UNIVERSITÀ DEGLI STUDI DI GENOVA

POLYTECHNIC SCHOOL

DIME

Department of Mechanical, Energy, Management and Transport Engineering

MASTER'S THESIS
IN
MANAGEMENT ENGINEERING

Exploring the Capabilities of Large Language Models in Optimizing Supply Chain Operations

Advisor:

Prof. Eng. Flavio Tonelli

Co-Advisor:

Dott. Eng. Antonio Giovannetti

Student:

Valeria Carolina Stufano

March 2024

Exploring the Capabilities of Large Language Models in Optimizing Supply Chain Operations

Abstract

This Master's Thesis investigates the integration of Generative AI, with a focus on Large Language Models (LLMs), within the domain of Supply Chain Management (SCM) to navigate modern market complexities and enhance operational efficiencies. The current SCM environment is marked by increased complexity due to factors such as global competition, heightened consumer expectations, and intricate patterns of customer demand. The shift towards digitalization in supply chains highlights the critical role of data management, demanding enhanced analytical capabilities and cross-functional expertise to ensure efficient operations and decision-making. Given the escalating challenges in SCM, the research emphasizes the necessity for innovative and disruptive solutions. Among these, artificial intelligence, and more specifically Generative AI, emerges as a pioneering innovation, offering a strategic advantage in navigating the evolving landscape of SCM. Focusing on LLMs, including OpenAI's GPT-3.5 Turbo and Mistral-7B-Instruct-v0.2, the study explores their application across key SCM areas: knowledge management, forecasting, supplier relationship management, and customer service enhancements.

The methodology centers on deploying these LLMs as a new layer within a structured framework between humans and data driven operations, utilizing the LangChain toolkit and Python's Pandas library to demonstrate their practical utility in analyzing complex supply chain datasets. The results underscore LLMs' significant potential to streamline SCM processes, indicating notable improvements in data-driven decision-making and predictive analytics.

By systematically evaluating the impacts and applicability of different LLMs, the thesis contributes to the understanding of AI's transformative impact on supply chain operations, offering insights into optimizing strategies and enhancing overall supply chain resilience and sustainability. This work lays the groundwork for future advancements in AI-enhanced supply chains, addressing both practical applications and emerging challenges.

Esplorare le Capacità dei Large Language Models nell'Ottimizzare le Operazioni della Supply Chain

Sommario

Questa tesi di laurea esamina l'integrazione della Generative AI, con un focus sui Large Language Models (LLMs), nel dominio del Supply Chain Management (SCM) per navigare le complessità del mercato moderno e potenziare le efficienze operative. L'attuale ambiente SCM è caratterizzato da una maggiore complessità a causa di fattori quali la concorrenza globale, aspettative dei consumatori elevate e schemi complessi della domanda dei clienti. Il passaggio verso la digitalizzazione nelle catene di fornitura evidenzia il ruolo critico della gestione dei dati, richiedendo capacità analitiche migliorate e competenze interfunzionali per garantire operazioni efficienti e processi decisionali. Data l'escalation delle sfide nell'SCM, la ricerca sottolinea la necessità di soluzioni innovative e disruptive. Tra queste, l'intelligenza artificiale, e più specificamente l'AI Generativa, emerge come un'innovazione pionieristica, offrendo un vantaggio strategico nel navigare il panorama in evoluzione della SCM. Concentrandosi sugli LLMs, inclusi il GPT-3.5 Turbo di OpenAI e il Mistral-7B-Instruct-v0.2, lo studio esplora la loro applicazione in aree chiave dell'SCM: gestione della conoscenza, previsione, gestione delle relazioni con i fornitori e miglioramento del servizio clienti.

La metodologia si concentra sul dispiegamento di questi LLMs come un nuovo strato all'interno di un quadro strutturato tra esseri umani e operazioni basate su dati, utilizzando il toolkit LangChain e la libreria Pandas di Python per dimostrare la loro utilità pratica nell'analizzare complessi dataset della catena di fornitura. I risultati sottolineano il significativo potenziale degli LLMs nell'ottimizzare i processi di SCM, indicando notevoli miglioramenti nel processo decisionale basato sui dati e nell'analisi predittiva.

Valutando sistematicamente l'impatto e l'applicabilità di diversi LLMs, la tesi contribuisce alla comprensione dell'impatto trasformativo dell'AI sulle operazioni della catena di fornitura, offrendo intuizioni su come ottimizzare le strategie e potenziare la resilienza e la sostenibilità complessive della catena di fornitura. Questo lavoro getta le basi per futuri avanzamenti nelle catene di fornitura potenziate dall'AI, affrontando sia le applicazioni pratiche che le sfide emergenti.

Table of Contents

1. Introduction

In global supply chain management, adapting to the modern market's complexities, such as risk management and globalization, is crucial. These challenges, from environmental disruptions to financial instabilities, require innovative solutions. Enter artificial intelligence, especially Large Language Models (LLMs), transforming traditional practices into data-driven, efficient systems. This Master Thesis explores the integration of LLMs to address supply chain challenges, aiming to revolutionize processes through enhanced knowledge management, accurate forecasting, and improved supplier relationships. The journey into AI-enhanced supply chains presents an innovative frontier, albeit with new challenges and ethical considerations, mapping a future of optimized, resilient, and sustainable supply chains.

The primary objective of this thesis is to examine how Generative AI, such as Large Language Models (LLMs) (Mistral and OpenAI's GPT-3.5 Turbo are taken in consideration for the experimentation in this Thesis) , can improve supply chain management processes. This involves developing chatbots and advanced Agents using these LLMs and LAM (Lagre Action Models) to delve into specific areas such as knowledge management, forecasting, supplier relationship management, customer relationship management, and manufacturing. The research aims to assess the potential of LLMs to optimize and streamline these crucial supply chain operations.

This thesis systematically investigates the integration and impact of Generative AI (GenAI) within supply chain management, meticulously laid out over several comprehensive chapters. The journey commences with an initial exploration of the broader context of Artificial Intelligence (AI): starting from its historical roots, moving through its various definitions, and advancing towards the evolution and increasing relevance of Large Language Models (LLMs). This foundation sets the stage for a detailed discourse on AI's classifications—spanning capabilities, functionalities, and learning processes—and delves into the core AI techniques, including Machine Learning (ML), Deep Learning (DL), and particularly Generative AI.

The discourse then narrows down to the intricate web of supply chain management. Beginning with an overview of the current state of supply chains, the discussion expands to illustrate each critical component from sourcing to distribution, highlighting modern challenges like risk management and globalization. The narrative further progresses to articulate how AI technologies, especially GenAI and LLMs, are revolutionizing supply chain facets, including knowledge management and forecasting. McKinsey (2023, 2024) and Accenture (2023) highlight GenAI's transformative impact, estimating its value creation in supply chain improvements to be between $3.5 trillion and $4 trillion. This innovation is set to enhance operational efficiency, cut costs, and generate new revenue streams. Accenture (2023) highlights the substantial productivity savings that derive from GenAI, approximately 20%, underscoring GenAI's critical role in transforming and economically elevating global supply chain operations.

A methodological chapter meticulously outlines the experimental design adopted for the study, segueing into the practical implications of LLMs in analyzing and optimizing supply chain operations. Following this, the thesis presents a nuanced comparative analysis of LLM performance, culminating in a rich discussion on the broader implications and challenges encountered within the context.

Structured to progress from a general understanding to specific applications, each section is aligned with the central aim of this Master Thesis: to elucidate the transformative potential of AI and LLMs in enhancing supply chain efficiency. This narrative approach ensures a holistic and comprehensive exploration, marrying theoretical underpinnings with practical implementations to showcase the significant role of GenAI in modern supply chain dynamics.

The methodological approach of this research focuses on detailing the strategies and techniques used to investigate the efficacy of Large Language Models (LLMs) within the domain of supply chain management. This encompasses the experimental design and specific methodologies adopted, from the utilization of innovative frameworks like LangChain to the integration of advanced data analysis tools. Through a systematic exploration, the research assesses how LLMs, particularly through examples such as OpenAI's GPT-3.5 Turbo and the open-source Mistral-7B-Instruct-v0.2 model, can enhance various aspects of supply chain operations.

By examining both theoretical concepts and practical implementations, this research aims to illustrate how LLMs, particularly GPT-3.5 Turbo and Mistral-7B-Instruct-v0.2, can optimize supply chain operations by enhancing decision-making, operational efficiency, and insight into complex datasets. While showcasing significant advancements, the research also underscores the importance of choosing the right LLMs tailored to specific supply chain needs, advocating for a balanced integration that aligns with business strategies and operational requirements.

While implementing Large Language Models (LLMs) offers significant advantages for supply chain optimization, it also presents challenges. This study underscores the difficulties associated with data complexity and the scalability of models, critical factors to consider for effective integration and utilization within supply chain contexts. The research uncovers specific obstacles, such as the necessity for massive datasets, considerable computational resources, and the intricate process of fine-tuning LLMs to specific needs. Additional real-world challenges include the intricacies of model selection, computational resources constraints, and scalability issues in practical SCM applications. These findings stress the critical need for tailored model selection and the balancing of computational efficiency against operational requirements.

2. Introduction to Artificial Intelligence

2.1. Definition of Artificial Intelligence

As articulated by Russell and Norvig (2020), in the exploration of Artificial Intelligence (AI), various perspectives have historically been pursued. The definition of intelligence has been approached from different angles, ranging from fidelity to human performance to a more abstract, formal concept known as rationality, characterized by doing the "right thing." Moreover, the focus of inquiry spans from internal thought processes and reasoning to external manifestations of intelligent behavior.

This dual perspective, considering human vs. rational and thought vs. behavior, results in four distinct combinations, each garnering support and research efforts. The methodologies employed diverge as well: the pursuit of human-like intelligence involves empirical scientific methods related to psychology, while the rationalist approach combines mathematics and engineering, intertwining with disciplines such as statistics, control theory, and economics.

Let us delve into the four approaches more comprehensively, as elaborated by Russell and Norvig.

2.1.1. Acting Humanly: The Turing Test Approach

Alan Turing's Turing test, proposed in 1950, serves as a thought experiment aiming to circumvent the philosophical ambiguity of whether a machine can think. It posits that a computer passes the test if a human interrogator, through written interaction, cannot distinguish between responses from a person or a computer. The necessary capabilities include natural language processing, knowledge representation, automated reasoning, and machine learning.

Other researchers have proposed the total Turing test, which extends the requirements to interaction with the real world, necessitating computer vision, speech recognition, and robotics.

These six disciplines compose most of Artificial Intelligence.

2.1.2. Thinking Humanly: The Cognitive Modeling Approach

To assert that a program thinks like a human, understanding human thought processes becomes imperative. This understanding can be derived through introspection, psychological experiments, and brain imaging. The interdisciplinary field of cognitive science unites AI computer models with experimental psychology techniques to formulate precise and testable theories of the human mind.

Once we thoroughly understand the mind, we can turn that understanding into a computer program. If the program's actions mimic human behavior, it suggests that some of the program's processes could be similar to those in humans.

The domain of cognitive science combines AI computer models with experimental methods from psychology to formulate precise and verifiable theories about the human mind.

Modern distinctions between algorithmic performance and human modeling have accelerated the development of both AI and cognitive science.

2.1.3. Thinking Rationally: The “Laws of Thought” Approach

The ancient philosopher Aristotle initiated the codification of "right thinking" through syllogisms, which provided patterns for irrefutable reasoning processes. The consequential development of logic in the 19th century facilitated precise notations for statements about objects and the relations among them, allowing for programs to theoretically solve any solvable problem described in logical notation by the mid 1960s. The logicist tradition in AI aims to build intelligent systems based on logical programs.

However, conventional logic relies on certain knowledge, a condition seldom met in reality. Probability theory addresses this gap, allowing rigorous reasoning with uncertain information. Despite this, rational thought alone does not generate intelligent behavior. For that, we need a theory of rational action. Rational thought, by itself, is not enough.

2.1.4. Acting Rationally: The Rational Agent Approach

An agent, defined as something that acts, is expected to operate autonomously, perceive its environment, persist over time, adapt to change, and pursue goals. A rational agent seeks the best or expected outcome under uncertainty. The skills required for the Turing test align with rational agent capabilities.

The rational-agent approach proves advantageous for its generality and scientific development. It surpasses the "laws of thought" approach and facilitates scientific derivation of agent designs achieving provable rationality. The criterion for rationality is defined mathematically and universally applicable.

This approach has prevailed throughout AI's history, adapting logical foundations initially and later incorporating probability theory and machine learning.

In essence, AI has concentrated on constructing agents that "do the right thing," with the definition of the right thing contingent on specified objectives. This pervasive paradigm can be aptly denominated the standard model.

2.2. Classification of Artificial Intelligence

Artificial Intelligence (AI) has drastically transformed the digital landscape, heralding a new era of technological capabilities. AI is categorized into various classifications based on the system's ability to emulate decision-making, learning from historical data, and achieving a degree of self-awareness. See the classification in Fig. 2.1.

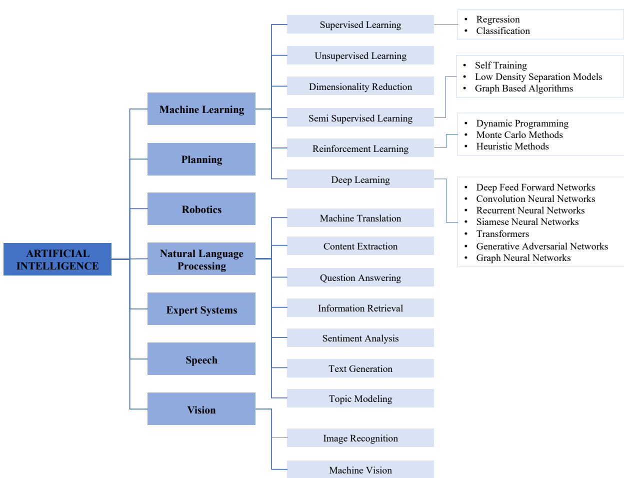


Fig. 2.1. Subsections of Artificial Intelligence

2.2.1. Based on Capabilities

- **Narrow (Weak) AI**

Weak AI, also referred to as Narrow AI, is designed to perform a single or narrow set of tasks. It operates under predefined rules and constraints and does not possess understanding or consciousness beyond its specific application. Examples of Narrow AI include voice assistants like Siri or Alexa, spam filters, and expert systems used in fields such as medicine or customer service. These systems excel in their respective domains but lack the capability to perform tasks beyond their programming. Weak AI systems are characterized by their reliance on symbolic systems or machine learning algorithms to process data and make decisions based on specific input patterns, without the ability to learn for the sake of learning or engage in tasks outside their programmed scope (Finlay, 2018; Rose, 2018).

- **General (Strong) AI**

Strong AI, or General AI, is a type of AI that exhibits broad intelligence across a wide range of environments and problems, similar to human cognitive abilities. Unlike Narrow AI, General AI has the potential to understand, learn, adapt, and implement knowledge across various domains without being confined to a single specialized task. This level of AI would be capable of self-awareness, emotional understanding, and creative problem-solving. It would not only match human intelligence but could potentially surpass it, enabling machines to perform any intellectual task that a human can. Although still theoretical and not yet realized, General AI represents the ultimate goal of AI research, aiming to create machines that can act as lifelike, equally intelligent assistants to humans in everyday life. Strong AI encompasses the ability to learn new languages, understand emotions, and engage in human-like social interactions (Finlay, 2018; Rose, 2018).

- **Superintelligent AI**

"Superintelligence" suggests a future where artificial intelligence significantly surpasses human cognitive abilities, including reasoning and decision-making. This level, known as Artificial Super Intelligence (ASI), encompasses capabilities far beyond the human brain's, transcending the limitations set by our biological neurons. ASI represents a pinnacle of machine intelligence, where not only can machines emulate human thought processes, but they may also vastly exceed them in efficiency and scope. The concept underscores the immense potential of AI to go beyond human tasks, introducing an era where machines could undertake complex challenges effortlessly, tasks currently deemed unfeasible for humans. The progression towards ASI encapsulates the notion that machines might one day not only replicate but also surpass human intellect, marking a transformative leap in what artificial intelligence can accomplish (Reddy, 2020).

2.2.2. Based on Functionality

The classification of AI systems can be divided into four distinct categories based on their capabilities and resemblance to human cognition, as detailed by Hassani et al. (2020):

- **Reactive Machines**

Reactive machines represent the initial phase in AI development. These systems, exemplified by IBM's Deep Blue, are designed to perform specific tasks and respond to a set sequence of inputs. Their operations are not influenced by past interactions or experiences, meaning they lack the capability to learn or adapt over time. They operate based on the present data and predefined algorithms without any memory of past events.

- **Limited Memory**

Building on reactive machines, limited-memory machines incorporate past data into their decision-making process. These AI systems can use historical information to improve future actions. This category encompasses most current AI implementations, such as chatbots, virtual assistants, and autonomous vehicles. They represent a significant step forward in AI technology, capable of adapting their responses based on accumulated experiences.

- **Theory of Mind**

Theory of mind AI remains largely theoretical and represents an advanced stage of AI development that aims to understand and interpret the emotional states and needs of humans. Unlike reactive or limited-memory systems, these AI models aspire to discern the intentions, desires, and beliefs of human beings they interact with, enabling a more nuanced and informed response mechanism.

- **Self-aware AI**

Self-aware AI marks the zenith of AI research, proposing a future where machines possess consciousness akin to humans. Such systems would not only understand but also be aware of their own existence and the emotions of others, mirroring human-like self-awareness. This stage of AI is purely hypothetical and represents the ultimate goal of AI development.

2.2.3. Based on Learning Process

- **Machine Learning**

Machine Learning (ML) is a primary subset of AI, characterized by systems' ability to learn and make predictions or decisions from data, bypassing the need for explicit programming. This process, incorporating techniques such as data mining, helps in identifying patterns and extracting valuable insights using statistical methods. The significance of ML is evident across various digital platforms and embedded technologies, making it a fundamental element for contemporary business models and operations due to its flexibility and growth potential (Sharma & Jain, 2022).

Natural Language Processing (NLP) represents the intersection of human language and computer understanding, enabling seamless communication between the two. By utilizing ML, NLP technologies can interpret and produce human language, transforming spoken dialogue into text and vice versa. This facilitates a range of services from automatic grammar checks in word processors to sophisticated language translation tools, showcasing the versatility and critical role of NLP in improving human-machine interactions (Sharma & Jain, 2022).

In the domain of Automation and Robotics, AI has redefined productivity and efficiency standards. Integrating ML, neural networks, and advanced algorithms, automation has surpassed traditional manufacturing methods, providing solutions that perform repetitive, high-volume tasks without fatigue. Robotic Process Automation (RPA), as an example, demonstrates AI's capability in enhancing operational workflows, leading to more cost-effective and time-efficient production processes (Sharma & Jain, 2022).

- **Deep Learning**

Deep learning, a subset of machine learning techniques, employs complex algebraic circuits as hypotheses, characterized by adjustable connections. The term "deep" denotes the multi-layered structure of these circuits, enabling intricate computation pathways from input to output through numerous stages. This methodology has become predominant in fields such as object recognition in images, language translation, both recognition and synthesis of speech, and image creation. It also has significant applications in reinforcement learning (Russell & Norvig, 2020).

- **Reinforcement Learning**

Reinforcement Learning (RL) involves an agent that learns by interacting with its environment to achieve a goal. In this framework, the agent is rewarded for beneficial actions and penalized for undesirable ones, guiding it to make better decisions over time. This learning method is similar to how a player learns to improve in a game: by understanding which actions lead to winning (rewards) and which to losing (punishments). The agent needs to determine which actions were critical for the outcome and adjust its strategy to increase the likelihood of receiving future rewards. This type of AI is commonly used in various fields, including gaming, robotics, and navigation systems (Russel & Norvig, 2020).

- **Supervised Learning**

Supervised Learning is characterized by the use of labeled datasets to train algorithms to classify data or predict outcomes accurately. The training data include input-output pairs, where the output is known, allowing the AI to learn the mapping function from the input to the output. Over time, the algorithm can apply this function to new, unseen data to make accurate predictions or classifications. This learning process is analogous to a student learning from a teacher where the correct answers are provided; the student's job is to learn the relationship between the questions and the correct answers. Supervised learning is prevalent in applications such as image and speech recognition, medical diagnosis, and spam detection (Russel & Norvig, 2020).

- **Unsupervised Learning**

In Unsupervised Learning, AI systems learn to identify complex patterns and relationships within a dataset without any external guidance or labeled outcomes. This type of learning is akin to a scenario where an individual tries to make sense of a situation by identifying patterns and organizing information based on their intrinsic properties. The most common unsupervised learning task is clustering, where the system attempts to group data points with similar features. Unsupervised learning algorithms are crucial in data mining, anomaly detection, and customer segmentation applications (Russel & Norvig, 2020).

- **Semi-supervised Learning**

Semi-supervised Learning falls between supervised and unsupervised learning. In this approach, the AI is trained on a limited set of labeled data supplemented by a large amount of unlabeled data. This method is particularly useful when acquiring a fully labeled dataset is expensive or laborious. Semi-supervised learning is beneficial in scenarios where the additional unlabeled data can provide a more comprehensive view of the underlying structure of the data, helping to improve learning accuracy. It is widely used in language processing, web content classification, and image recognition where labeling large sets of data may be impractical (Russel & Norvig, 2020).

- **Self-supervised Learning**

Self-supervised Learning is a subset of unsupervised learning where the system generates its own supervisory signal from the input data. Typically, the algorithm predicts part of the data from other parts, thereby learning features that can be used for a designed task. This approach is particularly effective in scenarios where large quantities of unlabeled data are available. Self-supervised learning enables AI systems to learn representations from the data itself, providing a foundation for subsequent tasks like classification or anomaly detection. It is increasingly used in areas such as natural language understanding and computer vision (Russel & Norvig, 2020).

2.2.4. Additional AI Concepts and Technologies

- **Fuzzy Logic** introduces an approach to computing based on "degrees of truth" rather than the usual "true or false" (1 or 0) Boolean logic on which the modern computer is based. This methodology allows for more nuanced decision-making, reflecting the complexity of human reasoning and effectively handling the uncertainties of real-world situations. Fuzzy logic's practical applications range from consumer electronics to advanced control systems, illustrating its utility in diverse scenarios where binary logic is insufficient (Sharma & Jain, 2022).
- **Expert Systems**, an early success of AI, employ the specific knowledge and experience of human experts to offer guidance or make decisions. The efficiency of these systems is directly linked to the quantity and quality of the knowledge they hold. Their use in applications like search engines, which help in identifying spelling or grammatical inaccuracies, underscores the contribution of expert systems to problem-solving and information dissemination (Sharma & Jain, 2022).

By examining these classifications and methodologies, one gains a comprehensive understanding of AI's diverse and evolving landscape. The trajectory of AI and ML technologies showcases a move towards increasingly sophisticated systems capable of tackling complex, real-world challenges, marking a significant step in the journey of digital innovation.

2.3. A Brief History of AI: From its Origins to the Present Day

The introduction of the term field “Atificial Intelligence” dates back to 1956 when John McCarthy organized a workshop, laying the foundation for subsequent pioneering work by McCarthy, Minsky, Rochester, Shannon, Samuel, Selfridge, Solomonoff, Newell, Simon, and others (McCarthy et al., 2006; Solomonoff, 1985). Alan Turing, in his article "Computing Machinery and Intelligence," proposed the idea of designing computers capable of autonomous learning (Turing, 1950), although Charles Babbage, an English mathematician and inventor, conceived the idea of a programmable machine nearly a century before Turing's seminal work. Babbage's conceptualization of the Analytical Engine in the 1830s established the theoretical foundations for the programmable computers that would eventually embody AI systems. The late '90s witnessed a surge in AI's prominence, driven by advancements in computing power, the internet's data gathering capabilities, and statistical techniques enabling solutions derived from vast datasets (Dash, McMurtrey, Rebman & Kar, 2019).

Technological progress in the last two decades has seen the emergence of AI as a powerful force, with technologies such as Cognitive Computing, Computer Vision, Context-aware Computing, Natural Language Processing, Predictive Analytics, Machine Learning, Reinforcement Learning, Supervised Learning, Unsupervised Learning, and Deep Learning providing a conceptual framework for processing input and making informed decisions.

In Fig. 2.2, a summarized timeline of the history of AI is depicted.

Fig. 2.2. Timeline of the history of AI

2.3.1. The Emergence of Artificial Intelligence (1943-1956)

The foundational work marking the inception of what is now recognized as artificial intelligence (AI) was carried out by Warren McCulloch and Walter Pitts in 1943. Drawing inspiration from Nicolas Rashevsky's mathematical modeling work and incorporating insights from the physiology of neurons, propositional logic by Russell and Whitehead, and Turing's theory of computation, they proposed a model of artificial neurons. These neurons, conceptualized as being either "on" or "off," demonstrated the potential to perform computations, implement logical connectives, and even exhibit learning through suitably defined networks.

In 1950, Harvard undergraduates Marvin Minsky and Dean Edmonds constructed the first neural network computer, SNARC, simulating a network of 40 neurons. Minsky continued his exploration of universal computation in neural networks at Princeton, garnering attention and, reportedly, support from John von Neumann.

The landscape of early AI initiatives included checkers-playing programs independently developed in 1952 by Christopher Strachey at the University of Manchester and Arthur Samuel at IBM. However, Alan Turing's influential vision, presented in his 1950 article "Computing Machinery and Intelligence," encompassed concepts like the Turing test, machine learning, genetic algorithms, and reinforcement learning.

John McCarthy, in 1955, orchestrated a pivotal event by convening a workshop at Dartmouth College in 1956. Attended by notable figures like Claude Shannon, Nathaniel Rochester, Allen Newell, and Herbert Simon, the workshop aimed to explore the possibility of creating machines that could simulate human intelligence. Although the Dartmouth workshop did not yield immediate breakthroughs, it laid the groundwork for subsequent developments in the field.

Simon and Newell presented their Logic Theorist (LT), a mathematical theorem-proving system, claiming to have invented a computer program capable of non-numerical thinking. Despite significant achievements, such as proving theorems from Principia Mathematica, the reception of their work, including the rejection of a paper coauthored by Newell, Simon, and Logic Theorist, reflected the complexities and challenges in the early pursuit of AI.

The foundational concepts and early milestones in AI's history, such as the work of McCulloch and Pitts, Turing's contributions, and the Dartmouth workshop, draw from the detailed account by Russell and Norvig (2020).

2.3.2. Early Optimism and High Expectations (1952-1969)

During the 1950s, the intellectual community largely embraced the notion that "a machine can never do X," as delineated by Turing. In response, AI researchers embarked on a series of demonstrations challenging various aspects of X, focusing particularly on tasks deemed

indicative of human intelligence, such as games, puzzles, mathematics, and IQ tests. John McCarthy humorously referred to this period as the "Look, Ma, no hands!" era.
Building on the success of the Logic Theorist (LT), Newell and Simon introduced the General Problem Solver (GPS). Unlike LT, GPS was explicitly designed to emulate human problem-solving procedures. This program embodied the "thinking humanly" approach, marking a shift toward models of cognition that mirrored human reasoning processes. The triumph of GPS and subsequent AI programs led to the formulation of the physical symbol system hypothesis, positing that a system, human or machine, exhibiting intelligence must manipulate data structures composed of symbols.
Nathaniel Rochester and his colleagues at IBM produced noteworthy AI programs, including Herbert Gelernter's Geometry Theorem Prover (1959), a precursor to modern mathematical theorem provers. Arthur Samuel's pioneering work on checkers in 1956, utilizing reinforcement learning, demonstrated that computers could surpass human instruction, a concept later echoed by contemporary systems like TD-GAMMON and ALPHA GO.
In 1958, John McCarthy's contributions included defining the high-level language Lisp and proposing a conceptual framework for AI systems based on knowledge and reasoning, exemplified by the hypothetical Advice Taker. Marvin Minsky, joining MIT in 1958, pursued a different path, emphasizing practical program development over formal logic. McCarthy's vision at Stanford focused on logic-based methods, with the resolution method, a complete theorem-proving algorithm for first-order logic, playing a pivotal role.
At MIT, Minsky supervised projects exploring limited domains, or "microworlds," addressing problems that seemingly required intelligence to solve. Notable examples include James Slagle's SAINT program (1963), Tom Evans's ANALOGY program (1968), and Daniel Bobrow's STUDENT program (1967).
In 1966, one of the first chatbots, ELIZA, was introduced, simulating a conversation as a psychotherapist, and offering an early glimpse into the potential for AI in natural language processing and human-computer interaction. Despite its rudimentary capabilities, ELIZA marked a significant step in AI development by attempting to understand and emulate human conversation.
The blocks world, a microworld featuring solid blocks on a tabletop, became a focal point for various AI projects at MIT. Simultaneously, work building on the neural networks of McCulloch and Pitts flourished, with contributions from Shmuel Winograd, Jack Cowan, Bernie Widrow, and Frank Rosenblatt.
This period's optimism and significant achievements, including the development of the Logic Theorist and General Problem Solver, are based on analyses presented by Russell and Norvig (2020).

2.3.3. A Reality Check (1966-1973)

From the beginning, AI researchers were not shy about making predictions of their coming successes. The following statement by Herbert Simon in 1957 is often quoted:
"It is not my aim to surprise or shock you—but the simplest way I can summarize is to say that there are now in the world machines that think, that learn and that create. Moreover, their ability to do these things is going to increase rapidly until—in a visible future—the range of problems they can handle will be coextensive with the range to which the human mind has been applied" (Simon, H., 1957).
However, Simon's predictions, such as a computer becoming a chess champion within 10 years, didn't fully materialize until 40 years later. Early AI systems, despite promising

performances on simple tasks, encountered significant challenges with more complex problems.
Two primary reasons contributed to the early failure of AI. Firstly, many AI systems relied on "informed introspection," mimicking human approaches without a comprehensive analysis of the task itself. Secondly, there was a lack of understanding of the inherent complexity in the problems AI aimed to solve.
Early problem-solving systems often employed a strategy of trying various combinations until a solution was found, which worked well for small-scale problems. However, the illusion of limitless computational power shattered when faced with larger problems. The notion that scaling up only required faster hardware and larger memories proved overly optimistic.
The failure to grasp the challenge of "combinatorial explosion" was a significant critique of AI, as highlighted in the Lighthill report (Lighthill, 1973), which formed the basis for the decision by the British government to end support for AI research in all but two universities.
Another obstacle emerged due to fundamental limitations in the structures generating intelligent behavior. Minsky and Papert's book "Perceptrons" (1969) demonstrated that perceptrons, a simple form of neural network, could only represent a limited set of functions. This limitation, though not applicable to more complex networks, led to a decline in funding for neural-net research until a resurgence in the late 1980s and 2010s with the development of new back-propagation learning algorithms (Minsky & Papert, 1969; Lighthill, J., 1973).
The challenges and recalibrations in AI's journey, highlighted by the Lighthill report and Minsky and Papert's critique of perceptrons, reflect insights from Russell and Norvig (2020).

2.3.4. Expert Systems (1969-1986)

The depiction of problem-solving that emerged during the initial decade of AI research portrayed a general-purpose search mechanism attempting to connect elementary reasoning steps for comprehensive solutions. These approaches, often termed weak methods, were general but struggled to scale to larger or more challenging problem instances. The alternative, employing more potent, domain-specific knowledge, facilitated more extensive reasoning steps, adept at handling typical cases within narrow areas of expertise.
An early exemplar of this approach was the DENDRAL program, developed at Stanford to infer molecular structure from mass spectrometer data. DENDRAL strategically utilized well-known patterns in mass spectrometry, reducing the number of possible structures significantly. Notably, DENDRAL marked the inception of knowledge-intensive systems, relying on large sets of specialized rules rather than first principles (Feigenbaum et al., 1971).
The subsequent major endeavor, the MYCIN system, addressed the diagnosis of blood infections, outperforming junior doctors with its approximately 450 rules. Unlike DENDRAL, MYCIN acquired its rules through extensive expert interviews, incorporating a calculus of uncertainty called certainty factors to reflect the uncertainty in medical knowledge.
The first commercially successful expert system, R1, commenced operations at Digital Equipment Corporation, saving an estimated $40 million annually by 1986. Its success spurred the widespread adoption of expert systems across major corporations, highlighting the significance of domain knowledge in real-world problem-solving.
In the domain of natural language understanding, researchers like Eugene Charniak and Roger Schank argued for the necessity of general knowledge about the world and a method to utilize that knowledge for robust language understanding. The development of expert

systems led to the creation of various representation and reasoning tools, ranging from logic-based systems like Prolog to structured approaches inspired by Minsky's frames (1975).
The AI industry flourished in the 1980s, with the Japanese "Fifth Generation" project and the U.S. Microelectronics and Computer Technology Corporation. Despite ambitious goals, these projects did not achieve the expected impact.
The AI industry experienced a boom in the 1980s, growing from a few million dollars in 1980 to billions in 1988. This growth included hundreds of companies specializing in expert systems, vision systems, robotics, and dedicated software and hardware. However, this period was followed by the "AI winter," characterized by the failure of many companies to deliver on extravagant promises, revealing the challenges of building and maintaining expert systems for complex domains.
The rise of expert systems and their impact, illustrated through DENDRAL and MYCIN, are discussed in the context provided by Russell and Norvig (2020).

2.3.5. The Return of Neural Networks (1986-Present)

In the mid-1980s, the back-propagation learning algorithm, initially developed in the early 1960s, experienced a resurgence as at least four different groups independently reinvented it. This algorithm found applications in various computer science and psychology learning problems. The excitement generated by these developments, disseminated in the collection "Parallel Distributed Processing" (Rumelhart and McClelland, 1986), marked the emergence of connectionist models, viewed by some as direct competitors to symbolic and logicist approaches.
Connectionist models, championed by figures like Geoff Hinton, posed a challenge to the notion that humans manipulate symbols at a logical level. Hinton referred to symbols as the "luminiferous aether of AI," suggesting that connectionist models, with their ability to form internal concepts in a fluid and imprecise manner, might better suit the complexities of the real world. These models exhibited the capability to learn from examples, adjusting parameters to enhance performance on future tasks.
The resurgence of neural networks and the development of back-propagation learning algorithms are outlined based on the comprehensive review by Russell and Norvig (2020).

