Deloitte.

Gen AI in Manufacturing

POV

Deloitte

Table of Content

Here's a summary of the document

This Point of View (PoV) marks the first in a series of the current and potential future adoption of Gen AI's across manufacturing, while also detailing the evolution of its capabilities. While it may be particularly suited for audiences with some technical background, it also illuminates the "art of the possible" for a broader audience, spotlighting how these evolving capabilities can optimize production processes. Look out for upcoming installments that will explore the strategic and business value of Gen AI, turning today's insights into tomorrow's competitive advantage.

Here are some key takeaways from this POV

- Nearly 65% of manufacturing firms plan to invest in Gen AI over the next three years, targeting areas like production, supply chain management, and customer interactions.
- Early adopters are focusing on high-impact use cases such as design optimization and quality assurance, but face challenges with low data quality and the high cost of fine-tuning Large Language Models (LLMs).
- The Short-term adoption might involve increased usage of fine-tuned Smaller Language Models (SLMs), which offer 2x better contextual understanding than LLMs, and can be used for tasks like generating 3D designs from text, optimizing operational instructions, and managing pricing strategies.
- The Long-term adoption might involve usage of Compound Agent Systems, impacting multiple activities across a typical manufacturing value chain.
- To harness the full potential of Gen AI, it is crucial to establish a strong data foundation and align people and processes with the technology.

1. Introduction

1.1 Background on Gen AI in wider industry

Global spending on Gen AI-enabled solutions is anticipated to nearly reach $200 billion, with a compound annual growth rate (CAGR) of 29% over the next four years. The financial services sector is expected to lead this growth, driving 20% of the total global investments in Gen AI[1].

1.2 Gen AI in Manufacturing

The manufacturing sector is increasingly investing in a diverse array of industrial metaverse technologies to achieve their Smart Factory objectives. This strategic move aims to enhance competitive advantage by optimizing asset efficiency and generating additional revenue streams through cross-selling, among other benefits. On average, approximately 65% of firms plan to implement these technologies within the next 1 to 3 years, spanning areas such as production, supply chain management, and customer interactions[2].

However, the current level of Gen AI expertise and readiness for scaling in the manufacturing sector appears to be lower compared to industries like financial services[3]. As a result, the manufacturing sector is likely to adopt Gen AI capabilities cautiously. In the short term, investments will be focused on use cases that directly enhance production efficiency or boost customer loyalty, while aiming to secure a competitive edge in the long term.

1.3 Purpose of this Point of View (PoV)

The POV serves three primary purposes:

01. To highlight the diverse use cases across various segments of the manufacturing value chain that are currently in the early adoption phase, as well as those expected to be adopted in the short and long-term horizon.

02. To demonstrate the systemic evolution in the development and deployment of Gen AI use cases and solutions:

    A. Initially, the discussion will be on the early adoption phase, which primarily involves enhancing existing Large Language Models (LLMs), without many customizations or fine-tuning, for specific use cases as part of experimentation or proof-of-concept initiatives.

    B. In the short term, innovations will likely center on the use of fine-tuned Smaller Language Models (SLMs), for faster processing in specialized performance scenarios[4], or LLMs, for large-scale applications. The key measure for success will be the ease of deployment and scaling.

    C. In the long term, the focus will shift towards developing compound AI systems, also known as agents, which use the ability of both SLMs and LLMs. The emphasis will be on creating multi-agent systems that impact multiple stages of the entire manufacturing value chain, by executing multiple use cases together. However, the ability to create and manage such systems will depend on various factors, some of which are touched upon in this POV.

03. To set the stage for a wider series of discussions regarding the adoption of Gen AI in manufacturing contexts.

    A. While this POV is anchored on the development of this technology's usage up to and around the present day, subsequent articles will focus on the skepticism and apprehension that has risks delaying the embracing of this transformation. This will include an analysis of how the Gen AI manufacturing applications introduced in this article can be linked to topical issues present in the industry, and how this impacts the technology's business value.

2. The Early Adopters

2.1 Overview

The dawn of Gen AI in manufacturing marks a new era of industrial revolution. This subset of AI, capable of creating new content, is poised to redefine the manufacturing landscape. From enhancing existing technologies to pioneering novel applications, the early adoption of Gen AI is laying the groundwork for a more efficient, innovative, and sustainable future.

