APPLYING GENERATIVE AI AND LARGE LANGUAGE MODELS IN BUSINESS APPLICATIONS

Lappeenranta–Lahti University of Technology LUT

Bachelor’s Programme in Industrial Engineering and Management, Bachelor's thesis

2023

Marko Korhonen

Examiner: Leena Tynninen *(University Lecturer)*

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT Teknis-luonnontieteellinen

Tuotantotalous

Marko Korhonen

Generatiivisen tekoälyn ja laajojen kielimallinen hyödyntäminen liiketoimintasovelluksissa

Tuotantotalouden kandidaatintyö

Vuosiluku 2023

41 sivua, 10 kuvaa

Tarkastaja(t): Leena Tynninen, Yliopisto-opettaja

Avainsanat: Generatiivinen AI, laajojen kielimallien hyödyntäminen liiketoimintasovelluksissa

Tutkielmassa keskitytään tutkimaan, miten generatiivinen tekoäly ja suuret kielimallit (LLM) voidaan integroida liiketoimintasovelluksiin. Työssä tarkastellaan mahdollisia hyötyjä, haasteita ja mahdollisuuksia, joita näiden teknologioiden käyttö liiketoimintakontekstissa tarjoaa.

Tutkimusmenetelmänä käytetään narratiivista kirjallisuuskatsausta, joka tarjoaa laajan yleiskatsauksen tekoälystä liiketoiminnassa ja kattaa sen kehityksen, sovellukset ja eettiset näkökohdat. Tämä lähestymistapa on valittu sen vuoksi, että se tarjoaa kokonaisvaltaisen käsityksen aiheesta.

Keskeisiä havaintoja ovat muun muassa tekoälyn merkittävä potentiaali liiketoimintaprosessien automatisoinnissa ja tehostamisessa, tiedon laadun ja mallien koulutuksen merkitys, eettiset näkökohdat sekä tarve tekoälysovellusten avoimuuteen ja selitettävyyteen. LLM:ien avulla voidaan muun muassa tuottaa tekstiä, tiivistää tietoa ja osallistua keskusteluun, mikä tarjoaa merkittäviä mahdollisuuksia liiketoiminnan kehittämiseen. Tutkielmassa käsitellään myös tekoälyn integrointia SAP:n liiketoimintasovelluksiin ja korostetaan strategista lähestymistapaa generatiivisen tekoälyn käyttöönotossa.

Jatkotutkimusta varten tutkielmassa ehdotetaan syventymistä tiettyihin sovellustapauksiin, perustamismallien ja LLM:ien teknisiin näkökohtiin sekä käytännönläheisempien liiketoimintaongelmien ja -ratkaisujen tutkimista. Tähän sisältyy tekoälylähtöisten innovaatioiden potentiaali eri liiketoiminta-aloilla ja tekoälyteknologioiden jatkuva kehitys.

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT School of Engineering Science

Industrial Engineering and Management

Marko Korhonen

Applying Generative AI and Large Language Models in Business Applications

Bachelor's thesis

2023

41 pages, 10 figures

Examiner(s): Leena Tynninen *(University Lecturer)*

Keywords: Generative AI, Large Language Models in Business Applications

The thesis focuses on exploring how Generative AI and Large Language Models (LLM) can be integrated into business applications. It examines the potential benefits, challenges, and opportunities of using these technologies in a business context.

The research method employed is a narrative review, which provides a broad overview of AI in business, covering its evolution, applications, and ethical considerations. This approach is chosen for its ability to offer a holistic understanding of the subject matter.

Key findings include the significant potential for LLMs in automating and enhancing business processes, the importance of data quality and model training, ethical considerations, and the need for transparency and explainability in AI applications. LLMs can be used to produce text, summarise information and participate in discussions, among other things, offering therefor significant opportunities to developing business processes. The thesis also discusses the integration of AI in SAP's business applications, emphasizing the strategic approach to employing Generative AI.

For future research, the thesis suggests delving deeper into specific application cases, the technical aspects of foundation models and LLMs, and the exploration of more practical business problems and solutions. This includes the potential for AI-driven innovations in various business sectors and the continuous evolution of AI technologies.

Table of contents

1 Introduction

1.1 Background

In the modern era, artificial intelligence (AI) has rapidly integrated into numerous sectors, revolutionizing business processes, and redefining traditional business models. As the commercial and social implications of AI expand, it is imperative for business stakeholders and decision makers to understand its evolution, current applications, and future prospects. This literature research provides an overview of the significance of AI in business, tracing its evolution and emphasizing the need for comprehensive understanding through a narrative review approach. Businesses today are not just leveraging AI for operational efficiencies but are embedding it into the core of their strategies. Whether it's streamlining processes through Cloud ERP, enhancing human capital management, or optimizing customer relationships, AI-powered business processes are becoming a norm rather than an exception. (Tikare, 2023)

1.2 Research objective, research questions, and research method

The primary objective of this research is to explore the applications, benefits, and potential challenges of integrating Generative AI and Large Language Models in contemporary business applications.

Research tries to answer following **three major research questions**:

1. How does the integration of AI in business applications influence the efficiency and effectiveness of organizational processes?

2. What are the challenges on business-specific training for AI?

3. What role do ethical standards and data privacy norms in AI development play in fostering stakeholder trust and ensuring sustainable business practices?

A narrative review method will be employed in this study. Unlike systematic reviews that focus on addressing specific research questions, a narrative review provides a broader

overview of a topic area. (Salminen, 2011) This method is particularly suitable for this study as it aims to offer a holistic understanding of AI in business, encompassing its evolution, applications, and ethical considerations.

1.3 Research Scope

The research is delimited such that practical application cases are not covered, and the study does not delve deeply into the technical concepts and operational logic of foundation models and large language models. The research is targeted at business decision-makers or non-technical stakeholders who are not interested in in-depth technical details. With this delineation, aim is to keep the content of the work relevant and understandable for the target audience. A deep dive into the technical concepts of foundation models might require more time and resources than are available. The aim to keep the objective of the work more practical than theoretical. The final project may seek to give business decision-makers or stakeholders an understanding of how generative artificial intelligence and language models can be applied to practical business problems, without a deep technical immersion. Research emphasizes more on the business aspects in the final work, such as application opportunities, benefits, challenges, and potential strategies.

1.4 Structure of the Thesis

This bachelor's thesis is structured methodically to offer readers a comprehensive understanding of the application of generative AI and large language models in business, with a specific emphasis on SAP Business AI.

Chapter two (2) discusses the diverse applications of Artificial Intelligence within the business domain. There is an overview of how AI has been integrated into various business operations and emphasis is given to the business-specific uses of AI, highlighting the key areas where AI has made significant difference. Additionally, the chapter explores industry-centric AI applications, demonstrating the tailored uses of AI in various industry sectors. Concluding the section, the discourse shifts to the pivotal role AI plays in shaping modern business, emphasizing its transformative impact on business ecosystems.

In this chapter three (3), readers will be introduced to the fundamental concepts of generative AI. From basic definitions to various techniques and algorithms, this chapter aims to demystify generative AI. A case study is included to showcase popular implementations of generative AI.

This chapter four (4) exclusively focuses on large language models and their significance in business. After an introduction to LLMs, the chapter investigates their capabilities, features, and applications in various business scenarios, highlighting both opportunities and challenges.

Given the importance of SAP Business AI in this research, chapter five (5) takes a closer look at its offerings. It starts with an overview, followed by specific use cases and applications in business. The chapter concludes by discussing the benefits and limitations of SAP Business AI in a business context.

2 Introduction to Artificial Intelligence

Artificial Intelligence, today a household name, began as a vision, an aspiration to create machines that could mimic the cognitive functions of the human brain. In the past, AI was often perceived as a distant state, seemingly more pivotal for large-scale projects or enterprises but detached from the individual's daily life. This perception was widespread, and many stated that AI, though significant in certain arenas, held little relevance in peoples personal and professional daily lives. (Allardice, 2023)

2.1 Early Pioneers and Their Contributions

Perhaps the most known pioneer on artificial intelligence applications is the Turing Test. It was a vision of one man, the brilliant English mathematician and computer pioneer Alan M. Turing. During World War II, Turing directed a secret group that developed computing equipment powerful enough to break the code the Germans used for military communications. Introduction to Turing's ground-breaking paper "Computing Machinery and Intelligence" in which he speculated that by the year 2000, it would be possible to program a computer so that an “average interrogator will not have more than 70 percent chance” of distinguishing the computer from a person “after five minutes of questioning”. (Epstein, Roberts and Beber, 2008)

The Dartmouth Summer Research Project on Artificial Intelligence, held at Datrmouth College in 1956, is regarded as the official ‘‘birthplace’’ of AI. John McCarthy a pivotal role, as he was one of the main organizers and also the person who coined the term "Artificial Intelligence" to describe this new field of study. The Dartmouth conference aimed to bring together researchers interested in exploring the possibilities of machines simulating intelligence. McCarthy's vision was to investigate the idea that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. The proposal set out a bold, optimistic agenda that machine intelligence could be achieved in the course of a two-month project with about a dozen participants. (Kline, 2011)

