaluation score for each criterion. Textual records are categorized and structured to be input into the LLM for evaluation.

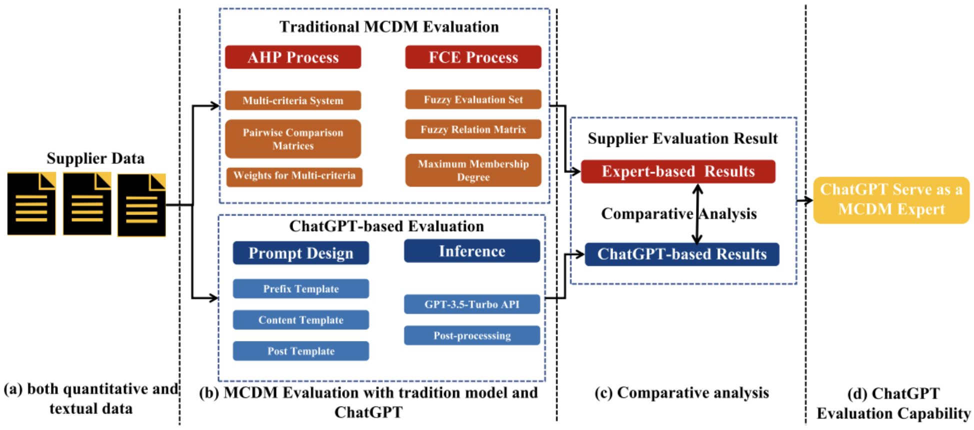

Figure 3 The overall architecture of ChatGPT-based MCDM Method (Wang & Wu, 2024)

The traditional MCDM approach uses a combination of the AHP and FCE methods (Figure 3 (b)). The AHP process begins with creating a hierarchical structure of evaluation dimensions and criteria. This involves constructing judgment matrices for criteria and sub-criteria, where elements represent the relative importance of each criterion. The eigenvectors corresponding to the maximum eigenvalue of these matrices are used to determine the relative weights of criteria and sub-criteria. The FCE method defines membership functions for each sub-criterion to determine the supplier's degree of membership to various evaluation grades. A fuzzy evaluation matrix is constructed for each supplier, incorporating these membership functions, and the aggregate scores are computed to derive a comprehensive evaluation grade for each supplier.

The LLM-based methodology leverages ChatGPT to perform supplier evaluations. This process begins with designing prompts to guide the model (Figure 3 (b)). The prompt design includes a prefix prompt that establishes the role and context for the evaluation, the main content prompt that provides specific textual records related to each criterion for a supplier, and a post-prompt that instructs the model to output evaluation grades for each criterion and the overall supplier score. For each supplier and sub-criterion, the textual description is inputted into ChatGPT, which predicts the evaluation grade based on the description and outputs a grade vector for each supplier. Post-processing techniques ensure consistency and relevance of the model's output by applying methods like regular expression extraction and synonym dictionaries to align the results with predefined evaluation categories.

The study conducts an experimental and comparative analysis between traditional MCDM methods and the ChatGPT-based approach. Using AHP and FCE methods, the traditional evaluations involved calculating criteria and sub-criteria weights and constructing fuzzy evaluation matrices for each supplier. These matrices were then used to derive aggregate scores and final evaluation grades. For the ChatGPT-based approach, the model was prompted with structured inputs to evaluate each criterion for the suppliers. Various techniques, such as Chain-of-Thought prompting, demonstration-based in-context learning, and voting ensemble methods, were tested to enhance the model's performance.

The performance of ChatGPT-based evaluations was compared to traditional methods using metrics like accuracy, precision, recall, and F1-score. The results indicated a high level of alignment between ChatGPT's evaluations and those of human experts. The best configuration of ChatGPT achieved an F1-score of 74.05%, demonstrating its capability to perform multi-criteria decision-making tasks effectively. The Chain-of-Thought approach improved logical consistency, while the demonstration-based learning and voting ensemble methods enhanced robustness and accuracy. The voting ensemble method showed the highest accuracy, suggesting the benefit of aggregating multiple model predictions to capture diverse perspectives. This analysis underscores ChatGPT's potential as a reliable and efficient tool for supplier evaluation in SCM.