At present, early adopters are predominantly dedicated to investigating the potential of LLMs in manufacturing for select use cases with significant impact. In many instances, these use cases already employ AI models or algorithms to attain desired business outcomes. However, companies are making limited investments in fine-tuning LLMs due to the substantial costs associated with this process. To facilitate the contextual knowledge necessary for these use cases, Retrieval Augmented Generation (RAG) is employed. However, for more than 42% of early adopters, low Data Quality continues to be a major obstacle to enabling Gen AI using RAG[5]. In cases where there is scarcity of high-quality training data sets, organizations are likely to incorporate third-party datasets to augment their existing internal data[6].

Current examples where Gen AI is being used commercially include generation of novel product designs based on predefined constraints and objectives, prediction of equipment failures, and detection of defects and anomalies in products.

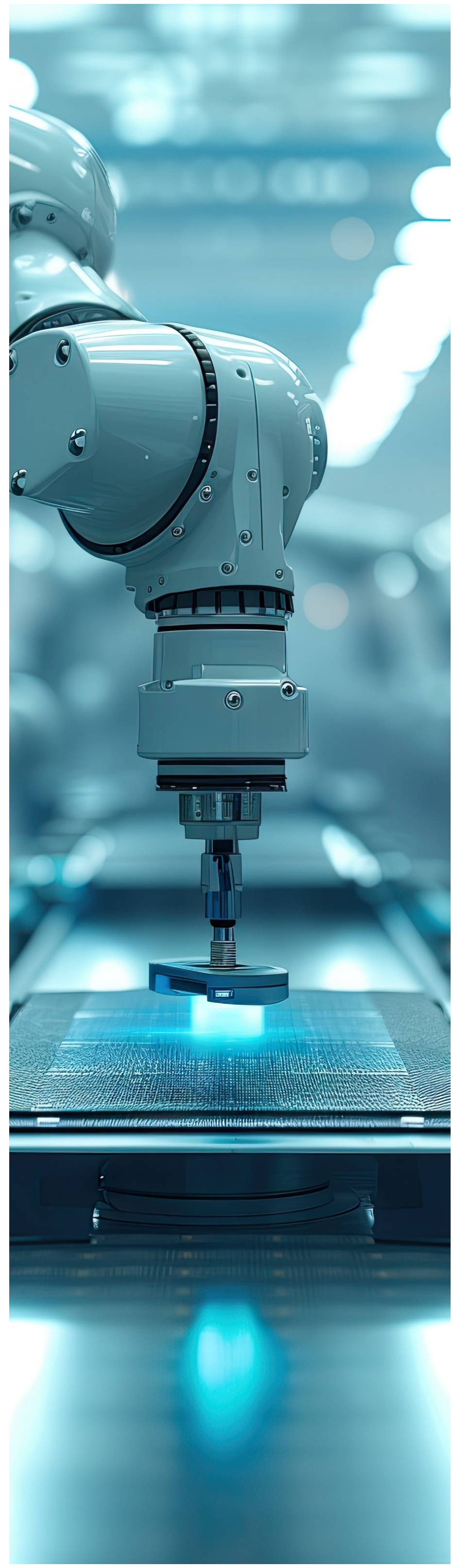
The early adopters are laying the initial groundwork for effectively harnessing this capability while mitigating risks. As these early adopters gain confidence in the potential of Gen AI and the industry gains more clarity on scaling this technology, a revolution in manufacturing is inevitable.

2.2 Value Chain Use Cases

Within the manufacturing value chain, the early adoption of Gen AI is predominantly observed in the operations stage, with a stronger emphasis on the pre-production stage. However, there is a growing momentum in the production stage as well. Here are a few commercial instances where this capability is being utilized:

Deloitte Smart Factory Uses: The Deloitte Smart Factory, located in Montreal, has been designing, developing, and implementing various innovations that incorporate machine learning and AI in a manufacturing environment. Initially focused on artificial intelligence, the factory is now expanding its use of Gen AI across several functions:

- Linear programming (LP) has long been a standard approach for scheduling in manufacturing plants. At the Smart Factory, the team enhances LP-based scheduling by incorporating real-time information to quickly respond to changes such as machine breakdowns, supply chain disruptions, or sudden shifts in demand. This also includes using AI and machine learning to develop dynamic scheduling systems that can adapt in real time. For example, if a critical machine breaks down, the AI can instantly reschedule tasks, allocate resources, and adjust workflows to minimize downtime. Additionally, the data collected can be utilized by a Gen AI system to conduct scenario analyses that help mitigate future scheduling risks thus supporting increased efficiency, reduced operational costs, and improved production timelines.