2.3.6. Probabilistic Reasoning and Machine Learning (1987-Present)

The limitations of expert systems' brittleness prompted a shift toward a more scientific approach in AI, incorporating probability over Boolean logic, machine learning over hand-coding, and experimental results over philosophical claims. This era favored building on existing theories, relying on rigorous theorems, and demonstrating relevance to real-world applications.
The adoption of shared benchmark problem sets became common for showcasing progress, including repositories like the UC Irvine repository, competitions like the International Planning Competition, and datasets such as LibriSpeech, MNIST, ImageNet, and COCO. This period marked a departure from AI's early isolationism, recognizing the value of integrating machine learning with information theory, uncertain reasoning with stochastic modeling, and search with classical optimization and control.
The field of speech recognition exemplifies this shift. In the 1970s, diverse ad hoc approaches were attempted, but in the 1980s, hidden Markov models (HMMs) dominated. HMMs provided a rigorous mathematical framework, enabling researchers to build on decades of mathematical results from other fields. This shift contributed to the widespread

application of speech technology and character recognition in industrial and consumer domains.
The year 1988 played a pivotal role in connecting AI with other fields, notably through Judea Pearl’s "Probabilistic Reasoning in Intelligent Systems," which ushered in a new acceptance of probability and decision theory in AI. This period also witnessed Rich Sutton’s work connecting reinforcement learning to the theory of Markov decision processes (MDPs), significantly influencing AI planning research and finding applications in robotics and process control.
The AI field's renewed emphasis on data, statistical modeling, optimization, and machine learning led to the reintegration of subfields like computer vision, robotics, speech recognition, multiagent systems, and natural language processing. This reunification brought substantial benefits in both practical applications, such as the widespread deployment of practical robots, and a more cohesive theoretical understanding of AI's core problems.
The shift towards probabilistic reasoning and the integration of machine learning into AI, as described by Russell and Norvig (2020), marks a significant evolution in the field.

2.3.7. Big Data (2001-Present)

The advent of substantial computing power and the emergence of the World Wide Web have given rise to an era characterized by vast datasets, often termed "big data." These datasets encompass trillions of words, billions of images, hours of speech and video, along with extensive genomic, vehicle tracking, clickstream, and social network data.
This era has spurred the development of learning algorithms tailored for immense datasets. Often, a significant portion of examples in these datasets lacks labels, requiring specialized learning algorithms. Yarowsky’s (1995) work on word-sense disambiguation is illustrative, showcasing that with large datasets, algorithms can achieve high accuracy, even with unlabeled examples. Banko and Brill (2001) emphasized the substantial performance gains obtained by increasing dataset size compared to algorithmic tweaks.
In computer vision tasks, such as image restoration, the availability of millions of images proved transformative. Hays and Efros (2007) demonstrated improved hole-filling in photographs by blending pixels from similar images, achieving superior results with vast datasets. The ImageNet database's tens of millions of images ignited a revolution in computer vision.
The convergence of big data and machine learning marked a turning point, restoring commercial appeal to AI. Notably, IBM’s Watson triumphing over human champions in the Jeopardy! quiz game in 2011 underscored the impact of big data on AI's public perception.
The era of big data and its implications for AI, including the development of algorithms for large datasets, are based on discussions from Russell and Norvig (2020).

2.3.8. Deep Learning (2011-Present)

Deep learning, a form of machine learning utilizing multiple layers of adaptable computing elements, traces its roots back to experiments in the 1970s. Convolutional neural networks, a type of deep learning, achieved some success in handwritten digit recognition in the 1990s (LeCun et al., 1995). However, it was in 2011 that deep learning methods gained significant traction, initially in speech recognition and subsequently in visual object recognition.
The turning point occurred during the 2012 ImageNet competition, where a deep learning system from Geoffrey Hinton’s group demonstrated substantial improvement over previous systems, mainly relying on handcrafted features. Since then, deep learning has surpassed

human performance in various vision tasks, with similar strides seen in speech recognition, machine translation, medical diagnosis, and game playing. Notably, ALPHAGO's victories over leading human Go players were facilitated by a deep network for the evaluation function.

These breakthroughs have reignited interest in AI across diverse sectors, captivating students, companies, investors, governments, and the general public. The trend is marked by frequent reports of new AI applications achieving or surpassing human performance, fueling speculations about accelerated success or the prospect of a new AI winter.

Deep learning's efficacy relies heavily on robust hardware capabilities. Standard computer CPUs, performing at 10^9 or 10^10 operations per second, are eclipsed by specialized hardware like GPUs, TPUs, or FPGAs, capable of executing between 10^14 and 10^17 operations per second. Additionally, the success of deep learning hinges on abundant training data and specific algorithmic strategies.

The advancements and impact of deep learning, from its roots to its role in surpassing human performance in various tasks, are drawn from the insights of Russell and Norvig (2020).

2.4. Fundamental Techniques and Types of AI

2.4.1. Machine Learning

Machine learning encompasses a vast array of techniques aimed at deriving insights and making predictions based on historical data. This field fundamentally operates on the principle of either minimizing a cost function or maximizing a reward, utilizing a diverse set of functions, parameters, and weights to construct predictive models (Anderson & Coveyduc, 2020). As an interdisciplinary science at the intersection of statistics, artificial intelligence, and computer science, machine learning, also known as predictive analytics or statistical learning, has become ubiquitous in modern technology. Its applications range from personalized content recommendations on digital platforms to sophisticated image recognition systems, illustrating its critical role in enhancing user experiences across various online services (Müller & Guido, 2017).

The processes and algorithms in machine learning draw parallels with human cognitive abilities, where the identification of patterns and correlations within data sets enables the forecasting of outcomes for new, unseen scenarios. This learning mechanism is reminiscent of human observation and experience, where accumulated knowledge progressively sharpens our decision-making skills. Such an analogy underscores machine learning's capability to evolve and adapt, thereby improving its predictive accuracy over time (Finlay, 2018). The genesis of machine learning can be traced back to the 1950s, a period characterized by groundbreaking advancements in artificial intelligence. The discipline is defined by its focus on using computational methods to leverage past experiences for future performance enhancement or accurate prediction. It encompasses various types of learning, including supervised, where models predict outcomes based on labeled examples; unsupervised, which involves discovering hidden patterns in data without predefined labels; and reinforcement learning, where models learn optimal behaviors through trial and error to achieve specific goals (Akerkar, 2019).

Diving deeper into machine learning's techniques, we find that it is broadly categorized into four main areas: classification, clustering, association learning, and numeric prediction. Each area serves distinct purposes, such as classifying documents into categories, clustering similar documents, uncovering relationships between different features, and predicting numerical values. These methodologies facilitate the organization, analysis, and interpretation of large data sets, enabling both the discovery of new knowledge and the application of such insights in various domains, from natural language processing to autonomous systems (Anderson & Coveyduc, 2020; Müller & Guido, 2017; Finlay, 2018; Akerkar, 2019).

2.4.2. Deep Learning

The conceptual foundation of deep learning traces back to initial attempts at simulating the brain's neuronal networks, leading to the development of what are now known as neural networks. Despite their name, the actual similarity of these networks to biological neural structures is minimal (Russell & Norvig, 2020).

Akerkar (2019) describes deep learning as employing multiple hierarchical layers for processing data in a non-linear fashion, allowing simple concepts at lower levels to inform more complex concepts at higher levels. Deep learning methods, through their multi-layered nonlinear information processing capabilities, are adept at tasks involving large volumes of complex data, such as understanding natural language, image processing, and other tasks that mimic human cognitive functions. Pioneering efforts by entities like Google, under the leadership of Andrew Ng, have underscored the potential of deep learning in handling such complex datasets with efficiency and a human-like understanding.

Deep learning is distinguished by its layered processing approach, favoring a hierarchical model where each layer of processing builds upon the previous one, often through unsupervised pre-training. This methodology is especially effective in managing and interpreting large-scale data by organizing it into layers based on various attributes such as time or nature. The technique is broadly divided into three architectural categories: generative, discriminative, and hybrid, each with its unique approach to data processing and network training. The generative model focuses on layer-by-layer pre-training, the discriminative model integrates outputs for deeper analysis, and the hybrid model combines features of both to enhance deep learning capabilities (Akerkar, 2019).

Fig. 2.3 illustrates the hierarchical relationship among Artificial Intelligence (AI), Machine Learning, and Deep Learning, highlighting how each field encompasses and builds upon the other.

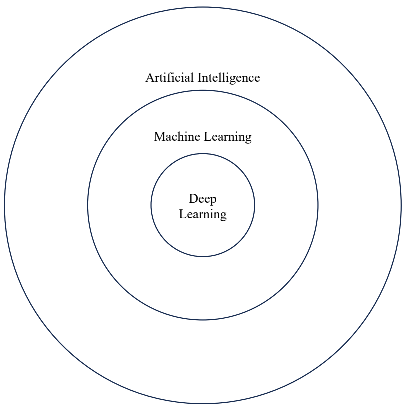


Fig. 2.3. Venn Diagram Describing How Deep Learning Relates to AI

Mukhamediev et al. (2022) expand upon the deep learning landscape, showcasing its versatility and expanding capabilities. With architectures such as Convolutional Neural

Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs), deep learning transcends traditional data processing methods, offering innovative solutions for real-world applications. CNNs are pivotal in image processing tasks, employing filters to detect patterns and features within images. RNNs excel in handling sequential data, making them ideal for speech recognition and natural language processing. GANs, on the other hand, introduce a novel framework where two networks, the generator and discriminator, work in opposition to generate new, synthetic instances of data that are nearly indistinguishable from real data.

These deep learning components form a comprehensive ecosystem that addresses a wide range of AI challenges, from the recognition of complex patterns in vast datasets to the creation of content that mimics real-life artifacts. The integration of these architectures enables the development of sophisticated AI systems capable of learning, adapting, and performing tasks that were once deemed exclusive to human intelligence.

Deep learning continues to be the fastest-growing segment within AI, propelled by its ability to employ deep neural networks for end-to-end problem-solving. This approach reduces preliminary data processing requirements, as networks independently deduce patterns and significant features directly from input data. However, the efficacy of deep learning networks depends heavily on having substantial training data and the correct architectural choices for the neural networks (Mukhamediev et al., 2022).

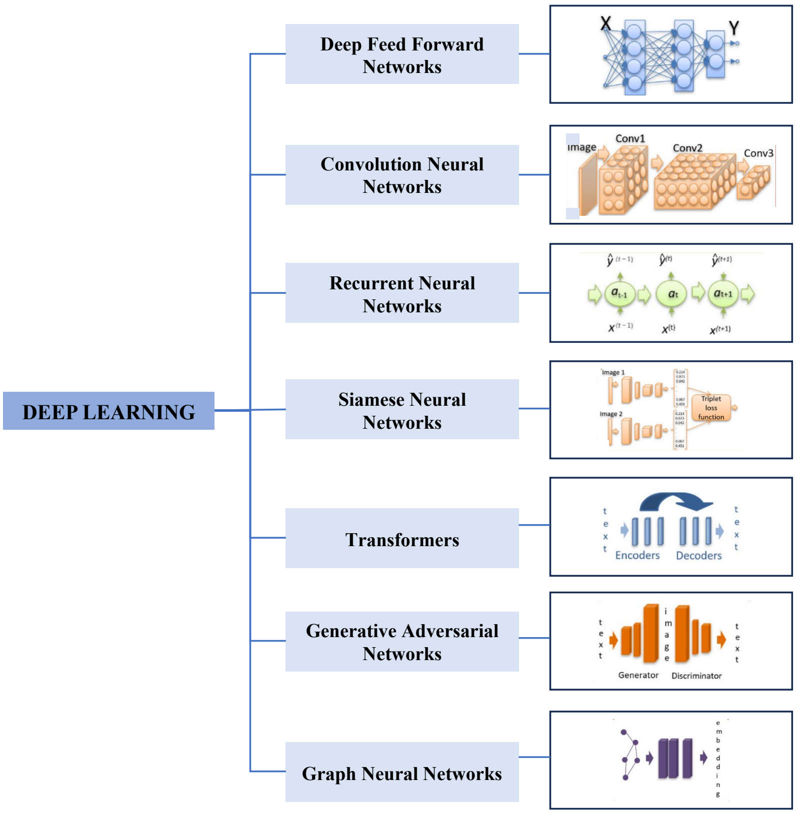


Fig. 2.4. Deep networks

2.4.3. Neural Networks: MLP, LSTM, GRU

Rajendra Akerkar (2019) delves into the intersection of Artificial Intelligence and business, spotlighting the transformation of biological neural networks into their artificial counterparts for machine learning. These artificial neural networks (ANNs) serve as the foundation for deep learning, which stacks multiple layers of ANNs to process data in complex ways.

A neural network operates through a network of input-output units, each linked by connections with specific weights. It mimics the structure of biological neurons, where the output of one neuron becomes the input for others, creating a web of interconnected neurons. The architecture of a neural network, determined by the number and arrangement of neurons, plays a crucial role in its functionality. Neurons are categorized based on their roles into input, hidden (or working), and output neurons, facilitating the flow of signals through the network.

The development of a neural network is an ongoing process, involving changes in neuron connections, states, and weights over time. This evolution is divided into three main phases: architectural updates that shape the network's structure, computational changes that adjust neuron states based on inputs, and adaptive adjustments that fine-tune connection weights for optimal performance. While these phases offer a structured approach to understanding neural network dynamics, they simplify the continuous and concurrent nature of changes within a biological nervous system.

Neural networks excel in pattern recognition and noise tolerance, outperforming other classifiers in handling complex and noisy data sets. However, their outputs, often in the form of symbolic patterns, can be challenging to interpret compared to other classification methods. Their intricate connectivity allows for the analysis of high-order variable interactions and correlated data, making them highly effective for a range of business applications. These include sales forecasting, industrial process control, customer research, data validation, risk management, and targeted marketing, showcasing their widespread utility in solving real-world business challenges.

Neural networks are the backbone of many current artificial intelligence (AI) systems, inspired by biological neural networks. They involve units known as neurons, which process information and are interconnected by weights, influencing the signal strength between neurons. Neural networks can vary from simple structures like the perceptron, with a single layer of input and output, to complex multi-layer networks, often referred to as deep neural networks. In these structures, multiple layers of neurons process the input sequentially, where each layer's output becomes the input for the next layer, enhancing the network's ability to learn from data (Aggarwal, 2018).

A **Multi-Layer Perceptron** (MLP) consists of an input layer, multiple hidden layers, and an output layer. Unlike single-layer networks, where computations are straightforward, MLPs perform complex transformations using hidden layers. The hidden layers allow for the processing and transformation of inputs into a form that the output layer can use. This architecture enables the network to learn from data in a deeper and more nuanced manner, making it capable of handling complex, non-linear problems. They are known as feed-forward networks as information moves forward from input to output. The structure is defined by the number of layers and the type of nodes, usually fully connected. MLPs can be used for various predictions, adapting the loss function accordingly, like cross-entropy for classification and squared loss for regression tasks. They are also subject to overfitting, particularly when they have a large number of parameters compared to the dataset size (Aggarwal, 2018).

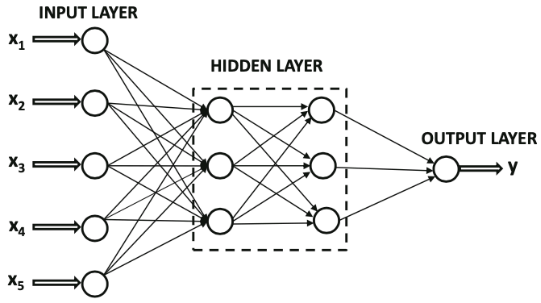


Fig. 2.5. Basic architecture of a feed-forward network with two hidden layers and a single output layer

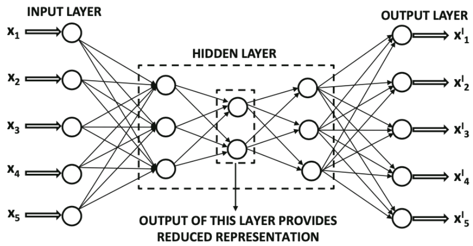


Fig. 2.6. An example of an autoencoder with multiple outputs

Long-Short Term Memory (LSTM) networks, a special kind of Recurrent Neural Networks (RNNs), address the issue of learning long-term dependencies. Traditional RNNs suffer from vanishing and exploding gradients, making it hard to retain information over many time steps. LSTMs introduce cell states and structured gates, allowing information to flow across many time steps without alteration. This architecture enables the preservation and careful modification of information, thus facilitating learning over sequences of considerable length. LSTMs are widely used in sequence prediction problems, like language modeling and translation, due to their efficiency in managing sequence information over time (Aggarwal, 2018).

Gated Recurrent Units (GRUs) simplify the LSTM architecture by combining the input and forget gates into a single update gate and merging the cell and hidden states, This results in fewer parameters and a simpler model structure while maintaining the ability to manage long-term dependencies. Despite their simplified structure, GRUs perform similarly to LSTMs in many tasks and offer advantages in terms of computational efficiency and simplicity. They adapt the flow of information by using reset and update gates, allowing each unit to retain or discard information based on the relevance to the task. They are particularly effective in scenarios where LSTM's extended capabilities are not strictly needed, providing a more efficient alternative without significantly compromising performance. GRUs achieve a balance between the ability to model long-term dependencies and computational efficiency, making them a popular choice in the design of recurrent neural networks (Aggarwal, 2018).

2.4.4. Prediction Models: ARIMA

The ARIMA model, which stands for Autoregressive Integrated Moving Average, is a time series forecasting method that combines autoregressive (AR) features, differencing (Integrated part), and moving average (MA) features. The process can be denoted as

ARIMA(p,d,q), where p, d, and q are non-negative integers that stand for the order of the autoregressive part, the degree of differencing, and the order of the moving average part, respectively (Shumway & Stoffer, 2017).
In this context, "autoregressive" refers to the use of past values in the regression equation for the time series. Specifically, an AR(p) model is formulated as in equation (2.1):

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + W_t \qquad (2.1)$$

where:

$\phi_1, \phi_2, \dots, \phi_p$ are parameters of the model,

X_t is the value of the time series at time t, and

W_t is white noise.

The "integrated" part of ARIMA indicates that the data have been differenced d times to induce stationarity. Differencing is the process of computing the differences between consecutive observations. This process makes the series stationary, which is a requirement for the AR and MA components of the model to be applicable. The notation in equation (2.2) denotes d-th order differencing, where B is the backshift operator.

$$(1 - B)^d X_t \qquad (2.2)$$

The "moving average" component involves modeling the error term as a linear combination of error terms at various times in the past. An MA(q) model has the form in equation (2.3):

$$W_t = \theta_1 W_{t-1} + \theta_2 W_{t-2} + \cdots + \theta_q W_{t-q} + \epsilon_t \qquad (2.3)$$

where:

W_t is the white noise at time t and

$\theta_1, \theta_2, \dots, \theta_q$ are the parameters of the model.

The ARIMA model is then a combination of these components, expressed as equation (2.4):

$$\Phi(B)\, \nabla^d\, X_t = \Theta(B)\, W_t \qquad (2.4)$$

where:

$\Phi(B)$ and $\Theta(B)$ are the polynomials in the backshift operator B for the AR and MA parts, respectively,

and ∇^d represents the differencing operator.

Forecasting with ARIMA involves estimating the parameters of the model from historical data and then using the model to predict future values. Diagnostics such as the ACF and PACF plots, as well as criteria like AIC or BIC, are used to identify the best-fitting ARIMA model among different combinations of p, d, and q (Shumway & Stoffer, 2017).

2.4.5. Generative AI

Feuerriegel et al. (2023) discuss the capabilities and impact of Generative AI, a subfield of artificial intelligence, that is fundamentally about creating new content. It encompasses computational methods capable of producing new and significant content from existing data, such as text, images, or audio.

The authors provide the following classification based on the type of output:

- **Text generation:** Generative AI in text generation involves models like GPT-4 and LLAMa 2, which are built to create textual content. These models process and generate text, effectively simulating human writing styles. They are utilized in various applications, from creating conversational responses in chatbots to generating narrative content. The emphasis here is on their capability to produce coherent, contextually relevant, and diverse text forms, making them instrumental in automating and personalizing written communication.
- **Image/Video generation:** In the area of visual content, GenAI utilizes models such as Stable Diffusion and DALL-E 2 to generate images and videos from textual descriptions or other inputs. These systems are pioneering in translating abstract concepts into detailed visual representations, transforming the way we generate and interact with graphical content. They serve various purposes, from artistic creation to commercial product visualization, enabling a broad spectrum of visual creativity and innovation.
- **Speech/Music generation:** This category includes models capable of transforming text or other forms of input into audio, such as speech or music. Examples include MusicLM for generating musical compositions and VALL-E for speech synthesis. These models have profound implications for entertainment, accessibility, and digital communication, offering new ways to produce and interact with sound and music based on textual or other data inputs.
- **Code:** AI models like Codex and AlphaCode specialize in generating programming code from natural language descriptions. This type of generative AI is revolutionizing the field of software development by providing tools that can understand human instructions and translate them into functional code, enhancing developer productivity and code quality. These systems are becoming increasingly integral to modern programming environments, assisting in everything from routine coding tasks to complex problem-solving.

Generative AI operates through a process known as generative modeling, which has notable mathematical distinctions from discriminative modeling, the latter of which is commonly used in data-driven decision support. While discriminative modeling focuses on classifying data points into predefined classes by learning the boundaries between them, generative modeling aims to understand and replicate the actual distribution of data. This enables generative models to create entirely new, synthetic samples of data, such as new observation-target pairs or new observations for a given target, drawing from the learned data distribution patterns.

The practicality of generative AI models lies in their ability to learn from and replicate complex patterns within data, often using deep neural networks as their underlying machine learning architecture. These models form part of a larger generative AI system that includes the infrastructure for data processing and user interfaces, facilitating interaction and broader application across various domains. The practical applications of generative AI systems are diverse and impactful, encompassing areas such as search engine optimization, content generation, and code generation, where they solve real-world challenges and foster innovation across multiple fields. These systems and their applications represent significant steps forward in leveraging AI to enhance decision-making and creative processes in business and other sectors.

This way, Generative AI can also be classified based on the used technique, which Feuerriegel et al. (2023) outline as follows:

- **Generative Adversarial Networks (GAN):** Generative Adversarial Networks employ a unique architecture comprising two competing neural networks, termed the generator and the discriminator. In this setup, the generator aims to create data samples from a latent space, while the discriminator evaluates whether these samples are indistinguishable from actual data. The adversarial interaction between these networks ensures that the generated samples increasingly resemble the real data distribution, enhancing the model's ability to generate realistic outputs. This architecture is particularly beneficial for tasks requiring high-quality data generation like image and video synthesis. See Fig. 2.7 for a representation of how GANs work:

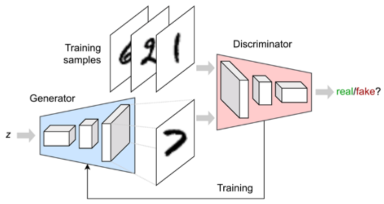


Fig. 2.7. Generative Adversarial Network (GAN)

- **Variational Autoencoders (VAE):** Variational Autoencoders are designed to compress data into a latent space and then reconstruct it, learning the essential characteristics of the data distribution. Unlike conventional autoencoders, VAEs introduce a probabilistic approach to the encoding-decoding process, enabling them to handle the variability and complexity inherent in datasets effectively. This aspect makes VAEs suitable for generating new data instances, such as images or text, and applications like anomaly detection where understanding data distribution is crucial.
- **Diffusion Probabilistic Models:** These models are based on the concept of simulating the diffusion process, where data gradually transitions from a meaningful state into random noise. By learning to reverse this process, diffusion probabilistic models can generate data that closely mimics the original dataset. Their applications extend across various fields, including image generation and enhancement, where they contribute to creating detailed and lifelike visuals from abstract inputs.
- **Transformers:** Transformers revolutionize the handling of sequential data through the self-attention mechanism, which allows the model to weigh and prioritize different parts of the input data. Unlike traditional sequential models that process data in order, transformers assess all input simultaneously, making them highly efficient for language-related tasks. Their ability to understand context and semantics in text data has made them foundational for modern natural language processing and generation tasks.

GenAI is transforming fields that hinge on creativity, innovation, and knowledge processing, enabling applications previously thought impractical for automation. This way, it can also be classified based on application:

- **Creative Arts**: This revolutionary technology is widely recognized for its role in artistic domains, mimicking writers or illustrators to create novel works. Historically, the belief was that tasks involving creativity, such as writing poems, designing fashion, or composing music, were exclusive to humans. However, generative AI has drastically changed this perception by producing content—text, images, or audio—that often cannot be distinguished from human-generated work (Feuerriegel et al., 2023).

- **Scientific Research**: GenAI has significant potential in transforming scientific research by enhancing productivity and fostering innovative discoveries. By integrating AI tools, researchers can streamline various aspects of their work, making it easier to navigate the complex landscape of their fields. This integration not only leads to groundbreaking discoveries but also improves learning outcomes in educational settings. The application of GenAI in research offers a mix of opportunities and challenges, prompting the need for a balanced perspective on its use. It is crucial to engage the chemistry community and beyond in discussions about the responsible integration of AI to fully leverage its potential while addressing associated ethical and practical challenges (Alasadi & Baiz, 2023).
- **Education:** GenAI is revolutionizing teaching and learning experiences by providing personalized learning opportunities and adaptive materials. AI-driven tools are used to tailor educational content to individual student needs, identifying strengths, weaknesses, and learning preferences. This personalization leads to more effective learning outcomes and a deeper engagement with the material. Moreover, GenAI can offer real-time feedback and assessment, helping educators identify student struggles early and provide necessary support. Additionally, it has the potential to overcome language barriers, making education more inclusive for non-native English speakers by facilitating efficient translation and comprehension of instructional materials (Alasadi & Baiz, 2023).
- **Content Generation**: Generative AI is reshaping content generation in marketing and e-commerce, introducing the capability to automate the creation of personalized content. This technology enables the development of marketing materials that cater to individual preferences and characteristics, such as creating distinct sales slogans for different personality types. The use of GenAI extends beyond mere content personalization; it encompasses automating various marketing and media tasks. This includes writing news articles, summarizing content for different formats, generating multimedia elements like thumbnail images, and adapting written content into formats accessible for visually or hearing-impaired individuals. Moreover, GenAI's application in recommender systems highlights its potential to enhance the personalization and effectiveness of information dissemination, ensuring content is better suited to each recipient's unique needs and capabilities. This evolving domain presents numerous research opportunities to further explore and optimize GenAI's role in delivering targeted and effective marketing strategies (Fueuerriegel et al., 2023).
- **Programming and Software Development**: Generative AI is significantly influencing programming and software development by automating repetitive tasks, thus potentially transforming the industry's traditional practices. This automation wave, powered by AI advancements, extends to various stages of the software development lifecycle, including code generation, documentation, and testing. The rise of large language models (LLMs) and AI-driven tools like GitHub Copilot represents a leap towards more integrated and automated software development processes. These tools assist developers by providing suggestions, automating routine tasks, and enhancing productivity, leading to a shift in how software is conceptualized, created, and maintained. The growing adoption of such technologies prompts a reevaluation of roles within the software development ecosystem, emphasizing the necessity for professionals to adapt to the changing landscape where

AI and human expertise coexist to optimize efficiency and innovation (Sauvola et al., 2024).

- **Business and Finance**: Generative AI has the potential to significantly impact the business and finance sectors. This technology can automate various tasks traditionally performed by humans, such as content creation, customer service, and code generation, potentially reducing costs and fostering innovation and growth. For instance, AI-enabled language translation has already shown economic benefits. In the context of business and finance, generative AI systems could enhance the speed and quality of code development, boost creativity among artists, and transform user-generated content platforms. The adoption of generative AI may lead to new business models and innovations, changing work patterns, organizational structures, and management practices. As such, the integration of generative AI in business processes and its implications for economic policy and competitive advantage require careful examination and adaptation of existing theories and frameworks (Fueuerriegel et al., 2023).

Brühl (2023) examines the burgeoning field of Generative Artificial Intelligence (GAI), which has captivated public interest, particularly with the advent of ChatGPT in November 2022. This innovation highlights the capabilities of GPT (generative pretrained transformer) systems, setting a new benchmark for ease of use and effectiveness in AI technology.

At the heart of GAI technology, particularly in applications like ChatGPT (a collaboration between OpenAI and Microsoft), are large language models (LLMs) such as GPT-3 and GPT-4. These models are paralleled by competitors like Google's Bard, which relies on its foundational model LaMDA.

The effectiveness of these GAI systems is significantly influenced by the scale of training data and the complexity of the neural network, as demonstrated by GPT-3's training on 570 GB of text and its capability to optimize up to 175 billion parameters. These advancements are supported by the evolution of computer hardware, particularly GPUs, which allow for the high-speed parallel processing essential for the operation of these sophisticated machine learning models.

Brühl (2023) outlines some of the various applications of Generative AI, as shown in Tab. 2.1:

Tab. 2.1. Use cases of Generative AI

Modality	Use Cases (selection)	Examples
Text (Chatbots)	Customer communication	Customer care, sales, marketing
	Data Analytics	Customer segmentation, profiling
	Virtual Assistants	Sales support, technical support
	Editing	Publishing services, translations

Tab. 2.1. Continued

Modality	Use Cases (selection)	Examples
Image/Videos	Image recognition	Face recognition, cyber security services
	Image generation	Advertising, marketing, PR
	Video generation	Media production, PR
	Cross media	E-commerce, E-learning
Code	Code quality check	Code audits for critical applications
	Code generation	Software development
	Code optimization	Software engineering
	Prototyping	Product development

As discussed in this chapter, generative AI serves as intelligent support in various practical domains. The implications of this technology are far-reaching, with projections suggesting that generative AI could significantly increase global GDP by 7% and potentially displace 300 million knowledge work jobs. This advancement presents a dual-edged sword, offering extraordinary opportunities and substantial challenges that must be addressed to ensure its responsible and sustainable use (Feuerriegel et al., 2023).

2.4.6. Large Language Models

Feuerriegel et al. (2023) delineate that large language models (LLMs) are types of neural networks specifically designed for processing and generating textual data. LLMs are essentially deep learning networks designed to grasp the nuances of human language across various contexts by training on extensive text datasets. These systems employ "transformer" models, a sophisticated type of LLM enhanced with an "attention mechanism." This mechanism enables the model to identify and focus on the most relevant parts of the text, understanding word associations and their meanings in context more effectively and efficiently than traditional methods. Unlike older models that process words sequentially, transformers analyze all words in a text concurrently and adjust their focus dynamically to improve task comprehension and execution (Brühl, 2023). LLMs are characterized by a trio of distinct features. Firstly, these models utilize expansive sequential neural networks, such as transformers that incorporate attention mechanisms. Secondly, they undergo an initial training phase using self-supervision which involves auxiliary tasks that facilitate natural language representation learning while avoiding overfitting issues, like predicting the subsequent word in a sequence. Thirdly, this preliminary training phase leverages vast datasets comprising text sources, such as Wikipedia or datasets that span multiple languages. Subsequently, practitioners can adapt these models for particular applications, such as responding to queries or generating language, by fine-tuning them with specialized datasets. LLMs have recently seen a significant evolution, now encompassing models with billions of parameters. Notable examples of these extensive LLMs include BERT, introduced by Google in 2018, and GPT-3, developed by OpenAI in 2020, which boast approximately 340 million and 175 billion parameters, respectively.

2.4.7. Why LLMs Have Gained Popularity Recently

The significant surge in the application of Large Language Models (LLMs) in recent times, as analyzed by Raiaan et al. (2024), is primarily driven by advancements in computational technology and an abundance of data for model training. Initially limited by rule-based and statistical approaches, LLMs evolved significantly with the advent of neural network methodologies, setting the stage for more nuanced language understanding and production.

Historically, the journey of LLMs began decades ago, transitioning from simple rule-based systems to statistical models and eventually to the sophisticated neural network-based models seen today. This evolution was significantly propelled by the introduction of artificial neural networks in the mid-20th century, with subsequent decades witnessing the rise of neural language models, particularly after the mid-2000s with the advent of word embeddings like Word2Vec and GloVe.

The transformative era for LLMs truly began in the 2010s with the development of models such as the recurrent neural network language model (RNNLM), which improved the processing of sequential data. Google's launch of the Google Neural Machine Translation (GNMT) model in 2015 marked a significant milestone, enhancing translation tasks considerably over previous models. However, the real breakthrough came with the development of the Transformer model in 2017, leading to the creation of advanced models like BERT and GPT, which utilized self-attention mechanisms to better understand language nuances.

These developments were underpinned by the significant increase in computational power, especially through GPUs, enabling the training of larger models and processing of extensive datasets. Additionally, the explosion of data available for training, primarily from the internet, has allowed LLMs to learn from a vast array of linguistic patterns, further enhancing their capabilities (Khan Raiaan et al., 2024).

Fig. 2.8 presents an overview of the historical development of Large Language Models, as outlined by the authors.