3.3.3. Use Cases Analysis

In the context of SCM, LLMs have been gaining traction for their potential to enhance various SCM functions. Two key papers provide a foundation for understanding how LLMs can be integrated into SCM: "Can ChatGPT Serve as a Multi-Criteria Decision Maker? A Novel Approach to Supplier Evaluation" and "UniTime: A Language-Empowered Unified Model for Cross-Domain Time Series Forecasting."

As shown in Table 1, both papers highlight the transformative potential of LLMs in SCM. The first paper addresses the gap in automating supplier evaluation processes, reducing the reliance on human experts, and demonstrating the potential of LLMs in MCDM tasks. The second paper, "UniTime: A Language-Empowered Unified Model for Cross-Domain Time Series Forecasting", fills the gap in cross-domain time series forecasting by proposing a model that generalizes across multiple domains, thus enhancing predictive performance and operational efficiency.

Table 1 Comparative Analysis of Large Language Model Applications in Supply Chain Management

Study	Methodology	Key Findings	Gaps Addressed
Can ChatGPT Serve as a Multi-Criteria Decision Maker?	Traditional MCDM (AHP, FCE) vs. ChatGPT	ChatGPT's evaluations align closely with human experts, demonstrating efficiency and cost-effectiveness	Automates supplier evaluation, reduces reliance on human experts.
UniTime: A Language-Empowered Unified Model	Cross-domain time series forecasting with Language-TS Transformer	Advances forecasting performance and zero-shot transferability, adaptable to various domains	Generalizes across multiple domains, enhancing predictive performance

4. CONCLUSIONS AND FUTURE WORKS

The integration of LLMs into SCM represents a significant advancement in the field, offering the transformative potential to enhance decision-making, demand forecasting, supplier evaluation, and overall operational efficiency.

This study contributes to the field of SCM by demonstrating the practical applications and benefits of LLMs across various functions. The comprehensive review and case studies provide valuable insights into how LLMs can enhance operational efficiency, decision-making, and resilience in supply chains. The integration of LLMs into SCM processes highlights the potential for significant cost savings, improved supplier management, and more effective logistical planning. Furthermore, the thesis emphasizes the importance of addressing challenges such as user adoption, hallucinations, privacy concerns, and AI advancements to ensure the successful and safe implementation of LLMs in SCM.

To further advance the integration of LLMs in SCM, future research should focus on the following areas:

- Develop fine-tuning methods that address the nuances of supply chain operations, ensuring that LLMs can adapt to diverse SCM environments.
- Explore the integration of LLMs with emerging technologies such as digital twins and IoT to enhance real-time data collection and analysis.
- Assess the potential of combining LLMs with blockchain technology to improve transparency and security in supply chain transactions.
- Research privacy-preserving techniques to ensure the secure use of sensitive supply chain data.
- Develop ethical guidelines and frameworks for the responsible deployment of LLMs in SCM, addressing issues such as bias and fairness.
- Investigate the scalability of LLM-based solutions in large-scale supply chain networks.
- Study the factors influencing user adoption of LLM-based solutions in SCM, including user training and support mechanisms.

By addressing these areas, future research can further unlock the potential of LLMs in SCM, driving innovation, efficiency, and resilience in supply chain operations. The findings of this thesis provide a solid foundation for continued exploration and practical implementation of LLMs in the evolving landscape of supply chain management.

