

Grid Dynamics white paper

Automotive industry leveraging GenAI and LLMS

www.griddynamics.com



Grid Dynamics

Contents

Introduction	3
Hybrid search	5
Language-based assistants	7
LLMOps	8
About Grid Dynamics	10

Introduction

The integration of Generative AI (GenAI) and Large Language Models (LLMs) is ushering in a new era for the Automotive Industry. By leveraging the immense potential of GenAI and LLMs, automakers can harness vast amounts of data to drive advancements in autonomous vehicles, energy efficiency, and safety standards. The synergy between Artificial Intelligence (AI) and natural language processing is reshaping the Automotive sector, paving the way for unprecedented innovation and efficiency.

On November 30, 2022, OpenAI launched ChatGPT and stunned the public. However, the transformer architecture that powers ChatGPT and other LLMs had been introduced 5 years earlier in 2017, and the first Generative Pre-trained Transformer (GPT) model was released in 2018. Since then, the transformer architecture has become the dominant network architecture for LLMs and other natural language processing tasks, at least for the time being.



In the Automotive Industry, where a vast amount of knowledge resides across the trinity of Product Lifecycle Management (PLM), Enterprise Resource Planning (ERP), and Manufacturing Execution Systems (MES), the interest in LLMS and how to leverage them have been key topics from 2023 onwards.

In the domain of LLMS are Foundation Models (FMs), developed by a handful of research labs, and designed to incorporate an extensive breadth of human knowledge across various domains. While the emergence of transformers and FMs brings inherent risks associated with any new technology, organizations have been actively collaborating with Grid Dynamics to explore how these powerful language models can drive operational excellence across various aspects of their businesses. The applications are broad and span accelerating digital transformation and adding value, improving productivity and supply chains.

The Automotive Industry is witnessing a paradigm shift with the integration of GenAI and LLMS, leading to a plethora of innovative use cases. One prominent application lies in personalized customer interactions, where GenAI-powered chatbots and virtual assistants offer tailored assistance, enhancing the overall customer experience [1].

Moreover, LLMS facilitate seamless communication between vehicles and their users, enabling intuitive voice commands and natural language understanding [2]. Beyond customer interactions, these technologies revolutionize design processes, allowing for rapid prototyping and simulation through natural language input, thereby accelerating product development cycles [3]. Additionally, predictive maintenance emerges as a critical domain benefiting from GenAI and LLMS, as they analyze vast datasets to forecast potential issues, optimizing vehicle performance and reliability [4].

LLMS are opening new possibilities for digital product and enterprise capabilities in the Automotive Industry. One of the unique strengths of LLMS lies in their ability to summarize text and generate content, making them particularly valuable for processing unstructured data. Some 80% of all new data in enterprises is unstructured [5]. In this article, we single out two specific capabilities: hybrid search and language-based assistants.