The perceptron, developed by Frank Rosenblatt in 1957 can be considered as the birth of of neural networks. As a foundational model for artificial neural networks, the perceptron was initially considered as a machine not just for pattern recognition, but with the potential to learn from its experiences and previous iterations. It was structured to mimic the basic functionality of a biological neuron: receiving inputs, processing them, and producing a single binary output. Although the early perceptions had limitations, which most relevant was that it could only solve linearly separable problems, the concept fostered a new way of thinking about machine learning, and leading to the creation of more sophisticated algorithms and the evolution of neural networks. (Sejnowski, 2018)

Herbert Simon and Allen Newell's development of the Logic Theorist marks a milestone in the field of artificial intelligence in 1955. Logic Theorist program can be considered as first artificial intelligence program. Program was designed with the capacity to mimic human problem-solving skills, especially on the area of symbolic logic. The program was developed in the mid-1950s and presented for the public in 1956. It was engineered to simulate the process of human thought, particularly the ability to reason and prove mathematical theorems. The Logic Theorist worked by applying a series of heuristic techniques to solve problems, much as a human would employ intuitive judgments to guide their reasoning. Herbert Simon and Allen Newell's approach to creating the Logic Theorist was groundbreaking. Their work was founded on the hypothesis that human problem-solving was a symbolic process, and thus could be emulated by a machine through the manipulation of symbols. The program utilized a set of rules and a decision-making process to select and apply these rules in order to construct proofs of logical statements. (Harris, Paul and Pal, 2022)

2.2 Elements of Artificial Intelligence

Artificial intelligence is the science and technology of simulating human intelligence. Artificial intelligence was developed for machines and robots that, by analyzing data sets, acquire problem-solving skills, like those of humans and living objects. Artificial intelligence is a concept that is based on simulating human intelligence. A machine that copies human intelligence and performs a task like a human is called a machine with elements of artificial intelligence. (Hrischev and Shakev, 2022) This is accomplished by the development of

complex software systems that perform certain tasks like the human brain. These systems collect and analyze data using predefined mathematical algorithms, generate useful information from that data, and then use that information to make a final decision and fix the problem with a flexible approach and adaptive solutions.

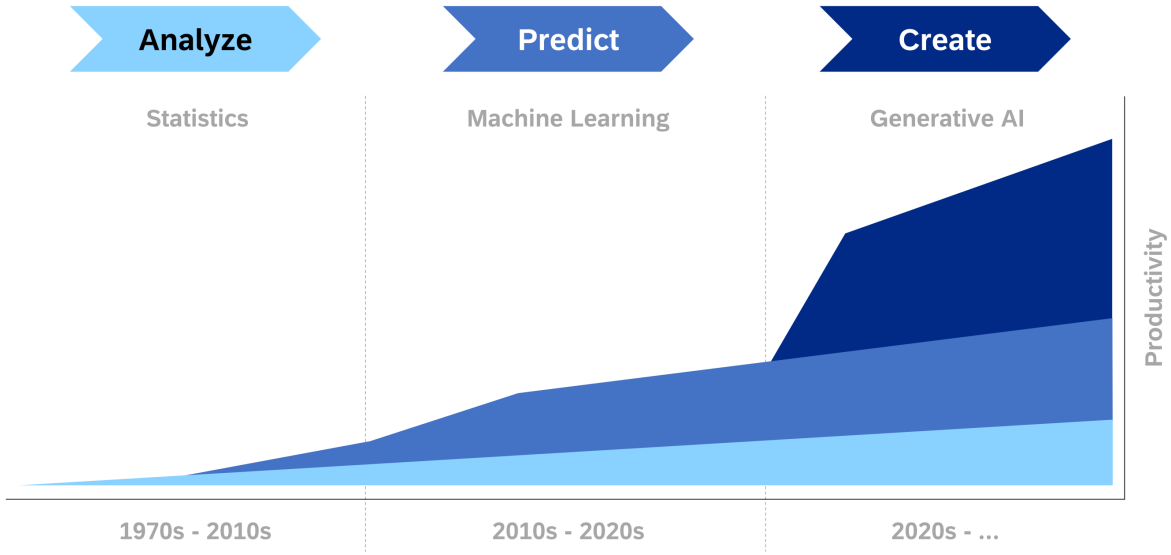

Figure 1. Evolution of Artificial Intelligence (Allardice, 2023)

The foundation of artificial intelligence is the strong development of mathematical and algorithmic methods and the revolution in information technology and computing.

Artificial intelligence is developing in five main interrelated directions:

• Machine learning (ML)

• Neural Networks (NN)

• Deep Learning (DL)

• Natural Language Processing (NLP)

• Robotics

Machine Learning

In 1959, the term "Machine Learning" was coined by the scientist Arthur Samuel. Essentially, it refers to acquiring knowledge from past experiences. The artificial system learns from provided examples and can extrapolate from them once the learning phase is

concluded. In machine learning, algorithms construct a model using training data. This implies that the system doesn't just retain examples but identifies patterns within the training data. Consequently, the machine learning algorithm can predict unfamiliar behavior patterns. (Hrischev and Shakev, 2022) Machine learning involves knowledge within databases, retrieving data, and subsequently fetching pertinent information as needed. It's evident that machine learning is intertwined with database operations and the use of algorithms tailored for sophisticated data analysis.

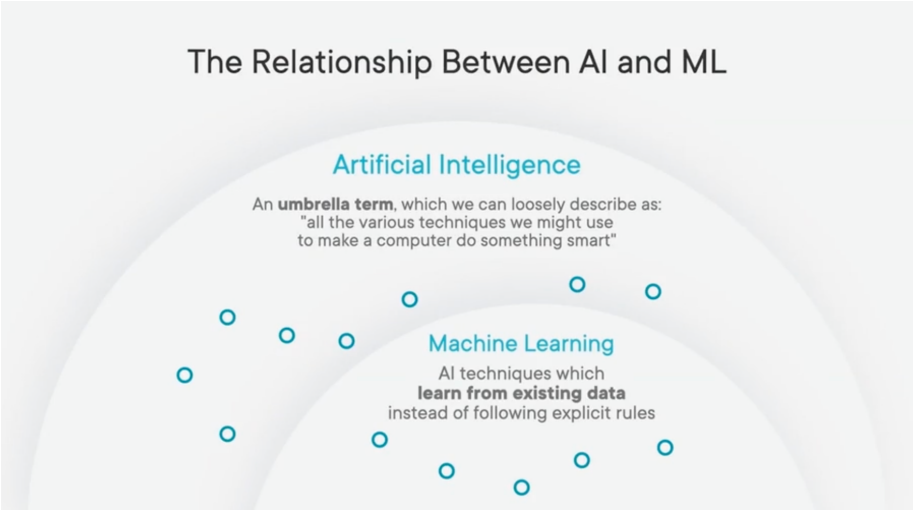

Figure 2. The relationship between AI and ML (Allardice, 2023)

Machine learning, a sub-branch of AI, is dedicated to crafting algorithms that enhance their task performance using feedback and experience. This evolving field is generally split into three learning methods: Supervised, Unsupervised, and Reinforcement learning. Supervised learning is about prediction. Imagine teaching a machine using a labelled dataset, like photos marked as 'cats' or 'dogs'. The machine uses this data to make educated guesses on new, unlabeled photos. The effectiveness of its guesses, or its "classifier", is determined by its accuracy rate in pinpointing correct labels. A classifier that gets every prediction right is considered perfect. The process entails training the machine on one set of data and testing its learned knowledge on another. While the preparation phase can be lengthy, once the classifier is set up, it can swiftly label new data.

Unsupervised learning, on the other hand, delves deeper into understanding data's inherent structures. Instead of predicting labels, it explores patterns and relationships within the data. It's like giving the machine a jigsaw puzzle without the picture on the box, letting it figure out how pieces fit together. Lastly, there's Reinforcement learning, which operates on trial

and error. Think of a robot navigating a room: it learns by moving around and getting feedback, like sensing proximity to a desired spot. The robot is "rewarded" for actions bringing it closer to the target and "penalized" for those leading it astray. Over time, it learns the best actions to reach its goal. (Bartneck et al., 2021)

Neural networks

Neural networks in AI are like simplified versions of human brain's wiring. Imagine human brain as a big web of tiny units called neurons, connected by links named synapses. In AI, this web is made of math-based units that act like our neurons. These AI units get information as numbers, add them up, and use a formula to decide how excited they should be. This decision then gets passed to the next units in the web. So, in simple terms, a neural network in AI is like a series of layers where each layer has these units passing messages to the next. (Hrischev and Shakev, 2022)

Deep Learning

Deep Learning is like teaching a computer to think using a system that imitates our brain's structure. Think of it as a multi-layered web. The term "deep" means it has multiple layers. In this web, the first layer takes in the data, the middle layers process it in ways we can't directly see, and the final layer gives us the computer's interpretation or decision. This setup helps the computer recognize patterns in the data and make smart choices. (Hrischev and Shakev, 2022)

Neural Language Processing

Natural Language Processing (NLP) sits at the intersection of artificial intelligence and language studies. Target of NLP is to bridge the communication gap between humans and computers. On one hand, NLP allows systems and various applications to transform complex data into understandable human language format. This is useful in tasks like summarizing large volumes of text. NLP lets computers grasp and process our spoken or written words, leading to innovations like voice commands and auto-generated content. Essentially, while some NLP tools help computers understand and manage our documents, others enable them to recognize and even produce human-like speech. (Hrischev, Shakev 2022)

Robotics

Robotics stands at the forefront of innovation, mirroring the capabilities of both man and machine. This field investigates creating automated machines that can execute tasks typically done by humans, which related closely to the world of artificial intelligence. While the word 'robot' often associate images of tangible, mechanical beings, the term is expanding to encompass specialized digital applications. Whether streamlining manufacturing or enhancing data management, robots are reshaping the way we approach tasks and processes. (Hrischev and Shakev, 2022)

2.3 Modern applications of Artificial Intelligence

The year 2022 can be marked as pivotal moment in the landscape of artificial intelligence, serving as turning point for significant advancements and the widespread adoption of AI technologies that reshaped industries and public interpretations of machine intelligence.