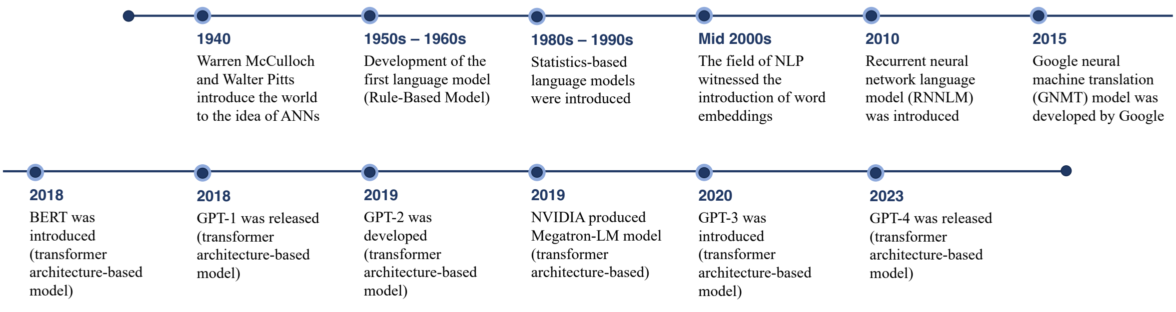


Fig. 2.8. Brief history of LLMs

2.5. Current Applications of AI in Different Sectors

Artificial Intelligence (AI) is significantly altering the landscape of business operations across various sectors. By harnessing AI, businesses are now able to engage in sophisticated customer relationship management (CRM) practices. Utilizing regression analysis and clustering techniques, companies can segment their customers into distinct groups based on demographics and purchasing history. This segmentation allows for targeted marketing strategies that are likely to yield higher conversion rates (Akerkar, 2019).

Moreover, AI systems excel in detecting outliers and potential fraud by analyzing patterns that deviate from the norm. This capability is immensely beneficial not only in financial sectors for fraud detection but also in healthcare and pharmaceutical research for identifying anomalies. Another notable application is in demand forecasting, where businesses leverage AI to predict product sales, enabling them to optimize inventory and marketing efforts more effectively (Akerkar, 2019).

In addition to these applications, AI has revolutionized the maintenance and operation of machinery in manufacturing and energy sectors. Predictive analytics, powered by AI, can forecast when machines and components will likely require maintenance, thereby reducing unexpected downtime and increasing operational efficiency. Furthermore, AI is instrumental in developing personalized recommendation engines, enhancing user engagement across various digital platforms like streaming services and e-commerce websites. These engines analyze past user behaviors and similarities among users to suggest relevant content or products (Akerkar, 2019).

In human resources, AI facilitates improved hiring processes and employee retention strategies. By analyzing data from HR systems, companies can optimize their recruitment processes and identify strong candidates who may have been overlooked. Additionally, AI can predict employee turnover and potential conflicts, allowing for preemptive measures to enhance workplace harmony (Akerkar, 2019).

Deep learning, a subset of AI, is delving deeper into business analytics, offering a richer understanding of consumer engagement and behavior. For example, in e-commerce, deep learning algorithms analyze the consumer journey to predict purchasing decisions before they occur. This predictive capability enables personalized shopping experiences, improving conversion rates and customer satisfaction. Moreover, deep learning is instrumental in advancing technologies such as self-driving cars and image recognition, showcasing the extensive application of AI beyond traditional business tasks (Akerkar, 2019).

On a broader scale, AI applications extend into robotics, natural language processing (NLP), and the Internet of Things (IoT). Robotics combined with AI leads to machines that can adapt and perform tasks in ever-changing environments. In contrast, NLP has transformed how machines understand and respond to human language, making interactions more intuitive. The IoT, enhanced by AI, connects a myriad of devices, enabling them to learn from user behaviors and improve efficiency and personalization. For instance, smart health monitors use AI to analyze data and provide insights into potential health issues, exemplifying how AI can lead to proactive healthcare solutions (Rose, 2018).

Both Akerkar (2019) and Rose (2018) underline the transformative impact of AI and deep learning across industries, from enhancing customer engagement strategies to revolutionizing product development and operational efficiency. AI's ability to analyze vast datasets and learn from patterns is paving the way for innovative solutions and smarter business practices, thereby reshaping the future landscape of multiple sectors.

3. The Supply Chain and its Complexity

3.1. Definition and Components of the Supply Chain

The concept of a supply chain encompasses a comprehensive network of participants who contribute either directly or indirectly to satisfying a customer's demand. According to Chopra & Meindl (2016), the supply chain is not limited to manufacturers and suppliers; it also includes transporters, storage facilities, retailers, and the customers themselves. Within an organization, the supply chain touches upon all functions that take part in processing and responding to a customer's request, including areas such as product innovation, marketing, day-to-day operations, distribution, financial management, and customer relations.
For instance, let's consider the scenario where a customer enters a Walmart store with the intention of buying detergent. This action activates the supply chain, starting with the customer's need. The Walmart store they visit replenishes its inventory, which might come from a warehouse holding finished goods or directly from a distributor, with the transportation often outsourced to a third-party service. The distributor would have received its supply from a manufacturing firm, for example, Procter & Gamble (P&G). The manufacturing process at P&G would depend on raw materials sourced from several suppliers, who may in turn rely on further upstream suppliers. Packaging materials, for example, could be sourced from a company like Pactiv Corporation, which would have procured its raw materials from other suppliers. The structure and flow of this supply chain are visualized in the accompanying Fig. 3.1, with arrows representing the flow of the physical product.

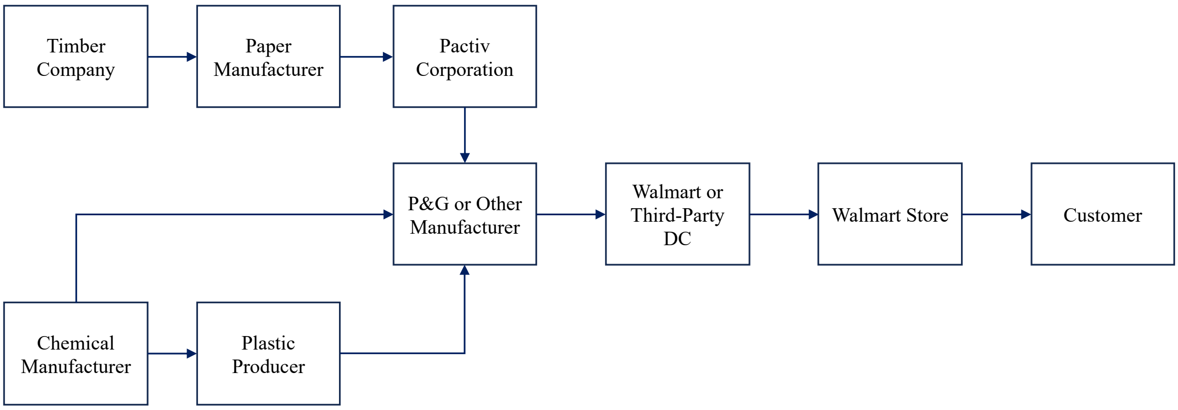


Fig. 3.1. Stages of a Detergent Supply Chain

The supply chain is a fluid system that involves the movement of information, goods, and finances between various segments. In the case of Walmart, the company not only offers the product but also shares pricing and availability details with the customer, who in turn transacts funds to Walmart. Walmart communicates sales data and restocking orders to its distributors or warehouses, which then organize the logistics to return inventory to the stores. Walmart settles the financials with the distributor post-replenishment. This distributor also communicates pricing information and provides delivery schedules to Walmart. In some cases, Walmart may send back materials for recycling. This type of exchange of information, materials, and finances is a characteristic of the entire supply chain network.

This concept extends beyond physical retail to e-commerce platforms like Amazon, where the supply chain includes the customer, Amazon's interface, the company's warehouses, and all the suppliers up the chain. Customers receive information about pricing, product diversity, and availability via the website. They select products, place orders, and make payments online, with the potential to check on order status subsequently. The upstream stages of the supply chain utilize the customer's order information to fulfill the requests, thereby involving an additional interchange of information, products, and finances across the supply chain's various levels.
These scenarios underscore the customer's critical role within the supply chain, which fundamentally exists to meet customer needs while generating profit. While the term 'supply chain' evokes an image of a linear flow from suppliers to manufacturers to distributors to retailers and finally to customers, it's essential to recognize the bidirectional flow of information, finances, and products. Moreover, the term 'supply chain' may misleadingly suggest that only one entity operates at each stage, whereas in practice, entities such as manufacturers may interact with multiple suppliers and distribute to various distributors, thereby forming a complex network. Therefore, terms like 'supply network' or 'supply web' may more accurately represent the intricate structure of most supply chains, as demonstrated in Fig. 3.2.

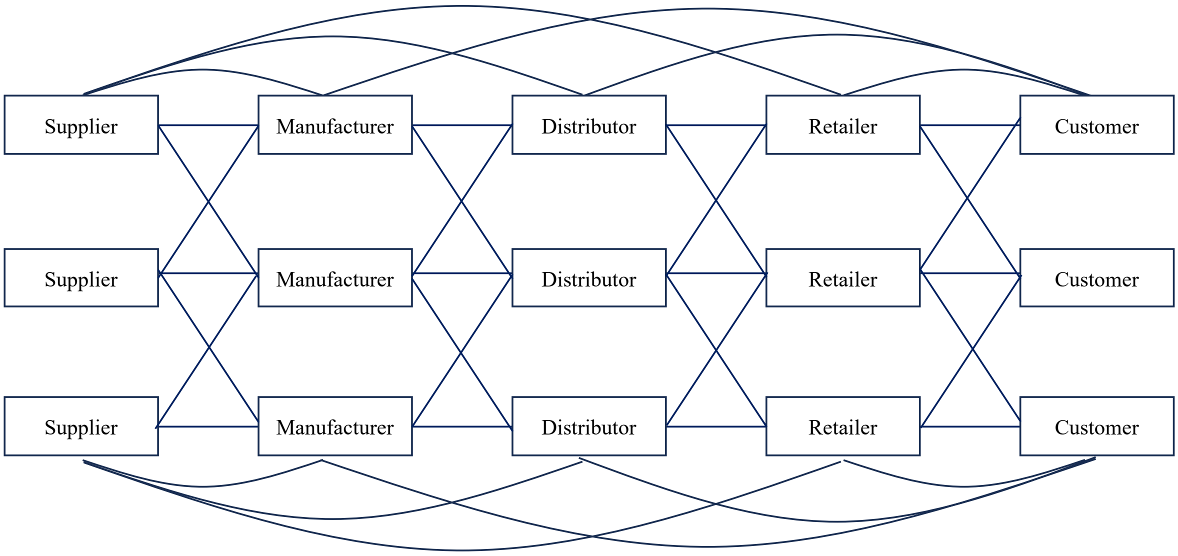


Fig. 3.2. Supply Chain Stages

Supply chains can be composed of several key stages, including customers, retailers, wholesalers/distributors, manufacturers, and raw material suppliers. Each stage is linked by the flow of products, information, and finances, often in both directions, and may be coordinated by the stages themselves or through intermediaries. The presence of each stage in the supply chain is not mandatory and depends on the customer's needs and the role each stage plays. For example, Dell has distinct supply chain models for different product lines: it adopts a build-to-order approach for servers, directly initiating production in response to customer orders without involving additional retailers or distributors. Conversely, Dell's consumer products like PCs and tablets are sold through retailers such as Walmart, adding a retailer stage to the supply chain compared to Dell's direct sales model for servers.

3.2. Key Elements: From Sourcing to Distribution

According to Basu and Wright (2008), supply chains can be examined through two primary lenses: the cycle view and the push/pull view, concepts originally delineated by Chopra and Meindl (2006).

The cycle view breaks down the supply chain into sequential stages, where each stage represents a process cycle at the interface between two successive supply chain stages, involving processes such as order generation, fulfillment, and reception. This view aligns closely with structures found in MRPII or ERP systems and is pivotal in managing inter-stage dependencies effectively.

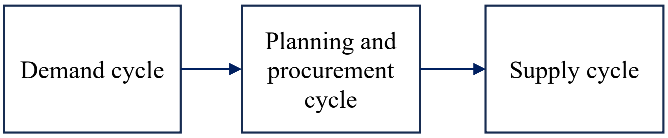


Fig. 3.3. Simplified process cycles in supply chain

The push/pull view distinguishes between processes initiated by actual customer orders (pull, associated with Lean Thinking or Lean Manufacturing) and those driven by anticipated orders (push). This distinction is crucial for inventory and production strategy, where pull processes typically align with demand-driven approaches like just-in-time manufacturing, while push processes are often associated with forecast-based planning.

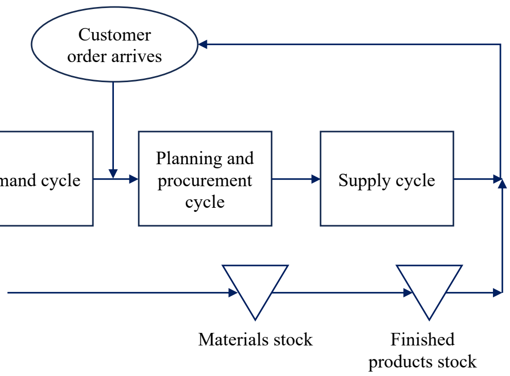


Fig. 3.4. Push process in a supply chain

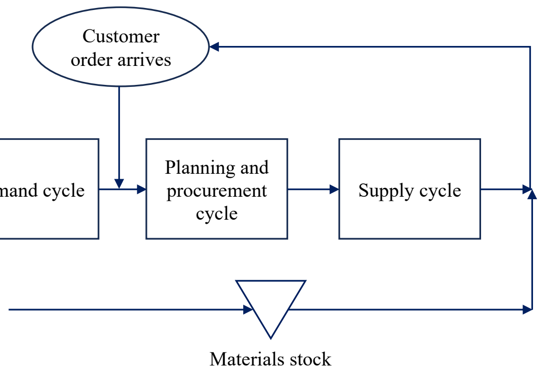


Fig. 3.5. Pull process in a supply chain

The framework presented by Basu and Wright (2008) for comprehensive supply chain management is structured around six fundamental components: 1) Customer Focus and Demand; 2) Resource and Capacity Management; 3) Procurement and Supplier Focus; and 4) Inventory Management. These elements collectively form the backbone of the model, which is depicted in Fig. 3.6 of their work.

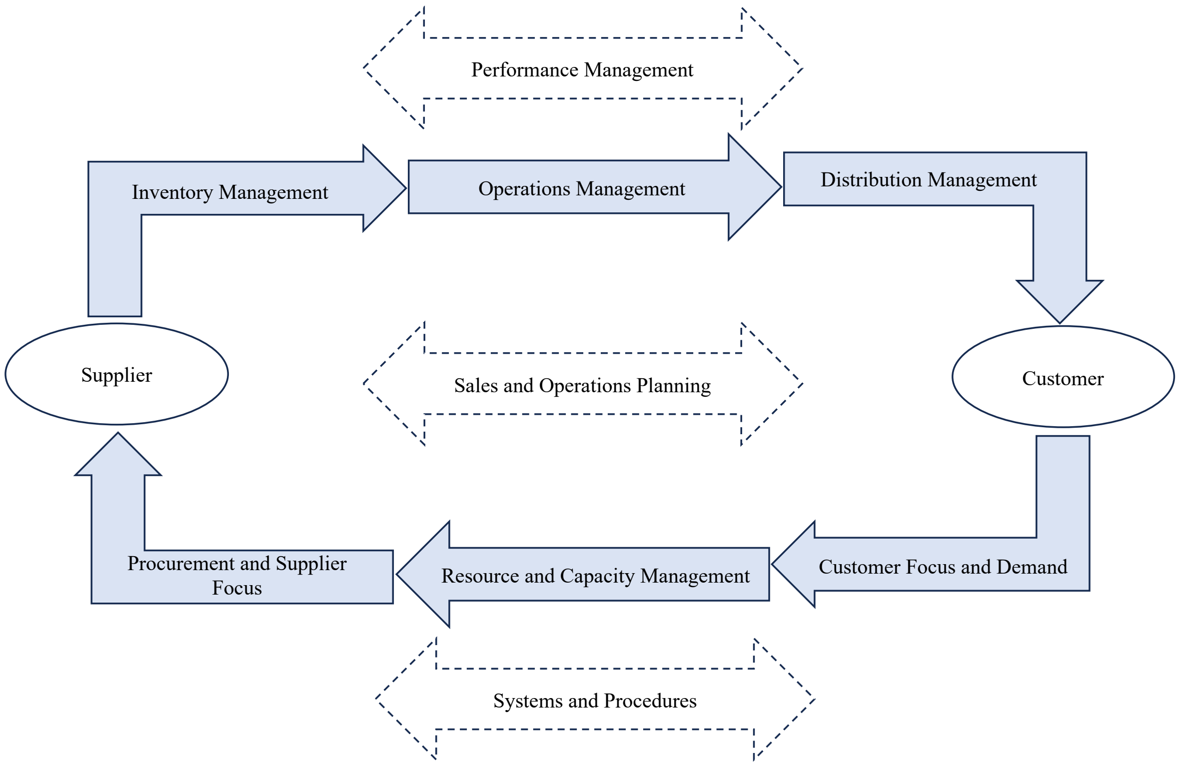


Fig. 3.6. Elements of the supply chain

3.2.1. Customer Focus and Demand

In supply chain management, customers are both the alpha and omega, initiating and concluding the series of transactions and interactions. As noted by Basu and Wright (2008), these pivotal stakeholders encompass a wide range of entities from consumers to wholesalers, and from distributors to retailers, all of whom are integral in generating the demand that drives market dynamics.

Understanding and predicting future customer needs forms the bedrock of supply chain strategy. Contrary to the common misconception that future demand can be overlooked, particularly in just-in-time (JIT) systems, Basu and Wright (2008) argue that a robust forecasting mechanism is essential. This is because both anticipatory (push) and reactive (pull) strategies within the supply chain rely heavily on the prediction of demand to ensure adequate resource allocation and capacity planning.

This approach is universally applicable, spanning across different sectors from manufacturing to services, and even extending into non-profit domains where demand can be particularly unpredictable. The intricacies involved in forecasting — influenced by factors such as historical demand, market trends, economic conditions, and promotional activities — underscore the inherent challenges in achieving perfect accuracy. However, as Basu and Wright (2008) reflect, drawing from Peter Drucker's philosophy, while the future

is inherently uncertain, strategic efforts to 'create it' through informed forecasting can significantly enhance the accuracy and reliability of supply chain decisions.
While it is universally acknowledged within the field, as Basu and Wright (2008) emphasize, that forecasting is fraught with challenges — including inevitable errors and the decreasing accuracy over longer time horizons — the practice remains indispensable. The process of estimating future demand, despite its imperfections, continues to be a cornerstone of effective supply chain management, enabling organizations to navigate the complex interplay of market forces and internal capabilities.

3.2.2. Resource and Capacity Management

Optimizing the supply capacity to meet demand punctually is a core aim of supply chain management, a notion underscored by Basu and Wright (2008). Given the real-world constraint of finite resources, ensuring timely customer satisfaction is pivotal. Capacity enhancements, encompassing various assets like machinery, warehouse space, transport logistics, stock levels, and human resources, come at a significant cost. Hence, supply chain managers are tasked with meticulously deciding on capacity thresholds and creating buffers to accommodate demand fluctuations. This could mean either scaling capacity or managing reserves of finished goods to mitigate the impacts of demand variability. Organizations might opt to maintain surplus capacity to cater to peak demand periods or cap it at a level reflecting average demand, thereby balancing the costs associated with excess inventory against the risk of lost sales.
Basu and Wright (2008) discuss several strategies for capacity optimization, highlighting the role of aggregate planning. This approach involves determining optimal levels of capacity, production, and inventory to maximize profits over a given planning horizon. Optimization strategies might be theoretical, employing mathematical models like linear programming, or practical, utilizing continuous cross-functional reviews such as sales and operational planning (S&OP).
Furthermore, the authors explain the evolution of Enterprise Resource Planning (ERP) from earlier systems like MRP and MRPII, illustrating its integration with S&OP. ERP involves a series of sequential processes supported by a unified database system, encompassing demand planning, rough-cut capacity planning, master operations scheduling, and more, typically facilitated by software like SAP R/3. The effectiveness of ERP systems, as Basu and Wright (2008) note, hinges on structured reviews conducted by planners, managers, and users, ensuring that all aspects of the supply chain are aligned and optimized.

3.2.3. Procurement and Supplier Focus

Supplementing internal capacity with external resources is a critical component of supply chain management, highlighting the 'make or buy' decision, a form of backward integration where organizations decide between purchasing materials or performing operations in-house versus outsourcing, a concept detailed by Reid and Sanders (2002). Basu and Wright (2008) expand on this by discussing the diverse aspects of supply chain procurement, such as acquiring packaging materials or outsourcing services like maintenance, which are essential for operational flexibility and efficiency.
The financial impact of procurement is significant, with external resource costs contributing to 60–90 percent of the cost of goods sold in manufacturing organizations, as Basu and Wright (2008) note, paralleling insights from Reid and Sanders (2002). This underpins the importance of strategic purchasing and supply management in enhancing service quality and

securing cost reductions, a necessity in the competitive landscape that demands continuous search for efficient suppliers and outsourcing opportunities.
Technological advancements, particularly the Internet, have transformed supply chain operations, fostering unprecedented levels of connectivity. Wright and Race (2004) emphasize how digital networks facilitate constant communication across the supply chain, reducing the need for large inventories and lowering transaction costs, benefits that Basu and Wright (2008) acknowledge as critical for improving supply chain performance and customer value.
Despite the growth of the professional service industry, Mitchell (1998) noted that purchasing teams have been slow to capitalize on cost reduction through outsourcing. However, the evolving dynamics of global supply chains have emphasized the growing importance of service level agreements and supplier partnerships, a trend supported by Wade's (2003) findings that a significant portion of procurement costs is attributed to external services.
Effective supplier selection, as Basu and Wright (2008) reiterate, aligning with Slack et al. (2006), involves evaluating potential partners on several key capabilities: technical, operational, financial, and managerial. The goal is to foster relationships that not only elevate supplier standards but also facilitate mutual learning and collaboration. This shift from rigid agreements to more dynamic partnerships underlines the importance of trust, commercial acumen, and efficient information exchange in achieving supply chain success.

3.2.4. Inventory Management

Inventories serve as a critical buffer in the supply chain, mitigating the uncertainties of supply and demand fluctuations. This is reflected in the three primary stages of inventory: input stocks, such as raw and packaging materials; in-process stocks, or semi-finished products; and output stocks, which are the completed goods (Basu & Wright, 2008). Wild (2002) expands on this by differentiating between consumed stocks, which are directly used and need frequent replenishment, and non-consumed stocks, such as capital equipment and labor, which require longer-term maintenance.
The allocation of inventories can be intentional or the unintended result of inadequate planning (Basu & Wright, 2008). While inventories are crucial for preventing production halts and loss of sales, they also represent a significant cost to businesses, encompassing capital, storage, handling, insurance, and risks such as damage, theft, and obsolescence. Conversely, insufficient inventory levels can lead to production and sales disruptions, underscoring the necessity of maintaining balanced stock levels as an insurance against supply chain variability.
According to Basu and Wright (2008), the strategic management of inventory is indicative of a supply chain's overall efficacy. Although stockpiling can temporarily improve customer service levels, it may conceal operational issues and is not financially sustainable due to the risks associated with cash flow and obsolescence. The optimization of inventory necessitates a balance between cycle stock and safety stock, influenced by factors such as ordering and transportation costs, supplier lead times, and demand variability (Basu & Wright, 2008).
In service industries, there is a noted disparity in inventory perception between operational managers and accountants (Basu & Wright, 2008). Grönroos (2000) highlights the distinct differences between services and tangible goods, such as intangibility and perishability, which challenge traditional inventory management. Despite this, service industries do manage consumable stocks like stationery but should place greater emphasis on non-consumable assets like databases and skilled personnel (Basu & Wright, 2008).

3.2.5. Operations Management

Operations management forms the foundational element of supply chain effectiveness, orchestrating the conversion of resources into finished goods or services through the strategic coordination of people, processes, and technology. This transformation is central to the flow of the supply chain, integrating inputs such as information, materials, and utilities to meet customer demands. While operations management is crucial, Basu and Wright (2008) observe that many standard texts on the subject tend to overlook its significant relationship with supply chain management.

Contrary to the narrow perception that operations are solely about physical transformations typical in manufacturing settings, operations span a diverse array of sectors. Basu and Wright (2008) argue against the misconception that activities such as sales, marketing, banking, insurance, health services, or charitable endeavors do not encompass operations management. In reality, any entity utilizing resources to create products or services is engaging in operations, highlighting the universal applicability of operations management principles across various types of organizations.

Historically, the discipline of operations management was associated exclusively with manufacturing industries during the 1960s. However, as Basu and Wright (2008) detail, the scope expanded significantly since the 1970s to include both manufacturing and service sectors, acknowledging that service operations can also be categorized into repetitive and non-repetitive types. This expansion allows for the application of manufacturing principles and techniques to the service sector, reflecting a more inclusive understanding of operations and process management. In their comprehensive approach, Basu and Wright (2008) assert that operations management is pertinent to all parts of an organization, thereby embracing a holistic view of the supply chain.

3.2.6. Distribution Management

Outsourcing distribution activities has become a common strategy for many organizations, which may inadvertently impact customer service due to a lack of internal distribution expertise. Basu and Wright (2008) stress that any failure in order fulfillment, be it related to quality, quantity, timing, or distributor conduct, ultimately falls back on the organization, not the third-party distributor. This places a spotlight on the growing issues of returns and reverse logistics within supply chain management.

Basu and Wright (2008) delineate distribution management into two primary segments: Physical distribution and strategic alliances, echoing the structure and concerns of enterprise resource planning (ERP) systems that focus on information flow and inbound logistics. The challenge lies in managing the physical movement of goods from production to end-user, mitigating demand and supply variability while balancing service levels against logistical costs. These costs encompass more than just transport; they also cover warehousing, insurance, and inventory financing, emphasizing the financial repercussions of excess stockholding.

Key components outlined by Basu and Wright (2008) in effective distribution management include strategy formation, warehouse operations, inventory control, and transportation planning. They also advocate for strategic alliances to foster an integrated supply chain, identifying critical partnerships such as third-party logistics (3PL), retailer-supplier partnerships (RSP), distributor integration (DI), and customer relationship management (CRM).

Additionally, Basu and Wright (2008) emphasize the importance of systems and procedures in harmonizing the structural elements of the supply chain. This encompasses adhering to external regulatory standards and internal quality benchmarks, managing financial performance without succumbing to short-sighted fiscal strategies, and leveraging information and communication technology (ICT) to facilitate seamless, real-time data sharing across the supply chain landscape.

3.3. Modern Challenges in Supply Chain Management

3.3.1. Risk Management

Risk management within the domain of supply chain logistics is a critical area of focus that requires diligent attention to a multitude of potential challenges and uncertainties. In the evolving landscape of global commerce, the scope and nature of risks that organizations must navigate have broadened considerably. In their comprehensive text, "Supply Chain Logistics Management" (6th ed.), Bowersox et al. (2024) provide a thorough analysis of these risks, which are no longer confined to the traditional concerns of demand fluctuations and lead time variability. Instead, modern supply chains encounter a spectrum of risks that span compliance, performance, environmental events, financial stability, and market segment dynamics, as illustrated in Tab. 3.1.

Tab. 3.1. Dimensions of Supply Chain Risk

Compliance	Performance	Environmental Events	Financial	Market Segment
Supplier code of conduct Supplier high-risk audits Restricted materials Certification	Achieving excellence Delivery Quality Audit results Capacity constraints	Natural disasters Labor disruptions Geopolitical risks Trade barriers Duties and tariffs Pandemics Terrorism Fires	Public companies Private companies Payment changes Bankruptcy Ownership changes Public press releases	Related industries Packaging Natural resources

- **Compliance Risks.** Compliance risks in supply chain management involve a multifaceted approach to adherence to laws and ethical standards. This includes implementing supplier codes of conduct that prohibit unethical practices such as bribery, as well as labor violations including forced or child labor. High-risk audits play a critical role in ensuring that supplier facilities prioritize worker safety and adhere to acceptable working conditions. The restriction of materials comes into play when sourcing involves materials that are considered scarce or whose extraction involves unethical labor practices. Furthermore, compliance extends to certifications ensuring that suppliers meet quality standards for non-genetically modified products or those that require specific purity certifications. The complexity here lies in the diversity and scope of regulations that vary by country and industry, necessitating robust compliance programs.

- **Performance Risks.** Performance risks relate directly to the capacity of suppliers to fulfill their commitments with respect to product quality and delivery timelines. Suppliers are expected to meet or exceed performance standards, and any shortfall in these areas poses a risk to the supply chain. The firm's reliance on suppliers for timely delivery of quality products at required capacities is crucial, and a breakdown in any part of this supply chain can lead to significant disruptions in production and service levels.

- **Environmental Risks.** Environmental risks encompass a wide array of external factors that are typically beyond the control of the firm. Natural disasters such as hurricanes, floods, or earthquakes can devastate infrastructure and disrupt supply chains. Socio-political factors like labor strikes, geopolitical tensions, or trade disputes can lead to unpredictability in supply chain continuity. Acts of terrorism pose a serious threat to both the physical infrastructure and the personnel within the supply chain. These risks require contingency planning and the development of robust risk management strategies to maintain supply chain resilience.

- **Financial Risks.** Financial risks are tied to the economic health of supply chain partners. A supplier's or customer's change in ownership can lead to renegotiation of contracts or even loss of business. Variability in payment terms can affect cash flow, while the risk of bankruptcy poses a significant threat to supply continuity. The financial stability of suppliers is critical to maintaining smooth operations, and financial risk management is essential to prepare for and mitigate these uncertainties.

- **Market Segment Risks.** Market segment risks are associated with the demand dynamics in interconnected industries. As industries experience their cyclical peaks and troughs, the demand for shared resources can fluctuate significantly. The example of the steel industry's demand impacting both the automotive and agricultural equipment industries illustrates how a surge in demand in one sector can lead to supply constraints in another. This inter-industry competition for resources can lead to increased prices and shortages, thereby affecting the firm's ability to produce goods.

3.3.2. Complexity Management

In the study of supply chain logistics, the concept of complexity management is essential for addressing the intricacies associated with product and process variations. Complexity management entails the administration of a multitude of product options, each distinguished by various attributes such as color, size, and packaging, known as stockkeeping units

(SKUs). The complexity is further influenced by the diversity in production processes, with each product potentially following different steps and sequences in the manufacturing process, leading to increased numbers of workstations and changeovers, negatively impacting economies of scale.
According to Bowersox et al. (2024), as companies increase their range of SKUs, there is an initial rise in revenue due to consumer attraction to a greater variety of products that appear more personalized. However, a point is reached where too many options can overwhelm consumers, causing a halt in revenue growth. Simultaneously, the costs linked with managing a large assortment of SKUs begin to mount significantly, due to increased setup, warranty, and procurement expenses, all of which can erode economies of scale.
The profit curve depicted in Fig. 3.7 follows a bell-shaped trajectory, increasing with the initial rise in SKUs, peaking, and then declining as the complexities and associated costs of managing numerous SKUs begin to surpass revenue increases. This illustrates that companies must find a sweet spot in their SKU assortment to optimize profits. Too few SKU options can lead to consumer dissatisfaction and reduced profits, while too many can inflate costs and similarly depress profitability. Thus, pinpointing the optimal number of SKUs for each product category becomes a strategic imperative for businesses aiming to maintain profitability while satisfying customer demands.

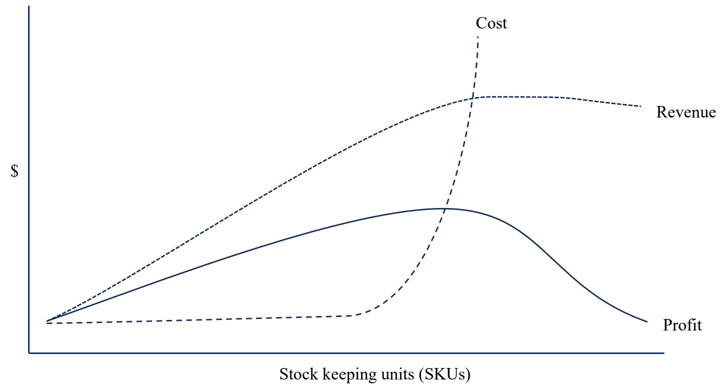


Fig. 3.7. Impact of SKUs on Revenue and Cost

3.3.3. Globalization

The phenomenon of globalization in supply chain management is critical due to the considerable gap between global demand and the capabilities of local supply. Bowersox et al. (2024) estimate that up to 90% of the world's demand is not met locally, which is accentuated by a population growth rate that adds over 200,000 individuals daily. This demographic trend points to a vast market opportunity, particularly in regions with contrasting economic statuses. Developed economies are seen as hubs for upscale consumer products along with value-added services, while developing nations, with their burgeoning populations and relatively lower purchasing power, present massive demand for essential goods and services. The authors note the significant potential in countries like India and China, where there is a growing need for fundamental products like food, clothing, and durable goods.
Global business engagement is also motivated by the pursuit of increased operational efficiencies, which can be realized through strategic sourcing of materials and components, tapping into labor advantages by manufacturing in developing countries, and leveraging

favorable tax conditions. Bowersox et al. (2024) discuss that firms typically embark on global operations by first engaging in import and export activities. The evolution of international business further involves establishing a local presence in foreign markets, ranging from franchising to setting up manufacturing and distribution facilities. This expansion is characterized by greater investment and managerial commitment.

The most advanced stage of international business, as described by the authors, is full-scale globalization, where a firm operates extensively within and across international borders. The logistics associated with such expansion differ significantly from domestic operations. They include longer order-to-delivery distances, complex documentation to comply with various international regulations, managing diverse work practices and local environments, and accommodating cultural differences in consumer behavior.

These insights underscore the strategic considerations that firms must account for when expanding globally, highlighting the importance of a sophisticated approach to international supply chain logistics to harness the potential of global markets.

3.4. The Importance of Optimization in the Supply Chain

Supply chain optimization is paramount to the success of contemporary industrial operations. The imperative for optimal supply chain design and the associated coordination across all supply chain entities emerges from the necessity to achieve seamless operations amidst large-scale and intricate supply networks, especially under conditions of uncertainty (Garcia & You, 2015).