BIBLIOGRAPHICAL REFERENCES

Agrawal, G., Kumarage, T., Alghami, Z., & Liu, H. (2023). *Can Knowledge Graphs Reduce Hallucinations in LLMs? : A Survey*. http://arxiv.org/abs/2311.07914

AhmadiTeshnizi, A., Gao, W., & Udell, M. (2023). *OptiMUS: Optimization Modeling Using MIP Solvers and large language models*. http://arxiv.org/abs/2310.06116

Ali Linkon, A., Shaima, M., Shohail Uddin Sarker, M., Nabi, N., Nasir Uddin Rana, M., Kumar Ghosh, S., Anisur Rahman, M., Esa, H., & Rahat Chowdhury, F. (2024). *Advancements and Applications of Generative Artificial Intelligence and Large Language Models on Business Management: A Comprehensive Review*. https://doi.org/10.32996/jcsts

Alkhaldi, F., Shehadeh, M., Abu-AlSondos, I. A., & Almazaydeh, L. (2023). Artificial Intelligence Applications and Roles in Supply Chain Sustainability. *2023 Asia Conference on Cognitive Engineering and Intelligent Interaction (CEII)*, 120–124. https://doi.org/10.1109/CEII60565.2023.00029

Amarasinghe, P. T., Nguyen, S., Sun, Y., & Alahakoon, D. (2023). *AI-Copilot for Business Optimisation: A Framework and A Case Study in Production Scheduling*. http://arxiv.org/abs/2309.13218

Babai, M. Z., Boylan, J. E., & Rostami-Tabar, B. (2022). Demand forecasting in supply chains: a review of aggregation and hierarchical approaches. *International Journal of Production Research*, *60*(1), 324–348. https://doi.org/10.1080/00207543.2021.2005268

Bahrini, A., Khamoshifar, M., Abbasimehr, H., Riggs, R. J., Esmaeili, M., Majdabadkohne, R. M., & Pasehvar, M. (2023). ChatGPT: Applications, Opportunities, and Threats. *2023 Systems and Information Engineering Design Symposium, SIEDS 2023*, 274–279. https://doi.org/10.1109/SIEDS58326.2023.10137850

Borucka, A. (2023). Seasonal Methods of Demand Forecasting in the Supply Chain as Support for the Company's Sustainable Growth. *Sustainability (Switzerland)*, *15*(9). https://doi.org/10.3390/su15097399

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). *Language Models are Few-Shot Learners*. http://arxiv.org/abs/2005.14165

Chen, J., Lin, H., Han, X., & Sun, L. (2023). *Benchmarking Large Language Models in Retrieval-Augmented Generation*. http://arxiv.org/abs/2309.01431

Chen, S., Hou, Y., Cui, Y., Che, W., Liu, T., & Yu, X. (2020). *Recall and Learn: Fine-tuning Deep Pretrained Language Models with Less Forgetting*. https://github.com/Sanyuan-Chen/RecAdam

Chen, Y., Biswas, M. I., & Talukder, M. S. (2023). The role of artificial intelligence in effective business operations during COVID-19. *International Journal of Emerging Markets*, *18*(12), 6368–6387. https://doi.org/10.1108/IJOEM-11-2021-1666

Chen, Z., Mao, H., Li, H., Jin, W., Wen, H., Wei, X., Wang, S., Yin, D., Fan, W., Liu, H., & Tang, J. (2023). *Exploring the Potential of Large Language Models (LLMs) in Learning on Graphs*. http://arxiv.org/abs/2307.03393

Chowdhury, M., Rifat, N., Latif, S., Ahsan, M., Rahman, M. S., & Gomes, R. (2023). ChatGPT: The Curious Case of Attack Vectors' Supply Chain Management Improvement. *IEEE International Conference on Electro Information Technology*, *2023-May*, 499–504. https://doi.org/10.1109/eIT57321.2023.10187385

DeepSeek-AI, :, Bi, X., Chen, D., Chen, G., Chen, S., Dai, D., Deng, C., Ding, H., Dong, K., Du, Q., Fu, Z., Gao, H., Gao, K., Gao, W., Ge, R., Guan, K., Guo, D., Guo, J., … Zou, Y. (2024). *DeepSeek LLM: Scaling Open-Source Language Models with Longtermism*. http://arxiv.org/abs/2401.02954

Deiva Ganesh, A., & Kalpana, P. (2022). Future of artificial intelligence and its influence on supply chain risk management – A systematic review. *Computers and Industrial Engineering*, *169*. https://doi.org/10.1016/j.cie.2022.108206