Firstly, 2022 witnessed an unforeseen level of innovation in AI models, particularly on the are of generative and foundational models. These models demonstrated remarkable abilities to generate human-like text, creating images and music, and solving complex problems across various domains. The sophistication of these algorithms also led into a democratization of AI tools, and making powerful AI capabilities accessible to a broader audience and enabling multiple creative and practical applications.

Secondly, the advancement and integration of AI in various industry domains and sectors became more evident. Healthcare, finance, education, and manufacturing industries began leveraging AI for improved efficiency, accuracy, and innovation. In healthcare, AI-driven diagnostic tools and personalized treatment plans became more prevalent, while in finance, AI algorithms enhanced predictive analytics for investments and fraud detection. The emergence technologies like Chat GPT created a fundamental change on the AI landscape. OpenAI’s ChatGPT showed its potential to reshape economies, redefine work paradigms. Such was the significant pull of AI that it became a household conversation, leaving many considering whether to embrace this technological miracle or tread with caution. (Allardice, 2023)

Modern AI's evolution did not happen in closed environment and isolation but was using more the power of ecosystems and software development communities. Tools like Chat GPT, Google Bard, and image generation software like Dali and Mid-Journey represent the AI-driven ecosystem's results. Furthermore, platforms tailored for developers, such as GitHub Copilot, and AI-integrated desktop applications from giants like Microsoft Office to Adobe Photoshop, have witnessed significant AI potential. Cloud computing platforms have played a significant role enabling and bringing complex AI models accessible for everyone. Cloud computing environments are able to hold massive amount of data points and host highly complex algorithms, which is mandatory for high quality results these new AI models can bring to its users. On last decade cloud computing technical progress has also enabled of more advanced natural language processing capabilities, epitomized by models that exhibited deep understanding and context awareness, putting AI into new frontiers of human-computer interaction. This led to AI becoming an integral part of daily life, as virtual assistants, chatbots, co-pilots and customer service AI became more nuanced and capable.

Finally, 2022 saw also a significant increase in investment and collaboration in the AI sector, even between major technology vendor and innovators. Both public and private funding for AI research and startups shoot up, driving innovation and encouraging partnerships between academic institutions, tech giants, and emerging AI ventures. This enormous amount of resources and collaborative efforts accelerated the pace of AI research and development, catalyzing breakthroughs and the implementation of AI solutions worldwide. The culmination of these developments in 2022 served as a harbinger for a new chapter in AI, one where artificial intelligence began to redefine the boundaries of technology's role in society. It set a new benchmark for what AI could achieve and laid the groundwork for future advancements that could further expand the capabilities and influence of artificial intelligence in the years to follow.

3 Large Language Models and Generative AI

3.1 Introduction to Large Language Models

In today's digital age, the rise and importance of large language models (LLMs) have become increasingly evident. These models, primarily generative AI, have the unique capability to produce new content, be it text, images, or videos, in direct response to user input. Specifically tailored for tasks related to natural language processing, the essence of LLMs lies in their expansive architecture, housing billions of parameters that can be adjusted during their learning phase.

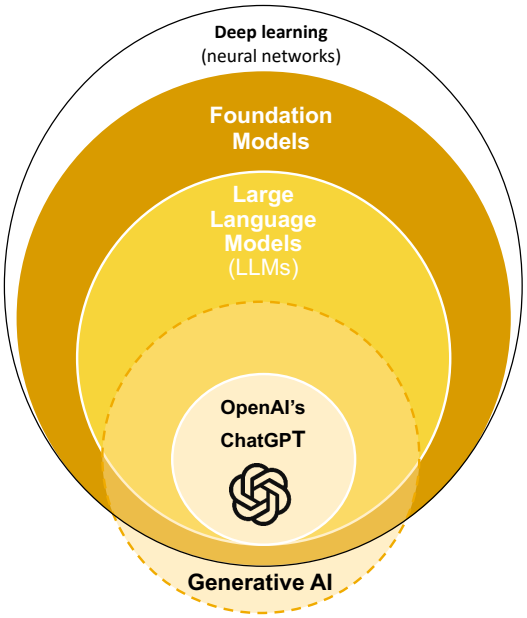

Figure 3. Large Language Models and Generative AI defined

Large language models are AI systems that are designed to process and analyze vast amounts of natural language data and then use that information to generate responses to user prompts. These systems are trained on massive data sets using advanced machine learning algorithms to learn the patterns and structures of human language. Foundation models are deep learning models, i.e., neural networks, trained on massive volumes of unlabeled data using a self-supervised learning objective. The models can be applied to a wide range of tasks and often

demonstrate surprising emergent capabilities. Generative AI (Gen AI) can create novel output in text, images, sound, or video based on simple user input called a “prompt”. While not all foundation models are generative, and some generative models are not foundation models, Gen AI is often used as a blanket term for foundation models and LLMs.

ChatGPT is a specific chat assistant based on GPT-3.5 and GPT-4 large language models created by the vendor OpenAI. The underlying structure of these models is a deep neural network known as the transformer. Its distinct design facilitates the understanding of sequential data such as text or speech. Notably, many LLMs start as foundation models, initially trained on massive datasets and later fine-tuned to execute specific tasks.

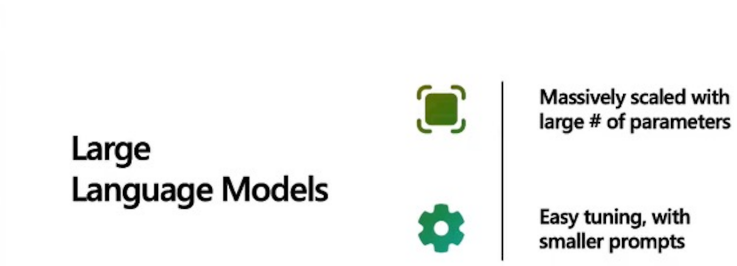

Figure 4. Large Language Model (Sun, 2023)

Large Language Models are constructed on a grand scale, utilizing a considerable number of parameters. Parameters are described as variables within the model that the system adjusts as it learns from data. Models are highly complex, designed to handle and interpret extensive datasets, enabling them to generate or understand language with high precision. Despite the model complexity, these models can be fine-tuned or calibrated in a straightforward manner by providing them with "smaller prompts". This means that even with minimal input or guidance, the models can adapt or be directed to perform specific language-related tasks, making them both powerful and accessible for various applications. (Sun, 2023)

3.2 Capabilities and Features of LLMs

LLM’s foundational capabilities of these models are multifaceted and include several key functions:

Natural Language Understanding (NLU): LLMs are adept at deciphering the meaning behind the text. This includes parsing sentences, recognizing the relationship between words, understanding idioms, and inferring implied meanings. Their understanding extends to complex constructs such as sarcasm, metaphor, and humor, to a certain extent.

Natural Language Generation (NLG): Beyond comprehension, LLMs excel at producing text. They can construct sentences, paragraphs, and entire documents that are syntactically correct and semantically coherent. The generative capability also includes the ability to maintain context over longer passages and to adjust the style and tone to suit various purposes or audiences.

Question Answering: LLMs can provide answers to questions by extracting information from their training data or by inferring answers using their understanding of language and the world. This makes them useful for applications such as virtual assistants and information retrieval systems.

Translation: These models are capable of translating text from one language to another, benefiting from their extensive exposure to multiple languages during training. They can capture idiomatic expressions and colloquialisms, often producing translations that are not only accurate but also natural sounding.

Summarization: LLMs can condense longer texts into shorter summaries while retaining the core message and essential details. This summarization capability is useful for digesting large volumes of information quickly.

Text Completion and Autoregeneration: Given a starting point or a partial text, LLMs can complete the text in a way that is contextually appropriate. This is helpful for writing assistance tools, email autocomplete functions, and creative writing prompts.