Supply Chain Networks (SCNs) encompass an intricate interplay between suppliers, manufacturers, distribution networks, and customers, emphasizing a unified goal of maximizing overall value. However, achieving this harmonization is fraught with challenges, especially when SC elements hold contrasting interests or when the network lacks integration (Matinrad et al., 2013). The dynamic business environment, compounded by advances in technology and the competitiveness of business models, necessitates corporations to collaborate efficiently, thus giving rise to complex supply chain management (SCM) systems. Within these systems, the optimization of multi-criteria problems, like cost minimization, service level enhancement, and lead time reduction, becomes essential for improving both individual and global performance (Matinrad et al., 2013).

The burgeoning emphasis on supply chain optimization is further justified by the evolving landscape marked by globalization and technological advancements, which have interconnected industries and economies unprecedentedly. Enterprise-Wide Optimization (EWO) and sustainability have emerged as significant areas within supply chain research, driven by the need to reduce costs, inventories, and address energy concerns comprehensively (Garcia & You, 2015). These areas present fertile ground for enhancing supply chain designs by incorporating advanced modeling and optimization practices that can substantially improve profitability and shareholder value.

However, these advancements usher in multifaceted challenges, notably in multi-scale, multi-objective, and multi-player dimensions of supply chain design. Addressing these challenges necessitates innovative computational methods and collaborative efforts across academic and industrial domains to devise efficient solutions that are scalable and applicable across various sectors. Significantly, the multi-scale nature of supply chains demands a harmonized approach to model and optimize operations across different spatial and temporal scales, encompassing the entire gamut from material sourcing to product delivery and returns (Garcia & You, 2015).

Moreover, the drive towards sustainability underscores the need for supply chain designs that not only cater to economic but also environmental and social objectives. The inclusion of life cycle-based optimization approaches represents a forward-thinking methodology in tackling the multi-objective challenges inherent in modern supply chains, aligning with global sustainability goals and consumer expectations (Garcia & You, 2015).

Supply chain optimization stands as a critical endeavor in today's business environment. It encapsulates a comprehensive approach that extends beyond mere cost reduction to include sustainability and resilience in supply chain designs. By addressing the inherent multi-scale, multi-objective, and multi-player challenges, companies can enhance their competitiveness, operational efficiency, and responsiveness to market changes and global challenges. The continued exploration and development in this domain are essential for creating adaptable, efficient, and sustainable supply chains capable of meeting contemporary and future demands.

4. Applications of AI in the supply Chain

4.1. General Overview: From Automation to Optimization

The advent of artificial intelligence (AI) marks a pivotal shift in the conduct of supply chain management, transcending its traditional confines. AI, as defined by Boute & Van Mieghem (2021), grants computers and robotic systems the ability to undertake tasks that, if performed by humans, would rely on cognitive abilities. This technological leap has been embraced across various business sectors, most notably since the turn of the last decade, where AI has gone from a niche concept to a cornerstone of modern IT architecture, as underscored by surveys from Accenture and McKinsey.

The integration of AI in supply chain management has been transformative, building upon a long-standing foundation of analytics and computer support. As described by Dash et al. (2019), AI acts not as a replacement for human intelligence but as an augmentation, enhancing our strategic and analytical capabilities. This synergy of human and artificial cognition is particularly evident in the advancements of mobile computing, cloud-based learning, and information processing algorithms, which have been leveraged to achieve near-perfect demand forecasting and optimize R&D, manufacturing, and marketing strategies.

Boute & Van Mieghem (2021) further elaborate on the digital transformation of logistics through the interconnectivity of assets via sensors and devices, coining this as the fourth industrial revolution or Industry 4.0. This revolution introduces a "digital control tower" concept, analogous to an airport control tower, offering real-time insights into the supply chain workflow, allowing for preemptive actions and decision-making based on predictive analytics. The historical data used in these systems are harnessed to develop sophisticated algorithms that reveal patterns and correlations beyond human detection.

In addition to the predictive analytics mentioned by Dash et al. (2019), Boute & Van Mieghem (2021) delve into the nuances of machine learning—a subset of AI—where algorithms learn to make decisions without explicit instructions. They categorize machine learning into supervised, unsupervised, and reinforcement learning. Each type carries distinct applications within logistics, such as demand forecasting, where supervised learning uses a vast array of data points to learn sales influences; or ETA predictions, where models classify transportation timeliness based on various real-time data inputs.

Furthermore, the power of unsupervised learning is harnessed to discover inherent data patterns, leading to customer and product segmentation, which can drastically improve logistical efficiency. Reinforcement learning, on the other hand, focuses on prescribing optimal decisions by learning from simulated sequences of actions and rewards, thus facilitating complex decision-making in logistics scenarios like multi-source replenishment or perishable inventory management.

The application of AI in logistics extends to the sphere of sustainability. Boute & Van Mieghem (2021) explore how AI can optimize operational parameters to not only enhance efficiency but also advance sustainability goals. For instance, improved demand forecasting accuracy can reduce waste by minimizing safety inventory requirements and returns from obsolete stocks. Similarly, control towers can optimize shipment bundling to reduce carbon footprints, illustrating AI's dual potential in operational and environmental domains.

The conversation around AI also broaches the topic of autonomous supply chains. Boute & Van Mieghem (2021) differentiate between automation and autonomy, explaining that while current AI implementations support task automation, they do not equate to full autonomy,

which would require no external intervention. This distinction resonates with the observations of Dash et al. (2019), who note that AI enhances human roles without supplanting them, particularly in strategic and creative endeavors.
As businesses evolve towards greater digitization, the role of humans within AI-supported systems is anticipated to shift. Boute & Van Mieghem (2021) foresee a future where human planners' responsibilities evolve, necessitating new competencies like AI literacy for managers. The increasing importance of data management in this new landscape could lead to the emergence of specialized roles such as chief data officers, responsible for overseeing data processes across organizations.
In conclusion, integrating the insights from Dash et al. (2019) with those of Boute & Van Mieghem (2021) offers a rich and detailed perspective on AI's multifaceted role in supply chain management. The narrative weaves together the technological advancements, the expansion of AI applications from automation to optimization, and the evolving human-machine dynamics, illustrating a future where AI is poised to reconfigure the very fabric of supply chain operations and strategies.

4.2. Benefits, Impact and Risks of AI in Supply Chain

4.2.1. Benefits

Artificial intelligence (AI) is revolutionizing supply chain management (SCM) by bringing unprecedented efficiency and strategic depth across its various segments. Stoychev (2023) articulates the versatility of AI, particularly machine learning (ML), in revolutionizing inventory management and forecasting. Complementing this, Barzizza et al. (2023) offer a comprehensive review that extends beyond ML to encompass the wider spectrum of AI, detailing performance advantages such as operational improvements, and informational advantages like enhanced transparency and decision-making. These insights lay the groundwork for understanding AI's transformative impact on SCM (Stoychev, 2023; Barzizza et al., 2023).

- **Inventory Management.** AI, through ML, is redefining inventory management, as outlined by Stoychev (2023), by enabling more accurate demand predictions, minimizing inventory costs, and mitigating sales loss due to stockouts. Barzizza et al. (2023) echo this sentiment, demonstrating how AI enhances operational efficiency and financial performance, citing substantial cost reductions and profitability increases as key outcomes of AI's data-driven approach to SCM.
- **Warehouse Management.** AI's influence extends to warehouse management, where it automates and optimizes operations, as Stoychev (2023) notes. Barzizza et al. (2023) detail how AI contributes to performance advantages by streamlining warehouse activities, improving production systems, and enabling better resource allocation and scheduling, which are crucial for adapting to market demands and operational performance within the SCM.
- **Logistics and Transportation.** In logistics and transportation, Stoychev (2023) highlights AI's role in enhancing efficiency and customer satisfaction through improved route optimization and delivery accuracy. Barzizza et al. (2023) further discuss AI's impact on time-related metrics such as lead time and time to market, enabling a more agile supply chain response to market changes and contributing to a faster, more responsive SCM system.

- **Production.** AI's applications in production include predictive maintenance, which, as Stoychev (2023) explains, is vital for reducing downtime and maintaining production flow. Barzizza et al. (2023) elaborate on AI's broader benefits, including the facilitation of internal process optimization and defect reduction in production lines, underscoring AI's contributions to both operational and innovation benefits in SCM.
- **Chatbots.** The rise of chatbots in SCM is not only a technological trend but also a strategic enhancement, according to Stoychev (2023). These AI-driven tools automate customer service tasks, enhancing efficiency and responsiveness. Barzizza et al. (2023) recognize these as informational benefits, where AI improves communication, facilitates the exchange of information, and assists in creating new business opportunities and customer value.
- **Security.** AI's role in bolstering supply chain security is critical, a point both Stoychev (2023) and Barzizza et al. (2023) agree on. Stoychev (2023) focuses on AI's capabilities in detecting fraud and cybercrime, while Barzizza et al. (2023) highlight the advantages of AI in promoting supply chain transparency and proactive risk management, key components in maintaining supply chain integrity.

The integration of AI into SCM, as discussed by Stoychev (2023) and supported by Barzizza et al., (2023) presents a compelling array of advantages. AI's diverse applications—from operational efficiencies to strategic decision-making enhancements—underscore its vital role in advancing SCM. Despite the complexity of implementation, the significant benefits AI offers make it a key driver for innovation and competitive advantage in today's dynamic business environment.

4.2.2. Impact

The application of Artificial Intelligence (AI) into supply chain and manufacturing processes is forecasted to generate an annual economic value between $1.3 trillion and $2 trillion (McKinsey, 2021). Early adopters have already experienced significant improvements, such as a 15% reduction in logistics costs, a 35% decrease in inventory levels, and a 65% enhancement in service levels compared to their counterparts (McKinsey, 2021). This introductory section lays the foundation for a detailed examination of AI's impact across different supply chain areas, emphasizing quantifiable benefits.

- **Inventory Management.** AI's implementation in inventory management has led to a 10-30% improvement in forecast accuracy, significantly reducing stockouts and carrying costs, and thereby enhancing customer satisfaction (Shoushtari et al., 2021). McKinsey's research supports this, highlighting that companies could see up to a 20% reduction in supply chain costs through AI applications (McKinsey, 2021).
- **Warehouse Management**. In warehouse management, AI has proven instrumental, as evidenced by Amazon's 20% efficiency improvement through AI automation in picking, packing, and shipping processes (Shoushtari et al., 2021).
- **Logistics and Transportation**. In the logistics and transportation sector, AI has been a game-changer. For example, Maersk's AI-driven demand forecasting for shipping containers reduced empty container repositioning costs by $1 billion annually, while Walmart's AI application in transportation network optimization led to a $500 million yearly cost reduction (Shoushtari et al., 2021).
- **Production**. The role of AI in production extends to improving scheduling and throughput, as seen in case studies from Deloitte (2023), where companies achieved

significant economic benefits, including millions in cost savings and improved operational efficiency, by leveraging AI for production planning and control.

- **Security**. AI's integration into security protocols has led to marked reductions in fraud and cybercrime incidents, with companies experiencing a decrease in financial losses and an acceleration in the detection and response times to threats. Moreover, AI's role in enhancing supply chain transparency has facilitated higher compliance rates with regulatory standards, reduced the time and resources required for audits, and minimized errors in documentation. Additionally, proactive risk management, another benefit highlighted by AI integration, has improved the identification and assessment of potential risks, reduced operational downtimes, and enhanced response times to supply chain disruptions. These improvements collectively contribute to a more secure, efficient, and transparent supply chain environment, leading to direct cost savings and reinforcing the overall integrity of supply chain operations (Stoychev, 2023; Barzizza et al., 2023)
- **Chatbots.** AI-powered chatbots in supply chain management facilitate real-time customer interactions, streamline order processing, and enhance order fulfillment efficiency. These chatbots can answer customer queries, process orders, and provide shipment tracking, significantly improving operational efficiency and customer satisfaction. This automation of repetitive and manual tasks associated with order processing and inventory management contributes to cost reduction and higher efficiency (Shoushtari et al., 2021).

4.2.3. Risks

The study by Rana et al. (2022) provides an in-depth analysis of the complexities introduced by AI-integrated Business Analytics (AI-BA) within firms. The authors identify several critical risks such as AI-BA opacity, which refers to the lack of transparency and understanding surrounding AI processes and outcomes. This opacity can lead to suboptimal business decisions, undermining a firm's operational efficiency and competitive edge. Additionally, the article discusses the amplification of technological and security risks due to poor data governance, quality, and training within AI-BA systems. These challenges collectively contribute to increased operational inefficiency, affecting sales growth and employee satisfaction, which are crucial components of supply chain management.

Ballamudi (2019) explores the implications of AI on management, emphasizing the transformation of managerial roles and the displacement of jobs due to automation. Ballamudi argues that the integration of AI into management requires a reevaluation of traditional managerial structures and the cultivation of new skills to handle AI-driven operations effectively. This transition presents a significant challenge for supply chain management, as it necessitates a shift in workforce dynamics and the re-skilling of employees to adapt to AI-enhanced processes.

The research conducted by Al Maqbali et al. (2021) specifically addresses the logistics sector in Oman, highlighting several risks pertinent to the adoption of AI in supply chains. High implementation costs, potential job losses due to automation, and the extensive time required for integrating AI systems are identified as major barriers. The study also points out the necessity for skilled personnel to manage AI technologies, alongside the associated security and privacy concerns, which are particularly relevant in the logistics and supply chain context, where sensitive and proprietary data are often involved. Moreover, the authors raise issues related to the ethical and legal challenges posed by AI, underscoring the importance

of developing robust frameworks to guide AI implementation in supply chains (Al Maqbali et al. 2021).
Lastly, Bhbosale et al. (2020) discuss general disadvantages associated with AI, such as the high costs involved in development and maintenance, the threat of unemployment due to automation, and the potential for reduced human interaction. While these concerns are broader, they have specific implications for the supply chain sector, including the risk of diminishing collaborative and creative human input essential for dynamic supply chain management (Bhbosale et al. 2020).
In synthesizing these perspectives, it becomes evident that while AI offers transformative potential for supply chains, it also introduces a range of risks that need to be carefully managed. The integration of AI into supply chain operations requires not only significant financial investment but also a strategic approach to workforce development, data governance, and ethical considerations. Addressing these challenges through comprehensive planning, stakeholder engagement, and continuous learning will be key to leveraging AI's benefits while mitigating its risks in the supply chain domain.

4.3. Areas of Logistics for Analysis and Optimization

4.3.1. Supplier Selection

In every sector, selecting the right suppliers is pivotal for a business's growth and innovation. The competitive landscape places a heightened emphasis on enhancing product quality to broaden market acceptance and appeal. Managers are tasked with devising and executing improvement strategies that align with budget constraints and product objectives. The intricacies of supplier selection and evaluation necessitate sophisticated decision-making methodologies, such as those offered by artificial intelligence (AI) techniques. These approaches not only streamline the selection process but also significantly contribute to a business's revenue generation, setting it apart from competitors. Industries face the challenge of evaluating suppliers across a spectrum of criteria, including product quality, delivery performance, cost, and reliability. This multifaceted evaluation often reveals a trade-off between cost and quality—where some suppliers may offer components at lower prices but with compromised quality, or high-quality items at a premium (Ahmad et al., 2020).
The evaluation and selection of suppliers are fundamental to supply chain management, especially in global contexts characterized by a broad spectrum of potential suppliers, evolving procurement regulations, shifting social policies, and changing customer preferences. This dynamic environment necessitates a decision-making process that is rapid, adaptive, and sophisticated, capable of handling uncertainties and satisfying multiple criteria and stakeholders (Zavala-Alcívar et al., 2020).
Recent studies have highlighted the increasing integration of artificial intelligence (AI) into supplier selection and management processes across various industries. The innovative applications of AI have been instrumental in transforming traditional supply chain operations into more efficient, data-driven systems:

- Evolutionary Algorithms and Genetic Algorithms (GAs): These AI techniques are particularly favored for supplier selection due to their robustness in dealing with complex decision-making scenarios. They simulate evolutionary processes to identify optimal supplier combinations based on multiple criteria, such as cost, quality, and delivery time, enhancing strategic sourcing decisions (Martinez-Soto et al., 2014; Lakshmanpriya et al., 2013).

- Artificial Neural Networks (ANNs): ANNs are deployed to predict supplier performance and assess risk by analyzing historical data and identifying underlying patterns. This predictive capability supports procurement professionals in making informed decisions, thereby reducing supply chain vulnerabilities (Sudarsanam et al., 2022; Asthana & Gupta, 2015).
- Fuzzy Logic Systems: These systems are employed to manage the ambiguity and uncertainty inherent in supplier evaluation. Fuzzy Logic provides a mathematical framework to handle imprecise information, such as supplier reliability and sustainability practices, enabling a more nuanced assessment of suppliers (Dargi et al., 2014; Gupta et al., 2015).
- Hybrid Models: Combining different AI techniques, hybrid models leverage the strengths of individual methods to offer comprehensive solutions. For instance, integrating ANNs with GAs can combine predictive analytics with optimization capabilities, offering a holistic approach to supplier management (Ahmad et al., 2020).

The practical application of AI in supplier selection and management has been illustrated through various case studies, showcasing tangible benefits across different sectors:

- Automobile Industry: In a notable case, an AI-based expert evaluation method employing Fuzzy AHP was applied to address supplier selection challenges. The study highlighted how AI could streamline decision-making processes, leading to enhanced supplier partnerships and operational efficiencies (Sudarsanam et al., 2022).
- Agri-Food Supply Chain: Research focused on sustainable supplier selection in the agri-food sector employed a combination of Fuzzy Logic and GAs. This approach allowed companies to evaluate suppliers based on environmental and social criteria, contributing to more sustainable and resilient supply chains (Zavala-Alcívar et al., 2020).
- Manufacturing Industry: Another study applied a novel method integrating ANNs and GAs for optimizing supplier selection in a manufacturing context. The approach demonstrated significant cost savings and improved supplier collaboration, underlining the potential of AI to address specific industry challenges (Lakshmanpriya et al., 2013; Kai et al., 2012).

These case studies affirm the transformative impact of AI on supplier selection and management, underscoring the potential for AI to enhance decision-making, foster sustainable practices, and improve overall supply chain resilience.

4.3.2. Customer Segmentation

Originating from the field of marketing, the strategy known as "market segmentation" involves organizing potential clients into specific categories or segments that share common needs and are likely to respond similarly to marketing efforts. This approach permits companies to cater to various customer groups, recognizing that each group may perceive the value of products or services differently based on their distinct experiences and perspectives. In the contemporary scope of artificial intelligence (AI), the emphasis is on applying sophisticated algorithms to extract insights from extensive data sets. The essence of AI is to train machines to learn autonomously, serve specific purposes, and devise solutions to everyday problems. The expansion of AI was notably propelled forward with the integration of machine learning techniques at the beginning of the twentieth century.

Utilizing AI in marketing endeavors allows companies to deliver a more personalized and digitalized consumer experience, tailoring interactions to meet individual customer needs, refine market segmentation, and develop more personalized relationships with consumers. As awareness of AI's capabilities grows among marketers and consumers, its application in enhancing sales strategies is anticipated to gain traction. Therefore, grasping how to effectively harness AI for business growth is crucial (Mandapuram et al., 2020).

Recent advancements in Artificial Intelligence (AI) have revolutionized the approach to customer segmentation and management across various sectors. These applications span from traditional retail settings to complex financial services, showcasing AI's versatility and impact:

- Machine Learning and Deep Learning: utilizing supervised and unsupervised learning algorithms, businesses can dissect vast customer datasets, uncovering hidden patterns and segmenting customers into precise, behaviorally similar groups. This granular segmentation allows for highly customized marketing strategies and product offerings (Tchelidze, 2019).
- Neural Networks and Fuzzy Systems: techniques like Artificial Neural Networks (ANNs) and Fuzzy Logic have been applied to tackle challenges in market segmentation, enabling businesses to categorize customers based on nuanced, often non-linear relationships within the data. Self-Organizing Maps (SOM), a type of ANN, are particularly noted for their efficiency in clustering tasks, providing a visual and intuitive grouping of market segments (Tiwari, Srivastava, & Gerac, 2020).
- Real-time Customer Data Processing: AI technologies enable the integration of real-time customer data, ranging from transaction histories to social media interactions. This dynamic data incorporation allows businesses to constantly update and refine customer segments, ensuring that marketing efforts remain relevant and timely (Chen & Zimbra, 2010).
- Automated Segmentation: AI-driven automated segmentation tools have shown significant improvements over traditional methods, enabling the scaling of segmentation efforts without additional human resources. These tools offer the advantages of speed, accuracy, and the ability to uncover previously unrecognized customer groups (Mandapuram et al., 2020).

The practical application of AI in customer segmentation has been illustrated through various case studies, showcasing tangible benefits across different sectors:

- Banking Sector Applications: banks employing AI for customer segmentation have experienced profound insights into consumer behaviors, enabling the creation of personalized financial products and communication strategies. For instance, AI algorithms have successfully identified distinct customer personas such as "High-Value Clients" or "Frequent Transaction Users," allowing for targeted service offerings and improved customer retention rates (Raiter, 2021).
- Retail and E-commerce Enhancements: in the retail domain, AI-driven segmentation has optimized marketing campaign performance, notably increasing customer engagement and conversion rates. By tailoring marketing messages to specific segments, retailers have reported enhanced customer satisfaction and loyalty, alongside improved ROI on marketing expenditures (Gutlapalli, 2017b; Mandapuram et al., 2020).
- Pharmaceutical Industry Innovations: pharmaceutical companies leveraging AI for customer segmentation have noted better alignment of products with customer needs.

Segmentation based on patient data and prescription patterns has led to more effective and personalized healthcare solutions (Mandapuram et al., 2020).

- Energy Sector Insights: in energy and utilities, AI has facilitated the segmentation of customers based on consumption patterns, aiding in demand forecasting and tailored energy solutions. This segmentation assists in optimizing energy distribution and developing targeted energy-saving programs (Article from AI for Energy Demand Forecasting, 2019).

The integration of AI into customer segmentation and management heralds a new era of data-driven marketing and customer service. By leveraging complex algorithms and large-scale data analysis, businesses can achieve unprecedented precision in understanding and catering to customer needs. Future research should continue to explore these technologies, focusing on the development of ethical AI practices and the exploration of emerging AI methodologies for market segmentation.

4.3.3. Demand Forecasting

Demand forecasting in contemporary supply chains encompasses a multifaceted set of challenges and contextual factors:

- Market Dynamics and Consumer Behavior: Demand forecasting is inherently complex due to the unpredictable nature of market conditions and consumer behaviors. These elements introduce significant variability and uncertainty in forecasting models, requiring sophisticated analytical approaches to predict future demand accurately (Syntetos et al., 2016).
- Data Quality and Integration: The effectiveness of demand forecasting heavily relies on the availability and quality of historical data. Challenges include data fragmentation, inconsistencies across different sources, and the lag in reflecting real-time market changes. These factors contribute to difficulties in creating accurate and actionable forecasts (Mitrea et al., 2009).
- Supply Chain Complexity: Modern supply chains are extensive and intricate, amplifying issues such as the bullwhip effect, where small variations in demand can lead to significant discrepancies in upstream supply chain forecasts. This phenomenon highlights the importance of accurate demand forecasting in minimizing inventory costs and ensuring supply chain efficiency (Syntetos et al., 2016).

The application of AI and ML in demand forecasting represents a significant advancement in tackling the aforementioned challenges:

- Neural Networks and Machine Learning: AI technologies, particularly neural networks and machine learning algorithms, offer substantial improvements over traditional forecasting methods. By capturing non-linear relationships and patterns in historical data, these models provide more nuanced and accurate demand predictions (Mitrea et al., 2009).
- Hybrid Forecasting Models: The integration of various forecasting methods, combining statistical, AI, and ML techniques, has proven to be beneficial. These hybrid models leverage the strengths of each approach, enhancing forecasting accuracy and reducing bias inherent in single-method forecasts (Makridakis et al., 2018).
- Natural Language Processing (NLP): Advanced NLP techniques are being explored to incorporate external data sources, such as social media sentiment, customer

reviews, and news articles, into demand forecasting models. This approach aims to capture the broader market and consumer trends that impact demand (Wood et al., 2016).

Several case studies have highlighted the practical applications and outcomes of AI-based demand forecasting:

- Retail Industry: The integration of AI in retail has facilitated more accurate demand predictions, leading to optimized inventory levels, reduced waste, and improved customer satisfaction. AI models have enabled retailers to respond more effectively to market trends and consumer preferences, driving increased sales and revenue (Anica-Popa et al., 2021).
- Energy Sector: In energy demand forecasting, AI and ML models have been employed to predict consumption patterns, supporting better planning and resource allocation. These models have been particularly effective in accommodating seasonal variations and peak demand periods, thereby enhancing grid stability and operational efficiency (Artificial Intelligence for Energy Demand Forecasting, 2020).
- Long Supply Chain Forecasting: Innovative NLP-based models have been tested to improve demand forecasting accuracy across extended supply chains. However, the applicability and effectiveness of these models in B2B contexts remain limited, suggesting an area ripe for further research and development (A Natural Language Processing Approach to Improve Demand Forecasting in Long Supply Chains, 2020).

In summary, demand forecasting is evolving rapidly with the integration of AI and ML technologies, offering promising solutions to traditional challenges. Despite notable advancements, the field continues to face issues related to data quality, model complexity, and industry-specific applicability. Future research should aim to address these challenges, further refine AI models, and expand the scope of case studies to fully realize the potential of AI in demand forecasting.

4.3.4. Dynamic Pricing

Dynamic pricing is a strategy that has garnered significant attention across various industries due to its potential to optimize revenue through real-time price adjustments based on market demand, supply conditions, and customer behavior. While this approach offers numerous benefits, it also presents several challenges:

- Consumer Perception and Trust: One of the paramount challenges is managing consumer perceptions of fairness and trust. Dynamic pricing can lead to significant price fluctuations, which consumers might perceive as unfair or exploitative, especially if prices rise sharply during periods of peak demand or emergencies. This negative perception can erode trust and loyalty, impacting long-term customer relationships (Aparicio & Misra, 2022).
- Regulatory and Ethical Issues: The implementation of dynamic pricing raises ethical and regulatory questions, particularly concerning price discrimination and consumer data privacy. The fine line between personalized pricing and discriminatory pricing practices is a contentious issue, necessitating clear regulations and ethical guidelines to ensure fair treatment of all consumers (Aparicio & Misra, 2022).
- Data and Privacy Concerns: Effective dynamic pricing relies heavily on data analytics, requiring access to vast amounts of consumer data. This raises concerns about data privacy and security, as businesses must ensure the protection of sensitive

customer information while leveraging data for pricing decisions (Aparicio & Misra, 2022).

- Complexity and Implementation: Developing and implementing dynamic pricing models involve sophisticated algorithms and real-time data analysis, posing significant challenges in terms of technological infrastructure, computational resources, and analytical expertise. Companies must invest in the right tools and talent to harness the full potential of dynamic pricing strategies effectively (Aparicio & Misra, 2022).
- Market Dynamics and Competition: Dynamic pricing strategies must also consider broader market dynamics and competitive actions. Businesses need to navigate not only their internal pricing strategies but also respond to competitors' pricing moves, market trends, and changes in consumer demand, adding layers of complexity to pricing decisions (Aparicio & Misra, 2022).

So, while dynamic pricing offers a promising avenue for revenue optimization and market responsiveness, it is accompanied by a range of challenges that businesses must address. Balancing pricing strategies with consumer perceptions, regulatory constraints, data privacy concerns, and operational complexities is crucial for the successful adoption of dynamic pricing models.

The application of AI in dynamic pricing represents a significant advancement in tackling the aforementioned challenges:

- Energy Sector: Q-learning algorithms have been implemented for dynamic pricing to optimize electricity consumption, catering to supply and demand fluctuations (Lu et al., 2018).
- Online Retail: Online coupons and perishable goods retail have seen the application of Q-learning for price adjustments based on inventory levels and consumer RFM (Recency, Frequency, Monetary value) status, enhancing efficiency in price changes and inventory management (Liu, 2021; Cheng, 2008).
- Competitive Pricing: Studies have applied Q-learning to analyze market outcomes when multiple firms simultaneously use dynamic pricing, allowing them to derive optimal strategies based on competitors' prices (Calvano et al., 2020b; Klein, 2021).

Several case studies have highlighted the practical applications and outcomes of AI-based dynamic pricing:

- Ride-Hailing Services: Uber and Lyft utilize surge pricing algorithms to adjust fares in real time based on immediate supply and demand conditions, leading to high-frequency price variations and optimizing the allocation of drivers (Hall et al., 2015).
- Airbnb Short-term Rentals: The study of Airbnb's pricing dynamics revealed that hosts face cognitive constraints in setting optimal prices, suggesting the need for AI tools to assist in more accurate price setting based on real-time market conditions (Huang, 2021).
- Online Grocery Shopping: An AI algorithm was employed to set dynamic prices, leading to significant price variability observed within days, contrasting with traditional retail pricing where variations are seen over longer periods (Aparicio et al., 2022).

These insights demonstrate the growing use of AI in dynamic pricing across different sectors, showing how algorithms can adapt prices in real time to match market conditions and consumer behavior. The applications and case studies illustrate the potential for AI to transform traditional pricing strategies, enabling businesses to respond more agilely to market demands while highlighting the need for careful consideration of consumer perceptions and ethical pricing practices.

4.4. Literature Review of the Most Used Techniques

Riahi et al. (2021) conduct a thorough exploration of Artificial Intelligence (AI) applications within supply chain management, elucidating AI's transformative impact across various supply chain processes. Their analysis underscores AI's capacity to not only enhance performance, lower costs, and minimize losses but also to fundamentally revolutionize supply chains, making them more adaptable, agile, and resilient. The authors categorize AI's benefits into four primary areas: analysis, modeling, control, and learning. This categorization highlights AI's prowess in uncovering insights within vast data sets, facilitating deep analyses across different supply chain operations, such as performance and resilience analysis or demand forecasting.

Moreover, Riahi et al. (2021) delve into AI's role in modeling, where it aids in pattern recognition to solve complex problems like routing, leading to optimization. The control aspect emphasizes AI's ability to leverage real-time data for decision-making, thereby managing critical operations such as risk and inventory control. In the domain of learning, AI significantly enhances operational performance by automating processes like demand forecasting and production planning, enabling quick data correlation and fostering a more mature supply chain.

The study also points to AI's instrumental role in achieving specific organizational goals and objectives within supply chains, suggesting that a strategic approach to AI adoption can streamline data structuring, facilitate the selection of appropriate AI techniques, and guide the overall strategy for AI integration. This strategic adoption, as the authors advocate, is key to harnessing AI's full potential in making supply chains more efficient and responsive to evolving demands and market conditions.

A detailed classification of AI techniques applied in supply chain management (SCM) was presented, highlighting the predominant use of specific algorithms across various studies. Among the techniques analyzed, Genetic Algorithms (GAs) emerged as the most frequently employed method, cited in fourteen papers. This preference can be attributed to GAs' effectiveness in mitigating both the bullwhip effect and cash flow bullwhip across supply chains.

The second most prevalent technique, Artificial Neural Networks (ANNs), was highlighted in five papers. ANNs were notably utilized for enhancing the responsiveness of logistics workflows and for proposing comparative forecasting methodologies to address uncertain customer demands in multi-level supply chain structures.

Despite the identification of these leading techniques, the review underscores a broad spectrum of AI methodologies employed across SCM studies, including Decision Trees, Intelligent Agents, Bio-inspired Algorithms, and Particle Swarm Intelligence. This diversity reflects the exploratory nature of AI application in SCM, where novel AI techniques are continuously being tested and implemented. Importantly, the review also notes that a significant number of papers provided overviews of various AI algorithms without necessarily implementing them directly within the studies, indicating a growing interest in and foundational exploration of AI's potential in SCM contexts.

Tab. 4.1 and Tab. 4.2 present the sources consulted by the researchers in their exploration of AI methodologies applied to demand forecasting and supplier selection, respectively. These areas represent key focal points within the scope of this study's investigation.

Tab. 4.1. AI for demand forecasting references

Algorithm	Key Contributions	Author(s)
ANN	Realized a daily demand predicting system in a supermarket using MLP by adding inputs that included previous demand, days' classification, and average demand quantities	Slimani et al. (2017)
ANN	Developed a forecasting model for retailers based on customer segmentation to improve performance of inventory	Bala (2012)
SVM	Investigated the applicability and benefits of ML techniques in forecasting distorted demand signals with high noise in supply chains	Carbonneau et al. (2007)

Tab. 4.2. AI for supplier selection references

Algorithm	Key Contributions	Author(s)
ANN	Proposed a new intelligent model to predict the performance rating of suppliers in the cosmetic industry	Vahdani et al. (2012)
Genetic algorithm	Presented a new intelligent model using a genetic algorithm to solve the suppliers' performance evaluation and prioritization problems	Fallahpour et al. (2017)
ANN	Described a framework using fuzzy logic and neural networks for handling supplier selection	Lau et al. (2002)

5. Generative AI and LLMs in the Supply Chain

5.1. Applications of Generative AI in the Supply Chain

In their study, Jackson et al. (2024) delve into the applications of Generative AI (GAI) within supply chain and operations management, presenting a capability-based framework for analysis and implementation. Their research, published in the International Journal of Production Research, articulates how GAI is not just revolutionizing supply chain management through enhanced demand forecasting but is also pivotal in reshaping distribution and transportation strategies, inventory management, and beyond.

In the subsequent sections, the various areas of Supply Chain Management (SCM) that can be enhanced through Generative Artificial Intelligence (GAI) are explored, as outlined by the researchers.

5.1.1. Demand Forecasting

In supply chain and operations management, Demand Forecasting stands as a crucial function, acting as a guiding principle for inventory, production, and distribution strategies. The advent of artificial intelligence (AI) has significantly enhanced this vital process in recent years. AI algorithms have been increasingly deployed for demand forecasting, utilizing extensive datasets to unearth complex patterns and forecast future needs. The continuous learning and adaptive capabilities of these AI models have led to heightened forecasting accuracy, thereby contributing to more effective and agile operations management. This improvement aids in optimizing the entire supply chain process (Jackson et al., 2024).