- GenAI in Operation: The Smart Factory team uses GenAI to enhance operations by leveraging its ability to scan maintenance manuals, support root cause analysis techniques such as the 5 Whys and fishbone diagram, and provide step-by-step guidance for operators in troubleshooting issues. These methods helped identify the key details and reasons behind equipment breakdowns, enabling the team to mitigate minor failures and reduce downtime. The Gen AI assistants enhance decision-making by reviewing existing work orders and maintenance strategies to support operational efficiency. These assistants can provide real-time insights and recommendations based on extensive data from various sources, including IoT devices, ERP, and MES systems.

Design Optimization: Toyota has begun using Gen AI to supplement its designing of new vehicles and components. The Toyota Research Institute (TRI) recently released a Gen AI technique to support its team of human designers. It allows them to add initial sketch designs and engineering constraints in a Gen AI system, resulting in faster optimization of vehicle designs to reconcile engineering considerations with stylistic factors[7].

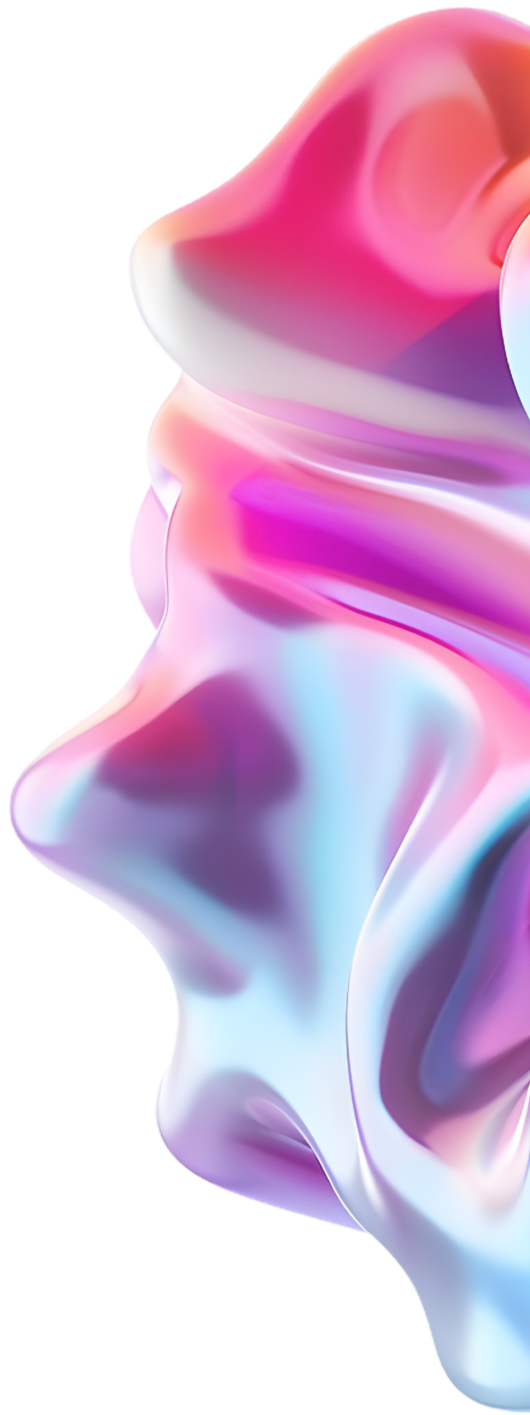
The technology allows for considerations like drag to be factored into designs specifying chassis dimensions such as ride height and cabin dimensions. A designer can also use text-based prompts around vehicle style (for example, “sleek” or “modern”) to request multiple design options conforming to input dimensional requirements.

This use case demonstrates the early adoption of Gen AI as an enhancement, or supplement, to existing manufacturing processes.