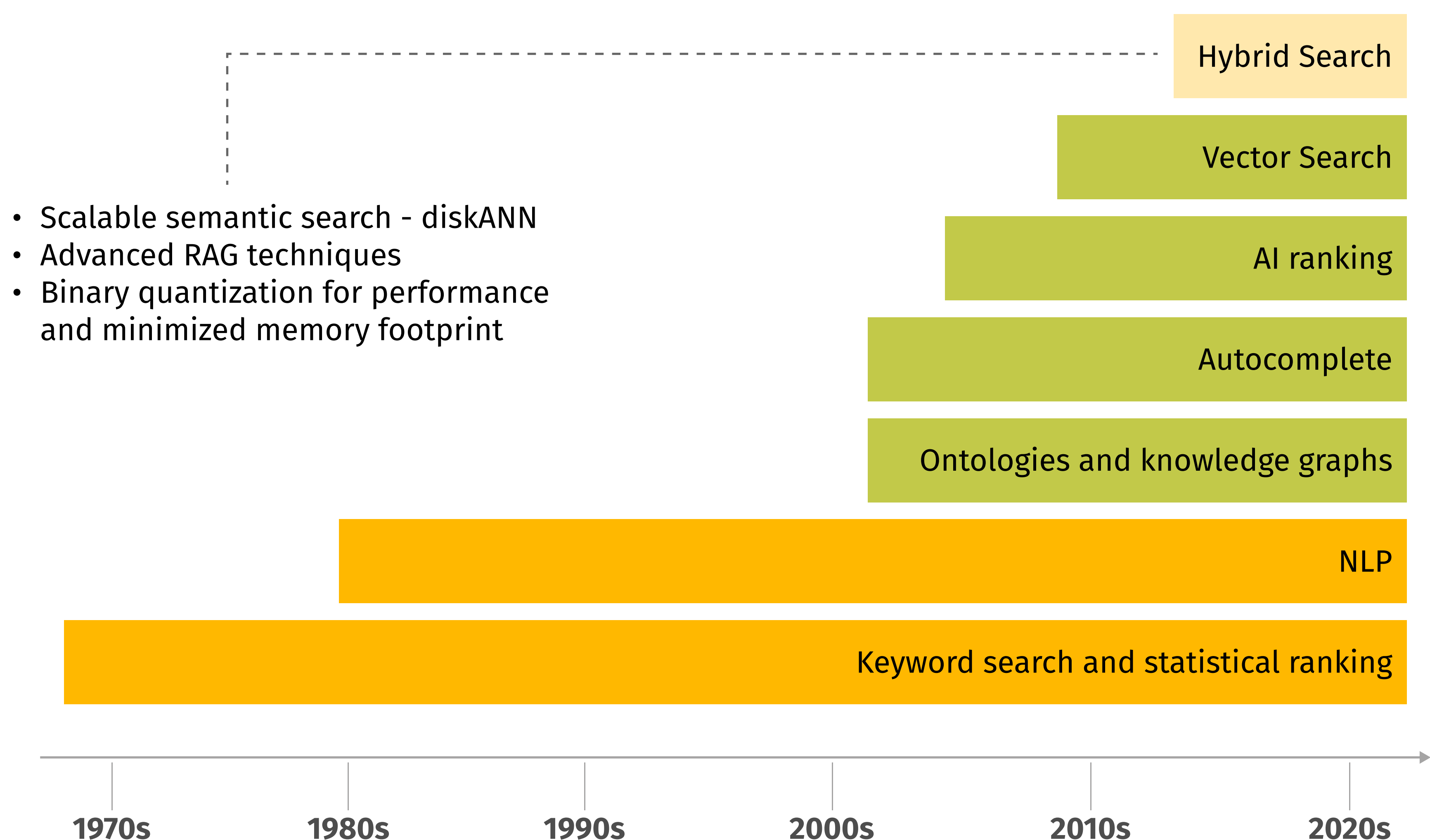


Hybrid search

Grid Dynamics has extensive experience in data-intensive systems, stemming from its early work with space-based architectures and technologies such as GigaSpaces, and later with ElasticSearch clusters powering faceted search on high-traffic websites. As search experts, we are now witnessing an evolution of these secondary databases. The traditional approach of lexical techniques is being expanded to handle embeddings or vector representations of language that can be understood by LLMs. This new representation has led to new types of databases that support the semantic algorithms for search. The combination of lexical and semantic techniques provides hybrid search capabilities, enabling use cases such as Retrieval-Augmented Generation (RAG) [6].

In RAG, the search component retrieves relevant information from a corpus, and the language model augments this retrieved information to generate a comprehensive response. This hybrid approach leverages the strengths of both traditional search and LLMs, enabling more accurate and contextually relevant search results and content generation.

Traditional search relies on the ability to identify and match specific keywords associated with the desired knowledge or information. However, these methods often fall short when semantically similar keywords or phrases are used instead of the exact terms. These limitations can lead to tribal knowledge issues, where users are unaware of the right way to formulate queries or lack the specific terminology.