Conversation and Dialogue: LLMs can engage in conversations by generating responses that are contextually appropriate and coherent over multiple exchanges. This enables them to function as chatbots or virtual agents capable of assisting with customer service, tutoring, and other interactive tasks.

Semantic Search: These models can assist in searching not just for keywords but for the meaning behind those keywords. This allows for more intuitive search capabilities that understand the user's intent and the context of the inquiry.

Pattern Recognition: LLMs are trained to recognize patterns in text, which enables them to identify trends, anomalies, or specific features within large datasets. This can be applied in data analysis, market research, and other areas where textual pattern recognition is valuable.

3.3 Overview of Generative AI

Generative AI is capable of synthesizing various content and forms, ranging from textual compositions to visual artwork. At its core, this capacity is orchestrated through a sequence of interconnected components and processes, each integral to the generative framework. Generative AI is consisting out of following key elements. (Kask, 2023)

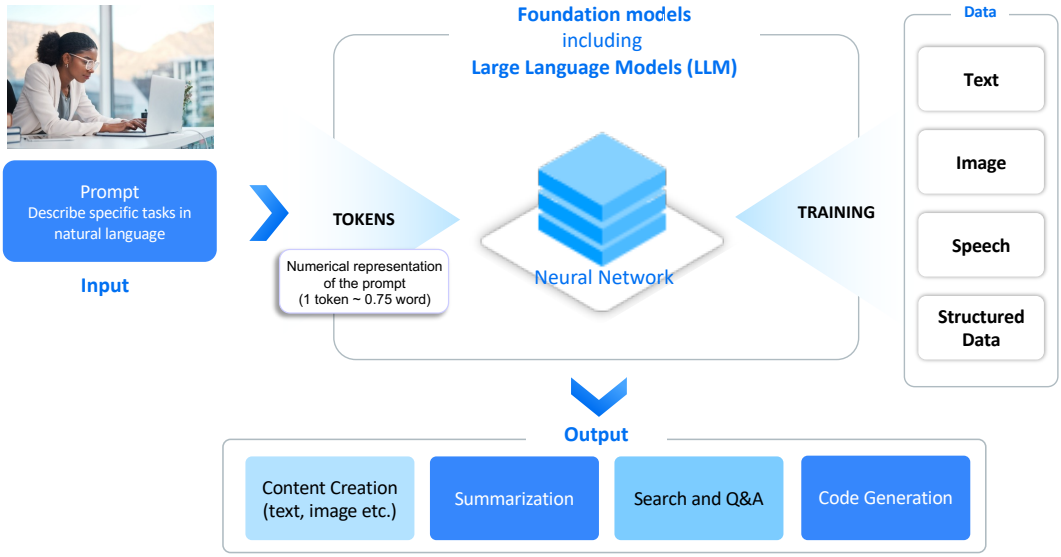

Figure 5. Simplified view of Generative AI (Tikare, 2023)

Prompt: Communication with these models is facilitated through prompts. A prompt serves as the user's query, which is then processed by the model to produce a completion or response. The effectiveness of this output is inseparable linked to the quality of the input prompt. Precise instructions can significantly enhance the quality of the generated content. Prompt engineering is focusing on refining prompts to optimize the output of generative AI models. The objective is clear: improve the accuracy, relevancy, and quality of the generated content. Achieving this, often involving a combination of instructions, context, examples, and output indicators. Each component plays a distinct role, guiding the model in

understanding the task, its context, the desired format, and the expected output. An engineer must possess a comprehensive understanding of language intricacies and the ability to convey instructions lucidly. In essence, they bridge the gap between user intent and AI capabilities, ensuring that the AI delivers results that align with the user's expectations.

Tokens: The mechanism of understanding and content generation within the AI operates through tokens. These tokens are essentially the fundamental units of data, like a kind of DNA of digital language, that the AI interprets, manipulates, and recombines to produce various outputs. In textual applications, tokens typically correspond to words or subword elements, which the AI assembles into meaningful sequences.

Foundation Models and Large Language Models: The intelligence of generative AI are realized by foundation models, which are extensive neural networks trained on vast and diverse datasets. Large Language Models represent a specialized category of foundation models, honed to comprehend and generate human language. These models form the bedrock of the AI's ability to engage in sophisticated linguistic tasks, from creative writing to generating programming code.

Training: The process of training is a critical phase where the AI system is exposed to a various of examples, enabling it to discern patterns, structures, and the intricacies of language or imagery. During this phase, the AI refines its internal parameters, enhancing its accuracy and reliability in generating outputs that closely align with the provided examples and learned contexts.

Various Outputs: Upon the culmination of training, the generative AI, equipped with the learned patterns and structures, is capable of producing a plethora of outputs. The diversity of these outputs is contingent upon the nature of the prompts supplied. For instance, a text-based generative AI may yield poetry, essays, or technical articles, while a visually oriented AI might generate paintings, design prototypes, or even simulate environmental aesthetics for virtual reality.

3.4 Large Language Models Opportunities and Challenges

Large Language Models represent a significant advancement in the field of artificial intelligence, offering a range of opportunities and presenting various challenges. These models

have the potential to transform numerous daily life challenges. One of the most notable opportunities provided by LLMs is their ability to understand and generate human-like text. This ability enables the creation of advanced chatbots and virtual assistants i.e. co-pilots that can offer more natural and engaging interactions. LLMs can assist in writing and summarizing large volumes of text, making information more accessible. Moreover, their predictive capabilities are invaluable where they can analyze documents for pertinent information, trends, and compliance risks.

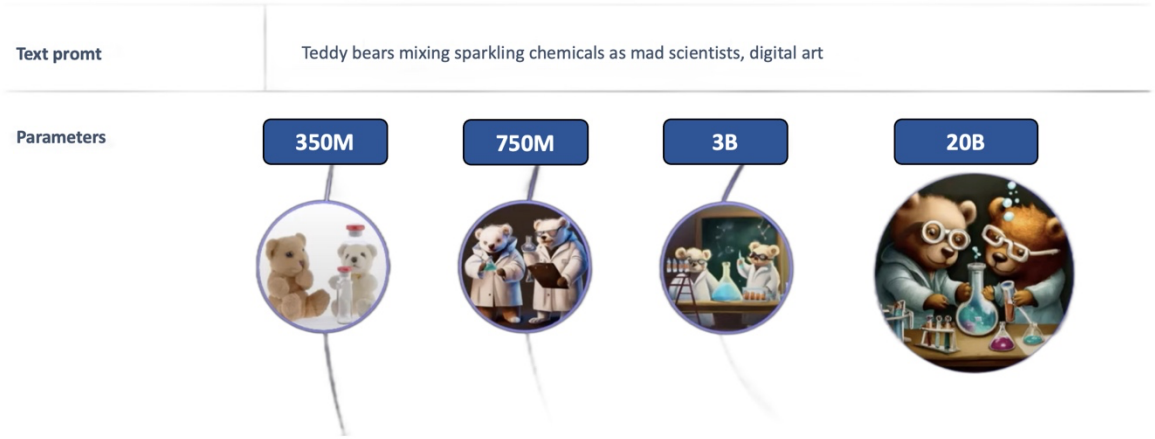

Figure 6. SAP Business AI framework (Sun, 2023)

LLMs are dependent on high-quality textual data to learn effectively. Gathering this data can be a significant task and requires extensive resources to collect, clean, and preprocess the information so that it's usable for training purposes. Additionally, the data must be kept current to ensure the model's relevance, adding to the ongoing data collection and curation burden. Like illustrated on the figure above, in order to get higher quality results, amount of data points straightly correlates how high-quality result model can present. More quality data points trained model has, better context trained model can present as a output.

On the model training side, the process is computationally intensive and time-consuming. Training a state-of-the-art LLM can take weeks or even months, utilizing powerful and expensive hardware, often with multiple GPUs or TPUs working in parallel. This not only leads to high financial costs but also has environmental implications due to the energy consumption and carbon emissions associated with operating such hardware at scale. Moreover, as the models grow in size – with some consisting of hundreds of billions of parameters –

the complexity of the training increases exponentially. This expansion requires advanced optimization algorithms and techniques to efficiently adjust the parameters and reduce the risk of overfitting, where the model performs well on the training data but poorly on unseen data. Another aspect of the challenge is the need for continuous monitoring and updating of the model. As language evolves and new data becomes available, LLMs must be retrained or fine-tuned to maintain their accuracy and relevance. However, each round of training demands a similar scale of computational resources, contributing to the ongoing operational challenges.

In conclusion, while Large Language Models hold great promise in revolutionizing the way we interact with machines and manage information, they also demand a cautious and thoughtful approach to mitigate the risks associated with their deployment. It is essential to navigate these challenges responsibly to fully realize the opportunities presented by this groundbreaking technology.

4 Artificial Intelligence in Business Applications

Artificial Intelligence has transcended its theoretical confines and has swiftly become an integral component on the various business domains. AI has already found its applications across various industries, revolutionizing processes, decision-making, and customer interactions.