Quality Assurance: In the production phase of the value chain, Bosch has begun utilizing Gen AI to enhance its existing methods of visual quality assurance and defect detection of its products. Whereas previously, these functions would be performed by a human with their naked eye, the company is now leveraging Gen AI-based image recognition.

An issue for Bosch in doing so, was the fact that the high-quality of their existing production process did not result in sufficient number of fault or issue scenarios that could be used as training images for the Gen AI system. As a result, they leveraged Gen AI to create “synthetic data”- essentially creating artificial images of defective components to then be used in Gen AI-based optical inspection.

At their Hildesheim plant in Germany, this technology allowed for the creation of over 15,000 artificial images based on two images for each of the six fault types that can occur in the production of an electric motor component. The earlier detection of faults is expected to reduce project lengths by up to six months[8].



3. Short-term Adoption

3.1 Overview

The next stage of Gen AI usage in manufacturing will focus on the development of short-term innovations that advance and further leverage the use of this technology in manufacturing processes. This will see mature early adopters of Gen AI enhance their existing use cases for more tailored and complex manufacturing processes, via increased usage of customized Smaller Language Models (SLMs)[3]. There will be significantly more focus and investments on fine-tuning the already deployed LLMs, to achieve even higher value from Gen AI. There is already evidence within aircraft manufacturing on more effective ways to fine-tune LLMs with component level contextual understanding, enabling it to perform 2x better, vs non-fined tuned LLMs, when used for predictive maintenance related use-cases[9].

Along the manufacturing value chain, in the short term, this could lead to increased adoption in entire operations (both pre-production and production stages), marketing & sales, logistics, and after- sales support.

For pre-production stages, this could mean generation of optimal design combinations, via automatic choice of the critical design parameters or constraints, with no human inputs. Whereas for production stage, this could mean a dynamic operational instruction generation as per current operator efficiency[10].

Short-term innovations in this capability also point to utility in the logistics stage of the value chain, from functions including the generation of dispatch documents, packaging, and routing of finished products. Gen AI also offers short-term opportunities in post-production functions including marketing and sales through the generation of targeted customer insights.

As this capability and its adoption in this sector matures, manufacturers will see Gen AI enable them to achieve incremental value from their smart factory initiatives.

3.2 Value Chain Use Cases

Operations: Gen AI can be adopted by manufacturing companies to streamline and optimize various stages of the production process, in a more supplemental rather than supportive manner as with the early adopter's stage.

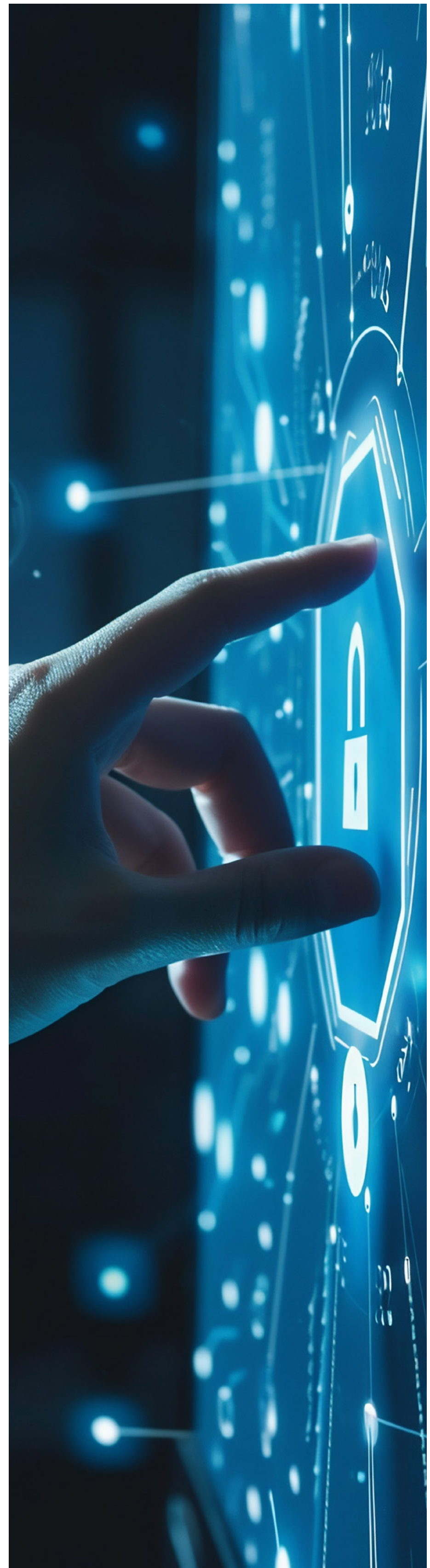
In the context of pre-production phase, there is ongoing research towards text-to-3D modelling capability of LLMs[11]. By simply describing a product concept via single or chain of prompts, Gen AI models will be able to generate accurate 3D models, significantly reducing the time and cost associated with traditional design methods. This allows designers to quickly iterate on designs and explore different options based on different variations of textual prompts.