In the context of semiconductor distribution, the uncertainty surrounding demand is a notable challenge. This is addressed through the application of deep reinforcement learning, illustrating the model's learning capability by utilizing data patterns to determine the most appropriate forecasting model for each product. This approach resonates with elements of supervised learning, propelled by a reward feedback mechanism. Central to the system's functionality is its Prediction capability, where it anticipates future demand patterns using historical data, potentially incorporating time series forecasting methods. The Interaction capability is showcased through the use of reinforcement learning, allowing the system to make informed decisions within a multifaceted supply chain environment. The system's ability to dynamically select forecasting models highlights its Adaptation capability, enabling strategy adjustments based on shifting demand patterns. The underlying Reasoning capability supports the system’s decision-making process in selecting the optimum demand forecast model for each product, showcasing strategic planning to mitigate the impacts of demand uncertainty.

Furthermore, a demand forecasting approach for intricate supply chains is explored through the use of a Long Short-Term Memory (LSTM) model, refined by a hybrid combination of genetic algorithms and scatter search. This study exhibits the LSTM model's superior performance in managing fluctuating demand, thereby aiding in the reduction of distribution costs within a physical internet supply chain network. The Learning capability of AI is demonstrated through the LSTM's use, a Supervised Learning method, which leverages historical demand data for future predictions. While the paper primarily highlights the

Prediction capability through the utilization of the LSTM model for demand forecasting, the Adaptation capability is apparent in the optimization of LSTM's hyperparameters via a hybrid genetic algorithm and scatter search, enabling the model to evolve and refine over time in response to new data and varying demand scenarios.
Lastly, the application of AI in predictive frameworks for solar energy management is considered, where an LSTM model is employed to analyze historical solar electricity generation data. This model embodies AI's Learning capability, especially in Supervised Learning, enabling precise future pattern forecasts. The focus on enhancing accuracy in solar electricity generation forecasts accentuates AI's Prediction capability, essential for improving grid reliability and efficiency within Smart Grid contexts. Though not explicitly detailed, elements of Adaptation and Reasoning likely contribute as the model adapts to diverse data inputs and intricate grid scenarios, optimizing solar energy management through strategic decision-making and ongoing learning (Jackson et al., 2024).

5.1.2. Distribution and Transportation Strategy

Jackson et al. (2024) present how it's possible to optimize the shortest path interdiction problem by leveraging AI's multifaceted capabilities. Through the integration of a Reinforcement Learning (RL) framework, the study explores how AI can dynamically interact with and adapt to the changing environment of SCOM. The RL model aids in understanding and optimizing variable scenarios, thereby enhancing strategic decision-making in distribution networks.

5.1.3. Inventory Management and Warehousing

In their study, Jackson et al. (2024) address key supply chain management challenges, specifically inventory distortion, by introducing No Code AI as an innovative, cost-effective solution for the retail sector. This approach empowers non-technical entities to develop machine learning models for better production and inventory management, aiming to alleviate common inventory issues and enhance sales through accurate demand forecasting. Furthermore, they explore the integration of AI in logistics simulations, highlighting how Natural Language Processing (NLP) via the GPT-3 Codex can creatively construct simulation codes from verbal inputs, thereby fostering a seamless Human-AI collaboration. This enables a more streamlined and strategic approach to logistics planning, underscoring the transformative potential of Generative AI in operational workflows.

5.1.4. Process Design

Jackson et al. (2024) also delve into advanced predictive modeling within manufacturing systems. They discuss the pivotal role of Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs) in advancing manufacturing analytics, utilizing the digital twin concept for enriched predictive insights. This detailed examination illustrates AI's capacity for learning from manufacturing data, improving process efficiency through advanced pattern recognition, and adapting predictive models to evolving environments, thereby fostering more informed strategic planning and operational optimization in the supply chain sector.

5.1.5. Production Planning and Control

The authors review an approach that revolutionizes production progress in bespoke manufacturing settings. This strategy, deeply rooted in the integration of big data and the Industrial Internet of Things (IIoT), employs advanced AI techniques such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTMs) networks. These technologies, known for their proficiency in learning from vast datasets, enable the nuanced prediction of manufacturing outcomes. By harnessing transfer learning, these AI models efficiently apply previously acquired insights to new tasks, enhancing their predictive performance while maintaining computational efficiency. This application not only illustrates AI's learning and adaptation capabilities but also its ability to perceive and interpret complex data structures, ultimately contributing to more accurate production scheduling and timely order delivery.

5.1.6. Sourcing Strategy

In the detailed analysis by Jackson et al. (2024), the authors explore applications of AI within the sourcing process.

Firstly, they detail Walmart's innovative approach to supplier negotiations within the expansive framework of Industry 4.0. Walmart's implementation of the AI-driven Pactum AI represents a significant leap forward, particularly in managing a large number of supplier relationships. The AI system, based on machine learning algorithms, is trained on predefined scripts crafted by Walmart's internal buyers, creating a supervised learning environment that tailors negotiation processes to the specific needs and contexts of Walmart's operations. This automated negotiation, facilitated by a text-based chatbot, not only streamlines interactions but also introduces a level of personalization previously unattainable in manual processes. The chatbot's ability to engage in complex negotiations, learn from each interaction, and adapt its strategies based on supplier feedback exemplifies the advanced capabilities of AI in learning, interaction, and adaptation.

Secondly, the study discusses Amazon Business's integration of AI into its procurement processes, highlighting a paradigm shift towards data-driven strategic planning and efficiency. The Amazon Business Spend Visibility tool, powered by machine learning, analyzes organizational purchasing patterns to deliver actionable insights, thereby reducing manual labor and enhancing decision-making processes. This tool exemplifies AI's learning and prediction capabilities by processing vast amounts of procurement data to forecast future buying behaviors and identify cost-saving opportunities. Moreover, Amazon Business utilizes AI to automate competitive bidding processes, showcasing AI’s interaction capability by streamlining sourcing and procurement, thereby enabling more efficient and cost-effective operations.

5.1.7. Risk Management

The authors discuss the integration of artificial intelligence (AI) as a pivotal solution to enhance the decision-making process, particularly noting its adaptability and data processing capabilities, which are essential in managing complex and dynamic supply chain environments.

The work further explores how AI, through its advanced learning, predictive, and reasoning functions, significantly optimized supply chain risk management, particularly benefiting small to medium-sized businesses confronted with rapidly shifting market dynamics. Jackson et al. (2024) reference the application of AI in these contexts to illustrate how technological advancements enable these businesses to remain resilient and responsive. They detail how employing methodologies like partial least squares-based structural equation modeling and artificial neural networks allows AI to extract actionable insights from intricate data arrays. This not only aids in comprehending and navigating through market volatilities but also in formulating cost-effective and resource-efficient strategies. The predictive nature of AI, as elucidated by the authors, is instrumental in forecasting a range of outcomes, thereby facilitating a more proactive approach to risk management within the supply chain domain.

5.2. Impact of Generative AI on the Supply Chain

In the changing world of supply chain management, the introduction of generative AI marks a significant change with significant economic effects. Insights from McKinsey (2023, 2024) and Accenture (2023) collectively underscore the transformative potential of GenAI, projecting a staggering value creation between $3.5 trillion to $4 trillion specifically within supply chain enhancements. This innovation not only promises to streamline operations but also catalyzes efficiency, reducing costs and spawning new revenue avenues. The numerical benefits are particularly striking, with predictions indicating that GenAI could automate up to 50% of activities in collaboration and management areas of supply chains, leading to significant increases in operational productivity. McKinsey (2024) further accentuates the broader economic benefits, estimating global gains ranging from $6.1 trillion to $7.9 trillion annually, underscoring the role of GenAI in augmenting traditional analytical capabilities and driving a 35–70% incremental economic impact over existing AI applications. Accenture's analysis (2023) corroborates these findings, highlighting the automation and augmentation potential in supply chain roles—forecasting an average productivity saving of almost 20%. This introductory overview paints a vivid picture of GenAI's integral role in reshaping supply chain dynamics, promising not only enhanced operational efficiency but also a significant economic uplift across the globe.
Tab. 5.1 presents the impact of GenAI throughout various sectors of the supply chain where it is implemented.

Tab. 5.1. Impact of Generative AI on different sectors of the supply chain

Area	Impact
Demand forecasting	McKinsey's (2024) report highlights that GenAI enhances forecasting accuracy by incorporating a wider range of unstructured data sources, such as market reports and social media, improving inventory management and reducing associated costs. This also minimizes stock discrepancies, thereby leading to cost saving. Accenture's (2023) findings reinforce this perspective, emphasizing GenAI's capability to merge and analyze extensive datasets, refining the precision and depth of demand forecasts. Similarly, EY's (2024) report acknowledges the growing utilization of AI for real-time demand models from vast sales data and market insights.

Tab. 5.1. Continued

Area	Impact
Distribution and Transportation Strategy	Furthermore, McKinsey's 2023 findings indicate that GenAI has the capacity to automate up to 50% of collaborative and management activities, streamlining routing, load planning, and scheduling processes, which in turn contributes to decreased transportation expenses and heightened delivery effectiveness. Accenture's 2023 insights align with this view, highlighting GenAI's role in refining route planning and freight management through the analysis of real-time data, including weather and traffic patterns. This results in more streamlined logistics operations and further reductions in transportation costs. EY's 2024 report expands on this, pointing out that GenAI can revolutionize logistics network design by considering a multitude of factors like warehouse locations and demand patterns, leading to faster delivery times, lower costs, and elevated service levels. Specifically, EY notes the use of GenAI for dynamic optimization of last-mile delivery routes, leading to significant efficiency improvements, fuel savings, and enhanced customer satisfaction. A case in point involves one of the largest logistics companies in the US, which experienced a 30% boost in workforce productivity and notable operational cost reductions by utilizing a GenAI-powered platform for warehouse route optimization.
Inventory Management and Warehousing	According to McKinsey's 2024 insights, GenAI is transforming inventory management processes by leveraging extensive data sources to enhance the accuracy of stock levels and diminish instances of overstocking and stockouts. This application of GenAI leads to a more streamlined approach, optimizing inventory to match demand more precisely and thus reducing associated costs. Accenture's 2023 findings complement these insights, showcasing how GenAI applications demonstrate operational efficiency, cost savings, and improved service levels.
Process Design	According to McKinsey's 2024 findings, GenAI significantly enhances model-based systems engineering, accelerating the development of new designs and models. This acceleration facilitates the creation of more efficient supply chain processes. Accenture's 2023 analysis complements these insights by illustrating how GenAI can automate and refine supply chain processes. By identifying inefficiencies and proposing improvements, GenAI contributes to more efficient operations, shorter lead times, and enhanced adaptability to market shifts.

Tab. 5.1. Continued

Area	Impact
Production Planning and Control	McKinsey's 2024 insights reveal that GenAI not only augments tasks such as procurement but also significantly boosts the efficiency and productivity of production planning activities. This is further supported by their 2023 report which suggests that the application of GenAI could potentially lead to a 50% increase in productivity. This substantial improvement is achieved through the automation of routine tasks and the sophisticated analysis of large datasets, resulting in enhanced production scheduling, elevated efficiency, and considerable waste reduction. Accenture's 2023 analysis complements these findings by demonstrating how GenAI contributes to a more harmonized balance between demand and manufacturing capacity. By optimizing production schedules and improving resource allocation, GenAI facilitates a higher level of operational efficiency and ensures stricter adherence to production timelines. Similarly, EY's 2024 report highlights GenAI's pivotal role in planning production and scheduling, which includes adapting to customer changes, assessing production capacities, and prioritizing order sequences. GenAI's capabilities in creating more effective production plans and resource distribution are instrumental in reducing bottlenecks and refining production efficiency.
Sourcing Strategy	McKinsey's 2024 report underscores how GenAI elevates sourcing and procurement by automating routine tasks and enriching the contract negotiation process. This leads to enhanced supplier selection and noticeable cost reductions. Accenture's 2023 insights delve deeper into GenAI's transformative potential in sourcing, detailing how it automates critical facets of supplier selection and contract negotiations, thereby facilitating more strategic and informed decision-making processes. This is based on GenAI's ability to analyze complex supplier data and market trends. EY's 2024 report provides a practical perspective, revealing how GenAI has been operationalized in vendor negotiations, significantly streamlining the tender process and reducing the emotional biases typically associated with face-to-face negotiations. A notable case involved a U.S. retailer where 65% of suppliers showed a preference for negotiating with GenAI systems over human employees.

Tab. 5.1. Continued

Area	Impact
Risk Management	McKinsey's 2024 insights highlight GenAI's utility in identifying irregularities, such as those in cold chain management crucial for pharmaceuticals, and its capacity to auto-populate compliance documents, thereby minimizing risks and bolstering compliance measures. This functionality directly contributes to safer, more reliable supply chain practices, potentially averting costly breaches and losses. Furthering this discussion, McKinsey's 2023 analysis elaborates on GenAI's ability to process and synthesize complex data sets. This capability significantly advances risk identification and mitigation, leading to more resilient supply chains and reduced financial setbacks. Accenture's 2023 report aligns with these findings, emphasizing how AI, particularly GenAI, bolsters supply chain resilience by pinpointing potential risks and vulnerabilities. The proactive analysis of varied data sources allows GenAI to anticipate disruptions and propose appropriate countermeasures, thereby solidifying supply chain integrity and operational continuity. EY's 2024 commentary provides additional depth, illustrating how GenAI transcends traditional risk management tools by offering tailored risk assessments, scenario simulations, and on-demand mitigation strategies. This forward-thinking approach enables supply chain planners to address and neutralize risks more dynamically and effectively. It is noted that approximately 40% of supply chain organizations are already leveraging this technology, reflecting its growing importance and the tangible benefits it brings in mitigating supply chain risks.

5.3. Applications of LLMs in the Supply Chain

In the evolving landscape of supply chain management (SCM), the integration of artificial intelligence (AI), particularly through Large Language Models (LLMs), marks a significant shift from traditional operational frameworks (Hendriksen, 2023). This chapter delves into the nuanced roles and impacts of LLMs within SCM, examining how they contribute to enhancing operational efficiencies and streamlining processes.

The research by Frederico (2023) and (Wang et al., 2023a) highlight the advent of ChatGPT, one of the most prominent LLMs in recent times, within the SCM domain. The authors' insights, although emerging from a preliminary stage of research, shed light on the model's burgeoning capacity to refine and innovate multiple dimensions of supply chain activities.

5.3.1. Knowledge Management

In the domain of Knowledge Management within supply chain management (SCM), Large Language Models (LLMs) play a crucial role. Srivastava et al. (2024) illustrate that by capturing and storing pertinent supply chain data, LLMs significantly improve the accessibility of information. This enhancement aids in bolstering decision-making processes and fostering the development of strategic initiatives. Additionally, these systems are instrumental in implementing proactive strategies to mitigate risks and ensure compliance, thereby contributing to overall risk reduction. Srivastava et al. (2024) further note that LLMs promote cost savings by enabling more efficient operations and informed decision-making. They highlight the user-friendly nature of these systems, with interfaces and automated summaries that enhance collaboration and communication among SCM professionals.

In the specific context of manufacturing, Wang et al. (2023b) discuss how Industrial-GPT, an application of LLMs, is being utilized. This model integrates industrial datasets with extensive domain knowledge, offering customized solutions that cater to the unique needs of the manufacturing sector.

The broader impact of LLMs on SCM, as identified by Hendriksen (2023), signifies a paradigm shift towards more data-driven decision-making processes. This transformation underscores the critical role of data in modern SCM strategies.

Moreover, the potential of ChatGPT in transforming knowledge management practices within the manufacturing industry is significant (Wang et al., 2023a). ChatGPT addresses the challenges linked with accessing and managing vast arrays of multi-modal knowledge from varied disciplines, such as material science, physics, and computer science. Its capabilities in text synthesis could potentially revolutionize the way knowledge is stored, accessed, and leveraged, providing swift and adaptable solutions to complex informational needs.

5.3.2. Risk and Security Management

In the analysis of Risk and Security Management within Supply Chain Management (SCM), Srivastava et al. (2024) detail the instrumental role of Large Language Models (LLMs) in enhancing the safety and efficiency of supply chains. They outline how LLMs process a comprehensive range of data sources to forecast and monitor potential SCM risks, thus empowering proactive risk mitigation strategies. This capability is crucial, as it allows managers to promptly address and reduce the impact of unexpected disruptions.

Furthermore, Srivastava et al. (2024) highlight the importance of real-time tracking and surveillance in protecting sensitive data and ensuring the operational integrity of supply chains, thereby bolstering supply chain resilience. The insights gleaned from the analysis of risk factors and vulnerabilities through LLMs facilitate the establishment of robust and adaptable supply chain networks while maintaining adherence to regulatory standards.

The integration of LLMs into SCM, particularly for risk and security objectives, necessitates a synergistic partnership with data scientists and cybersecurity professionals. Srivastava et al. (2024) also discuss how advanced LLMs refine SCM's risk assessment and security protocols by delving into both internal and external data sets, such as historical contract data, performance records of suppliers, and prevailing market trends.

The utility of LLMs extends to the identification and implementation of necessary corrective actions to prevent crises and fortify security policies. Additionally, the adoption of the ADO (Antecedents, Decisions, Outcomes) framework provides a systematic methodology to

assess and enhance the efficacy of LLM-driven risk and security management strategies in SCM, ensuring comprehensive protection against supply chain threats.
In parallel, Frederico (2023) explores the capacity of ChatGPT in identifying less risky supply chain sources, underscoring the evolving landscape of supply chain risk management facilitated by LLM technologies. This highlights the broader applicability and potential of LLMs, including ChatGPT, in mitigating supply chain vulnerabilities and enhancing operational security.

5.3.3. Forecasting

In the context of Forecasting within Supply Chain Management (SCM), LLMs have emerged as pivotal tools in enhancing the accuracy and effectiveness of various forecasting activities. Hendriksen (2023) acknowledges the potential of tools like ChatGPT in crucial SCM tasks, such as data analysis and inventory management, with a particular emphasis on forecasting sales demand. Additionally, Srivastava et al. (2024) detail how LLMs efficiently process a wide array of data, including historical information, to support demand forecasting and inventory management initiatives.
Expanding on this, Li et al. (2023) underscore the significant improvements brought by LLMs in supply chain forecasting. By employing natural language processing (NLP) and text analysis on diverse data sources, LLMs contribute to refining demand forecasting, sales predictions, and inventory management. The focus on leveraging LLMs to enhance precision and data-driven predictions is highlighted as essential for effective supply chain operations.
Teo (2020) offers a more focused perspective in his thesis, "A Natural Language Processing Approach to Improve Demand Forecasting in Long Supply Chains." Teo (2020) explores the application of modern NLP techniques within SCM, particularly through the NEMO model, aiming to forecast commodity demand without necessitating information sharing between downstream companies. His comparative analysis reveals that despite significant forecast errors encountered by all models, NEMO surpasses traditional forecasting methods like the ARIMA model and XGBoost in tracking actual data volatility and demand prediction, showing approximately 20% improvement in terms of Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).
Teo's (2020) work also dives into the utilization of NLP for bolstering demand forecasting accuracy, especially within B2B supply chains known for their length and complexity. He elaborates on the burgeoning role of sentiment analysis, a facet of NLP, traditionally leveraged by B2C companies but less adopted in B2B settings due to the nature of industry interactions. The thesis outlines the challenges faced in demand forecasting for B2B entities, the process of data collection and processing, and the integration of empirical sales data with NLP methodologies to enhance forecasting outcomes.
Finally, Teo's (2020) exploration of the NEMO model and contemporary NLP techniques proposes a novel approach to surmounting the perennial issues faced in B2B supply chain demand forecasting. By marrying modern NLP techniques with deep learning models, Teo (2020) demonstrates how advanced NLP can extract actionable insights from extensive textual datasets, thereby improving the reliability and accuracy of demand forecasts in sectors marked by complex supply chain structures. This innovative approach underscores the transformative potential of NLP and LLMs in redefining forecasting practices within intricate supply chain ecosystems.

5.3.4. Supplier Relationship Management

In supplier relationship management, Large Language Models (LLMs) play a pivotal role in refining supplier selection processes and improving communication channels. Hendriksen (2023) underscores the capacity of LLMs like GPT-4 in evaluating supplier profiles against set criteria, vital for supplier choice. Similarly, Frederico (2023) highlights ChatGPT's utility in streamlining supplier communication, boosting customer satisfaction and operational efficiency. In their study, Li et al. (2023) delve deeper, showcasing LLMs' application in sentiment analysis, contract evaluations, and optimizing supply chain communications, aiming to bolster collaborative and efficient supplier relationships. Srivastava et al. (2024) complement these findings by demonstrating how these models aid in informed decision-making regarding supplier selection and contract negotiations, leading to cost reductions and heightened supplier dependability. These diverse applications signify LLMs' transformative potential in managing and enhancing supplier relations within supply chains.

5.3.5. Customer Relationship Management

The implementation of Large Language Models (LLMs) like ChatGPT in customer service has been transformative, enhancing communication methods and providing tailored problem-solving strategies. According to Hendriksen (2023), these advancements lead to more personalized and effective interactions between businesses and their customers. Frederico (2023) further notes the significant impact of ChatGPT in streamlining customer communications, which not only increases satisfaction but also enhances process efficiency and reduces operational costs.

Specifically, within logistics and retail sectors, ChatGPT's application has been particularly noteworthy. Frederico (2023) highlights its use in creating efficient chatbots that improve customer service in logistics, and its role in elevating customer service levels in retail by providing robust support for marketing and workforce management, including training and guidance.

This leap in customer service capabilities is evidenced by companies such as Instacart, Salesforce, and Zalando integrating ChatGPT into their operations (Fosso Wamba et al., 2023). For instance, Instacart's ChatGPT plugin personalizes the shopping experience, aiding in recipe discovery and streamlining the ordering process. Salesforce's incorporation of ChatGPT into its Slack application is designed to improve the sales journey, particularly for B2B customers. Zalando's launch of a fashion assistant powered by ChatGPT aims to enhance the user experience on its platform by assisting in product selection and simplifying the shopping process (Fosso Wamba et al., 2023).

These examples underscore the broad and impactful adoption of ChatGPT and related technologies in enhancing customer service across the supply chain and retail industries, showcasing a shift towards more responsive, efficient, and tailored customer interactions.

5.3.6. Manufacturing

In the manufacturing sector, the introduction of the Industrial-Generative Pre-trained Transformer (Industrial-GPT) marks a substantial evolution in intelligent manufacturing systems (IMS). Developed to harness domain-specific expertise, Industrial-GPT significantly enhances manufacturing processes, encompassing equipment fault diagnosis, working status prediction, and product quality control. By tapping into industrial datasets and applying specific knowledge, Industrial-GPT delivers tailored solutions that meet the unique needs of the manufacturing industry. Additionally, this advancement brings forth the concept of Model as a Service (MaaS), offering a flexible and effective model to augment these capabilities (Wang et al., 2023b).

The application of Industrial-GPT in manufacturing operates through a methodical approach that includes the continuous collection and analysis of data to pinpoint anomalies (Data Space Reasoning), the utilization of scenario-based descriptions and domain expertise for task analysis (Initial Cognition by ChatGPT), and the employment of specific Industrial-GPT models to make informed decisions (Industrial-GPT Decision Making). Furthermore, it involves the execution of these decisions via an execution engine adhering to set rules (Scenario Task Execution) and the ongoing refinement of the knowledge base and decision-making processes to align with the evolving manufacturing conditions (Evaluation and Optimization).

This paradigm shift facilitated by Industrial-GPT exemplifies a move towards more adaptive, intelligent, and efficient manufacturing operations. Alongside, ChatGPT has emerged as a significant tool in redesigning human-machine collaboration within the manufacturing sector, simplifying the interaction between engineers and manufacturing systems. By offering a more straightforward, unified interface, ChatGPT substantially reduces cognitive and operational burdens, leading to improved system coordination and manufacturing efficiency (Wang et al., 2023a).

Moreover, the potential of LLMs like ChatGPT extends beyond manufacturing processes to include roles in predictive maintenance within supply chain management, highlighting their versatility in enhancing not only production workflows but also logistics operations (Frederico, 2023). This holistic application of LLMs underscores a broader trend towards digitization and automation, setting a new standard for innovation in manufacturing and supply chain management.

5.3.7. Shipment Tracking

Advancements in technology, particularly through the utilization of ChatGPT, have significantly enhanced operational capabilities, notably in the area of tracking and monitoring shipments. Frederico (2023) highlights that ChatGPT contributes considerably to automating these routine tasks, fostering a more data-driven approach to supply chain processes. By streamlining these operations, ChatGPT not only improves efficiency but also plays a pivotal role in reducing operational overheads and enhancing overall supply chain performance. This integration of ChatGPT into tracking systems exemplifies the practical application of AI in modernizing supply chain activities, underscoring the transition towards more autonomous and informed operational frameworks.

5.3.8. Routing

In modern logistics, LLMs such as ChatGPT has been identified as a valuable tool for optimizing delivery routes by leveraging shipping data, thereby enhancing logistical operations (Frederico, 2023). Similarly, Srivastava et al. (2024) emphasize the role of Large Language Models (LLMs) in refining delivery strategies, considering a range of factors including traffic and weather conditions to ensure more efficient route planning. Integrating insights from both sources provides a comprehensive view on the evolving use of AI in streamlining logistical frameworks, contributing significantly to the operational efficiency and flexibility of supply chain management.

5.3.9. Case Studies

Li et al. , (2023), affiliated with Microsoft Research and Microsoft Cloud Supply Chain, elaborate on the deployment of OptiGuide within Microsoft Azure's supply chain, focusing on server fulfillment. OptiGuide exemplifies the application of LLMs, specifically GPT-4, in tackling complex optimization problems, thereby facilitating decision-making processes. Azure's supply chain management involves a multifaceted approach to meet the growing demand in the cloud industry, including demand forecasting, strategic foresight, and hardware semantic search. The Intelligent Fulfillment System (IFS) plays a crucial role in this ecosystem, optimizing server assignments and shipments. The case study details how OptiGuide aids in navigating these complexities, making it easier for planners to understand and execute supply chain decisions.

Feedback from users, both planners and engineers, has been overwhelmingly positive, highlighting OptiGuide's ability to demystify the underlying optimization logic and support critical "what-if" scenarios. This not only grants planners more autonomy but also significantly reduces the engineering team's workload. Preliminary evaluations of OptiGuide show more than 90% accuracy, demonstrating its effectiveness in translating complex supply chain optimization problems into actionable insights.

This comprehensive exploration within Microsoft Azure's supply chain underscores the transformative potential of LLMs in enhancing supply chain operations, showcasing a practical application that bridges the gap between advanced computational models and intuitive, user-friendly interfaces.

Fosso Wamba et al. (2023), provide in their study a list of examples of use cases in related O&SCM fields, which can be seen in Tab. 5.2.

Tab. 5.2. Examples of use cases of LLMs in O&SCM

Use case	Context	Use of LLMs
DHL (Logistics company)	Intention to adopt ChatGPT	The company is in the process of identifying and understanding this technology's potential in logistics and supply chains. The company is convinced about ChatGPT's potential to automate processes to support efficiency improvement. Also, DHL believes that Gen-AI/ChatGPT can be widely used in warehouse operations and in the driver's cabin.
Instacart (Grocery delivery and pick-up service)	Creation of a plugin in collaboration with OpenAI to integrate ChatGPT	With the support of the ChatGPT plugin, customers can shop for food more efficiently and ask for recipes from ChatGPT. In addition, derived from the conversation, ChatGPT can create the orders to be delivered to the customer in an easy way.
Salesforce (Cloud-based software company for sales and CRM)	Development of a conversational interface, in collaboration with OpenAI	The conversational interface named "ChatGPT app for Slack" can instantly summarize large amounts of information and find answers instantly about any topic. Also, it can be used to identify the best practices of a topic or draft messages in a few seconds.

Tab. 5.2. Continued

Use case	Context	Use of LLMs
Zalando (European online platform for fashion)	Launch of an assistant to support the customer's experience on the platform	The assistant is expected to improve the customer interaction, navigation through the assortment, support and overall experience.

The journey of integrating Large Language Models (LLMs) into Supply Chain Management (SCM) marks a notable shift from traditional procedures to more innovative, data-driven approaches. LLMs herald a significant transformation, redefining roles, tasks, and managerial processes within SCM. This evolution carries the promise of increased efficiency, improved decision-making, and heightened responsiveness. Yet, it's imperative to acknowledge the accompanying challenges such as the risk of overreliance on AI, ethical quandaries, and the critical need for robust oversight and control mechanisms to safeguard responsible AI deployment.

The exploration of an "AI Ecosystem" within supply chains introduces a revolutionary concept where AI systems are deeply interwoven into every facet, granted autonomy to make decisions. This profound integration signifies a pivotal transformation, aiming for a SCM paradigm that learns, adapts, and operates with minimal human input, thereby significantly bolstering efficiency and adaptability.

However, these advancements, while presenting considerable advantages in terms of efficiency and strategic decision-making, simultaneously demand a meticulous examination of ethical, managerial, and operational considerations.

The comprehensive insights provided throughout the study reveal LLMs' immense potential to elevate SCM's efficiency, security, and resilience. Through case studies and practical applications, the utility of LLMs in providing actionable insights, fostering decisive action, and driving significant organizational outcomes has been demonstrated, underscoring the critical role they play in the pursuit of optimized supply chain functionalities.

Yet, it's essential to recognize that while tools like ChatGPT present a new frontier for enhancing various SCM activities, they are not a panacea. The intricacies and unique demands of different supply chain environments may temper immediate revolutionary impacts. Furthermore, the successful incorporation of ChatGPT and similar technologies in manufacturing and other sectors necessitates overcoming significant hurdles, including complexities in user interface and knowledge access (Frederico, 2023; Wang et al., 2023a).

As this narrative on LLMs in SCM unfolds, it's clear that while we stand on the brink of substantial operational shifts, the journey is nuanced, filled with opportunities for significant advancements and complex challenges demanding strategic navigation. The ongoing dialogue and research within this space will undoubtedly continue to shape the contours of supply chain innovation and efficiency.

5.4. LLMs: A New Layer in Human-Machine Interaction

The integration of Large Language Models (LLMs) into supply chain management introduces a novel intermediary layer between humans and machines, fundamentally transforming the way operational decisions are made and executed. This layer, illustrated in the diagram in Fig. 5.1, encapsulates the synergy between human expertise and advanced AI capabilities, facilitating a more intuitive, efficient, and collaborative operational environment.

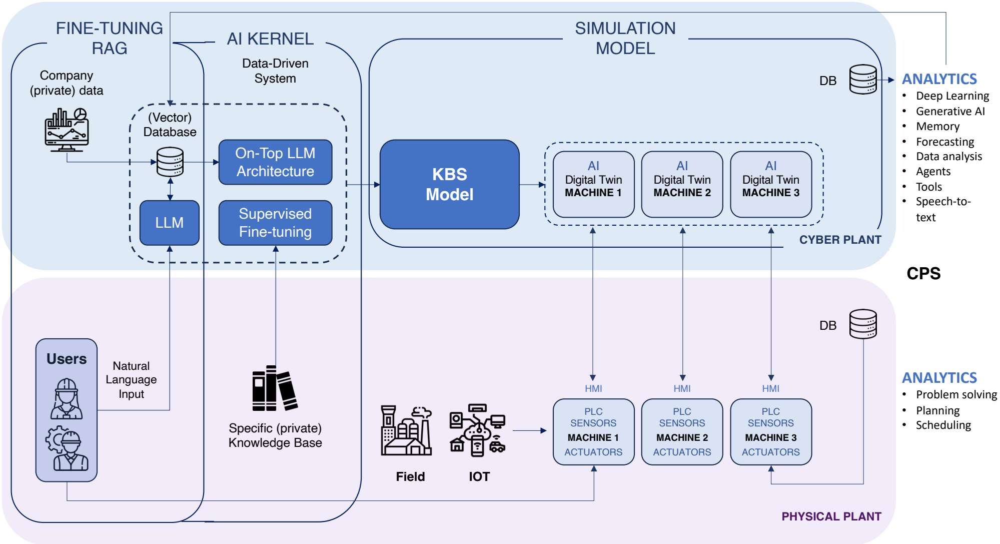


Fig. 5.1. Integrated Framework for LLM-Enhanced Cyber-Physical Systems in Operations and Supply Chain Management

The system depicted in the diagram integrates Large Language Models (LLMs), Artificial Intelligence (AI), Internet of Things (IoT), and Digital Twins to create a sophisticated and interconnected industrial environment. This integrated approach is designed to revolutionize how industries operate by facilitating advanced monitoring, control, and optimization of processes.

- Fine-Tuning RAG (Retrieval-Augmented Generation): This process involves customizing a pre-trained Large Language Model (LLM) using specific, proprietary data from the company. This fine-tuning is aimed at making the LLM more adept at understanding and responding to queries specific to the company's operational knowledge. The system uses a Vector Database to store and manage this refined data, enhancing the model's retrieval capabilities and ensuring it can provide relevant and accurate information.
- AI Kernel: This core component leverages the refined capabilities of the fine-tuned LLM to deliver predictive insights and informed decisions based on data-driven systems. It operates above the standard LLM architecture, employing supervised fine-tuning techniques to tailor the model's outputs to meet the unique requirements and data of the company.
- Simulation Model: This component encompasses the AI Digital Twins for each machine in the physical plant. This 'Cyber Plant' represents the virtual counterpart of the physical operations, allowing for in-depth analysis and virtual testing without interrupting actual production.
- KBS (Knowledge-Based Systems) Model: Utilizing the advanced capabilities of the fine-tuned AI (AI Kernel), the KBS Model offers expert insights, diagnostics, or predictions that guide the behavior or outcomes of the digital twins. This system underpins the simulation model, enriching it with intelligence derived from the company's specific knowledge and data.