4.1 The Evolution of AI in Business

AI in business was mostly experimental, used in limited ways to automate simple tasks. These early applications were primarily focused on streamlining processes and reducing human error. However, as technology advanced, AI's capabilities expanded, leading to its integration in more complex business functions.

The next phase saw AI becoming more sophisticated, with the introduction of machine learning and deep learning technologies. These advancements enabled AI to not only perform tasks but also to learn and improve over time. Businesses started to harness AI for data analysis, gaining insights that were previously unattainable. This era marked a significant shift as AI began to influence decision-making processes and strategy formulation. The current state of AI in business is characterized by its widespread adoption across various sectors. AI is no longer seen as just a tool for efficiency; it has become integral to driving innovation and creating competitive advantages. From customer service with chatbots and AI-driven personalization to predictive analytics in finance and supply chain management, AI's applications are diverse and impactful.

Moreover, the evolution of AI in business is not just about technological advancements but also involves changes in organizational culture and business models. Companies are increasingly recognizing the need for AI literacy among their workforce and are adopting more data-driven approaches in their operations. Looking ahead, the future of AI in business promises even greater integration and innovation. With the advent of technologies like quantum computing and the increasing emphasis on ethical AI, the potential for AI to revolutionize business practices is immense.

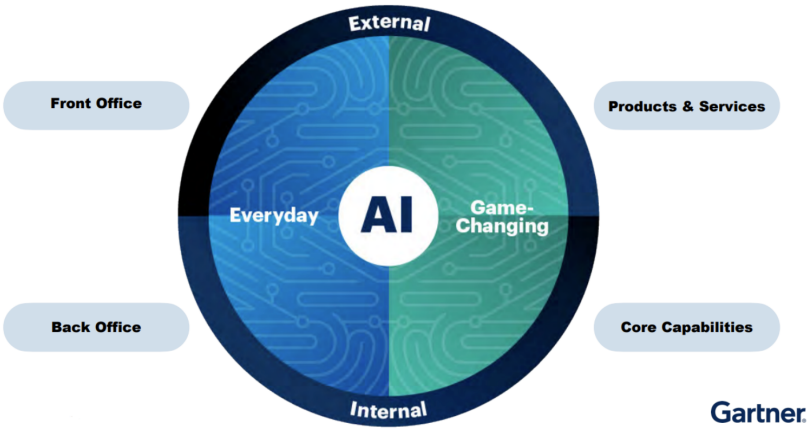

Figure 7. AI Opportunity Radar (Scheibenreif Don and Mesaglio Mary, 2023)

One way of looking various AI opportunities in businesses, is to classify requirements on two (2) different categories. AI related requirements can be either “Everyday” of “Game changing” requirements. (Scheibenreif Don and Mesaglio Mary, 2023)

Everyday AI is primarily designed to boost productivity, acting as a digital partner to help workers accomplish their tasks more quickly and effectively. At present, a significant majority (77%) of Chief Information Officers (CIOs) and technology leaders globally are paying attention to how everyday AI can benefit their operations. However, it's important to understand that the novelty and appeal of everyday AI are expected to diminish rapidly. These AI tools, once considered innovative, will soon become standard and widely accessible to all. As a result, they won't offer a long-term competitive edge to businesses. Essentially, incorporating everyday AI is rapidly becoming a basic requirement, not an extraordinary advantage.

Game-changing AI stands out because it's all about sparking creativity and innovation. This kind of AI doesn't just help end or business users do things faster or better in the usual ways. Instead, it opens up entirely new possibilities. It can lead to the creation of brand-new products and services that we haven't seen before, thanks to AI's unique capabilities. Or, it can change the way we come up with these new ideas and products, by adding new core abilities that we didn't have access to before. With game-changing AI, we're looking at machines that can totally shake up the way businesses operate and even transform entire industries.

4.2 Key Business Applications of AI

Artificial Intelligence is revolutionizing various business sectors. Demands and business requirements varies a lot based on sectors, but common trend which is applicable for all the business sectors, is that AI applications are brining most value for improving operational efficiency and productivity. Business sectors are seeking how to improve their overall productivity, but also how to streamline existing business operations.

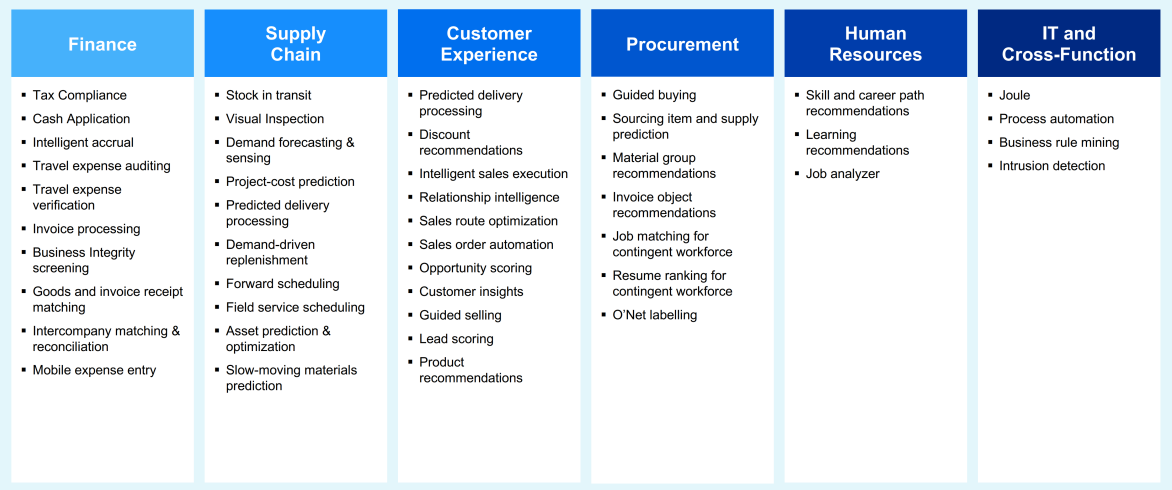

Figure 8. Generative AI powered scenarios across various business functions (Allardice, 2023)

Here below are couple typical examples how AI is implemented and what are the business use cases and patterns AI is applied to:

Sales and Customer Service: AI is revolutionizing sales and customer service. This includes AI-driven chatbots that can handle customer queries instantly, predictive sales analytics which anticipate future sales trends, and AI-powered personalized shopping experiences that tailor product recommendations to individual customer preferences.

Human Resources (HR): AI's role in HR is increasingly prominent, especially in recruitment. It's being used to screen resumes to identify the most suitable candidates efficiently, facilitate video interviewing processes, and employ predictive analytics to foresee employee performance, helping in making informed hiring decisions.

Finance: In the finance sector, AI is a game-changer. It's used for detecting fraudulent activities more accurately, powering robo-advisors for investment and financial advice, and

enabling predictive financial modeling to forecast future financial scenarios and market trends.

Supply Chain and Manufacturing: AI significantly contributes to supply chain and manufacturing efficiency. It's used for optimizing inventory management by predicting stock requirements, conducting predictive maintenance to pre-empt equipment failures, and forecasting demand to ensure supply chain efficiency.

Marketing: In marketing, AI is crucial for data-driven strategies. It helps in creating more effective marketing campaigns by analyzing vast amounts of data, segmenting customers into distinct groups for targeted marketing, and utilizing predictive analytics to anticipate future market trends and consumer behaviors.

4.3 Requirements of AI in Modern Business Environment

On the contemporary business landscape, the integration of Artificial Intelligence applications is essential for maintaining competitive advantage, driving innovation, and enhancing operational efficiencies. However, for these applications to be effective, they must meet certain fundamental requirements.

AI applications must be highly **relevant** to the specific business processes and industry in question. This means the AI should be tailored to address precise business needs, whether it’s improving customer service, optimizing supply chain logistics, or personalizing marketing efforts. Relevancy also implies that the AI application should be capable of processing industry-specific data and learning from it to provide actionable insights. For example, in healthcare, AI should comply with medical protocols and use patient data to improve diagnoses and treatment plans. (Tikare, 2023)

Reliability in AI refers to the consistent performance of the application over time. Businesses require AI applications that they can trust to function correctly, make accurate predictions, and automate processes without frequent errors. On the business domains using AI as decision maker or even as a copilot, have always business impact or consequences on decisions made – either by AI application or human. The system should be robust against variations in data quality and be capable of operating effectively under different conditions.

AI must also exhibit a high degree of transparency and explainability, enabling users to understand and trust the results and decisions it makes. (Tikare, 2023)

AI applications must be **resilient**, meaning they should be designed to handle and recover from disruptions, such as changes in market dynamics, data streams, or operational environments. Resilient AI systems are able to adapt to new data, learn from unforeseen situations, and continue to operate effectively even in the face of challenges like cyber attacks or system malfunctions. Additionally, they should be sustainable and maintain performance without requiring excessive maintenance or retraining, which can be costly and time-consuming. (Tikare, 2023)

5 Exploring SAP Business AI

After the launch of OpenAI's ChatGPT has sparked significant impact how business applications are going to get developed and used. Companies are swiftly moving to capitalize on the potential efficiencies offered by generative AI (GenAI). SAP's clients depend on SAP itself to integrate GenAI into SAP product offering and infrastructures. SAP can achieve this by infusing their services with GenAI capabilities and melding them with SAP's data expertise. Typically SAP clients utilize SAP for their most crucial operations, expecting not only robust solutions but also a commitment from SAP to address ethical, data privacy, and security issues. This has an impact what is SAP's strategic approach to employing GenAI. SAP has a history of integrating AI into its business applications, however recent advancements have unlocked extensive new possibilities that surpass previous capabilities. As new use cases and technologies emerge, SAP needs to revisit their AI strategy regularly and refine it.