In the production phase, Gen AI can be used to optimize manufacturing schedules. Customized SLMs can generate optimal daily or weekly schedules that minimize downtime, reduce costs, and ensure timely delivery of products by incorporating fine-tuning based on a manufacturing plant's historical data, manufacturing processes, bill-of-materials, assembly line capacity, and demand forecasts. Additionally, SLMs can generate detailed operational instructions for manufacturing workers, improving efficiency and reducing errors. These instructions can be tailored to individual workers' skill levels and experience, ensuring that they have the information they need to perform their tasks effectively[10].

Marketing and Sales: By analyzing vast datasets of historical sales data, market trends, competitor pricing, and customer behavior, SLMs or LLMs can provide valuable insights to inform pricing strategies. It can also help study the impacts of upstream raw materials costs, in order to recommend strategic pricing considerations for contracts. By continuously analyzing market conditions, Gen AI models can identify optimal pricing points to maximize revenue and profitability.

This capability can also be used to create highly effective marketing campaigns. By analyzing customer data, preferences, and behaviors, Gen AI can generate recommendations on marketing strategies for companies, incorporating detailed product specifications[10].

After-Sales Support: At the final stages of the manufacturing process, Gen AI has the potential to replace existing human-reliant functions including training and demonstrations for team members, document generation, communicating with clients and ensuring completion before transit. As language models mature, this capability can be increasingly relied on by companies to support and outright handle interactions with customers including the receipt of products. This can also entail support with the customer's installation of products[10].



4. Long-term Adoption

The long-term adoption of Gen AI within the manufacturing industry will be influenced by the tangible value derived from early adoption, the standardization of Machine Learning Operations (MLOps), and the expansion of LLMs' capabilities. These factors will shape the industry's approach to Gen AI and pave the way for its successful integration into manufacturing processes.

4.1 Factors Influencing the Long-term Adoption

Looking ahead to 2026 and beyond, the rate of adoption of Gen AI within the manufacturing industry, as well as the implementation of various use cases at different maturity levels, will be influenced by several key factors.

01

Evaluation of Early and Short-term Adoption's Tangible Value

According to a recent report by Deloitte [3], companies across various sectors are increasingly focused on the value that Gen AI can bring and how it can align with their broader corporate strategies. It is highly likely that in 2026, companies will assess the actual value derived from different Gen AI technologies and processes, using this evaluation to create a roadmap for further investments.

02

Standardization of MLOps for Managing Custom AI Systems

Managing Machine Learning Operations (MLOps) is already a complex task due to the diverse ways in which a single end-to-end Gen AI use case can be designed and optimized [12]. Manufacturing companies will likely base their future investments on the maturity of their internal MLOps teams, which will be responsible for managing and scaling the already implemented Gen AI use cases. To assess this maturity level, standardized and measurable processes must be established. Early indications suggest that custom AI systems are likely to provide higher accuracy and value compared to single large language models [12]. Therefore, effective management of AI systems is crucial to justify continued high investments and support initiatives related to the industrial metaverse, a strategic initiative that can eventually help manufacturers integrate virtual 3D environments into their operations , through the added value of Gen AI [2].

03

Expansion of LLMs from Semantic Search to Reasoning

Significant research efforts are underway to assess and enhance the reasoning capabilities of LLMs. For example, popular platforms like arXiv see an average of 30+ publications daily on this topic [13]. This research focus has the potential to shift the use case horizon of LLMs beyond semantic search-based outcomes, allowing them to provide outputs for complex strategic problems without extensive training datasets [14].