Dong, Q., Li, L., Dai, D., Zheng, C., Wu, Z., Chang, B., Sun, X., Xu, J., Li, L., & Sui, Z. (2022). *A Survey on In-context Learning*. http://arxiv.org/abs/2301.00234

El Raoui, H., Oudani, M., & El Hilali Alaoui, A. (2020). Coupling Soft Computing, Simulation and Optimization in Supply Chain Applications: Review and Taxonomy. *IEEE Access*, *8*, 31710–31732. https://doi.org/10.1109/ACCESS.2020.2973329

Fosso Wamba, S., Guthrie, C., Queiroz, M. M., & Minner, S. (2023a). ChatGPT and generative artificial intelligence: an exploratory study of key benefits and challenges in operations and supply chain management. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2023.2294116

Fosso Wamba, S., Guthrie, C., Queiroz, M. M., & Minner, S. (2023b). ChatGPT and generative artificial intelligence: an exploratory study of key benefits and challenges in operations and supply chain management. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2023.2294116

Fosso Wamba, S., Queiroz, M. M., Chiappetta Jabbour, C. J., & Shi, C. (Victor). (2023). Are both generative AI and ChatGPT game changers for 21st-Century operations and supply chain excellence? *International Journal of Production Economics*, *265*. https://doi.org/10.1016/j.ijpe.2023.109015

Giabbanelli, P. J. (2023). *GPT-Based Models Meet Simulation: How to Efficiently Use Large-Scale Pre-Trained Language Models Across Simulation Tasks*. http://arxiv.org/abs/2306.13679

Hendriksen, C. (2023). Artificial intelligence for supply chain management: Disruptive innovation or innovative disruption? *Journal of Supply Chain Management*, *59*(3), 65–76. https://doi.org/10.1111/jscm.12304

Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). *LoRA: Low-Rank Adaptation of Large Language Models*. http://arxiv.org/abs/2106.09685

Hugos, M. (2011). *ESSENTIALS of Supply Chain Management*. www.wiley.com.

Jackson, I., Ivanov, D., Dolgui, A., & Namdar, J. (2024). Generative artificial intelligence in supply chain and operations management: a capability-based framework for analysis and implementation. *International Journal of Production Research*. https://doi.org/10.1080/00207543.2024.2309309

Jackson, I., Jesus Saenz, M., & Ivanov, D. (2024). From natural language to simulations: applying AI to automate simulation modelling of logistics systems. *International Journal of Production Research*, *62*(4), 1434–1457. https://doi.org/10.1080/00207543.2023.2276811

Jackson, I., & Rolf, B. (2023). Do Natural Language Processing models understand simulations? Application of GPT-3 to translate simulation source code to English. *IFAC-PapersOnLine*, *56*(2), 221–226. https://doi.org/10.1016/j.ifacol.2023.10.1572

Jeong, C. (2023). A Study on the Implementation of Generative AI Services Using an Enterprise Data-Based LLM Application Architecture. In *Advances in Artificial Intelligence and Machine Learning; Research* (Vol. 3, Issue 4). https://www.oajaiml.com/

Jia, W., Wang, W., & Zhang, Z. (2023). From simple digital twin to complex digital twin part II: Multi-scenario applications of digital twin shop floor. *Advanced Engineering Informatics*, *56*. https://doi.org/10.1016/j.aei.2023.101915

Kalajdzievski, D. (2024). *Scaling Laws for Forgetting When Fine-Tuning Large Language Models*. http://arxiv.org/abs/2401.05605

Kalyan, K. S., Rajasekharan, A., & Sangeetha, S. (2021). *AMMUS : A Survey of Transformer-based Pretrained Models in Natural Language Processing*. http://arxiv.org/abs/2108.05542

Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). *Scaling Laws for Neural Language Models*. http://arxiv.org/abs/2001.08361

Kashem, M. A., Shamsuddoha, M., Nasir, T., & Chowdhury, A. A. (2023). Supply Chain Disruption versus Optimization: A Review on Artificial Intelligence and Blockchain. *Knowledge*, *3*(1), 80–96. https://doi.org/10.3390/knowledge3010007