5.1 Vision and high-level strategy

According to analyst companies, like IDC and Gartner, SAP is one of the frontrunners of integrating Artificial Intelligence into the business landscape. Currently, over 24,000 SAP cloud customers are utilizing upwards of 130 AI functionalities within their business applications and their technology platform Business Technology Platform (BTP). SAP's AI strategy, underpinned by GenAI, remains consistent. Approach is to embed GenAI directly into SAP's core business applications and BTP. SAP also aims to provide access to business oriented GenAI services through BTP, which can be used by customers and partners to enhance their own business applications. Furthermore, SAP is committed to delivering AI that is reliable and trustworthy. It's important to note that SAP is not looking to enter the general-purpose AI platform market. (Muller Jurgen, 2023)

The recent advancements in GenAI technology have opened doors to remarkable possibilities. They have the potential to elevate levels of automation to new heights, enable more natural interactions between humans and machines, and enhance cognitive and decision-

making abilities, allowing individuals to focus on creativity and critical tasks. Reimagining the ease of coding an extension to the existing business logic, simply by describing the problem in natural language or an application that intuitively understands how to integrate and configure itself within various settings. The applications of GenAI are vast and varied.

SAP's differentiation lies in its ability to leverage the unique business context provided by business data, process expertise, and in-depth industry knowledge accumulated from more than 400,000 customers. This enables SAP to apply GenAI in a manner that is both responsible and tailored to the needs of businesses. To facilitate this, SAP is planning to incorporate GenAI capabilities into their technology platform as foundation capabilities. Also, part of the SAP's AI strategy is to use of open-source tools and frameworks, such as LangChain and LlamaIndex, and will develop their own services as necessary to meet the specific needs of SAP clients and partners. (Marfo, 2023)

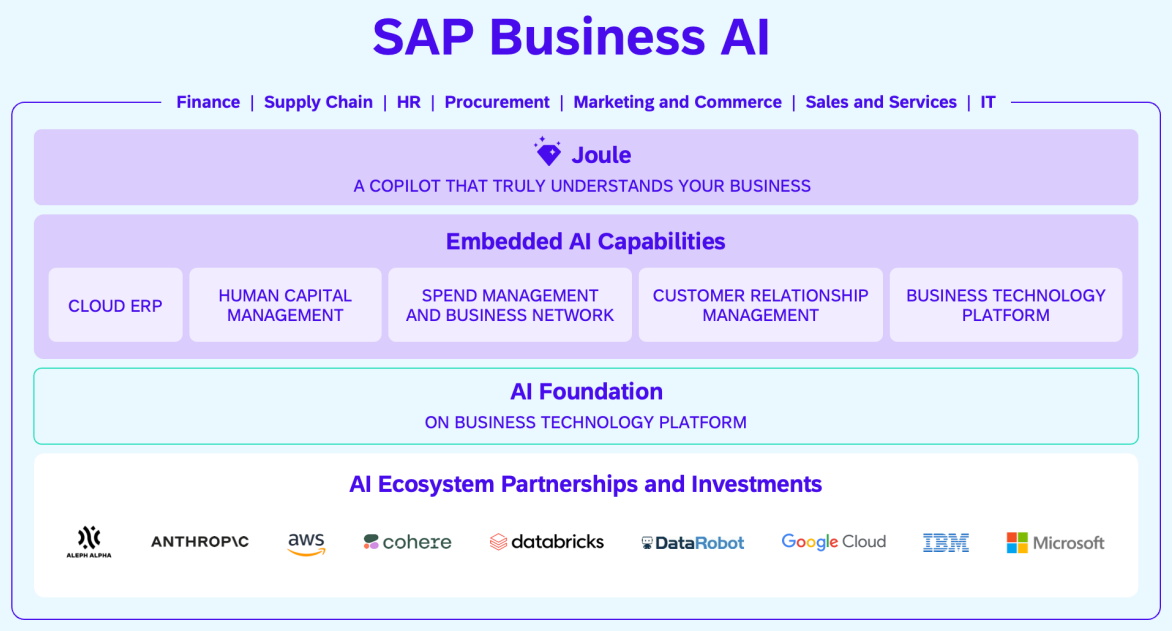

Figure 9. Infusing Generative AI into SAP's product portfolio and SAP´s technology platform (Muller Jurgen, 2023)

These capabilities will enable SAP development teams to infuse GenAI-powered process innovations across all SAP cloud business applications as well as into BTP itself. Likewise, SAP will enable their partners to benefit from GenAI made available via technology

platform, BTP. Short term, BTP will provide access to third-party, open-source, and own GenAI models provide a GenAI layer to ensure enterprise-grade GenAI that is used across the SAP ecosystem.

5.2 Use Cases & Requirements

Generative AI stands to revolutionize the application of artificial intelligence across numerous business sectors. These cutting-edge models enable even those with no technical background to address business challenges simply by articulating them in natural language. This development democratizes access to AI's transformative power, offering both enterprises and individuals the chance to leverage its potential without deep technical know-how. Foundation models, such as those in OpenAI's GPT series, are particularly noteworthy. They are crafted for broad applicability across a diverse range of tasks, negating the need for specialized re-training. Businesses employing these models can streamline operations, make more informed decisions, and automate mundane tasks, thus catalyzing growth and fostering innovation.

The versatility of general-purpose foundation models is another significant benefit. These models can be fine-tuned to address specific tasks in a relatively short span, enabling organizations to rapidly deploy bespoke AI solutions that align with their unique requirements. This adaptability is a game-changer, providing a fast track to harnessing the full potential of AI tailored to an organization's needs. (Kaupa, 2023)

1. **Content Generation**: Create articles, blog posts, social media content, mail drafts, product descriptions, or even poetry and stories based on a given theme or keywords.
2. **Question-Answering**: Build systems that answer questions based on a given context or knowledge base, extract valuable insights and information from large datasets or collections of text documents.
3. **Summarization**: Generate concise summaries of long articles, news, or research papers to help users grasp the main ideas quickly.
4. **Code Generation**: Given a natural language description, generate code snippets in various programming languages, or autocomplete code for developers.

5. **Creative Applications:** Assist with brainstorming ideas, generating names for products or companies, creating advertising slogans, writing song lyrics, generating image/video/voice.

The integration of Generative AI into SAP product portfolio and offering, brings benefits that promise to elevate the user experience, streamline content generation, and boost developer productivity. A primary advantage is the enhancement of the overall software and service experience for customers. Natural language interfaces simplify system navigation and facilitate access to necessary functionalities. This can revolutionize customer support by automating processes, leading to faster resolution of issues and heightened customer satisfaction. The ability to retrieve information conversationally empowers users to efficiently obtain the data they require, thereby enriching the user experience and boosting productivity. In the area of content creation and knowledge management, Generative AI serves as a powerful ally. It provides the capability to generate or refine diverse content types, such as marketing and sales materials, aiding businesses in effectively conveying their value proposition to customers. Furthermore, these models can assist in condensing and summarizing documents and relevant data e.g. to SAP ERP application - SAP S/4HANA. Biggest value embedding AI into SAP applications is to provide users with a concise overview and enabling them to make well-informed decisions.

Lastly, Generative AI stands to substantially accelerate and enhance the workflow of developers engaged with SAP products. Features like the generation of code from natural language descriptions and code auto-completion allow developers to operate with greater efficiency, consequently reducing the time to market for new features or enhancements. Automated documentation generation also plays a crucial role, ensuring developers have access to precise and current information, thereby streamlining the development lifecycle.

5.3 Architecture and implementation concepts

Architecture definition and implementation concepts play significant role on application development process. To infuse GenAI capabilities into SAP applications, there are certain aspects software development need to take into consideration to be aligned with SAP AI strategy. (Kaupa, 2023)

- accessing foundation models/LLMs
- selecting one or multiple foundation models/LLMs for use case, optimizing model usage
- engineering high quality prompts for the foundation models/LLMs to achieve the desired result
- provide trustworthy, responsible, and ethical AI applications

In the integration of Generative AI within development projects and initiatives, it's crucial to address financial considerations from the given customer business requirements, and underline in the Total Cost of Ownership (TCO) of built solution. A primary objective is to ensure that customers have the ability to anticipate and manage costs effectively. The main cost determinants in GenAI implementations encompass several factors: the choice of Large Language Models, the volume of input and output tokens required, the processes involved in creating and maintaining embeddings, and the expenses associated with training or fine-tuning the models. Additionally, substantial efforts may be necessary to navigate and resolve legal complexities.