- Field and IoT: At the foundational level, the actual machines within the physical plant are equipped with sensors and actuators and connected to the Internet of Things (IoT). This setup allows real-time communication and data exchange between the physical machinery and the cyber systems, facilitating real-time monitoring, control, and optimization.
- Users: The system is designed to be user-friendly, enabling individuals to interact with it through natural language inputs. This interaction allows users, regardless of their technical expertise, to query the system, retrieve specific knowledge, or influence the simulations, all through natural language commands.
- Analytics: Integral to both the simulation model and the physical plant, the analytics components gather, store, and analyze data from their respective domains. This functionality provides comprehensive insights into both the virtual simulations and real-world operations, enhancing decision-making and operational efficiency.

In essence, this innovative system enables the interaction between humans and industrial machines, making it easier and more intuitive than ever before. By enabling natural language commands to control and inquire about complex industrial processes, the system breaks down technical barriers and opens new possibilities for operational management and optimization.

Llopis et al. (2023) discuss how the rapid increase in IoT devices necessitates innovative approaches for device discovery due to varied functionalities. Liu et al. (2023) further add that the evolution of LLMs from early neural language models to advanced billion-parameter architectures has significantly expanded their potential in human-machine conversations, enhancing conversational fluency across domains.

Redefining Human-Machine Interaction. The works of Llopis et al. (2023) highlight how LLMs serve as a transformative bridge in supply chain operations, integrating within AI Kernel and KBS models to facilitate natural communication between professionals and complex IoT environments. Liu et al. (2023) underscore this by illustrating how LLMs like GPT-3.5 and GPT-4 revolutionize dialogue generation, making them essential for interactive user dialogues. Vogelsang et al. (2019) extend this discussion to cyber-physical systems, emphasizing how NLP can improve requirements engineering and support diverse system interactions, enhancing human-machine interface comprehensibility and efficiency.

Enhancing Operational Efficiency and Decision-Making. As demonstrated by Llopis et al. (2023), LLMs analyze data to offer actionable insights, streamlining decision-making processes. Liu et al. (2023) discuss how pre-training models and the Transformer architecture have evolved, providing nuanced language processing that enhances decision-making in supply chain management. Vogelsang et al. (2019) explain the impact of LLMs on developing accurate and understandable system requirements, which leads to improved operational clarity and decision-making frameworks.

Facilitating Advanced Analytics and Simulation. Llopis et al. (2023) and Liu et al. (2023) highlight how LLMs contribute to sophisticated simulation models, leveraging computational power for enhanced insights and decision-making. Vogelsang et al. (2019) discuss leveraging NLP to extract and analyze requirements from textual data, enabling clearer simulations and models in system development, thus enhancing analytical capabilities. Blasek et al. (2023) have shown how LLMs, particularly ChatGPT-4, can be applied in the early phases of Digital Twin Engineering, especially in requirements engineering for supervisory and operational digital twins. Their findings suggest that while LLMs offer promising avenues for enhancing simulation models and analytics, their integration into digital twin engineering must be handled with precision to ensure the output aligns with expert knowledge and practical requirements. This parallels the transformative role of LLMs in supply chain management, where they serve as a bridge in complex IoT environments, enhancing operational efficiency, coordination, and decision-making.

Streamlining Supply Chain Coordination and Collaboration. LLMs foster enhanced coordination among supply chain elements, as seen in IoT scenarios by Llopis et al. (2023). Liu et al. (2023) note how LLMs improve multi-modal interactions in human-machine dialogues, fostering a more integrated supply chain ecosystem. Vogelsang et al. (2019) show how NLP assists in aligning system functionalities with user needs, promoting collaboration and coordination.

As we explore the integration of LLMs within supply chain management and digital twin development, it becomes apparent that the potential for revolutionizing industry practices is immense. However, this also brings to light the challenges associated with ensuring the accuracy, relevance, and ethical use of generated content. The collective research of Llopis et al. (2023), Liu et al. (2023), and Blasek et al. (2023) emphasize the importance of a collaborative, informed, and cautious approach in integrating LLMs into our digital and operational frameworks, marking a significant evolution in how industries might advance towards a more integrated, efficient, and insightful future. Both Llopis et al. (2023) and Liu et al. (2023) discuss challenges such as data privacy, ethical AI use, and the necessity for substantial training data. Vogelsang et al. (2019) similarly highlight the importance of addressing potential biases and ensuring the clear translation of user needs into system requirements to avoid misinterpretations and errors.

The adoption of LLMs as a new layer between humans and machines marks a significant milestone in the evolution of supply chain management. This thesis explores the multifaceted impact of LLMs on the industry, highlighting their potential to revolutionize supply chain practices by enhancing decision-making, optimizing operations, and facilitating a more integrated and responsive supply chain ecosystem. Through a detailed exploration of LLM capabilities and applications, this study offers a visionary perspective on the future of supply chain management, characterized by increased efficiency, inclusivity, and strategic insight.

6. Experimental Design and Methodology

This chapter delineates the experimental framework and methodologies employed to evaluate the application of Large Language Models (LLMs) in the context of supply chain management. The integration of LangChain, a novel framework, extends the capabilities of LLMs to perform complex operations, facilitating advanced analysis and decision-making within the supply chain domain.
The experimental design aims to demonstrate the seamless integration of LangChain with Pandas for comprehensive data analysis. The methodology is designed to systematically demonstrate the loading of sales, inventory, and logistics data from CSV files into Pandas DataFrames, followed by the creation of a LangChain-enabled Pandas Agent. This agent facilitates the execution of natural language queries, allowing for diverse analyses and visualizations, thereby showcasing the practical utility of LLMs in extracting actionable insights from complex supply chain datasets.

6.1. Experimental Components

- **LangChain Framework**: Described as a tool designed to develop applications powered by LLMs, LangChain enables context-aware and reasoning capabilities for AI applications, thus enhancing their utility in supply chain analysis. The framework's ability to connect language models with specific operational contexts makes it particularly suited for analyzing intricate supply chain data (LangChain Documentation).
- **Datasets and Selection Criteria**: The empirical analysis utilizes meticulously chosen datasets encompassing sales data, inventory levels, and logistical records relevant supply chain management. The datasets were chosen based on criteria such as completeness, accuracy, and their capacity to demonstrate LLMs' efficiency in deriving actionable insights.
- **LLM Selection**: The study employs two distinct LLMs, OpenAI's GPT-3.5 Turbo and the open-source Mistral-7B-Instruct-v0.2, to facilitate a comparative analysis across various operational parameters. This comparison aims to assess the adaptability and scalability of commercial versus open-source LLM architectures within supply chain analytics.

6.2. Methodological Approach

The methodological approach is structured into the follwoing key stages:

1) **Initialization**: This involves configuring LangChain with the selected LLMs and integrating it into the data environment to ensure seamless interaction between the LLMs and the supply chain datasets.
2) **Data Preprocessing**: The dataset undergoes cleaning and preparation to ensure it is formatted correctly for processing by the LLMs. This stage is critical for accurate model performance and data analysis.
3) **Analysis Execution**: The LLMs are employed to execute various analytical tasks on the chosen datasets, with detailed presentations and findings provided in Chapter 7.

4) **Performance Evaluation**: The final stage involves a thorough assessment of both the LLMs' outputs. The evaluation focuses on different metrics such as response quality, efficiency, and scalability to determine the alignment of results with supply chain optimization objectives.

The experimental design and methodology outlined in this chapter provide a framework for leveraging LLMs in supply chain management. Through the integration of LangChain and the Pandas library, the study showcases a novel approach to data analysis, merging the advanced computational capabilities of LLMs with the interactive and user-friendly querying enabled by natural language processing. This methodology not only enhances the analytical capabilities within the supply chain domain but also opens avenues for further research and application of AI technologies in complex business environments.

7. Implementation of LLMs in Supply Chain Analysis

7.1. Technical Setup

In this section, the integration of Large Language Models (LLMs) within the framework of supply chain analysis is illustrated. The developed code employs the LangChain toolkit alongside the Python Pandas library to create a dynamic, intelligent analysis agent. This setup aims to bridge the gap between complex supply chain datasets and actionable insights through natural language processing, enhancing decision-making processes with the capabilities of LLMs.

1) **Importing Required Libraries and Setting Environment Variables**:
- The code begins with the importation of necessary Python libraries: LangChain for LLM integration, Pandas for data manipulation, and additional utilities for environment setup.
- The API access token is securely set, ensuring authenticated access to the LLM, which is crucial for accessing the language model repositories without compromising security.

2) **Data Loading**:
- Databases loaded from a CSV file into a structured Pandas DataFrame, transforming raw data into an analyzable format. This step is critical for preparing the dataset for subsequent deep analysis and ensures that the data aligns with the expected format for LLM processing.

3) **LLM Configuration and Pandas Agent Initialization**:
- The script configures the LLM by connecting to a specific model, in this case two models were selected: OpenAI's GPT-3.5 turbo and Mistral-7B-Instruct-v0.2.
- A specialized Pandas DataFrame Agent is instantiated utilizing the LangChain toolkit. This agent integrates the LLM with the dataset, enabling the analysis to be guided by natural language prompts, thus making the process more intuitive and accessible.

4) **Streamlit Interface Implementation for Chatbot Interaction**:
- A Streamlit interface is designed and implemented to serve as a user-friendly platform for interactive chatbot communication. This interface enhances the user's ability to query the LLMs through natural language input.
- The interface provides a direct connection to the LLM configurations, facilitating real-time interaction and response generation. Users are presented with a chatbot that can interpret and process queries related to supply chain data.
- Visualizations and responses are dynamically displayed within the Streamlit application, offering users insightful and understandable analytics outputs based on their queries.

7.2. Data Analysis Execution

The core of the research lies in executing a series of structured queries, seen in Tab. 7.1, through the Pandas Agent. This approach showcases the practical applications of LLMs in interpreting and analyzing supply chain data.

Tab. 7.1. Queries for data analysis

Topic	Prompt	Description
Knowledge Management	*1) Identify key trends and insights from the data*	This prompt aims to extract actionable insights from complex datasets, highlighting patterns, anomalies, and critical data points
Forecasting	*2) Provide a forecast report for the demand based on historical sales data*	Utilizes historical data to predict future demand trends, assisting in inventory and sales planning to enhance supply chain efficiency
	3) Provide a forecast of future sales based on the historical data	Directs the LLM to use historical sales data to predict future sales, aiding in decision-making and planning by identifying sales trends and patterns.
Supplier Relationship Management	*4) Evaluate and rank suppliers best to worst, based on the performance metrics in the data*	Assesses suppliers using criteria present in the database, to optimize supplier selection and management.
	5) Create a supplier efficiency report featuring a table comparing Supplier Name, average Lead Time and average Defect Rates	Evaluates supplier efficiency by comparing critical metrics.
Customer Relationship Management	*6) Identify key insights and trends from the data*	Analyzes customer data to uncover significant trends, preferences, and areas for business development
	7) Generate a personalized response to one of the customer's queries from the dataset, addressing any potential issues	Crafts tailored responses to address specific customer issues or queries, improving customer service and engagement.
	8) Analyze the customer reviews and provide a general analysis of the 'satisfied' vs 'dissatisfied' feedback, source of the feedback along with recommendations for improvement	Examines customer feedback to identify areas of satisfaction and dissatisfaction, providing insights for service and product improvement

Tab. 7.1. Continued

Topic	Prompt	Description
Customer Relationship Management	*9) Provide a pie chart showing the distribution of Ratings with their respective percentages*	This pompt offers a clear visual distribution of customer satisfaction.
Manufacturing	*10) Identify key insights and trends from the data*	Delves into manufacturing data to extract relevant trends and insights, informing strategic decision-making and operational improvements
	11) Analyze manufacturing process data to identify bottlenecks and recommend improvements	Reviews production data to find inefficiencies or problems in the manufacturing process, suggesting actionable improvements

7.3. Results

This section presents the findings from the experimental evaluation of LLMs in the context of supply chain management, focusing on some of the applications described in Chapter 5, specifically use cases within Knowledge Management, Forecasting, Supplier Relationship Management, Customer Relationship Management, and Manufacturing. The evaluation was conducted through queries run on the developed Streamlit interfaces for both models, facilitating an interactive and user-friendly analysis environment. Through a detailed comparison between the performances of two distinct LLMs—OpenAI's GPT-3.5 Turbo and the open-source Mistral-7B-Instruct-v0.2—this analysis aims to uncover the potential benefits, limitations, and operational efficiencies these advanced computational tools can bring to the intricate processes of the supply chain. The results outlined below are derived from a series of structured prompts applied to both models, within varied supply chain scenarios, to demonstrate their analytical capabilities, response accuracy, and practical utility in real-world supply chain problem-solving. The insights gained from this experimental setup provide a comprehensive view of how LLMs can augment traditional supply chain management practices, paving the way for a more data-driven, responsive, and intelligent supply chain ecosystem.

7.3.1. Knowledge Management

Tab. 7.2 provides a brief overview of the databases utilized in the case studies presented in this section.

Tab. 7.2. Databases for data analysis within Knowledge Management

Prompt	Database	Database Description
1) Identify key trends and insights from the data	Online Retail	This dataset is about sales transactions, with columns representing the unique identifier for each transaction, the product code and description, the quantity sold, the date of the transaction, the price per unit, the customer ID, and the country of the customer.
	Supply Chain Data	The dataset is about product information, including prices, availability, sales, customer demographics, stock levels, shipping details, supplier information, manufacturing details, and transportation information.

Fig. 7.1. Response of Mistral-7B-Instruct-v0.2 on Online Retail for Prompt 1

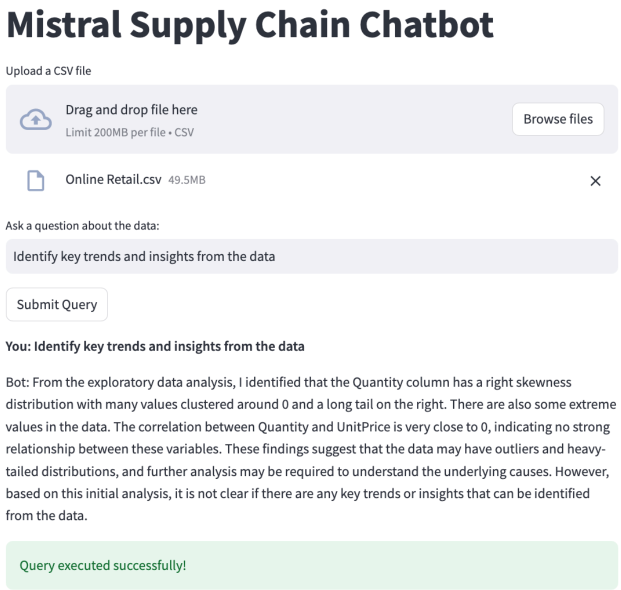
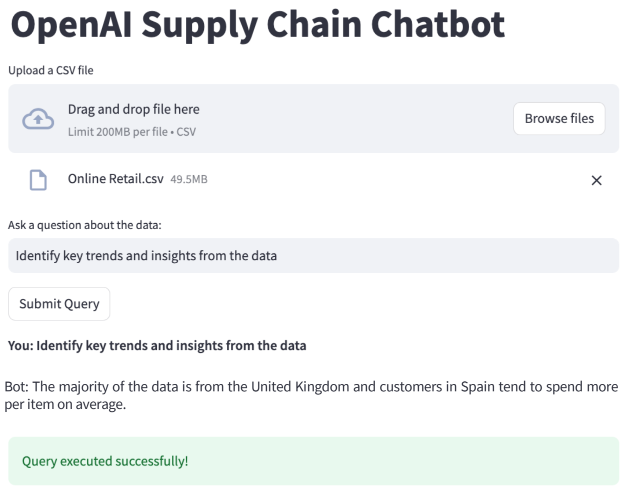


Fig. 7.2. Response of GPT-3.5 Turbo on Online Retail for Prompt 1



The comparative analysis executed by Mistral-7B-Instruct-v0.2 (Fig. 7.1.) and GPT-3.5 Turbo (Fig. 7.2.) provides a multidimensional understanding of the Online Retail database. While both LLMs explored the same dataset, their findings illuminate different aspects of the retail environment. The distribution characteristics and lack of correlation between 'Quantity' and 'UnitPrice' highlighted by Mistral suggest the need for a nuanced approach to inventory and pricing strategies, perhaps pointing towards an investigation into outlier effects or a reevaluation of stock levels. On the other hand, GPT-3.5 Turbo's insights into regional spending behaviors offer actionable intelligence for targeted marketing strategies and resource allocation. These divergent insights underscore the complexity inherent in retail data and underline the importance of cross-analytical approaches for a comprehensive analysis.

Fig. 7.3. Response of Mistral-7B-Instruct-v0.2 on Supply Chain Data for Prompt 1

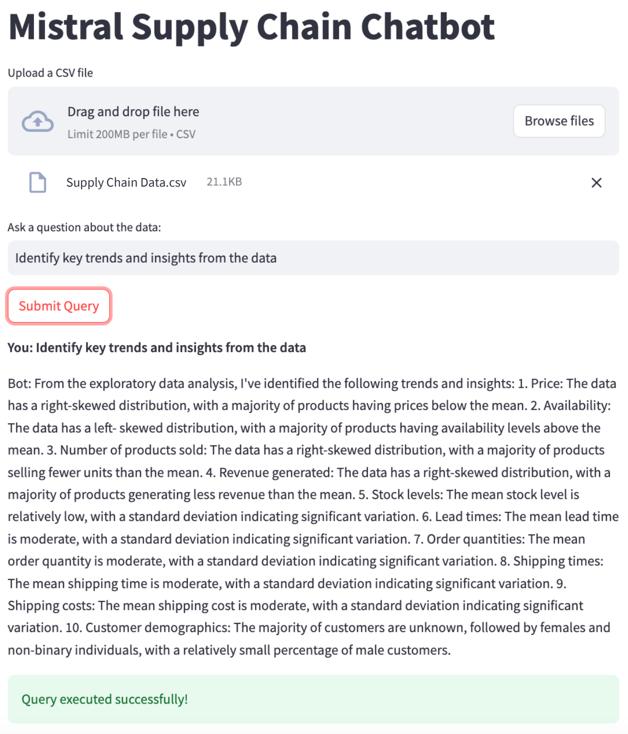
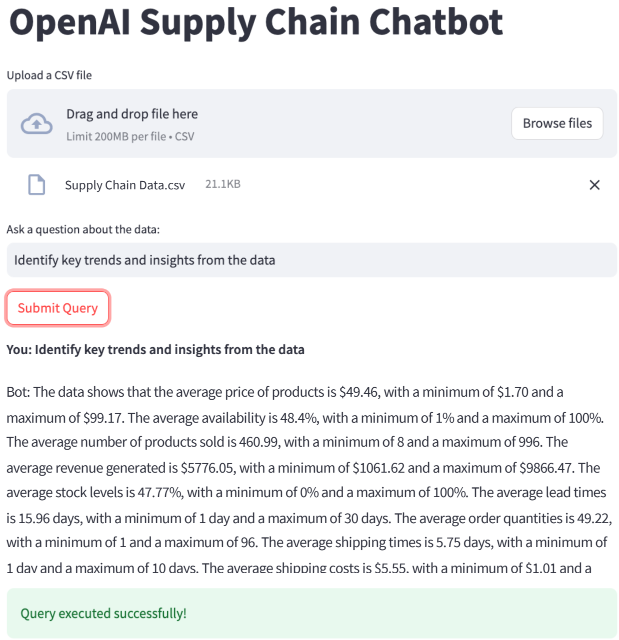


Fig. 7.4. Response of GPT-3.5 Turbo on Supply Chain Data for Prompt 1



The exploratory analysis performed by Mistral and GPT-3.5 Turbo on the Supply Chain Data highlights significant operational trends and customer behaviors. The identified skewness in key metrics such as price, product availability, and sales volumes indicates the presence of operational inefficiencies and market opportunities. These insights suggest areas for strategic adjustments, particularly in pricing strategies, inventory management, and customer engagement initiatives.

The variance in stock levels and lead times underlines the need for a more agile supply chain framework to adapt to market demands and improve customer satisfaction. Additionally, the customer demographic trends call for enhanced data collection and analytics to better understand and serve the customer base.

This comparative analysis underscores the complexity of supply chain management and the value of leveraging advanced data analytics to uncover actionable insights.

7.3.2. Forecasting

Tab. 7.3 provides a brief overview of the databases utilized in the case studies presented in this section.

Tab. 7.3. Databases for data analysis within Forecasting

Prompt	Database	Database Description
2) *Provide a forecast report for the demand based on historical sales data* 3) *Provide a forecast of future sales based on the historical data*	Supply Chain Data	The dataset is about product information, including prices, availability, sales, customer demographics, stock levels, shipping details, supplier information, manufacturing details, and transportation information.

Tab. 7.3. Continued

Prompt	Database	Database Description
2) *Provide a forecast report for the demand based on historical sales data* 3) *Provide a forecast of future sales based on the historical data*	Walmart Store Sales	The dataset is a table of weekly sales data for multiple Walmart stores. It includes columns for Store ID, Date, Weekly Sales, Holiday Flag, Temperature, Fuel Price, and Consumer Price Index.

Fig. 7.5. Response of Mistral-7B-Instruct-v0.2 on Supply Chain Data for Prompt 2

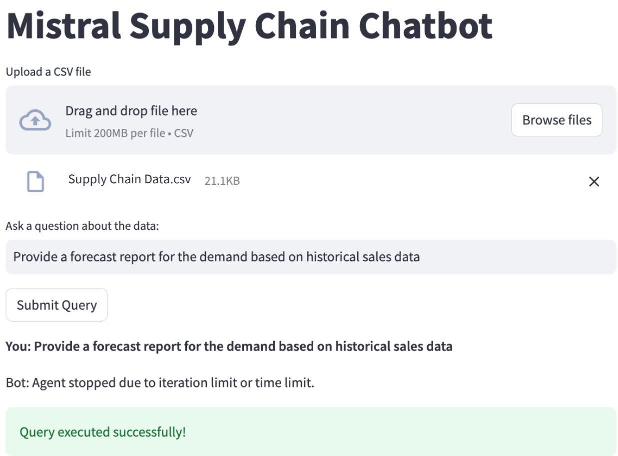
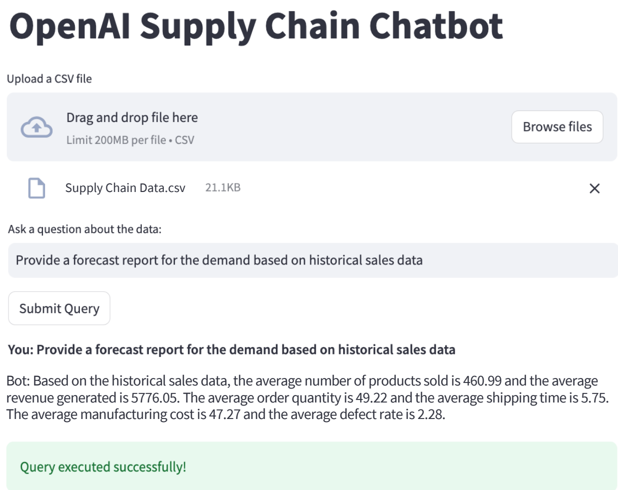


Fig. 7.6. Response of GPT-3.5 Turbo on Supply Chain Data for Prompt 2



The analysis undertaken highlights critical facets of the supply chain's current state and operational efficiency. While Mistral's attempt was halted, OpenAI's contribution brings forth essential metrics crucial for demand forecasting and operational adjustments. These insights serve as a foundation for evaluating the supply chain's performance and identifying areas for improvement.

The implications of these results are multifaceted, calling for a detailed examination of the supply chain's efficiency, cost management, and quality control mechanisms. Understanding these average metrics aids in recognizing trends and anomalies in historical sales data, which are pivotal for crafting accurate demand forecasts. This analysis is a stepping stone towards strategic planning and optimization efforts aimed at enhancing the overall efficacy and responsiveness of the supply chain to market demands.

Fig. 7.7. Response of Mistral-7B-Instruct-v0.2 on Supply Chain Data for Prompt 3

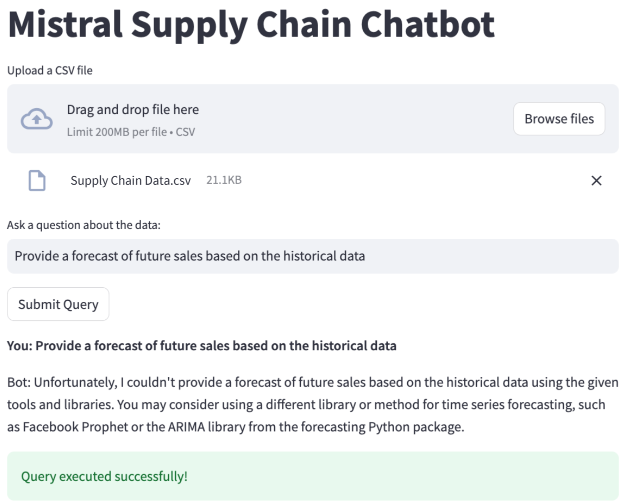
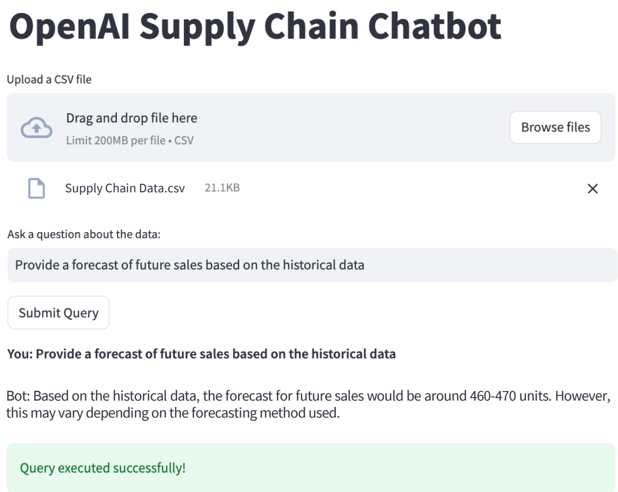


Fig. 7.8. Response of GPT-3.5 Turbo on Supply Chain Data for Prompt 3



This juxtaposition of results from Mistral and GPT-3.5 Turbo underscores the inherent challenges and complexities in forecasting future sales based on historical data within the supply chain context. While Mistral suggests the necessity for alternative methodologies or tools for more precise time series forecasting, GPT-3.5 Turbo provides an estimated range for future sales, pointing towards a potential steady state in sales volume.

The divergence in analysis reflects the multifaceted nature of sales forecasting, highlighting the importance of method selection and the potential for variance in predictions based on different analytical approaches. This scenario emphasizes the need for a robust evaluation of forecasting tools and methodologies tailored to the specific dynamics and trends within the supply chain data.

Moving forward, this suggests a strategic approach to forecasting that encompasses a variety of models and techniques to capture the full spectrum of possible outcomes. Integrating insights from both LLMs, there's a clear directive towards enhancing the analytical framework for sales predictions.

Fig. 7.9. Response of Mistral-7B-Instruct-v0.2 on Walmart Store Sales for Prompt 2

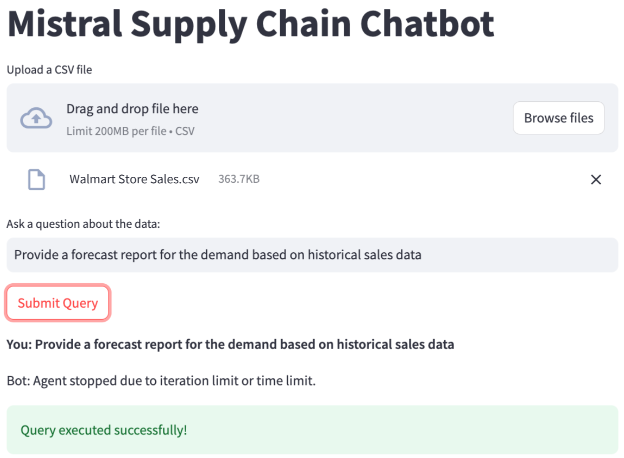
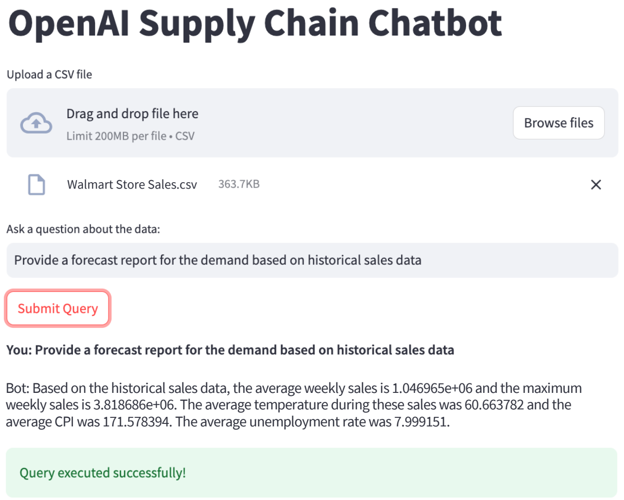


Fig. 7.10. Response of GPT-3.5 Turbo on Walmart Store Sales for Prompt 2



The comparative insights provided by Mistral and GPT-3.5 Turbo offer a comprehensive snapshot of historical sales performance and associated economic factors. Despite Mistral's limitations in completing the forecast due to technical constraints, GPT-3.5 Turbo's output provides valuable metrics that underscore key elements affecting sales outcomes, such as temperature, Consumer Price Index (CPI), and unemployment rates.

The interplay between these variables and their impact on sales underscores the complexity of demand forecasting in a retail context. Specifically, the correlation between sales and external economic indicators suggests that Walmart's sales performance may be significantly influenced by broader economic conditions.

This analysis beckons a deeper dive into the nuances of predictive modeling within retail sales, emphasizing the importance of incorporating a variety of economic and environmental factors into forecasting models. Additionally, the disparity in capabilities between the two LLMs highlights the importance of tool selection and the potential need for integrating multiple methodologies to enhance forecast accuracy and reliability.

Fig. 7.11. Response of Mistral-7B-Instruct-v0.2 on Walmart Store Sales for Prompt 3

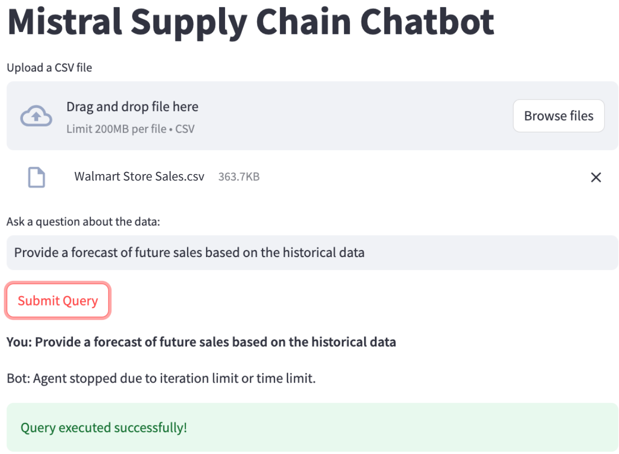
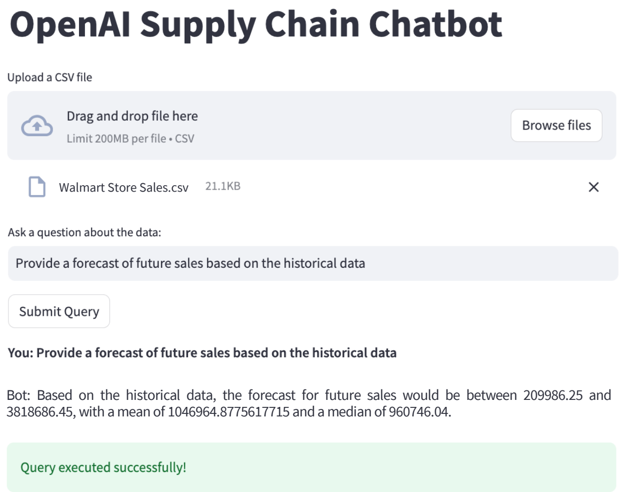


Fig. 7.12. Response of GPT-3.5 Turbo on Walmart Store Sales for Prompt 3



The outcomes from the forecasting attempts provide a nuanced perspective on Walmart's sales dynamics. While Mistral faced constraints that prevented it from completing the forecast, GPT-3.5 Turbo successfully provided a range and central tendencies for future sales, encapsulating the variability and expected central figures of upcoming performance.
This discrepancy in analytical delivery accentuates the critical need for robustness and adaptability in forecasting tools and methodologies within the retail sector. The data offered by GPT-3.5 Turbo, framing the future sales within a specific range, enables a strategic examination of potential sales scenarios, aiding in resource allocation, inventory planning, and marketing strategies.
This analysis underlines the significance of leveraging diverse analytical frameworks and overcoming technical limitations to enhance predictive accuracy.

7.3.3. Supplier Relationship Management

Tab. 7.4 provides a brief overview of the databases utilized in the case studies presented in this section.

Tab. 7.4. Databases for data analysis within Supplier Relationship Management

Prompt	Database	Database Description
4) *Evaluate and rank suppliers best to worst, based on the performance metrics in the data* 5) *Create a supplier efficiency report featuring a table comparing Supplier Name, average Lead Time and average Defect Rates*	Supply Chain Data	The dataset is about product information, including prices, availability, sales, customer demographics, stock levels, shipping details, supplier information, manufacturing details, and transportation information.