Kaswan, K. S., Dhatterwal, J. S., Malik, K., & Baliyan, A. (2023). Generative AI: A Review on Models and Applications. *2023 International Conference on Communication, Security and Artificial Intelligence, ICCSAI 2023*, 699–704. https://doi.org/10.1109/ICCSAI59793.2023.10421601

Keswani, G., Bisen, W., Padwad, H., Wankhedkar, Y., Pandey, S., & Soni, A. (2024). Abstractive Long Text Summarization using Large Language Models. In *Original Research Paper International Journal of Intelligent Systems and Applications in Engineering IJISAE* (Vol. 2024, Issue 12s). www.ijisae.org

Khan, S. A., Chaabane, A., & Dweiri, F. T. (2018). Multi-Criteria Decision-Making Methods Application in Supply Chain Management: A Systematic Literature Review. In *Multi-Criteria Methods and Techniques Applied to Supply Chain Management*. InTech. https://doi.org/10.5772/intechopen.74067

Kosasih, E. E., Papadakis, E., Baryannis, G., & Brintrup, A. (2023). A review of explainable artificial intelligence in supply chain management using neurosymbolic approaches. In *International Journal of Production Research*. Taylor and Francis Ltd. https://doi.org/10.1080/00207543.2023.2281663

Lai, Z., Wu, T., Fei, X., & Ling, Q. (2024). BERT4ST:: Fine-tuning pre-trained large language model for wind power forecasting. *Energy Conversion and Management*, *307*. https://doi.org/10.1016/j.enconman.2024.118331

Lambert, D. M., & Cooper, M. C. (2000). Issues in Supply Chain Management. In *Industrial Marketing Management* (Vol. 29). http://www.CLM1.org.

Li, B., Mellou, K., Zhang, B., Pathuri, J., & Menache, I. (2023). *Large Language Models for Supply Chain Optimization*. http://arxiv.org/abs/2307.03875

Li, Q., Zhang, L., & Mak-Hau, V. (2023). *Synthesizing mixed-integer linear programming models from natural language descriptions*. http://arxiv.org/abs/2311.15271

Lin, T., Wang, Y., Liu, X., & Qiu, X. (2022). A survey of transformers. *AI Open*, *3*, 111–132. https://doi.org/10.1016/j.aiopen.2022.10.001

Liu, C., Zhao, F., Qing, L., Kang, Y., Sun, C., Kuang, K., & Wu, F. (2023). *Goal-Oriented Prompt Attack and Safety Evaluation for LLMs*. http://arxiv.org/abs/2309.11830

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Computing Surveys*, *55*(9). https://doi.org/10.1145/3560815

Liu, X., Hu, J., Li, Y., Diao, S., Liang, Y., Hooi, B., & Zimmermann, R. (2024). UniTime: A Language-Empowered Unified Model for Cross-Domain Time Series Forecasting. *Proceedings of the ACM on Web Conference 2024*, 4095–4106. https://doi.org/10.1145/3589334.3645434

Liu, Y., He, H., Han, T., Zhang, X., Liu, M., Tian, J., Zhang, Y., Wang, J., Gao, X., Zhong, T., Pan, Y., Xu, S., Wu, Z., Liu, Z., Zhang, X., Zhang, S., Hu, X., Zhang, T., Qiang, N., … Ge, B. (2024). *Understanding LLMs: A Comprehensive Overview from Training to Inference*. http://arxiv.org/abs/2401.02038

Lopes, A. P., & Rodriguez-Lopez, N. (2021). A decision support tool for supplier evaluation and selection. *Sustainability (Switzerland)*, *13*(22). https://doi.org/10.3390/su132212387

Lu, S., Bigoulaeva, I., Sachdeva, R., Madabushi, H. T., & Gurevych, I. (2023). *Are Emergent Abilities in Large Language Models just In-Context Learning?* http://arxiv.org/abs/2309.01809

Malik, A. K., Garg, C. P., Singh, A., Jit, S., & Garg, C. P. (2010). Supply chain management: an overview. In *International Journal of Logistics and Supply Chain Management* (Vol. 2, Issue 2). https://www.researchgate.net/publication/284727581