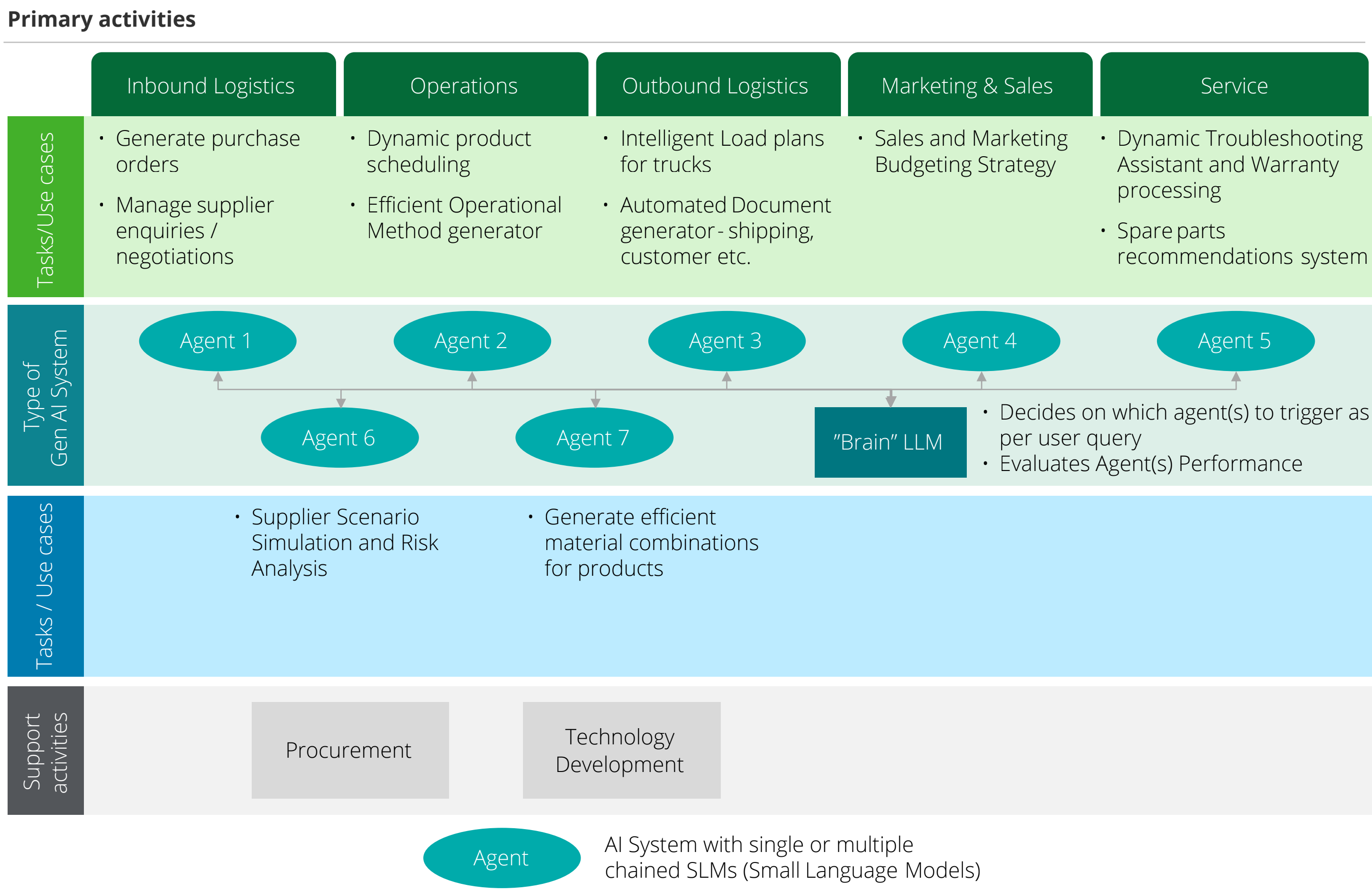
4.2 Compound Agent Systems

A Possible Long-term Adoption Use-case

Based on our findings from the Short-term Adoption (Section 3) and the trends influencing long-term adoption (Section 4.1), it is highly likely that in the long term there will be a collaborative relationship between SLMs and LLMs, leveraging the unique strengths of each to form a Compound Agent System. In this system, a single SLM will be capable of efficiently handling various activities and sub-activities within a specific business vertical or department, thanks to extensive training with domain or activity-specific datasets. Each of these individual SLMs can be labelled as an "Agent".