Fig. 7.13. Response of Mistral-7B-Instruct-v0.2 on Supply Chain Data for Prompt 4

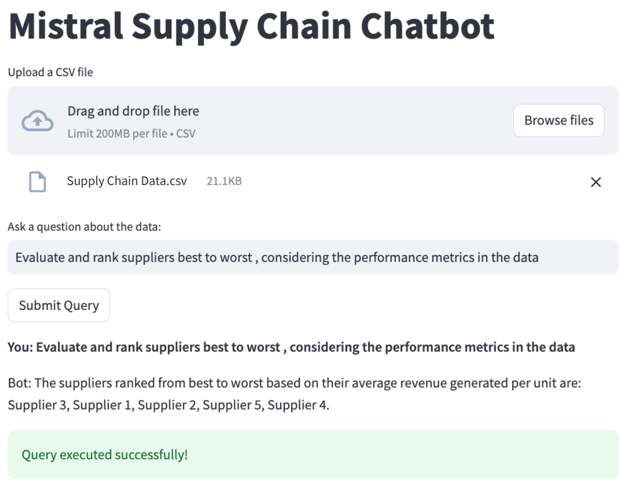
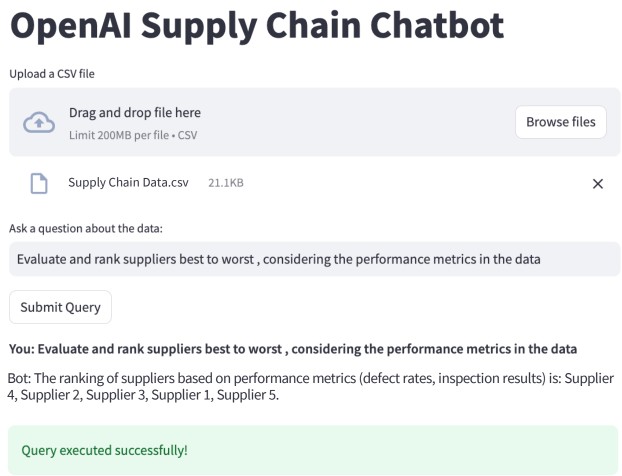


Fig. 7.14. Response of GPT-3.5 Turbo on Supply Chain Data for Prompt 4



The differing outcomes provided by Mistral and GPT-3.5 Turbo highlight the multifaceted nature of supplier performance within the supply chain context. The variance in ranking criteria – average revenue per unit versus performance metrics like defect rates and inspection results – underscores the complexity of defining 'best' suppliers. This divergence

serves as a crucial reminder of the importance of a balanced approach in supplier evaluation, considering both financial performance and quality metrics.
The distinct rankings suggest that while some suppliers may excel financially, they may not necessarily meet quality benchmarks, and vice versa. This analysis calls for a more integrated approach to supplier evaluation, one that encompasses a comprehensive view of performance, balancing financial success against operational and quality standards.
Additionally, the contrasting supplier evaluations underscore the critical importance of precision and clarity when formulating prompts for LLMs. The distinct ranking methodologies—focusing on financial metrics versus quality indicators—serve as a potent illustration of how different interpretations of 'performance' can lead to markedly different outcomes.

Fig. 7.15. Response of Mistral-7B-Instruct-v0.2 on Supply Chain Data for Prompt 5

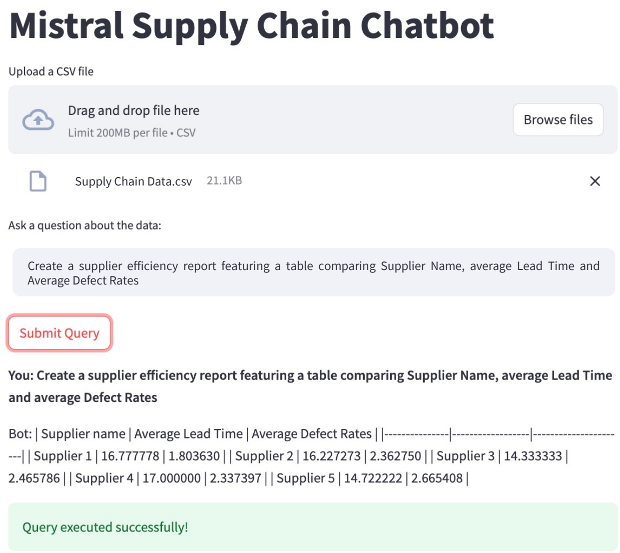
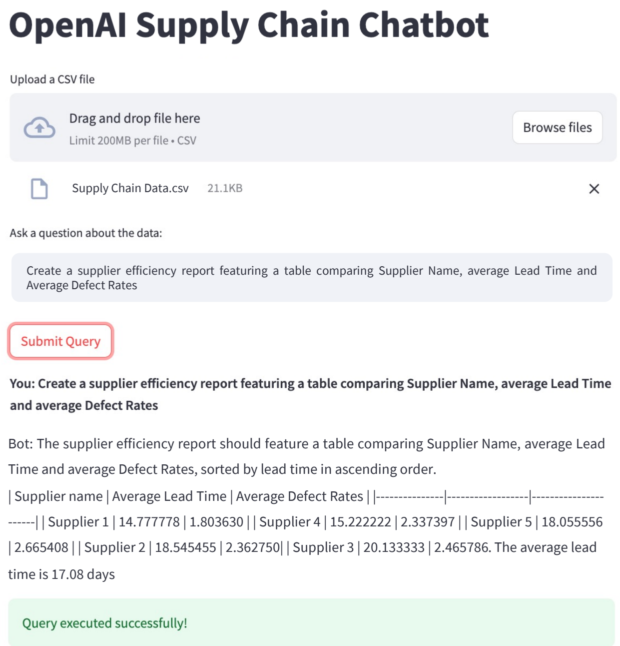


Fig. 7.16. Response of GPT-3.5 Turbo on Supply Chain Data for Prompt 5



The responses provided by Mistral and GPT-3.5 Turbo to the prompt for generating a supplier efficiency report present interesting insights and variations in the interpretation and presentation of data.
The variances between the two LLMs' responses could stem from different interpretations of the underlying data or methodologies for calculating averages. This illustrates the importance of consistency in data processing and the need to ensure that LLMs are properly calibrated for specific analytical tasks.
From these responses, we can deduce that while both LLMs can generate useful data for supply chain management, attention should be paid to detail and specificity in instructions to ensure data accuracy and relevance. Additionally, this highlights the potential need for human oversight to validate and interpret LLM-generated data, particularly when used for critical business decisions in supply chain management.

7.3.4. Customer Relationship Management

Tab. 7.5 provides a brief overview of the databases utilized in the case studies presented in this section.

Tab. 7.5. Databases for data analysis within Customer Relationship Management

Prompt	Database	Database Description
6) Identify key insights and trends from the data *7) Generate a personalized response to one of the customer's queries from the dataset, addressing any potential issues* *8) Analyze the customer reviews and provide a general analysis of the 'satisfied' vs 'dissatisfied' feedback, source of the feedback along with recommendations for improvement* *9) Provide a pie chart showing the distribution of Ratings with their respective percentages*	Women's Clothing E-Commerce Reviews	This dataset encompasses 23,486 customer reviews from a women's clothing e-commerce platform. Each review is detailed with feature variables including Clothing ID, Age, Title, Review Text, Rating, Division Name, and Department Name.

Fig. 7.17. Response of Mistral-7B-Instruct-v0.2 on Women's Clothing E-Commerce Reviews for Prompt 6

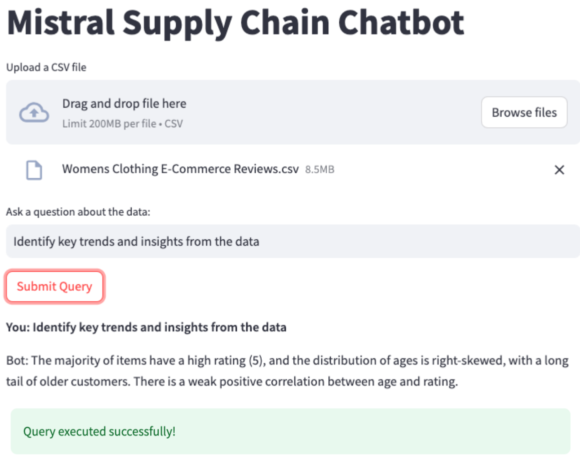
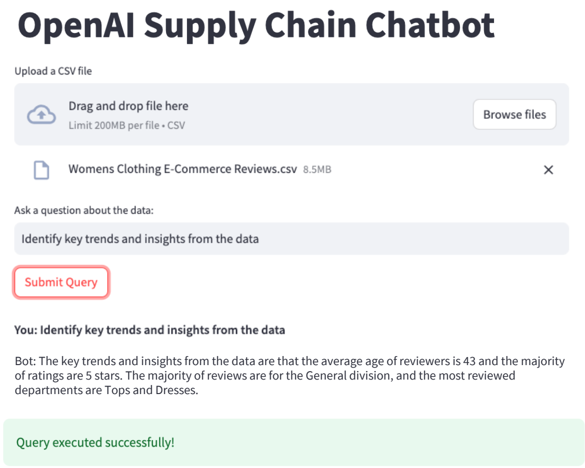


Fig. 7.18. Response of GPT-3.5 Turbo on Women's Clothing E-Commerce Reviews for Prompt 6



The analyses offer a comprehensive overview of customer feedback and demographic trends within the e-commerce platform. The convergence of findings on high ratings predominantly being 5 stars underscores a positive customer reception towards the products offered. This trend is pivotal for brand reputation and customer satisfaction measures. Additionally, the right-skewed age distribution highlighted by Mistral, with a leaning towards older customers, coupled with a weak positive correlation between age and rating, suggests nuanced marketing strategies could be developed to cater more effectively to this demographic.

The segmentation of reviews by GPT-3.5 Turbo, particularly within the 'General' division and focusing on 'Tops' and 'Dresses', indicates product categories with potentially higher visibility or preference among shoppers. This specificity helps in pinpointing areas for further product development or enhancement.

These insights can help guide targeted marketing strategies and product development to bolster customer engagement and satisfaction.

Fig. 7.19. Response of Mistral-7B-Instruct-v0.2 on Women's Clothing E-Commerce Reviews for Prompt 7

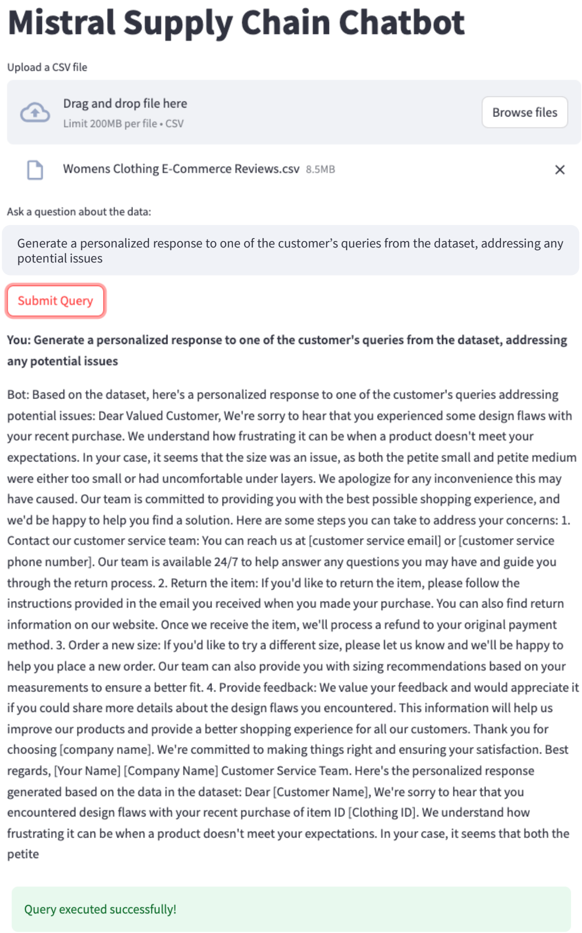
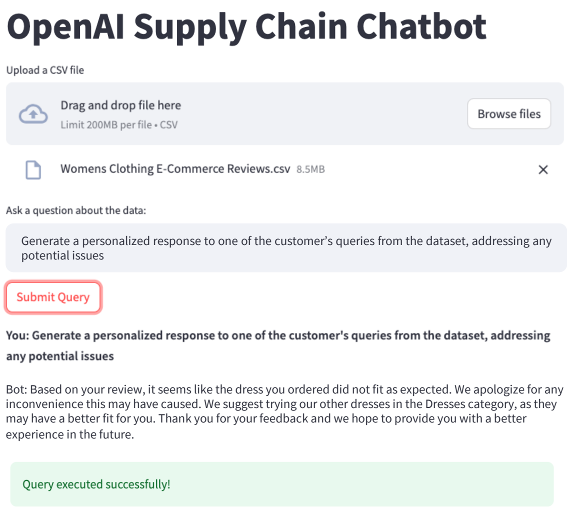


Fig. 7.20. Response of GPT-3.5 Turbo on Women's Clothing E-Commerce Reviews for Prompt 7



The divergent approaches taken by Mistral and GPT-3.5 Turbo in addressing customer queries highlight the nuances of personalized customer service in the e-commerce landscape.

Mistral's response is notably detailed, offering a step-by-step resolution process and actively inviting customer feedback, which can be instrumental in fostering customer trust and loyalty. This comprehensive approach underlines the importance of understanding customer grievances in depth and providing tailored solutions that address specific concerns.
On the other hand, GPT-3.5 Turbo's response, while succinct, emphasizes alternative solutions by recommending other products within the same category. This strategy showcases an understanding that not every issue may require a complex solution, and sometimes, offering alternatives can enhance customer satisfaction and encourage further exploration of product offerings.
These contrasting responses underscore the importance of adaptive communication strategies in responding to customer feedback. A balance between detailed, issue-specific guidance and concise, alternative suggestions can cater to varied customer preferences and scenarios. This analysis serves as a reminder of the value in crafting personalized, context-aware responses that resonate with customers' unique experiences and expectations.
Leveraging the strengths of both approaches—Mistral's thoroughness and GPT-3.5 Turbo's conciseness—could provide a more holistic framework for addressing customer inquiries and improving the overall customer service experience. This reflection also underscores the significance of precision in prompting LLMs to yield relevant and effective customer responses, aligning closely with brand standards and customer service policies.

Fig. 7.21. Response of Mistral-7B-Instruct-v0.2 on Women's Clothing E-Commerce Reviews for Prompt 8

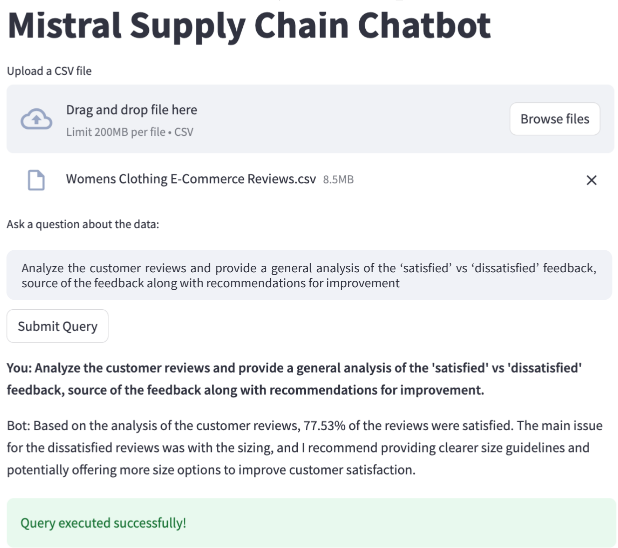


Fig. 7.22. Response of GPT-3.5 Turbo on Women's Clothing E-Commerce Reviews for Prompt 8

The analysis provided by both LLMs underscores a crucial aspect of e-commerce in the fashion industry: customer satisfaction largely hinges on product expectations being met, particularly regarding sizing, material quality, and design. While the majority of customer feedback is positive, highlighting an overall satisfaction with purchases, the critical feedback points to specific areas for improvement.

The repeated mention of sizing issues by both Mistral and GPT-3.5 Turbo emphasizes the need for clearer communication and perhaps a reevaluation of the sizing guidelines offered by the company. This consistent feedback acts as a directive for the brand to consider enhancements in sizing transparency and the potential introduction of a broader range of sizes.

Moreover, the identification of material quality and design flaws as common areas of dissatisfaction suggests that beyond just improving sizing guidelines, there's an imperative to delve deeper into product development processes.

This combined analysis not only sheds light on the critical areas for product and service improvement but also affirms the importance of actively listening to customer feedback. Implementing changes based on customer reviews can significantly elevate the customer experience and brand loyalty.

Fig. 7.23. Response of Mistral-7B-Instruct-v0.2 on Women's Clothing E-Commerce Reviews for Prompt 9

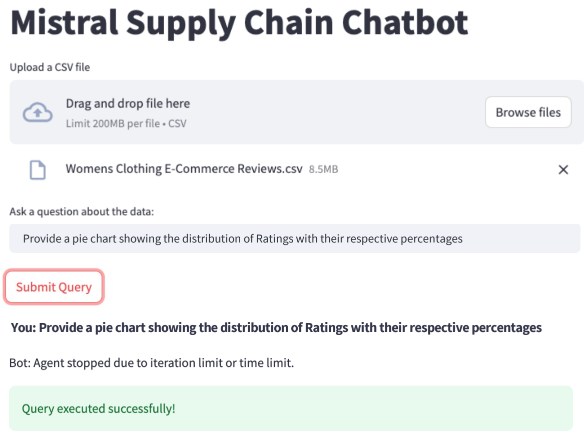
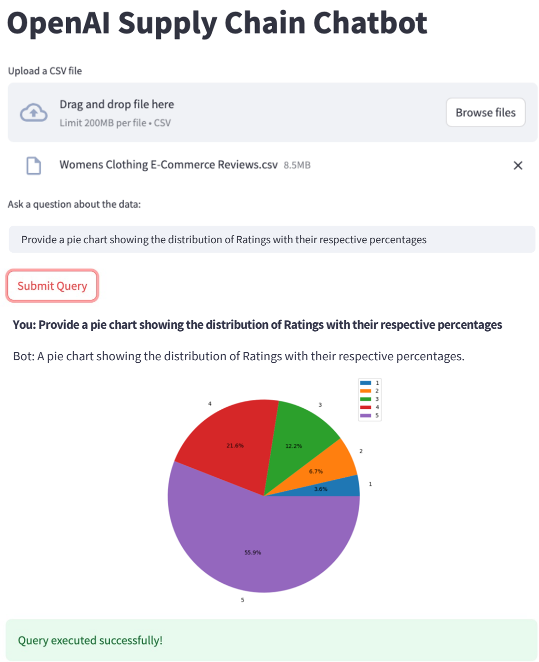


Fig. 7.24. Response of GPT-3.5 Turbo on Women's Clothing E-Commerce Reviews for Prompt 9



The provided pie chart illustrates the distribution of ratings, showcasing how customers have evaluated products or services. From the visual representation generated by GPT-3.5 Turbo, it's clear that most ratings fall into the highest category, indicated by '5', which occupies 55.9% of the pie. This suggests a predominantly positive customer response. The second largest segment is '4', representing 21.6%, which still reflects favorably on the perceived quality or satisfaction levels.

In contrast, the Mistral agent did not generate a response due to an iteration or time limit. This could indicate limitations in the system's capacity to handle certain data-intensive tasks or time-bound processes. The absence of a response from Mistral highlights a critical area of consideration for LLMs used in data analysis: the necessity for efficient data handling and processing capabilities within set time frames.

The success of GPT-3.5 Turbo in providing a pie chart with clear distribution percentages, compared to Mistral's failure to deliver, underscores the variability in performance between different LLMs. This difference emphasizes the importance of choosing the right tool for specific data visualization tasks and the potential need for contingency plans when certain models fail to produce the expected outcomes.

7.3.5. Manufacturing

Tab. 7.6 provides a brief overview of the databases utilized in the case studies presented in this section.

Tab. 7.6. Databases for data analysis within Manufacturing

Prompt	Database	Database Description
10) Identify key insights and trends from the data *11) Analyze manufacturing process data to identify bottlenecks and recommend improvements*	Process Sensor Data	The dataset is about a process' sensor readings, machine's status, and product count

Fig. 7.25. Response of Mistral-7B-Instruct-v0.2 on Process Sensor Data for Prompt 10

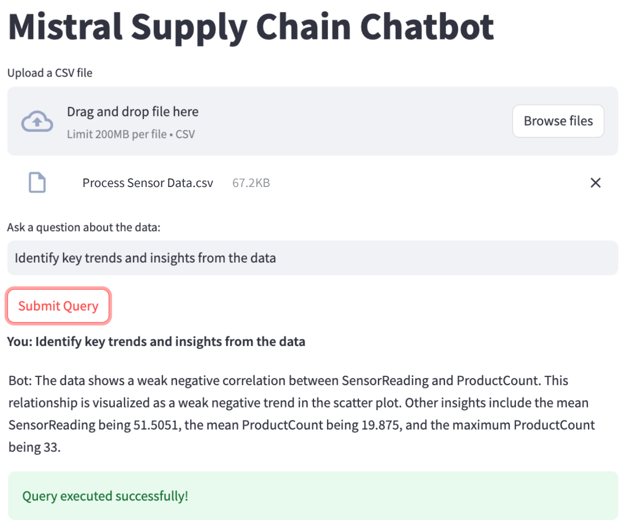
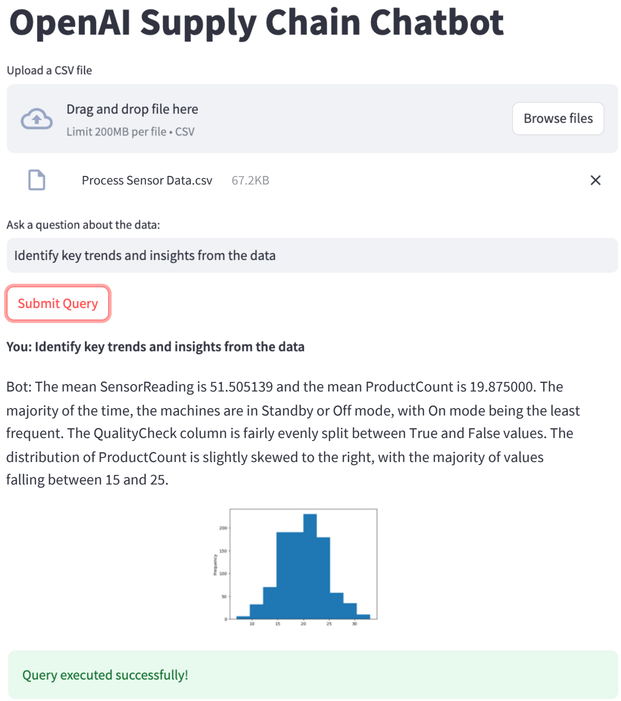


Fig. 7.26. Response of GPT-3.5 Turbo on Process Sensor Data for Prompt 10



The insights gleaned from both LLMs highlight significant aspects of the production process that could influence operational efficiency and product output. The identification of a weak negative correlation between SensorReading and ProductCount, as reported by Mistral, along with the operational modes detailed by GPT-3.5 Turbo, underscore the complexity and interplay of factors impacting the manufacturing environment.

The operational status of machines predominantly being in Standby or Off mode offers a key area for optimization to enhance productivity.

The distribution of the ProductCount, with a slight skew to the right, suggests that while production is generally consistent, there's room for improvement in maximizing output. Furthermore, the balanced distribution of QualityCheck outcomes indicates a stable quality control process but also points to the potential for reducing false positives or negatives to streamline operations.

These analyses serve as a foundation for targeted interventions aimed at improving sensor efficiency, machine usage, and overall production quality. Leveraging these insights could lead to strategic adjustments in operational protocols and maintenance schedules, aimed at optimizing production rates and enhancing product quality. The nuanced understanding of these key trends and insights affords an opportunity to refine and evolve the manufacturing processes for better efficiency and output.

Fig. 7.27. Response of Mistral-7B-Instruct-v0.2 on Process Sensor Data for Prompt 11

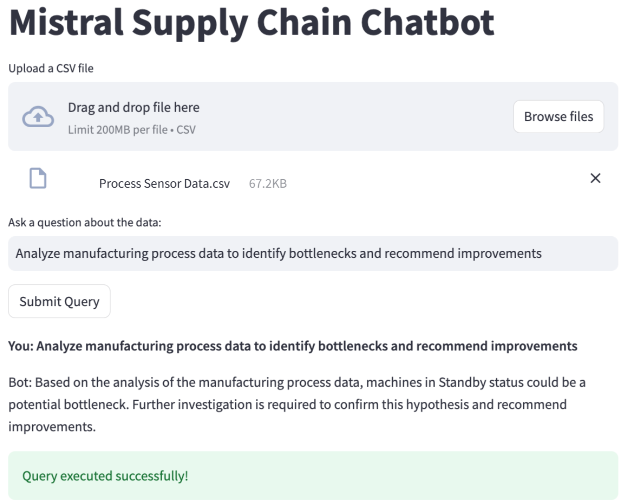
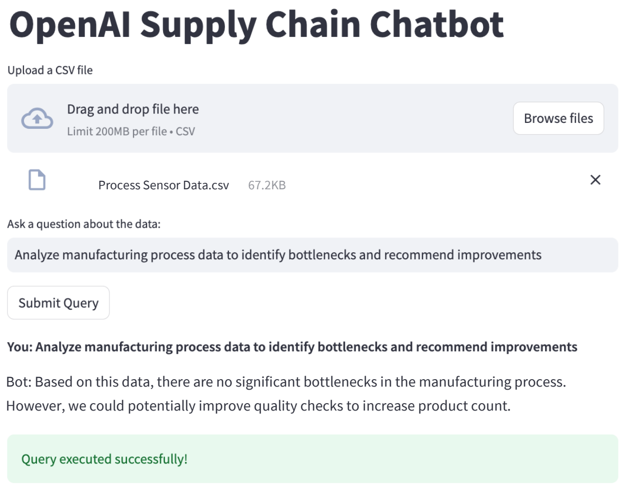


Fig. 7.28. Response of GPT-3.5 Turbo on Process Sensor Data for Prompt 11



The analysis from Mistral suggests that machines operating in Standby status might represent a bottleneck within the manufacturing process, indicating an area that warrants closer examination. This insight highlights the potential for increasing operational efficiency by reducing downtime or optimizing the transition between different operational modes.
On the other hand, GPT-3.5 Turbo's analysis does not identify clear bottlenecks but suggests an opportunity for improvement in the quality check process to potentially enhance product output. This discrepancy between the two analyses underscores the complexity of diagnosing manufacturing process inefficiencies and the importance of a multifaceted approach to data interpretation.
Both perspectives highlight the necessity of continual monitoring and analysis of process data to uncover latent inefficiencies and optimize production workflows. While Mistral directs attention to machine usage and standby times as possible areas for enhancement, GPT-3.5 Turbo points towards quality control processes as a lever for increasing efficiency and output.
The divergent findings emphasize the importance of a diversified analytical approach, ensuring that different aspects of the manufacturing process are scrutinized to identify and alleviate potential bottlenecks.

7.3.6. Discussion of Results

The methodological approach of this Master's Thesis facilitated the integration of LangChain with Pandas to undertake a comprehensive analysis of sales, inventory, and logistics data. This strategic methodology has enabled the natural language processing capabilities of LLMs, particularly through the Pandas Agent, to interpret and visualize complex datasets, thereby unearthing actionable insights pivotal for enhancing supply chain operations.

A focal point of thi research revolves around the comparative analysis between two distinct LLMs: OpenAI's GPT-3.5 Turbo and the open-source Mistral-7B-Instruct-v0.2. This comparative analysis has unveiled significant variations in their analytical competencies and application efficacies within different supply chain scenarios—from Knowledge Management to Forecasting, and Supplier Relationship Management to Customer Relationship Management.

The discrepancies in the performance of Mistral-7B-Instruct-v0.2 and GPT-3.5 Turbo were particularly illuminating. Mistral exhibited proficiency in delivering in-depth, nuanced analyses, making it highly suitable for scenarios demanding comprehensive evaluations, such as intricate supplier assessments or detailed manufacturing bottleneck analyses. Its ability to delve deep into datasets proved invaluable for tasks requiring extensive detail and precision.

On the contrary, GPT-3.5 Turbo demonstrated a propensity for broader, more generalized insights, excelling in swift data processing and trend identification, thus positioning it as an indispensable tool for overarching supply chain analyses and prompt decision-making processes. Its efficiency in generating forecasts and identifying macro-level trends underscores its utility in contexts where rapid, overarching insights are requisite for timely decision-making.

This divergence in LLM performances underscores the quintessential need for strategic selection based on specific analytical needs within the supply chain. The detailed evaluation and the contrasting outcomes from Mistral and GPT-3.5 Turbo elucidate the nuanced application spectrum of LLMs, suggesting that the choice of model should align with the task-specific requirements and objectives. The insights derived suggest a complementary approach, where the in-depth analytical strength of Mistral could be harmoniously combined with the agility and breadth of GPT-3.5 Turbo's insights, fostering a balanced analytical framework conducive to addressing the multifaceted challenges inherent in supply chain management.

In conclusion, the results from this experimental analysis not only spotlight the transformative potential of LLMs in revolutionizing traditional supply chain management practices but also pave the way for a nuanced understanding and application of these advanced computational tools. By leveraging the distinct capabilities of both Mistral-7B-Instruct-v0.2 and GPT-3.5 Turbo, supply chain professionals can harness a more holistic, efficient, and data-driven approach to navigating the complexities of modern supply chains, thereby enhancing operational efficiency, strategic decision-making, and ultimately, organizational performance.

7.4. Comparative Analysis of LLM Performance

In the pursuit of advancing natural language understanding and data analysis capabilities, this research delves into a comparative evaluation of two powerful Large Language Models (LLMs): OpenAI's GPT-3.5 turbo, representing a proprietary and private model, and Mistral-7B-Instruct-v0.2, an open-source language model. The utilization of LLMs has become

integral to modern data analysis, offering unprecedented potential in comprehending complex linguistic patterns and extracting meaningful insights from diverse datasets.
The private model, OpenAI GPT-3.5 turbo, stands as a testament to the cutting-edge advancements in natural language processing, harnessing vast resource. In contrast, Mistral-7B-Instruct-v0.2 exemplifies a commitment to openness and collaboration, making its architecture and parameters accessible to the public. This study aims to confront the performance of these two LLMs within the domain of data analysis when integrated with LangChain, employing a set of metrics to discern their strengths, limitations, and overall suitability for specific analytical tasks.
The comparative analysis of the two LLMs will be structured around a set of predefined criteria, as outlined in Tab. 7.7:

Tab. 7.7. Comparative Analysis Criteria

Criteria	Description
Response Quality	A qualitative assessment of the outputs produced by the LLMs in response to specific prompts.
Cost-effectiveness	The relative costs associated with the deployment and operation of the LLMs
Scalability	The LLMs' ability to handle increasing volumes of data and complexity of queries, defined by their respective token limits
Privacy and Security	Evaluation of the privacy and security risks associated with the use of the LLMs

7.4.1. Response Quality

The following sections will delve into the evaluation of these models across various dimensions. Metrics such as response quality, speed and cost will be studied and tested to provide an understanding of how these LLMs perform in data analysis processes. Through this comparative analysis, we attempt to contribute insights that inform the selection and deployment of LLMs in real-world data analysis scenarios, considering factors of both performance and accessibility.
The quality of responses generated by Large Language Models (LLMs) plays a pivotal role in determining their effectiveness across diverse domains. In this section, we compare the response quality of Mistral-7B and GPT-3.5 Turbo based on two notable research studies: "Performance of Large Language Models on Pharmacy Exam: A Comparative Assessment Using the NAPLEX" by Mirana Angel et al. (2023) and "SecQA: A Concise Question-Answering Dataset for Evaluating Large Language Models in Computer Security" by Zefang Liu (2023).
Mirana Angel et al. evaluated the reasoning abilities of GPT-3.5 Turbo and Mistral-7B, among other LLMs, using the North American Pharmacist Licensure Examination (NAPLEX).
GPT-3.5 Turbo, designed specifically for chat applications with over 154 billion parameters, demonstrated an accuracy of 67.6%. However, it left three questions unanswered, indicating room for improvement.
Mistral-7B-instruct, with 7 billion parameters, exhibited an accuracy of 39.1%. This model, while reasonably effective, left a significant number of questions unanswered.

In the specific context of the pharmacy exam, GPT-3.5 Turbo showcased a commendable accuracy of 67.6%, positioning it as a strong contender in answering questions related to pharmacy concepts. The study results highlight the varying performance of different LLMs, emphasizing the need to consider the specific requirements of the domain when selecting an LLM for a given application.

The research by Zefang Liu introduced SecQA, a dataset designed to evaluate LLMs' performance in computer security. In the 0-shot and 5-shot learning scenarios, Mistral-7B-instruct, which is an instruct fine-tuned version of the Mistral-7B generative text model, achieved an accuracy of 90.9% on SecQA v1 and 87.0% on SecQA v2. On the other hand, GPT-3.5 Turbo showcased high accuracy with 99.1% on both SecQA v1 and v2 datasets. These results underscore the nuanced understanding and application of computer security principles, with GPT-3.5 Turbo exhibiting exceptional performance across both datasets.

The comparative analysis reveals that the response quality of GPT-3.5 Turbo consistently outperforms Mistral-7B across different evaluation contexts. While GPT-3.5 Turbo excels in fundamental security concepts and demonstrates adaptability in the 5-shot setting, Mistral-7B exhibits limitations, particularly in specialized domains like pharmacy. These findings emphasize the importance of considering the specific requirements of the target domain when selecting an LLM for a given application.

These results, drawn from well-established evaluation methodologies, provide valuable insights into the strengths and limitations of Mistral-7B and GPT-3.5 Turbo, contributing to the broader understanding of LLM performance in diverse domains.

In our specific case study, the responses of both models were tested across various supply chain management scenarios, and a comprehensive analysis emerges that highlights the distinct capabilities and approaches of each model.

Mistral’s responses tend to provide deeper insights into potential operational issues and offer specific recommendations for addressing these issues. This suggests that Mistral may excel in identifying and diagnosing operational inefficiencies, making it particularly useful for detailed process analysis and improvement strategies.