It's also important to recognize that GenAI is not a universal fix for all scenarios. In some instances, alternative AI methodologies may not only prove to be more cost-effective but also provide outcomes that are more predictable and interpretable. Thus, it is essential for development teams to evaluate the suitability of GenAI on a case-by-case basis, weighing its benefits against other AI approaches to determine the most efficient and effective solution.

Large Language Model access

The landscape of Large Language Models is described by tight competition among various technology vendors, creating a highly dynamic market. This competitive trend is not just hanging on but also intensifying. A prime example of this is OpenAI’s GPT-4 release, which currently leads as the most advanced LLM in a multitude of tasks. However, this dominance was challenged in July 2023 when Anthropic released Claude-2. The new entrant appears to rival GPT-4 in terms of quality, potentially even surpassing it in e.g. coding tasks, while offering significant advantages such as lower cost and an expanded context window. For instance, the Aleph Alpha model, though considerably smaller, demonstrated superior performance in document content extraction tasks when fine-tuned with SAP data and hosted

on SAP's own dedicated business technology platform, was outshining OpenAI's offerings in this specific area. (Kaupa, 2023)

This rapid evolution in the LLM competition indicates that more vendors are entering the market, and open-source models might reach competitive standards, costs are expected to decline, and models will likely become more specialized for certain business tasks. In response to this ever-changing environment, technology vendors are likely to adopt a multi-vendor strategy to facilitate access to LLMs. This approach involves centrally contracting, validating, and providing access to various LLMs while managing several non-functional aspects such as security and metering. Despite this centralized approach to maintain and safeguard data protection and privacy disciplines. This entails refraining from transmitting personally identifiable information to third-party LLMs. GenAI capabilities are set to be utilized not only by business applications but also by infrastructure and software development tooling, including tasks like code generation, which forces technology vendors to centralize some of the GenAI shared capabilities for managing them properly.

In addition to providing access to cloud-based third-party LLMs such as those offered by major cloud vendors like Microsoft Azure, SAP is also planning to host and operate models from open sources and partners on their technology platform, including those from Aleph Alpha or Cohere. This diversified hosting strategy is particularly relevant for use cases with stringent data protection requirements or scenarios involving sensitive data that neither we nor our customers wish to risk exposing to any external LLM provider. Data protection and keeping critical business data on customer's ownership and management, is crucial for any successful GenAI implementation.

It is foreseen that in the future, there is the requirement for expanding of GenAI, including orchestration features to support the construction of LLM chains, agents, and ultimately, autonomous agents across the business application landscape.

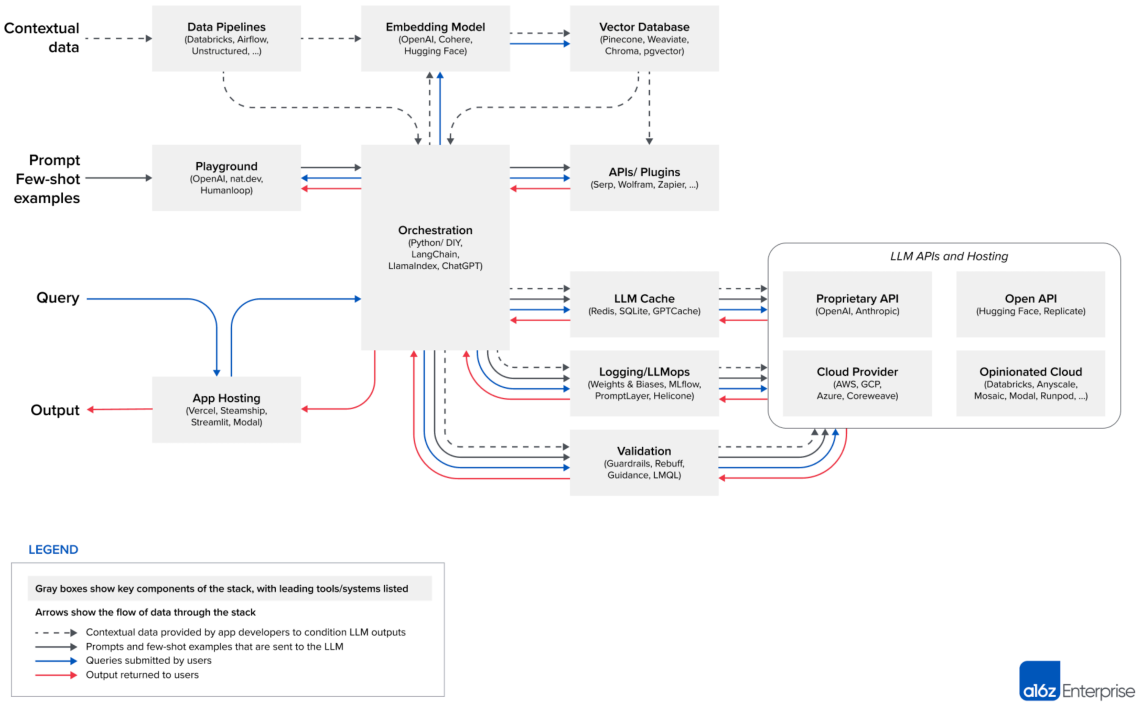

Figure 10. Emerging Architectures for LLM Applications (Borstein and Radovanovic, 2023)

Selecting suitable Large Language model

Incorporating a GenAI framework begins with a critical decision: identifying the Large Language Model (LLM) that aligns most effectively with the intended application and business use case. The market offers a numerous of LLM options, with their numbers and sophistication advancing rapidly. This is underscored by SAP's recent commitment to investing in several top LLM providers. Furthermore technology vendors like SAP has introduced playground for LLMs which serves as an experimental hub where developers can evaluate the compatibility of various LLMs through hands-on testing. Presently, GPT-4 is renowned for its superior output quality, though it's important to acknowledge its higher cost and comparatively slower response times. (Kaupa, 2023)

Prompt engineering in LLM deployment

The prompt engineering is at the heart of deploying the full potential of a Large Language Model. This intricate process involves the accurate construction of prompts that are finely tuned to extract the precise responses desired from an LLM, while steering clear of unproductive or illogical replies. The efficacy of a prompt hinges on its ability to clearly delineate

the task at hand. This involves detailing the expected format, moderating the style or tone of the generated content, and integrating context or limitations that may influence the response. Moreover, providing illustrative examples or models of the intended output can significantly sharpen the LLM's comprehension and output accuracy. (Kaupa, 2023) For business applications fluency in prompt engineering is not just a technical skill but a strategic imperative. It is a fundamental discipline that underpins the successful application of LLMs across a various of business scenarios. This knowledge becomes even more pivotal when complemented by a suite of supportive tools and advanced methodologies, such as embeddings, fine-tuning practices, or sophisticated orchestration strategies. It is through the mastery of prompt engineering that business users can truly leverage the formidable capabilities of LLMs to meet diverse and complex use case requirements.

Trustworthy AI

SAP places a strong emphasis on developing AI that is trustworthy, responsible, and ethical. As one of pioneers in this field, SAP was the first company to publish an AI Ethics policy, setting a high standard in the industry. This approach is not just a market differentiator but also a reflection of SAP's business values. (Kaupa, 2023)

Ethics: SAP's Global AI Ethics Policy aims to align with the UNESCO Recommendation on the Ethics of Artificial Intelligence and incorporate feedback from SAP’s AI Ethics Advisory Panel. This process tries to secure that Ethics Policies are followed on various product development initiatives, and that ethics principles are implemented. SAP has created AI Ethics assessment process and a comprehensive AI Ethics Handbook with detailed guidelines.

Data Protection and Privacy: SAP tries to actively address the risks associated with data leakage and the handling of personally identifiable information (PII). Typically, SAP business application hold customers business critical data, and therefor data protection is one of fundamental policies to be followed. This includes also contractual agreements with partners to ensure data privacy, aligning vendor agreements and implementing a mandatory legal intake process for GenAI use cases.

Security: GenAI introduces unique operational and development risks, necessitating robust controls to prevent data source tampering and to secure GenAI inputs/outputs. SAP seem to be focusing on educating its teams about these risks and implementing stringent security

protocols, including input/output validation, human-in-the-loop guidelines, and a centralized prompt engineering framework. Furthermore, SAP is exploring how GenAI can enhance its security processes, such as by detecting threats in logs or generating security documentation.

Explainability: The concept of Explainable AI (XAI) is particularly important for GenAI, as it contributes to both security and trust. XAI techniques can help detect data poisoning and AI hallucinations, and they are increasingly seen as a compliance necessity in light of regulations like the EU GDPR. This focus on explainability is not only a security measure but also a step towards more transparent and understandable AI operations.

6 Conclusion

6.1 Summary of Key Findings

The journey of AI and LLMs in the business world is a story of transformation. These technologies started off with basic tasks like data processing but have grown to become essential tools in modern business operations. This evolution is mostly thanks to advancements in how these models understand and process language, making them more useful for complex tasks.