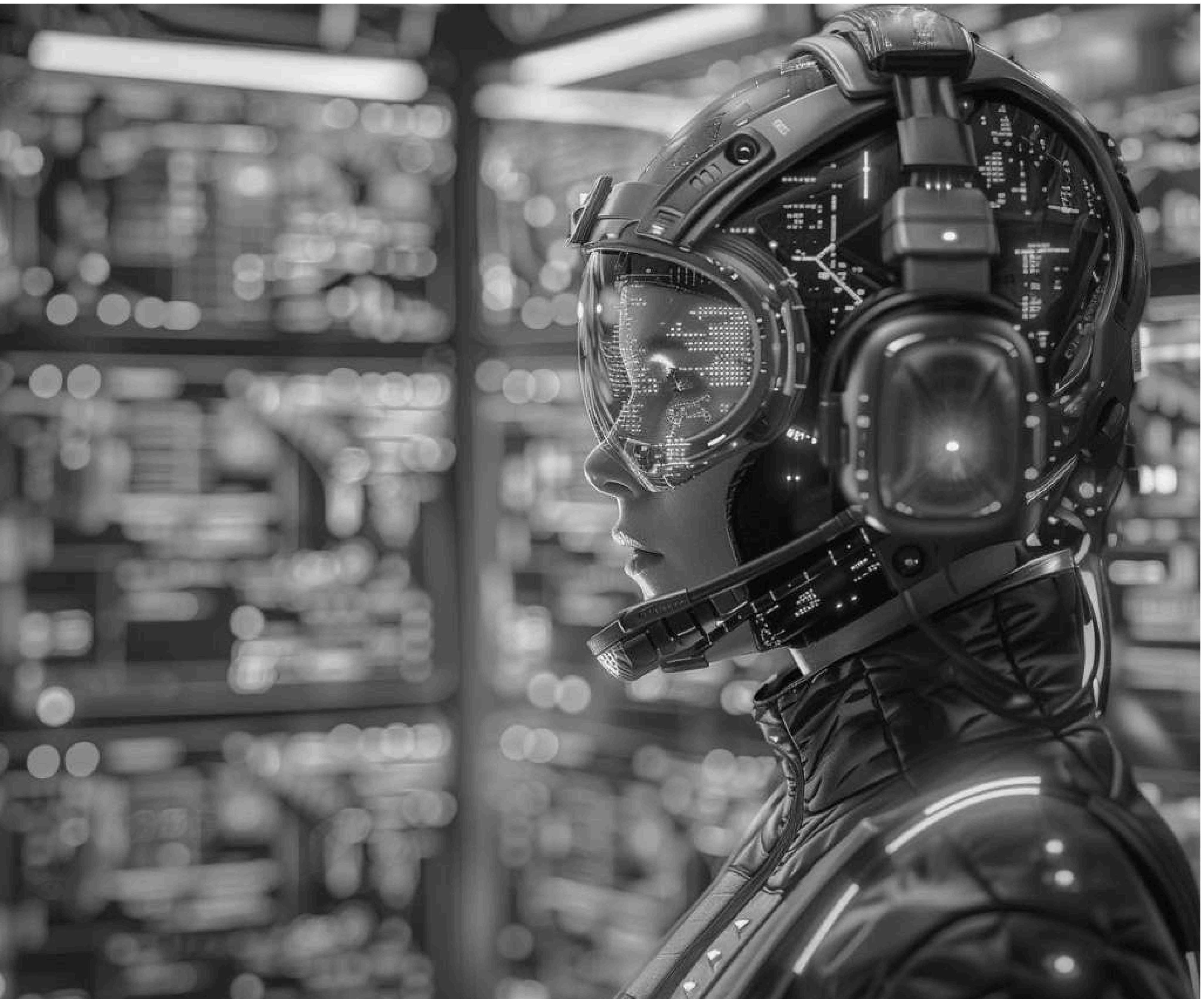
With the advent of hybrid search and the combination of lexical and semantic techniques, new types of databases have emerged to support the semantic algorithms required for this approach. The table on the previous page shows some of the vector databases we have come across in our discussions related to hybrid search capabilities.

The market for these vector databases is evolving, with established storage products adding requisite capabilities. By following a standardized lifecycle management process, downstream teams can align with and adopt these emerging technologies as they mature and prove their value in enabling hybrid search and related use cases.

Language-based assistants

In addition to hybrid search, the second capability that can extend digital solutions is the use of language-based assistants. This capability further expands the boundaries of what can be delegated to language models, and what aspects should be handled by traditional orchestration of services composed to achieve a particular goal. We see this at the very center of the language-based assistant stack.

Orchestration is essential as teams explore which generalization tasks can be delegated to a particular LLM, and which tasks require bespoke solutions. Two orchestration flows gaining traction are Chain of Thought and ReAct.