Manuj, I., & Sahin, F. (2011). A model of supply chain and supply chain decision-making complexity. *International Journal of Physical Distribution and Logistics Management*, *41*(5), 511–549. https://doi.org/10.1108/09600031111138844

Mustafee, N., Katsaliaki, K., & Taylor, S. J. E. (2021). Distributed Approaches to Supply Chain Simulation. In *ACM Transactions on Modeling and Computer Simulation* (Vol. 31, Issue 4). Association for Computing Machinery. https://doi.org/10.1145/3466170

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2023). *A Comprehensive Overview of Large Language Models*. http://arxiv.org/abs/2307.06435

Nayal, K., Raut, R., Priyadarshinee, P., Narkhede, B. E., Kazancoglu, Y., & Narwane, V. (2022). Exploring the role of artificial intelligence in managing agricultural supply chain risk to counter the impacts of the COVID-19 pandemic. *International Journal of Logistics Management*, *33*(3), 744–772. https://doi.org/10.1108/IJLM-12-2020-0493

OpenAI, :, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., … Zoph, B. (2023). *GPT-4 Technical Report*. http://arxiv.org/abs/2303.08774

Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2023). *Unifying Large Language Models and Knowledge Graphs: A Roadmap*. http://arxiv.org/abs/2306.08302

Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). *Pre-trained Models for Natural Language Processing: A Survey*. https://doi.org/10.1007/s11431-020-1647-3

Ribeiro, L. F. R., Schmitt, M., & Schütze, H. (2021). *Investigating Pretrained Language Models for Graph-to-Text Generation*. https://github.com/UKPLab/plms-

Richey, R. G., Chowdhury, S., Davis-Sramek, B., Giannakis, M., & Dwivedi, Y. K. (2023). Artificial intelligence in logistics and supply chain management: A primer and roadmap for research. In *Journal of Business Logistics* (Vol. 44, Issue 4, pp. 532–549). John Wiley and Sons Inc. https://doi.org/10.1111/jbl.12364

Rozanec, J. M. (2021). Explainable Demand Forecasting: A Data Mining Goldmine. *The Web Conference 2021 - Companion of the World Wide Web Conference, WWW 2021*, 723–724. https://doi.org/10.1145/3442442.3453708

Shatat, A., & Shatat, A. (2022). The Use of Artificial Intelligence Capabilities in Logistics and Supply Chain Management: Major Challenges and Key Benefits. *2022 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems, ICETSIS 2022*, 137–143. https://doi.org/10.1109/ICETSIS55481.2022.9888901

Shen, T., Jin, R., Huang, Y., Liu, C., Dong, W., Guo, Z., Wu, X., Liu, Y., & Xiong, D. (2023). *Large Language Model Alignment: A Survey*. http://arxiv.org/abs/2309.15025

Skórnóg, D., & Kmiecik, M. (2023). SUPPORTING THE INVENTORY MANAGEMENT IN THE MANUFACTURING COMPANY BY CHATGPT. *Logforum*, *19*(4), 535–554. https://doi.org/10.17270/J.LOG.2023.917

Su, X., Le, T., Bethard, S., & Howard, P. (2023). *Semi-Structured Chain-of-Thought: Integrating Multiple Sources of Knowledge for Improved Language Model Reasoning*. http://arxiv.org/abs/2311.08505

Svoboda, I., & Lande, D. (2024). *Enhancing Multi-Criteria Decision Analysis with AI: Integrating Analytic Hierarchy Process and GPT-4 for Automated Decision Support*.

Takala, J., Liu, Y., Imeri, S., Shahzad, K., & Ali, T. (2014). *Suppliers' Evaluation: An Empirical Study*. https://www.researchgate.net/publication/273573968

Tang, X., Zheng, Z., Li, J., Meng, F., Zhu, S.-C., Liang, Y., & Zhang, M. (2023). *Large Language Models are In-Context Semantic Reasoners rather than Symbolic Reasoners*. https://github.com/XiaojuanTang/ICSR.