LLMs are particularly good at understanding and generating language, translating between languages, summarizing large amounts of text, and recognizing patterns. These skills are not just automating routine tasks; they're also creating new ways to interact with customers and handle their needs. This has made businesses much more efficient and responsive. Introducing LLMs into business processes brings both big opportunities and significant challenges. While they can make things more efficient and lead to innovative solutions, there are also serious concerns about ethics, privacy, and the need for high-quality data to train these models properly. Finding the right balance here is key for using LLMs effectively and responsibly. When it comes to specific industries like healthcare, finance, and manufacturing, LLMs have shown their ability to adapt and provide tailored solutions. In healthcare, for example, they're being used for more accurate diagnoses and predictive analytics, while in customer service, technology providers have found a way automating and improving interactions with customers. Integrating LLMs into a business successfully means focusing on the quality of the used data, training the models well, and developing solutions that meet the specific needs of various industries.

We also can't ignore the ethical and legal implications of using AI and LLMs in business. First versions of the generative AI technologies need to be used in a responsible way, making sure they comply with data protection laws and ethical standards. This is crucial for maintaining trust and ensuring these technologies are used sustainably. This is magnified when operating on the highly volatile business environments. Finally, the future of AI and LLMs in business looks very promising. Applying these modern technologies could completely change how businesses operate, make decision-making processes better, and even affect the

future of work. As AI and LLMs continue to develop, they're likely to become even more influential, opening up new possibilities and challenges.

6.2 Final thoughts and reflections to research questions

Every technology has obstacles to overcome and Generative AI and Large Language Models are no different.

Research question - How does the integration of AI in business applications influence the efficiency and effectiveness of organizational processes?

One of the primary advantages of AI business applications is the automation of routine tasks. AI helps in taking over repetitive tasks like data entry and basic customer service queries, freeing human employees to focus on more complex and creative work. This shift not only enhances efficiency but also increases job satisfaction among company personnel. AI's ability to process large data volumes improves business decision-making. AI applications can analyze extensive datasets to identify business trends and patterns, leading to informed and effective decisions that might be beyond human analysts' capabilities. In customer service, AI-powered chatbots and virtual assistants provide a level of personalization and availability that significantly enhances the customer experience, offering immediate responses to customer queries around the clock.

Predictive analytics is another area where AI applications show clear benefits by utilizing historical data to forecast future outcomes. This capability is especially useful for example in inventory management, where AI's predictive insights can help in optimizing stock levels and reducing waste. Similarly, in supply chain management, AI can predict potential disruptions, optimize logistics routes, and manage supplier relationships more efficiently, resulting in streamlined operations. AI also plays a decisive role in analyzing employee productivity and engagement. By offering insights into potential improvements in workplace, AI tools help in boosting efficiency and creating a more satisfying work environment. In marketing and sales, AI enables high degrees of customization and personalization, analyzing customer data to create targeted campaigns and recommend tailored products or services, thus increasing the effectiveness of marketing strategies. Moreover, AI is a potential assistant for innovation and the identification of new business opportunities. With the use of AI, it can analyze

market trends and patterns, and AI can suggest new product or service opportunities and highlight areas ready for innovation.

Research question - What role do ethical standards and data privacy norms in AI development play in fostering stakeholder trust and ensuring sustainable business practices?

There are several legal, ethical and technical obstacles that needs point out here. Pre-trained LLMs provided by tech companies like Google are trained on massive amounts of text which is typically crawled from the web. The details of how this data was collected and processed are often not transparent. We need to assume that **legal and data protection regulations (copyright, licenses, GDPR) were not strictly followed**. Even if SAP is only using pre-trained LLMs created by others in its software products, there is a residual risk of legal and data protection incompliance. There is little experience the legal community yet on the risk regarding LLMs or how courts decide on these issues. LLMs have the potential to memorize large parts of their training data. If technology vendors, like SAP, wanted to create and open-source LLMs it would have to be **ensured that no confidential or personal data is accidentally leaked**. LLMs carry risks with regards to **ethical and trustworthy AI standards**. Existing unwanted biases and stereotypes existing in the training data can be learned by the model, for example stereotypes against women or minorities. LLMs, by their nature, are very large and difficult to deploy cost-effectively due to the high amount of compute necessary to run those models. The high computational costs in the creation of LLMs come with a **large carbon footprint and environmental concerns**. LLMs are still quite **brittle and subject to adversarial attack**.

Research question - What are the challenges on business-specific training for AI?

A technical problem with regards to LLMs is their fragile nature. At the end of the day, a LLM is a statistical model trained on vast corpora of text to produce probabilities over word sequences. As such, these LLMs are subject to biases in their training data and procedure and are susceptible to adversarial attacks where a malicious agent can find “trigger word sequences” that make the LLM produce gibberish or, even worse, unsavory content. There is active research probing the brittleness of LLMs and how to make them more robust. There are efforts to, for example, combine LLMs with knowledge graphs to ensure that these models have access to a source of ground truth. Current guardrails against these issues are

to deploy LLMs for tasks which are strictly constrained – e.g., for classification tasks, ranking tasks, or text-extraction tasks. Unconstrained tasks such as abstractive summarization or question-answering may introduce unanticipated risks and failure modes. A prominent example of such a failure mode occurred with Microsoft Tay, a chatbot which learned to produce increasingly racist and inflammatory language after interacting with the general public for only a few hours.

There has been interest lately in trying to make LLMs deployable at scale. **Beyond the constant progress towards faster chips and larger memories, there has also been a swell of research into model distillation, model compression, and architectural changes to reduce the size and increase the inference speed of a large language model.** The amazing progress in shrinking models and improving compute gives a strong indicator that deployment at scale for production-level tasks will soon be, if not already, within reach.

References

Allardice S (2023) *AI & Generative AI: Executive Briefing.* Available at: https://www.pluralsight.com/authors/simon-allardice.

Bartneck C, Lütge C, Wagner A and Welsh S, (2021) *An Introduction to Ethics in Robotics and AI.* Cham: Springer Nature.

Bernhardt C (2016) *Turing's Vision : The Birth of Computer Science.* Cambridge, Massachusetts: The MIT Press.

Borstein M and Radovanovic R, (2023) *Emerging Architectures for LLM Applications.* Available at: https://a16z.com/emerging-architectures-for-llm-applications/.

Cavallucci D, Livotov P and Brad S, (2023) *Towards AI-Aided Invention and Innovation : 23rd International TRIZ Future Conference, TFC 2023, Offenburg, Germany, September 12-14, 2023, Proceedings.* Cham: Springer.

Epstein R, Roberts G and Beber G, (2008) *Parsing the Turing Test Philosophical and Methodological Issues in the Quest for the Thinking Computer.* New York: Springer.

Foster D (2023) *Generative Deep Learning, 2nd Edition.* : O'Reilly Media, Inc.

Harris P, Paul M and Pal S (2022) ARTIFICIAL INTELLIGENCE FAMILY TO MODERN BUSINESS. *Internal Auditing* 37(5): 41-44.

Hodges A *Alan Turing and the Turing Test.* In: Anonymous Dordrecht: Springer Netherlands, 13-22.

Hrischev Rand Shakev N (2022) ARTIFICIAL INTELLIGENCE IN ENTERPRISE RESOURCE PLANNING SYSTEMS.

Kask S (2023) Generative AI in BTP .

Kaupa M (2023) Generative AI at SAP - Technology Strategy.

Kline RR (2011) Cybernetics, Automata Studies, and the Dartmouth Conference on Artificial Intelligence. *IEEE Annals of the History of Computing* 33(4): 5-16.

Marfo, M.K. (2023) *SAP Business AI Strategy Overview.*

Muller Jurgen (2023) *Executive Keynote with Juergen Mueller SAP TechEd in 2023.* Available at: https://www.youtube.com/watch?v=6BsvzygS5Hg.

Salminen, A. (2011) *Mikä Kirjallisuuskatsaus?* : Vaasan yliopisto.

Scheibenreif Don and Mesaglio Mary, (2023) *Gartner Decide Your AI Ambition.* Available at: https://www.gartner.com/en/newsroom/press-releases/2023-10-16-gartner-says-cios-must-prioritize-their-ai-ambition-and-ai-ready-scenarios-for-next-12-24-months.

Sejnowski TJ (2018) *3 The Dawn of Neural Networks.* In: Anonymous : MIT Press.

Shieber SM (2004) *The Turing Test : Verbal Behavior as the Hallmark of Intelligence.* Cambridge, Mass: Mit Press.

Sun W (2023) *Large Language Models and Copilot in Business Applications.* Available at: https://www.youtube.com/watch?v=w5AAgMw-eUA.

Taulli T (2023) *Generative AI: How ChatGPT and Other AI Tools Will Revolutionize Business.* Berkeley, CA: Apress L. P.

Tikare, A. (2023) *Business AI @SAP.* : SAP.

Yang X (2013) *Artificial Intelligence, Evolutionary Computing and Metaheuristics in the Footsteps of Alan Turing.* Berlin, Heidelberg: Springer Berlin Heidelberg