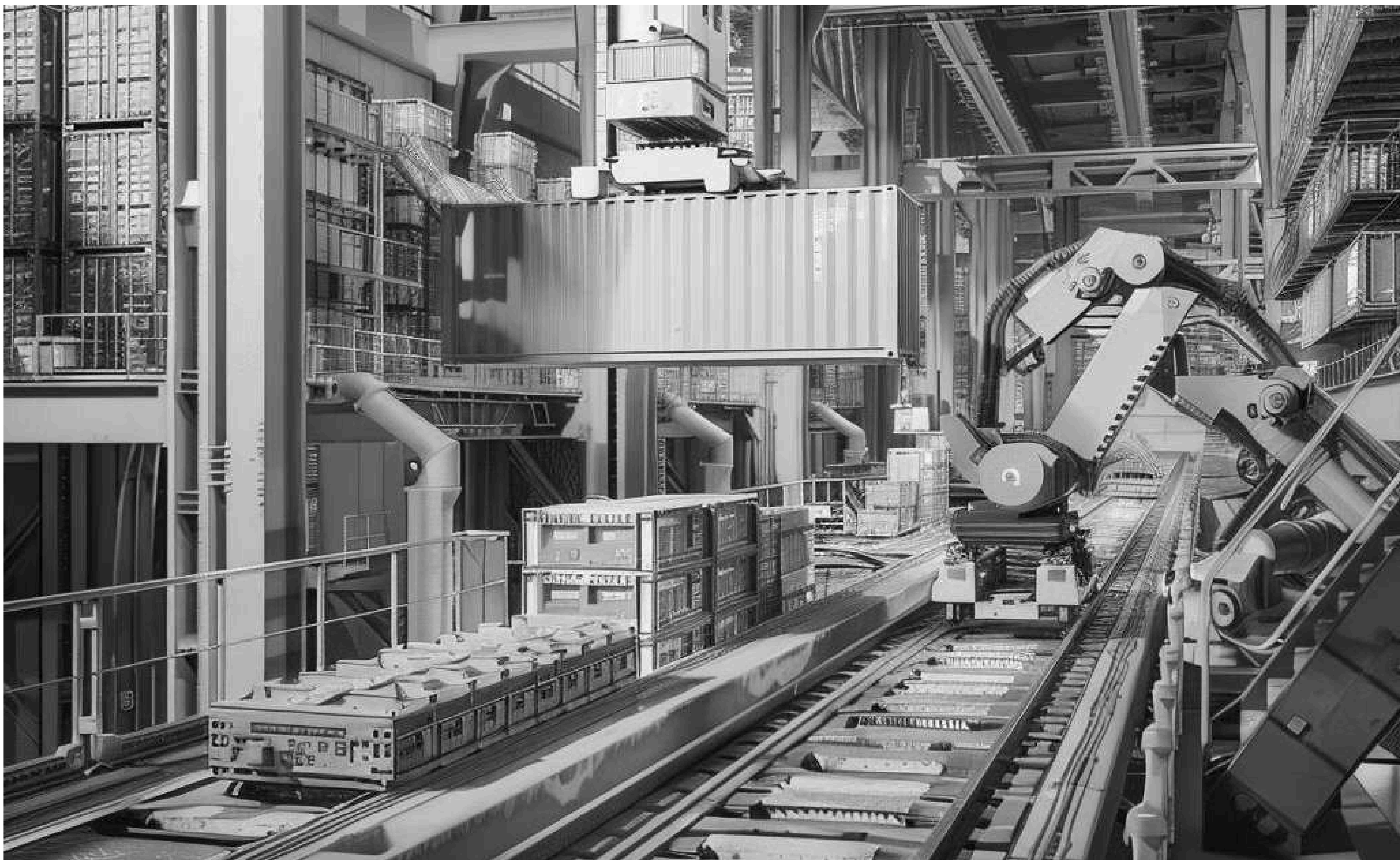


LLMOps

As with other data-intensive solutions in the applied machine learning space, platform aspects and the ability to sustain a delivery cadence with robust DORA metrics are essential. Grid Dynamics has outlined the functional and technical designs of an LLM platform in its blog, LLMOps blueprint for closed-source large language models [7]. We have specifically examined closed-source LLMs, and provided mappings to well-established and emerging frameworks, libraries, and products that can be used for implementation.

The ability to interact with data-intensive applications using natural language is a natural progression for modernizing many systems that were developed one or two decades ago. At Grid Dynamics, we have evolved from data-intensive systems for power users to wizards designed with UX concerns in mind. Implementing robust LLMOps practices, which encompass the processes, tools, and best practices for managing the entire life-cycle of LLM-powered applications, is crucial for organizations seeking to leverage these powerful models effectively and sustainably.

GenAI and LLMs undeniably mark a turning point in industrial services, paving the way for the future of the Automotive and Manufacturing industries. This powerful convergence of technologies represents a paradigm shift that will redefine boundaries and propel these sectors into uncharted territories of innovation.





