

# How Generative AI Is Changing How We Think Analytics?



Structured enterprise data may just be generative AI's next breakthrough area. We asked Nima Negahban, cofounder and CEO of Kinetica, to explain how the technology is being used in a variety of industries.

Generative AI tools such as ChatGPT are not limited to language applications alone; they can also be effectively applied to structured enterprise data. By integrating LLMs with structured enterprise data, organizations are beginning to ask new questions of their data and get answers immediately. This new paradigm is disrupting how enterprises think about managing and analyzing their data.

**Upside:** What are some generative AI use cases

from the field across healthcare, automotive, telecommunications, and the military?

**Nima Negahban:** Vector similarity search is a powerful generative AI technique that can be applied to tabular data to uncover new insights and patterns. By transforming tabular data into vectors, where each column or attribute represents a dimension in the vector space, we can measure the similarity between different patterns using distance metrics such as cosine similarity or Euclidean distance.

For the automotive industry, vector similarity search can be applied to analyze sensor data from vehicles, allowing auto manufacturers to detect

anomalies or patterns in vehicle performance, predict potential failures, and optimize maintenance schedules. It can also be used in autonomous vehicles to identify similar driving scenarios and enhance safety measures.

In the telecommunications sector, vector similarity search can aid in network optimization by comparing real-time data from network elements, such as routers and switches, with historical performance metrics. This can lead to improved network reliability, faster troubleshooting, and better overall user experience.

In defense, vector similarity search can support intelligence analysis by identifying patterns in data from various sources, including satellite imagery, radar, assets equipped with GPS, drone footage, and other sensor networks. This can assist in detecting potential threats, predicting enemy movements, and providing enhanced situational awareness for war fighters.

**How has the role of the GPU (graphics processing unit) expanded beyond neural networks to a broader variety of analytics?**

GPUs have proven to be instrumental in advancing generative AI beyond building machine and deep learning models. Their highly parallel architecture and massive computational power make them particularly well-suited for a variety of tasks involving intense mathematical operations.

In time-series analytics, for example, GPUs can

simultaneously process vast amounts of historical and real-time data where you have to compare different windows in time, making them ideal for tasks such as forecasting, anomaly detection, and pattern recognition. GPUs accelerate geospatial data processing and visualization, enabling tasks such as tracking objects and route optimization, making GPUs invaluable in urban planning, logistics, and environmental monitoring.

In graph analytics, which involves traversing and analyzing complex networks, GPUs provide a substantial advantage with their ability to perform parallel graph operations efficiently. Vector search analytics, where similarity search is performed across high-dimensional data vectors, benefits greatly from GPUs' capability to process large vector data sets using vectorized operations.

**How has generative AI brought about the end of data pipelines and tedious data engineering as a prerequisite to asking questions?**

Data pipelines are widely employed to address the performance limitations of databases, especially when dealing with large-scale or complex data sets. Databases, although efficient for storing and managing data, might struggle to handle the computational demands of complex data transformations, analytics, and machine learning tasks. Data pipelines act as intermediaries for offloading resource-intensive operations.

However, the downside is that questions must be

known in advance, and then pipelines are built over days or weeks, which then become part of the organization's tech debt. Building pipelines ultimately flies in the face of what people like about generative AI -- the ability to ask original questions and get the answers fast.

The recent breakthroughs in generative AI are enabled in large part by the GPU, which vectorizes operations. Vectorization allows for simpler data structures and reduces the need for building complex data pipelines by enabling efficient bulk operations on data arrays. With vectorized operations, data analysts can perform computations and transformations on entire arrays or matrices, eliminating the need to pre-process the data in advance. This leads to more concise and streamlined data structures.

Vectorization is enabled on NVIDIA GPUs through CUDA, which allows parallel code using threads and blocks, exploiting the GPU's massive parallel processing power to execute vectorized operations efficiently. Intel's Advanced Vector Extensions enables vectorization on their CPUs by supporting 512-bit vector registers and providing a wide range of SIMD (single instruction, multiple data) instructions, allowing for extensive parallelism and high-performance computation on compatible processors. Note that although pretty much all distributed analytics databases use SIMD to a degree, almost none of them vectorizes all of their

operations, meaning unless you use a natively vectorized database running on CPUs, you'll still need to resort to data pipelines.

### **Is generative AI any good at looking at structured data, or just the unstructured kind?**

Generative AI produces original output by learning patterns and generating novel outputs, as opposed to pre-programmed responses. It is often associated with generating content such as images, text, and videos, but generative AI techniques can also be applied to structured data to generate new data points that follow similar patterns or distributions as the existing data.

For example, in predictive maintenance for connected cars, generative AI leverages historical sensor data to not only identify patterns of potential faults but also discover novel insights that could be challenging or time-consuming to detect using traditional machine learning methods. The ability to generate diverse and representative data points enables the model to uncover hidden correlations and unique failure patterns. The data in this example is derived from car sensors that are highly structured.

### **What are the opportunities for using generative AI in conjunction with time-series, spatial, and graph analytics?**

By integrating generative AI with time-series, spatial, and graph analytics, organizations can uncover novel insights, improve predictive

accuracy, enhance data quality, and develop innovative solutions across a wide range of applications.

Generative models can be employed to detect anomalies in sensor and machine data. Any deviation from learned patterns can be flagged as anomalies, facilitating early detection of abnormalities or fraudulent activities. Combining time-series, spatial, and graph analysis with generative models can improve predictive maintenance strategies or hyper-accurate accident recreation by generating synthetic fault scenarios and simulating system behaviors under different conditions.

Generative models can be leveraged to impute missing data points in a time-series or spatial data set, ensuring continuity and enhancing the quality of subsequent analyses. Time-series generative models can predict future data points based on historical patterns, enabling better forecasting and decision-making. Similarly, in spatial analytics, generative models can simulate realistic scenarios, useful for urban planning, traffic optimization, and disaster management.

#### **What is your perspective on the proliferation of LLMs beyond ChatGPT?**

The proliferation of LLMs is a positive development for several reasons. Competition among various LLMs will drive innovation, leading to improvements in performance, accuracy, competitive pricing, and

better service offerings.

Different LLMs might be optimized for specific use cases or industries such as healthcare, finance, or telecommunications. Having multiple LLM options allows organizations to select the one that best aligns with their specific requirements and offers customization options to adapt to their environment. We have several clients that are driving us to support other LLMs like AWS' Bedrock and NVIDIA's NeMo LLMs for these reasons. Ultimately, customers will value choice in this increasingly diverse landscape of LLMs.

[Editor's note: Nima Negahban is Kinetica's chief executive officer. Early in his career, he pioneered advances in real-time analytics to find and track terrorists for the Department of Defense. This inspired Kinetica, which he co-founded and where he was the original developer of the Kinetica database that used GPUs to crunch large volumes of data faster and more efficiently. He continued to evolve Kinetica into a new class of distributed, vectorized, SQL database that incorporates graph, spatial and time-series analytics to provide instant results to questions that span multiple real-time data sets. Kinetica is used by leading telecommunications, automotive, financial services, logistics, and defense agencies. Negahban holds a B.S. in computer science from the University of Maryland.]