On the other hand, an LLM, with its advanced reasoning capabilities, will be equipped to make tactical execution decisions across the value chain. Acting as a central "Brain", the LLM will determine which "Agent(s)" to invoke and additionally monitor and evaluate their output quality. Below scenario exemplifies the functioning of a Compound Agent system, wherein a group of "Agents" collaboratively execute multiple digital or analytical activities across a typical manufacturing value chain.

4.2 Sample Use Case – A Compound Agent system impacting across the Manufacturing Value Chain (~ 10 use cases executed by one Compound Agent System)



5. Conclusion

This PoV has explored how GenAI is enabling to improvements in the manufacturing industry with real-world examples, including how companies are already using GenAI to enhance product design, improve quality control, and boost production efficiency. In the near future, the focus will likely shift to refining AI models for more specialized tasks, while the long-term vision involves creating integrated AI systems that can influence multiple stages of the manufacturing process.

To truly tap into these opportunities, it's crucial to build a strong foundation centered on data, cloud technologies, and advanced analytics, all while ensuring that people, processes, and technology are aligned.

Starting with the Foundations

- **Data Infrastructure**: The first step is to establish a robust data infrastructure. This involves conducting a thorough audit of all available data sources—like sensors on equipment, machine logs, and enterprise systems—and making sure this data is accurate and accessible. Implementing good data governance practices will help maintain data quality and address privacy and security concerns. For example, if you're aiming to predict equipment failures before they happen, you'll need reliable historical data on how your machines perform over time.

- **Cloud Computing**: Adopting cloud platforms can provide the scalability and computing power needed to run complex AI models without hefty investments in physical hardware. Cloud services from providers such as AWS, Azure, or Google Cloud offer specialized tools for AI and machine learning, making it easier to process large datasets and deploy AI solutions efficiently.

- **Advanced Analytics and AI Tools**: Utilizing advanced analytics and AI frameworks is key to turning raw data into actionable insights. Investing in machine learning algorithms, neural networks, and natural language processing models can help tailor AI applications to meet specific manufacturing needs.

Aligning People, Processes, and Technology

- **People**: The success of any AI initiative heavily depends on the people involved. Upskilling your current workforce and bringing in new talent with expertise in AI and data science are important steps. Offering training programs and fostering a culture that encourages continuous learning can empower your team to effectively use AI tools and contribute to ongoing innovation. This includes empowering team members to feel confident in asking the right questions to understand the organization's priorities when it comes to Gen AI.
- **Process**: Integrating GenAI into your operations might require rethinking existing workflows. It's important to identify where AI can add the most value—be it in optimizing designs, predicting maintenance needs, or enhancing quality checks. Establishing clear processes for deploying and monitoring AI models will help ensure they operate reliably and consistently across your organization.
- **Technology**: Keeping up with the fast-paced advancements in AI technology is essential. This means adopting the latest AI frameworks, leveraging cloud-based services, and exploring how AI can work in tandem with other emerging technologies like the Internet of Things (IoT) and edge computing. For instance, combining IoT sensors with AI models can enable real-time monitoring and predictive analytics, making your operations more agile and efficient.

The potential of GenAI to transform the manufacturing industry is immense; however, addressing safety concerns is crucial. The integration of AI technologies into manufacturing processes introduces new risks that must be managed to ensure safe and reliable operations. Creating a culture that prioritizes safety is fundamental to the successful adoption of AI in manufacturing. This involves training employees in the safe use of AI systems, encouraging them to report safety concerns, and promoting a proactive approach to identifying and mitigating risks. Leadership should emphasize the importance of safety in all AI initiatives and ensure that it is a core value throughout the organization.

Embarking on the journey with GenAI is both exciting and challenging. By focusing on building a solid foundation in data and technology, and by aligning your people and processes, you can position your organization to fully harness the potential of AI. This isn't just about implementing new tools—it's about creating an ecosystem where technology enhances human expertise, leading to smarter, more efficient, and more sustainable manufacturing practices.