Sources:

- [1] [Four ways conversational AI transforms digital experiences](#)
- [2] [Insights by Grid Dynamics](#)
- [3] [Transform your product design processes and personalization services with generative AI](#)
- [4] [From design to delivery: The role of artificial intelligence in the automotive industry](#)
- [5] [Why Unstructured Data Is Your Organization's Best-Kept Secret – GeekWire](#)
- [6] [RAG and LLM business process automation: A technical strategy](#)
- [7] [LLMOps blueprint for closed-source large language models](#)

About Grid Dynamics

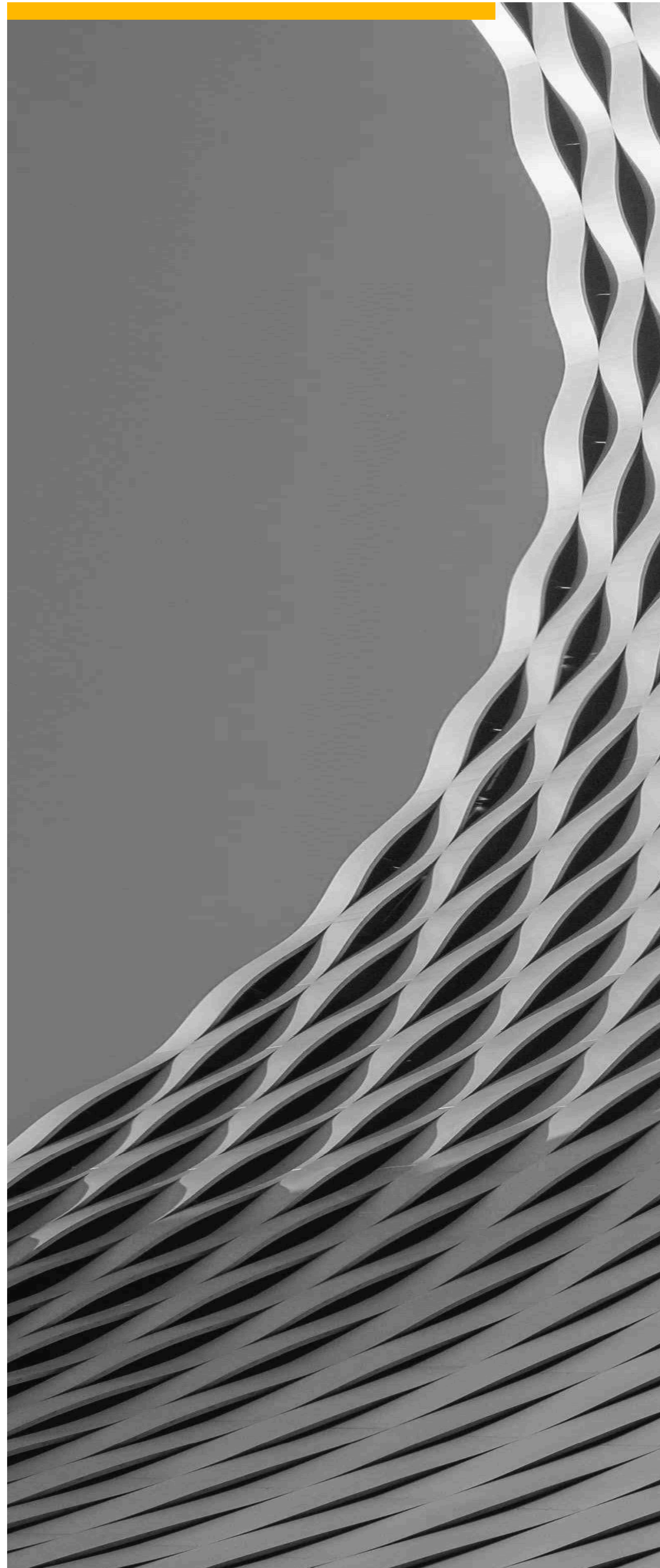
Grid Dynamics empowers enterprises to drive growth, boost efficiency, and transform their digital capabilities through expert technology consulting, platform engineering, AI, and advanced analytics.

Fusing technical vision with business acumen, Grid Dynamics (NASDAQ: GDYN) fuels a relentless drive toward sustainable digital transformation. As a forefront provider of technology consulting, platform and product engineering services, and bespoke software development, we draw from over 8 years of leadership in Enterprise AI, coupled with profound expertise in cloud, data, and advanced analytics.

Our commitment to engineering excellence, R&D leadership, a co-innovation ethos, globally efficient follow-the-sun delivery model, and an unwavering "whatever it takes" dedication to client success empower us to solve even the most complex enterprise challenges, ensuring profitable business outcomes and future-proof growth.

Founded in 2006, Grid Dynamics is headquartered in Silicon Valley and has a global talent pool of intellectually curious problem solvers in offices across the US, Mexico, Jamaica, the UK, Europe, and India.

LEARN MORE AT [GRIDDYNAMICS.COM](https://griddynamics.com) →





Grid Dynamics

trusted engineering partner for digital transformation

Grid Dynamics Holdings, Inc.

5000 Executive Parkway,
Suite 520 / San Ramon, CA
650-523-5000
www.griddynamics.com