Tay, Y., Dehghani, M., Tran, V. Q., Garcia, X., Wei, J., Wang, X., Chung, H. W., Shakeri, S., Bahri, D., Schuster, T., Zheng, H. S., Zhou, D., Houlsby, N., & Metzler, D. (2022). *UL2: Unifying Language Learning Paradigms*. http://arxiv.org/abs/2205.05131

Terzi, S., & Cavalieri, S. (2004). Simulation in the supply chain context: A survey. *Computers in Industry*, *53*(1), 3–16. https://doi.org/10.1016/S0166-3615(03)00104-0

Thangarasa, V., Gupta, A., Marshall, W., Li, T., Leong, K., Decoste, D., Lie, S., & Saxena, S. (2023). *SPDF: Sparse Pre-training and Dense Fine-tuning for Large Language Models*.

Tian, S., Jin, Q., Yeganova, L., Lai, P.-T., Zhu, Q., Chen, X., Yang, Y., Chen, Q., Kim, W., Comeau, D. C., Islamaj, R., Kapoor, A., Gao, X., & Lu, Z. (2023). *Opportunities and Challenges for ChatGPT and Large Language Models in Biomedicine and Health*. https://www.medrxiv.org/

Urlana, A., Kumar, C. V., Singh, A. K., Garlapati, B. M., Chalamala, S. R., & Mishra, R. (2024). *LLMs with Industrial Lens: Deciphering the Challenges and Prospects -- A Survey*. http://arxiv.org/abs/2402.14558

Usman Hadi, M., al tashi, qasem, Qureshi, R., Shah, A., muneer, amgad, Irfan, M., Zafar, A., Bilal Shaikh, M., Akhtar, N., Wu, J., Mirjalili, S., Al-Tashi, Q., & Muneer, A. (2020). *A Survey on Large Language Models: Applications, Challenges, Limitations, and Practical Usage*. https://doi.org/10.36227/techrxiv.23589741.v1

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*. http://arxiv.org/abs/1706.03762

Wang, X., & Wu, X. (2024). Can ChatGPT Serve as a Multi-Criteria Decision Maker? A Novel Approach to Supplier Evaluation. *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 10281–10285. https://doi.org/10.1109/ICASSP48485.2024.10447204

Wasserkrug, S., Boussioux, L., Hertog, D. den, Mirzazadeh, F., Birbil, I., Kurtz, J., & Maragno, D. (2024). *From Large Language Models and Optimization to Decision Optimization CoPilot: A Research Manifesto*. http://arxiv.org/abs/2402.16269

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022). *Emergent Abilities of Large Language Models*. http://arxiv.org/abs/2206.07682

Yahya, M., Breslin, J. G., & Ali, M. I. (2021). Semantic web and knowledge graphs for industry 4.0. In *Applied Sciences (Switzerland)* (Vol. 11, Issue 11). MDPI AG. https://doi.org/10.3390/app11115110

Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Yin, B., & Hu, X. (2023). *Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond*. http://arxiv.org/abs/2304.13712

Zhai, Y., Tong, S., Li, X., Cai, M., Qu, Q., Lee, Y. J., & Ma, Y. (2023). *Investigating the Catastrophic Forgetting in Multimodal Large Language Models*. http://arxiv.org/abs/2309.10313

Zhang, G., Hou, Y., Lu, H., Chen, Y., Zhao, W. X., & Wen, J.-R. (2023). *Scaling Law of Large Sequential Recommendation Models*. http://arxiv.org/abs/2311.11351

Zhang, Y., Chen, Z., Zhang, W., & Chen, H. (2023). *Making Large Language Models Perform Better in Knowledge Graph Completion*. http://arxiv.org/abs/2310.06671

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., … Wen, J.-R. (2023). *A Survey of Large Language Models*. http://arxiv.org/abs/2303.18223

Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G., Zhang, K., Ji, C., Yan, Q., He, L., Peng, H., Li, J., Wu, J., Liu, Z., Xie, P., Xiong, C., Pei, J., Yu, P. S., & Sun, L. (2023). *A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT*. http://arxiv.org/abs/2302.09419