On the other hand, GPT-3.5 Turbo appears to offer a broader perspective, identifying general areas for improvement. This could indicate that GPT-3.5 Turbo is more suited for high-level analysis and identifying overarching trends, rather than in-depth process diagnostics.

The difference in response between the two models underlines the importance of selecting the right LLM for specific analytical needs. While Mistral might be preferred for detailed process improvement and problem-solving tasks, GPT-3.5 Turbo may be more effective for strategic planning and identifying general trends.

A notable finding from the research is that Mistral was unable to execute time series analysis despite being applied to two distinct datasets. In contrast, GPT-3.5 Turbo successfully delivered demand forecasts for both databases, showcasing its comparative strength in predictive analytics.

Furthermore, this analysis underscores the significance of clear and precise prompting when working with LLMs, as the quality of responses greatly depends on how the queries are formulated. The variance in model responses also suggests the value of employing multiple LLMs in tandem to gain a more holistic view of the data and cover a broader spectrum of analysis.

In conclusion, the experimental results from the comparative study of Mistral and GPT-3.5 Turbo provide valuable insights into their respective strengths and limitations within the domain of supply chain management. This comparative analysis not only sheds light on the practical applications of LLMs in the supply chain but also guides future strategies for their deployment, ensuring a more informed, efficient, and responsive supply chain ecosystem.

7.4.2. Cost-Effectiveness

Open-source Large Language Models, such as Mistral and similar others, offer the research community and the general public access to robust language processing capabilities without incurring monetary expenses. These open-source models can be leveraged through various approaches, including local downloads, or utilizing Hugging Face's server infrastructure, as employed in this research. This accessibility contributes to the broadening access to advanced natural language processing technologies.

In contrast, proprietary LLMs, such as various versions offered by OpenAI, operate on a cost-per-token model. Specifically, the GPT-3.5 Turbo model incurs a monetary charge of $0.0010 per 1,000 tokens for input and $0.0020 per 1,000 tokens for output (OpenAI Documentation, 2024). This pricing structure may pose challenges to widespread adoption and accessibility due to associated costs. The contrast in accessibility highlights a fundamental difference in the economic models governing open-source and proprietary LLMs, impacting their availability and utilization within the research community and beyond.

7.4.3. Scalability

In this section, we explore and compare the scalability of GPT-3.5 Turbo and Mistral-7B-Instruct-v0.2. Scalability, in this context, is defined by the models' ability to process and analyze increasing volumes of data and the complexity of user queries, which is inherently linked to their respective token limits, also known as the context window.

The context window refers to the maximum number of tokens (words, characters, or pieces of information) that the model can consider at one time when generating a response. This parameter is crucial for understanding the breadth and depth of context the model can maintain, directly impacting its ability to handle complex queries and long conversations.

GPT-3.5 Turbo is designed with a context window of 4,096 tokens. This size dictates its capacity to understand and generate text based on the input provided within this token limit. The token limit reflects the model's ability to sustain lengthy dialogues or process extensive documents in a single instance, affecting its scalability and utility in applications requiring deep contextual understanding or handling extensive data sequences.

Mistral-7B-Instruct-v0.2, on the other hand, is trained with an 8k (8,000 tokens) context length, which signifies a larger context window compared to GPT-3.5 Turbo. This larger context allows the model to incorporate more information from the text into its responses, offering deeper contextual understanding and enabling it to handle more complex or longer queries effectively. Further reinforcing its architectural superiority, Mistral-7B-Instruct-v0.2 employs a Fixed Cache Size mechanism, this feature complements the sliding window attention by maintaining a consistent view of the most relevant information without overwhelming the model's memory. This balance ensures efficiency and relevance in the model's outputs.

In summary, when comparing the scalability of GPT-3.5 Turbo and Mistral-7B-Instruct-v0.2, it's essential to consider how the context window—defined by token limits—impacts each model's ability to process and respond to large and complex data sets. The larger context window of Mistral-7B could theoretically offer advantages in scenarios requiring deep, contextual understanding over an extensive narrative, whereas GPT-3.5 Turbo's smaller window may limit its effectiveness in these situations but could offer faster responses for shorter queries.

7.4.4. Privacy and Security

OpenAI mandates the execution of programs on their proprietary servers, raising potential apprehensions for enterprises seeking to leverage the linguistic model's capabilities for data analysis. This requirement entails exposing sensitive data to an external server, posing privacy and security concerns.
In contrast, open-source models like Mistral present a more flexible approach. Users have the option to execute their programs either on the Hugging Face server or locally download the model onto their machines. This dual functionality not only enhances flexibility but also offers a more reassuring choice regarding privacy and security, as it allows for local processing without the necessity of sharing sensitive information externally.

7.5. Challenges

Implementing Large Language Models (LLMs) encompasses a range of challenges and limitations that can significantly impact their effectiveness and raise ethical concerns. One of the primary issues is the data complexity and scale required for constructing LLMs. These models depend on massive datasets, typically harvested from the internet, to grasp the intricacies of human language. Managing these vast datasets is challenging, particularly in ensuring data quality, representation, and diversity. The extensive size of these datasets can also introduce biases and inaccuracies, undermining the model's integrity and fairness (Raiaan et al., 2024).
Another significant challenge is the substantial computational resources required for training LLMs. The need for powerful GPUs and considerable electrical power makes the training process not only costly but also environmentally taxing. This high energy consumption contributes to the carbon footprint, limiting the technology's accessibility and exacerbating environmental degradation (Raiaan et al., 2024).
Furthermore, fine-tuning LLMs to specific applications adds another layer of complexity. This process requires creating specialized datasets through extensive human labor for annotation, which is both time-consuming and financially demanding. Such complexity hinders the rapid adaptation of LLMs to new domains or languages (Raiaan et al., 2024).
The real-time responsiveness of LLMs also poses a significant challenge. The computational intensity required for processing large context windows can impede the models' ability to provide immediate feedback, which is critical for applications like conversational agents or chatbots. Delays in response can significantly diminish the user experience (Raiaan et al., 2024).
Biases and undesirable outputs are additional concerns since LLMs reflect the biases present in their training datasets. This reflection can perpetuate stereotypes, discrimination, or misinformation, raising ethical issues regarding fairness and the potential harm caused by biased outputs (Raiaan et al., 2024). Moreover, the deployment of LLMs encompasses various ethical considerations, including privacy, security, and the potential for misuse, challenging developers and policymakers to ensure responsible usage that does not infringe on individual rights or propagate harmful content (Hadi et al., 2023).
Fosso Wamba et al. (2023) highlight that the challenges extend beyond data and computational demands. The static knowledge of LLMs, bound by the data available at their last update, leads to potential obsolescence. Additionally, the nuanced nature of language

and context complicates the evaluation of LLMs, as current metrics may not fully capture their capabilities. The dynamic nature of language necessitates continuous updates to evaluation frameworks to ensure the models' relevance and accuracy.

The implementation of LLMs presents multifaceted challenges that require concerted efforts to address. These include managing the complexity and scale of data, mitigating the environmental and financial costs of computational demands, refining the fine-tuning process, enhancing real-time responsiveness, and addressing the ethical implications of biases and privacy concerns. Addressing these challenges is crucial for advancing the development and ethical deployment of LLMs in various applications (Fosso Wamba et al., 2023; Raiaan et al., 2024; Hadi et al., 2023).

The implementation of Large Language Models (LLMs) in various domains, particularly in operations and supply chain management (O&SCM), presents a series of notable challenges. According to Raiaan et al. (2024), one of the primary hurdles is the data complexity and scale required for LLMs, which often leads to issues with data quality and biases. This is critical as LLMs are trained on extensive datasets, predominantly sourced from the internet, making the management of data quality, representation, and diversity exceedingly challenging. Furthermore, the computational demands for training these models are substantial, raising concerns about environmental impact due to their significant energy consumption.

Moreover, the process of fine-tuning LLMs for specific applications is highlighted as a labor-intensive and expensive venture, necessitating specialized datasets and extensive human annotation. The models' sensitivity to tokenization and their struggles with real-time responsiveness further exacerbate their limitations, particularly in applications requiring instantaneous feedback, such as in SCM operations (Raiaan et al., 2024). Additionally, Fosso Wamba et al. (2023) emphasize the apprehensions surrounding security, privacy, and trust associated with LLM deployment in O&SCM. These concerns underscore the challenges of integrating such advanced technologies into existing systems and practices while maintaining ethical standards and protecting sensitive information.

Additional to the challenges identified in the literature concerning the implementation of LLMs in supply chain management, some specific challenges emerged through the development of this research. These experiences bring to light the tangible complexities faced when applying advanced computational models to real-world supply chain data.

- **Model Selection and Configuration**: During this research, initial considerations involved the use of LlaMa2, a Large Language Model developed by Meta. Complications arose due to licensing issues, leading to the exploration of alternative models. The initial deployment of Mistral-7B exhibited inconsistencies, with fluctuating responses to identical prompts. This prompted a transition to Mistral-7B-Instruct-v0.2, which demonstrated enhanced response stability and relevance, aligning more closely with the research objectives.
- **Computational Resources**: The research faced challenges regarding computational resources, particularly the decision against local hosting of the Mistral model. Utilization of Hugging Face's server, while a practical alternative, may have impacted result efficiency and data privacy. A localized deployment could have provided a more controlled environment, potentially improving processing speed and security.
- **Scalability and Adaptability**: The research encountered limitations in scalability and adaptability, particularly when conducting complex analyses on larger datasets or executing intricate tasks. The Mistral model, for instance, showed constraints in

performing predictive analyses integral to the study's aims. Furthermore, attempts to process multiple prompts concurrently often yielded less satisfactory outcomes, underlining the model’s limitations in managing complex, simultaneous tasks and extensive datasets efficiently.

8. Conclusions

This research has explored the extensive capabilities of Large Language Models (LLMs) in enhancing and optimizing supply chain management (SCM) processes. Through meticulous literature review and empirical analysis, the significant impact of LLMs has been demonstrated, particularly focusing on advanced models such as GPT-3.5 Turbo and Mistral-7B-Instruct-v0.2, within the SCM domain.

The integration of LLMs into SCM has been shown to revolutionize the industry by improving decision-making processes, increasing operational efficiency, and providing deep insights into complex data patterns. These models have enabled a paradigm shift from traditional supply chain management to more dynamic, predictive, and efficient operations. The case studies underscore the practical benefits of LLMs in real-world scenarios, showcasing their ability to solve intricate optimization challenges and enhance strategic planning.

The experimental design and methodology highlighted the use of LangChain integrated with the Python Pandas library, illustrating the potential of LLMs to simplify complex data analysis. This integration facilitates an intuitive, conversational approach to data exploration, significantly lowering barriers to complex data analysis and enabling supply chain professionals to extract valuable insights effortlessly.

However, the detailed comparative evaluation conducted between GPT-3.5 Turbo and Mistral-7B-Instruct-v0.2 has uncovered some variations in terms of performance metrics, scalability capabilities, cost-effectiveness, and privacy guidelines. This analysis accentuates the crucial necessity of meticulously selecting the most fitting Large Language Model, which aligns precisely with the specialized demands and strategic imperatives characteristic of the supply chain management landscape. These findings not only emphasize the importance of aligning technological capabilities with business objectives but also shed light on the need for a nuanced understanding of each model's strengths and limitations within specific operational contexts.

As we project into the future, the trajectory for implementing Large Language Models (LLMs) within supply chain management (SCM) encompasses several innovative and strategic directions designed to amplify their utility and effectiveness.

1) **Integration of Built-in Analytical Functions**: The development and integration of specialized built-in functions for data analytics within LLM frameworks represent a pivotal advancement. By embedding functions that can execute complex operations such as data prediction directly, LLMs can serve as a sophisticated interlayer between users and computational processes. This approach enhances the LLMs' role, wherein they facilitate interaction and translation of user queries into actionable commands, while the embedded functions perform the heavy lifting. This dual-layer structure aims to mitigate the inherent margin of error associated with LLM responses, thus ensuring higher quality and more reliable outcomes. By making the LLMs smarter and more precise, we can significantly improve the decision-making process in SCM.
2) **Model Fine-Tuning on Context-Specific Data**: Tailoring and fine-tuning LLMs using context-specific datasets from the supply chain domain can drastically enhance their applicability and effectiveness. By training these models on domain-specific language, terminologies, and scenarios, they become more adept at understanding and predicting industry-specific challenges and requirements. This customization allows LLMs to deliver more nuanced and relevant insights, directly impacting

supply chain strategies and operations. Fine-tuning ensures that the models are not just general-purpose tools but are sharply attuned to the unique dynamics and complexities of the supply chain environment. In real-world applications, LLMs could be further optimized by training them on company-specific datasets, yielding even more precise and impactful results.

3) **Multimodal LLMs:** The development of LLMs that can process and integrate multiple forms of data, including text, images, and audio, is a key research area. Multimodal models have the potential to significantly expand the applications of LLMs and improve their understanding of complex, real-world scenarios (Raiaan et al., 2024). To further bridge the gap between human interaction and machine processes, integrating speech-to-text recognition into the LLM ecosystem marks a significant leap towards accessibility and efficiency. This feature would enable supply chain professionals to conduct complex operations and analyses through voice commands, thereby streamlining workflows and making advanced data analytics more accessible to a broader range of users.
4) **Mixture of Experts**: Future research in supply chain and operations could significantly benefit from exploring the integration of "Mixtral of Experts" models, such as Mistral's recently developed Mixtral 8x7B. This model, a high-quality sparse mixture of experts, has shown promise in enhancing computational efficiency and accuracy in various benchmarks. Its unique architecture and the capacity to handle extensive token contexts could offer new pathways for optimizing supply chain processes and decision-making. Implementing such advanced models may lead to more precise forecasting, better resource allocation, and improved operational strategies.
5) **Enhancing Bias Mitigation**: Ongoing research is focused on developing methods to reduce biases within LLMs. This includes refining training datasets, employing debiasing algorithms, and establishing continuous evaluation frameworks to ensure fairness and mitigate the propagation of harmful stereotypes (Raiaan et al., 2024).
6) **Efficiency Optimization**: Researchers are exploring more efficient training methodologies, such as federated learning and knowledge distillation, to reduce the computational resources required. These approaches aim to make LLM training more accessible and environmentally friendly while maintaining or improving performance (Raiaan et al., 2024).
7) **Dynamic Context Handling**: Future LLMs are expected to handle longer context windows and manage dynamic contexts more effectively. This research direction aims to improve the models' understanding and generation capabilities, making them more versatile and applicable across a broader range of tasks (Raiaan et al., 2024).
8) **Interpretable AI**: Efforts are underway to enhance the interpretability of LLMs. By making models more transparent, users can better understand the reasoning behind AI-generated outputs, thereby increasing trust and enabling more informed decision-making (Raiaan et al., 2024).
9) **Empirical Evidence and Organizational Learning**: Fosso Wamba et al. (2023) underscore the need for empirical research to assess the impacts of LLMs within the O&SCM domain. They stress the importance of organizational learning to support the effective adoption of LLMs, advocating for a deeper comprehension of how these technologies can be utilized to enhance operational efficiencies and overcome barriers to implementation.

10) **Policies and Governance Frameworks**: he establishment of robust policies and governance frameworks is crucial for the ethical use of LLMs. Addressing privacy concerns and managing technological risks are imperative to navigate the challenges posed by these advanced technologies responsibly (Fosso Wamba et al., 2023).

These directions not only highlight the areas where LLMs can be improved and optimized but also underscore the interdisciplinary nature of AI research, encompassing technical, ethical, and societal dimensions.

In expanding on these future directions, the goal is to not only enhance the practical application of LLMs in the supply chain context but also to push the boundaries of what is currently possible. By integrating advanced analytical functions, incorporating speech-to-text capabilities, and refining models with focused, context-specific data, we can elevate LLMs from tools of convenience to indispensable assets in the SCM toolkit. These advancements promise a future where supply chain management is more intuitive, efficient, and aligned with the rapidly evolving demands of the global market.

Bibliography

Aggarwal, C. C. (2018). *Neural Networks and Deep Learning: A Textbook*. Springer, Cham.

Accenture. (2023). *Generative AI: Why smarter supply chains are here*. Accenture. Retrieved from https://www.accenture.com/ch-en/blogs/business-functions-blog/generative-ai-why-smarter-supply-chains-are-here

Accenture. (2023). *Supply chain networks in the age of generative AI: Turning promise into performance.* Retrieved from https://www.accenture.com/content/dam/accenture/final/accenture-com/document-2/Accenture-Supply-Chain-Networks-In-The-Age-Of-Generative-AI.pdf

Ahmad, I., Liu, Y., Javeed, D., Shamshad, N., Sarwr, D., & Ahmad, S. (2020). A review of artificial intelligence techniques for selection & evaluation. *IOP Conference Series: Materials Science and Engineering, 853*, 012055. The 6th International Conference on Electrical Engineering, Control and Robotics. https://doi.org/10.1088/1757-899X/853/1/012055

Akerkar, R. (2019). *Artificial Intelligence for Business* (1st ed.). Springer International Publishing., Cham.

Al Maqbali, K. H., Slimi, Z., & Balasa, A. (2021). The Pros and Cons of Artificial Intelligence use in the Logistics Sector in Oman. *European Journal of Business and Management Research, 6*(4), 197–208. https://doi.org/10.24018/ejbmr.2021.6.4.973

Alasadi, E. A., & Baiz, C. R. (2023). Generative AI in Education and Research: Opportunities, Concerns, and Solutions. *Journal of Chemical Education, 100*(9), 2965–2971. https://doi.org/10.1021/acs.jchemed.3c00323

Anderson, J. L., & Coveyduc, J. L. (2020). *Artificial Intelligence for Business: A Roadmap for Getting Started with AI* (1st ed.). John Wiley & Sons, Inc., Hoboken.

Angel, M., Xing, H., Patel, A., Alachkar, A., & Baldi, P. (2023). Performance of Large Language Models on Pharmacy Exam: A Comparative Assessment Using the NAPLEX. *bioRxiv*. https://doi.org/10.1101/2023.12.06.570434

Anica-Popa, I., Anica-Popa, L., Rădulescu, C., & Vrîncianu, M. (2021). The Integration of Artificial Intelligence in Retail: Benefits, Challenges and a Dedicated Conceptual Framework. *Amfiteatru Economic, 23*(56), 120-136. https://doi.org/10.24818/EA/2021/56/120

Aparicio, D., & Misra, K. (2022). *Artificial Intelligence and Pricing*. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4149670

Asthana, N., & Gupta, M. (2015). Supplier selection using artificial neural network and genetic algorithm. *Int. J. Indian Culture and Business Management, Vol. 11, No. 4, 201.* https://doi.org/10.1504/IJICBM.2015.072428

Ballamudi, V. K. R. (2019). Artificial Intelligence: Implication on Management. *Global Disclosure of Economics and Business, 8(2)*. https://doi.org/10.18034/gdeb.v8i2

Barzizza, E., Biasetton, N., Ceccato, R., & Salmaso, L. (2023). Big Data Analytics and Machine Learning in Supply Chain 4.0: A Literature Review. *Stats, 6*, 596–616. https://doi.org/10.3390/stats6020038

Basu, R., & Wright, J. N. (2008). *Total supply chain management* (1st ed.). Elsevier, Amsterdam.

Bhosale, S., Pujari, V., & Multani, Z. (2020). Advantages And Disadvantages of Artificial Intelligence. *Aayushi International Interdisciplinary Research Journal (Special Issue No.77)*. https://www.researchgate.net/publication/344584269_Advantages_And_Disadvantages_Of_Artificial_Intellegence

Blasek, N., Eichenmüller, K., Ernst, B., Götz, N., Nast, B., & Sandkuhl, K. (2023). Large language models in requirements engineering for digital twins. In CEUR Workshop Proceedings (Vol. 3645) ceur-ws.org. http://ceur-ws.org/Vol-3645/dte1.pdf

Bowersox, D. J., Closs, D. J., Cooper, M. B., & Bowersox, J. C. (2024). *Supply Chain Logistics Management* (6th ed.). McGraw Hill, New York.

Boute, R., & Van Mieghem, J. (2021). AI in Logistics and Supply Chain Management. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3862541

Calvano, E., Calzolari, G., Denicolò, V., & Pastorello, S. (2020). *Artificial Intelligence, Algorithmic Pricing, and Collusion*. American Economic Review. https://doi.org/10.2139/ssrn.4149670

Chen, H., & Zimbra, D. (2010). Artificial Neural Networks in Customer Segmentation. *Journal of Marketing Research*.

Cheng, Y. (2008). *Dynamic Pricing in Perishable Goods Markets: A Q-Learning Approach*. IEEE International Conference on Wireless Communications, Networking and Mobile Computing. https://doi.org/10.1109/WiCom.2008.2786

Chopra, S., & Meindl, P. (2016). *Supply Chain Management: Strategy, Planning, and Operation* (6th ed.). Pearson Education Limited, London.

Dargi, A., Anjomshoae, A., Galankashi, M. R., Memari, A., & Tap, M. B. (2014). Supplier selection: A fuzzy-ANP approach. *Procedia Computer Science, 31*, 691–700. https://doi.org/10.1016/j.procs.2014.05.317

Dash, R., McMurtrey, M., Rebman, C., & Kar, U. K. (2019). Application of Artificial Intelligence in Automation of Supply Chain Management. *Journal of Strategic Innovation and Sustainability, 14*(3). https://doi.org/10.33423/jsis.v14i3.2105

Deloitte Consulting LLP. (2023). *Managing a modern supply chain: Utilizing AI to combat complexity*. Deloitte. Retrieved from https://www2.deloitte.com/us/en/pages/operations/articles/ai-in-modern-supply-chain-management.html

EY. (2024). *How generative AI in supply chain can drive value*. EY. Retrieved from https://www.ey.com/en_gl/insights/supply-chain/how-generative-ai-in-supply-chain-can-drive-value

Feuerriegel, S., Hartmann, J., Janiesch, C., et al. (2023). Generative AI. *Business Information Systems Engineering, 66*, 111–126. https://doi.org/10.1007/s12599-023-00834-7

Finlay, S. (2018). *Artificial Intelligence and Machine Learning for Business: A No-Nonsense Guide to Data Driven Technologies* (3rd ed.). Relativistic,

Fosso Wamba, S., Queiroz, M. M., Chiappetta Jabbour, C. J., & Shi, C. (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

Frederico, G. F. (2023). ChatGPT in supply chains: Initial evidence of applications and potential research agenda. *Logistics, 7*(2), 26. https://doi.org/10.3390/logistics7020026

Garcia, D. J., & You, F. (2015). Supply chain design and optimization: Challenges and opportunities. *Computers and Chemical Engineering, 81*, 153–170. https://doi.org/10.1016/j.compchemeng.2015.03.015

Gutlapalli, S. S. (2017b). Segmentation of Bank Consumers for Artificial Intelligence Marketing. *International Journal of Contemporary Financial Issues 1(1), pp. 39–54.*

Hadi, M. U., Tashi, Q. A., Qureshi, R., Shah, A., Muneer, A., Irfan, M., et al. (2023). A survey on large language models: Applications, challenges, limitations, and practical usage. *TechRxiv*. https://doi.org/10.36227/techrxiv.23589741.v1

Hall, J. V., Kendrick, C., & Nosko, C. (2015). *The Effects of Uber's Surge Pricing: A Case Study*. University of Chicago Booth School of Business.

Hassani, H., Silva, E. S., Unger, S., TajMazinani, M., & Mac Feely, S. (2020). Artificial Intelligence (AI) or Intelligence Augmentation (IA): What Is the Future? *AI, 1(2), 143–155.* https://doi.org/10.3390/ai1020008

Hendriksen, C. (2023). Artificial intelligence for supply chain management: Disruptive innovation or innovative disruption? *Journal of Supply Chain Management, 59*(1), 65–76. https://doi.org/10.1111/jscm.12304

Huang, Y. (2021). *Pricing Frictions and Platform Remedies: The Case of Airbnb*. SSRN Electronic Journal.

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

Lakshmanpriya, C., Lavanpriya, C., & Sangeetha, N. (2013). Vendor selection in the manufacturing industry using AHP and ANN. *The SIJ Transactions on Industrial, Financial & Business Management, 1*(1), 29–34.

Li, B., Mellou, K., Zhang, B., Pathuri, J., & Menache, I. (2023). Large Language Models for Supply Chain Optimization. *Microsoft Research & Microsoft Cloud Supply Chain*. https://doi.org/10.48550/arXiv.2307.03875

Liu, X. (2021). *Dynamic Coupon Targeting Using Batch Deep Reinforcement Learning: An Application to Livestream Shopping*. Working Paper.

Liu, Z. (2023). SecQA: A Concise Question-Answering Dataset for Evaluating Large Language Models in Computer Security. *Georgia Institute of Technology*. https://doi.org/10.48550/arXiv.2312.15838

Liu, Y., Chen, X., & Yan, R. (2023). Unleashing the Power of Large Models: Exploring Human-Machine Conversations. *Proceedings of the 22nd China National Conference on Computational Linguistics. Volume 2: Frotier Forum pages 16-29.*

Llopis, J.A., Fernández-García, A.J., Criado, J., Iribarne, L., Ayala, R., & Wang, J.Z. (2023). A deep learning model for natural language querying in Cyber–Physical Systems. *Internet of Things 24 (2023) 100922.* https://doi.org/10.1016/j.iot.2023.100922

Lu, R., Hong, S. H., & Zhang, X. (2018). A Dynamic Pricing Demand Response Algorithm for Smart Grid: Reinforcement Learning Approach. *Applied Energy*.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS ONE, 13(3), e0194889.* https://doi.org/10.1371/journal.pone.0194889

Mandapuram, M., Gutlapalli, S. S., Reddy, M., & Bodepudi, A. (2020). Application of Artificial Intelligence (AI) Technologies to Accelerate Market Segmentation. *Global Disclosure of Economics and Business, 9*(2), 141. https://doi.org/10.18034/gdeb.v9i2.662

Martinez-Soto, R., Castillo, O., & Aguilar, L. T. (2014). Type-1 and type-2 fuzzy logic controller design using a hybrid PSO-GA optimization method. *Information Sciences, 285*, 35-49.

Matinrad, N., Roghanian, E., & Rasib, Z. (2013). Supply chain network optimization: A review of classification models solution techniques and future research. *Uncertain Supply Chain Management, 1*(1), 5. https://doi.org/10.5267/j.uscm.2013.05.003

McCarthy, J., Minsky, M.L., Rochester, N., & Shannon, C.E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. *AI Magazine, 27*(4).

McKinsey & Company. (2024). *Generative AI in operations: Capturing the value*. McKinsey Operations Practice. Retrieved from https://www.mckinsey.com/capabilities/operations/our-insights/generative-ai-in-operations-capturing-the-value#/

McKinsey & Company (2021). *Succeeding in the AI supply-chain revolution*. McKinsey & Company. Retrieved from https://www.mckinsey.com/~/media/mckinsey/industries/metals%20and%20mining/our%20insights/succeeding%20in%20the%20ai%20supply%20chain%20revolution/succeeding-in-the-ai-supply-chain-revolution.pdf

McKinsey & Company (2023). *The economic potential of generative AI: The next productivity frontier*. McKinsey & Company. Retrieved from https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier#introduction

Mistral. (2023). Mixtral of experts. Retrieved from https://mistral.ai/news/mixtral-of-experts/

Mitchell, V.W. (1998). Buy-phase and buy-class effects on organisational risk perception and reduction in purchasing professional services. *Journal of Business and Industrial Marketing*, 13 (6), 461–478.

Mitrea, N., Lee, T. C., & Wu, J. H. (2009). Forecasting with Neural Networks: An Application Using Bankruptcy Data. *Information Systems Management, 26*(3), 231-242. https://doi.org/10.1080/10580530903018488

Mukhamediev, R.I.; Popova, Y.; Kuchin, Y.; Zaitseva, E.; Kalimoldayev, A.; Symagulov, A.; Levashenko, V.; Abdoldina, F.; Gopejenko, V.; Yakunin, K.; et al. Review of Artificial Intelligence and Machine Learning Technologies: Classification, Restrictions, Opportunities and Challenges. *Mathematics* 2022, *10*, 2552. https://doi.org/10.3390/math10152552

Müller, A. C., & Guido, S. (2017). *Introduction to Machine Learning with Python* (1st ed.). O'Reilly Media, Inc., Sebastopol.

Raiaan, M., Hossain, M. S., Fatema, K., Fahad, N., Sakib, S., Mim, M. M. J., Ahmad, J., Ali, M. E., & Azam, S. (2024). A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues, and Challenges. *IEEE Access*, *PP*(1), 1. https://doi.org/10.1109/ACCESS.2024.3365742

Raiter, O. (2021). New Approaches to Psychographic Consumer Segmentation. *International Journal of Market Research*.

Rana, N. P., Chatterjee, S., Dwivedi, Y. K., & Akter, S. (2022). Understanding dark side of artificial intelligence (AI) integrated business analytics: assessing firm's operational inefficiency and competitiveness. *European Journal of Information Systems, 31*(3), 364-387. https://doi.org/10.1080/0960085X.2021.195562

Reddy, P. P. (2020). Artificial Superintelligence: An AI That Makes Better AI's Recursively. *EasyChair Preprint, (4077).*

Reid, R.D. and Sanders, N.R. (2002). *Operations Management*. John Wiley & Sons, Hoboken.

Riahi, Y., Saikouk, T., Gunasekaran, A., & Badraoui, I. (2021). Artificial intelligence applications in supply chain: A descriptive bibliometric analysis and future research directions. *Expert Systems with Applications, 17.* https://doi.org/10.1016/j.eswa.2021.114702

Rose, D. (2018). *Artificial Intelligence for Business: What You Need to Know About Machine Learning and Neural Networks* (1st ed.). Chicago Lakeshore Press, Chicago.

Russell, S. J., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson Education, London.

Sauvola, J., Tarkoma, S., Klemettinen, M., Riekki, J., & Doermann, D. (2024). Future of software development with generative AI. *Automated Software Engineering, 31*(26). https://doi.org/10.1007/s10515-024-00426-z

Sharma, D. K., & Jain, M. (2022). *Data Analytics and Artificial Intelligence for Inventory and Supply Chain Management* (1st ed.). Springer, New York.

Slack, N., Chambers, S., Johnston, R. and Betts, A. (2006). *Operations and Process Management*. FT Prentice Hall, New Jersey.

Shoushtari, F., Ghafourian, E., & Talebi, M. (2021). Improving performance of supply chain by applying artificial intelligence. *International Journal of Industrial Engineering and Operational Research, 3*(1). Retrieved from http://ijieor.ir

Shumway, R. H., & Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples* (4th ed.). Springer, New York.

Solomonoff, R.J. (1985). The Time Scale of Artificial Intelligence: Reflections on Social Effects. *Human Systems Management, 5(2), 149–153.*

Srivastava, S. K., Routray, S., Bag, S., & Gupta, S. (2024). Exploring the Potential of Large Language Models in Supply Chain Management: A Study Using Big Data. *Journal of Global Information Management, 32*(1). https://doi.org/10.4018/JGIM.335125

Stoychev, V. (2023). The potential benefits of implementing machine learning in supply chain management [Master's thesis, Technische Hochschule Ingolstadt]. *Repository of Technische Hochschule Ingolstadt.* https://opus4.kobv.de/opus4-haw/frontdoor/index/index/docId/3748

Sudarsanam, S. K., Neelanarayanan, V., Umasankar, V., & Saha, I. (2022). Application of AI-based expert evaluation method in an automobile supplier selection problem. *Materials Today: Proceedings*.

Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply Chain Forecasting: Theory, Practice, Their Gap and the Future. *European Journal of Operational Research, 252*(1), 1-26. https://doi.org/10.1016/j.ejor.2015.11.010

Tchelidze, L. (2019). Artificial Intelligence for Energy Demand Forecasting. *IEEE Transactions on Sustainable Energy*.

Teo, W. W. J. (2020). *A natural language processing approach to improve demand forecasting in long supply chains* [Master's thesis, Massachusetts Institute of Technology].

Tiwari, R., Srivastava, S., & Gerac, R. (2020). Investigation of Artificial Intelligence Techniques in Finance and Marketing. *Procedia Computer Science*.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind, 59*(236), 433–460.

Vogelsang, A., Hartig, K., Pudlitz, F., Schlutter, A., & Winkler, J., (2019): Supporting the Development of Cyber-Physical Systems with Natural Language Processing: A Report. *Joint Proceedings of REFSQ-2019 Workshops, Doctoral Symposium, Live Studies Track, and Poster Track.* https://doi.org/10.14279/depositonce-8276.2

Wade, S.D.(2003), Managing your ‘services spend, *Today's Services*.

Wang, H., Liu, M., & Shen, W. (2023b). Industrial-generative pre-trained transformer for intelligent manufacturing systems. *IET Collaborative Intelligent Manufacturing*. https://doi.org/10.1049/cim2.12078

Wang, X., Anwer, N., Dai, Y., & Liu, A. (2023a). ChatGPT for design, manufacturing, and education. *Procedia CIRP, 119*, 7-14. https://doi.org/10.1016/j.procir.2023.04.001

Wild. R. (2002). *Operations Management*. Thomson, London.

Wood, L. C., Reiners, T., & Srivastava, H. S. (2016). Think Exogenous to Excel: Alternative Supply Chain Data to Improve Transparency and Decisions. *International Journal of Logistics, 19*(4), 272-290. https://doi.org/10.1080/13675567.2016.1267126

Wright, J.N. and Race, P. (2004). *The Management of Service Operations*. Thomson Learning, Boston.

Zavala-Alcívar, A., Verdecho Sáez, M.-J., & Alfaro Saiz, J.-J. (2020). Assessing and selecting sustainable and resilient suppliers in agri-food supply chains using artificial intelligence: A short review. In *IFIP Advances in Information and Communication Technology (Vol. 598*, pp. 501–510). https://doi.org/10.1007/978-3-030-62412-5_41