6. References

1 Worldwide Spending on Artificial Intelligence Forecast to Reach $632 Billion in 2028, According to a New IDC Spending Guide. (2024). IDC: The Premier Global Market Intelligence Company. https://www.idc.com/getdoc.jsp?containerId=prUS52530724

2 Exploring the industrial metaverse. (2023, September 12). Deloitte Insights; Deloitte. https://www2.deloitte.com/us/en/insights/industry/manufacturing/industrial-metaverse-applications-smart-factory.html?id=us:2sm:3li:4diUS176121:5awa:6di:MMDDYY:author&pkid=1011188#final

3 Deloitte's State of Generative AI in the Enterprise Quarter three report. (2024). https://www2.deloitte.com/content/dam/Deloitte/us/Documents/consulting/us-state-of-gen-ai-q3.pdf

4 Robles, A. (2024, August 26). Fine-Tuning Small Language Models: Cost-Effective Performance for Business use cases. Encora. https://www.encora.com/insights/fine-tuning-small-language-models-cost-effective-performance-for-business-use-cases

5 Woodie, A. (2024, February 6). Data quality top obstacle to GENAI, informatica survey says. BigDATAwire. https://www.bigdatawire.com/2024/02/05/data-quality-top-obstacle-to-genai-informatica-survey-says/

6 Panorama Consulting Group. (2024, October 28). The most coveted external data sources for business Decision-Making. Panorama Consulting Group. https://www.panorama-consulting.com/the-most-coveted-external-data-sources-among-ceos/

7 Toyota Research Institute Unveils New Generative AI Technique for Vehicle Design. (2024). Toyota. https://pressroom.toyota.com/toyota-research-institute-unveils-new-generative-aitechnique-for-vehicle-design/

8 Generative AI in manufacturing — out of the old, emerges the new. (2024). Bosch. https://www.bosch.com/stories/ai-image-recognition-production/

9 Wang, P., Karigiannis, J., & Gao, R. X. (2024). Ontology-integrated tuning of large language model for intelligent maintenance. CIRP Annals. https://doi.org/10.1016/j.cirp.2024.04.012

10 Bansal, S. & Kulkarni, N.D. Review of Generative Artificial Intelligence Use Cases Applicable to Manufacturing Industry. (2024). International Journal of Science and Research (IJSR). https://www.ijsr.net/archive/v13i1/SR24118230026.pdf.

11 Rios, T., Menzel, S., & Sendhoff, B. (2023, July 3). Large language and Text-to-3D models for engineering design optimization. arXiv.org. https://arxiv.org/abs/2307.01230

12 Seita, D., & Ghodsi, M. Z., Omar Khattab, Lingjiao Chen, Jared Quincy Davis, Heather Miller, Chris Potts, James Zou, Michael Carbin, Jonathan Frankle, Naveen Rao, Ali. (n.d.). The Shift from Models to Compound AI Systems. The Berkeley Artificial Intelligence Research Blog. https://bair.berkeley.edu/blog/2024/02/18/compound-ai-systems/

13 Artificial Intelligence authors/titles "new.AI." (n.d.). Arxiv.org. https://arxiv.org/list/cs.AI/new

14 Hao, S., Gu, Y., Ma, H., Jiahua, J., Wang, Z., Daisy, Wang, Z., Hu, Z., Uc, & Diego, S. (2023). Reasoning with Language Model is Planning with World Model (pp. 8154–8173). https://aclanthology.org/2023.emnlp-main.507.pdf

Contact

Francis Lemieux
Partner
Montreal
frlemieux@deloitte.ca

Anthony Hamer
Partner
Toronto
anhamer@deloitte.ca

Gerald Faustino
Partner
Montreal
gfaustino@deloitte.ca

Vivek Wadhwana
Senior Manager
Toronto
vwadhwana@deloitte.ca

Valerie Judd-Rohal
Senior Manager
Montreal
vjuddrohal@deloitte.ca

Contributors

Gunpreet Rekhi
Manager
Toronto
grekhi@deloitte.ca

Abhishek Kumar
Senior Consultant
Toronto
abhishekxyzkumar@deloitte.ca

Harsh Raj
Senior Consultant
Toronto
harraj@deloitte.ca

Sebastian Voermann
Consultant
Calgary
svoermann@deloitte.ca

Deloitte.

MAKING AN IMPACT THAT MATTERS since 1845