

# AI-Empowered Health Analytics

Gen AI, LLM's, and higher-quality data are transforming the future of healthcare today.

eBOOK





Before the recent explosion of Large Language Models (LLMs), other technologies under the “AI” umbrella were used in Healthcare for a growing number of use cases, from imaging diagnostics to operations to drug discovery. Generative AI (Gen AI) and LLMs, in particular, are poised to further this trend into previously unaddressable use cases. This ebook shares Kythera Labs’ insights on how cutting-edge AI technologies like Gen AI and LLMs can be leveraged to unlock valuable insights from Healthcare data. Here’s what we cover:

- AI technologies and applications
- Potential challenges and risks
- Building a Healthcare data analytics LLM
- Using high quality data and selecting the right use case

LLMs are perfectly suited for generating impressive and often amazing natural language outputs as a response to natural language prompts. Many of us are already using AI applications like ChatGPT, Gemini, Perplexity, or even Gmail’s Smart Compose. These tools are like virtual assistants, helping with basic research and content creation. Unlike traditional software, they don’t require training or complex technical setups – you can use them with minimal effort. As long as people follow basic guidelines for protecting sensitive data, these hosted AI platforms can be used immediately to dramatic effect.



**Generative AI** is a branch of artificial intelligence that excels at creating original content. It achieves this by ingesting vast datasets – text, images, code, or music – and identifying the underlying patterns and relationships that connect them. This learning process, also referred to as training, allows generative AI to autonomously produce novel outputs, mimicking the style, themes, and characteristics found within its source training data.



## Large Language Models

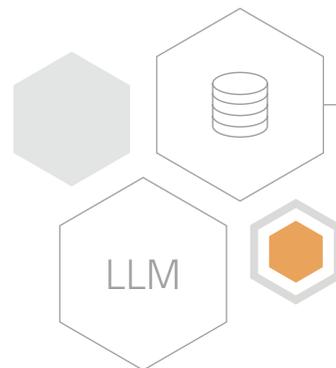
(LLMs) are a type of artificial intelligence that are adept at processing and understanding language. These models are trained by analyzing massive datasets of text allowing them to grasp the nuances of language, including grammar, syntax and cultural references. Through this extensive training, LLMs become proficient at various tasks, from translating languages to writing original content (even software code).

## A Game-Changer in Healthcare

Across industries, the real game-changer lies in using Gen AI to unlock the insights hidden within the vast amounts of data generated daily to uncover marketing intelligence and inform strategic business decisions. Within Healthcare, provider organizations have used claims data and analytics to support strategy and business intelligence for well over a decade. Deploying LLMs and Gen AI will dramatically enhance and accelerate their ability to discover answers from the data.

LLMs are particularly adept at learning patterns in language from large corpuses of documents and data, including claims datasets, then generating analytical queries of that data. With recent advancements, LLMs can now take a natural language prompt, produce the appropriate query, and deliver a natural language output without the user ever writing a line of code. Using natural language to interact with and query healthcare data empowers non-technical users to unlock insights from the data that was often only accessible to users with advanced skills in data science and analytics.

While the lure of using fast-acting hosted AI platforms, like Chat GPT and others, may be strong, it is important to recognize that they run on 3rd party architectures, requiring data (or sensitive metadata) to be supplied as part of a prompt, which can leave an organization's infrastructure open to the possibility of exposure. The terms and conditions of these sites may allow these third parties to use all the data supplied for further training and tuning of their models, making it difficult to maintain the security of sensitive data. Establishing a native capability within an organization to host LLMs privately can reduce this risk.

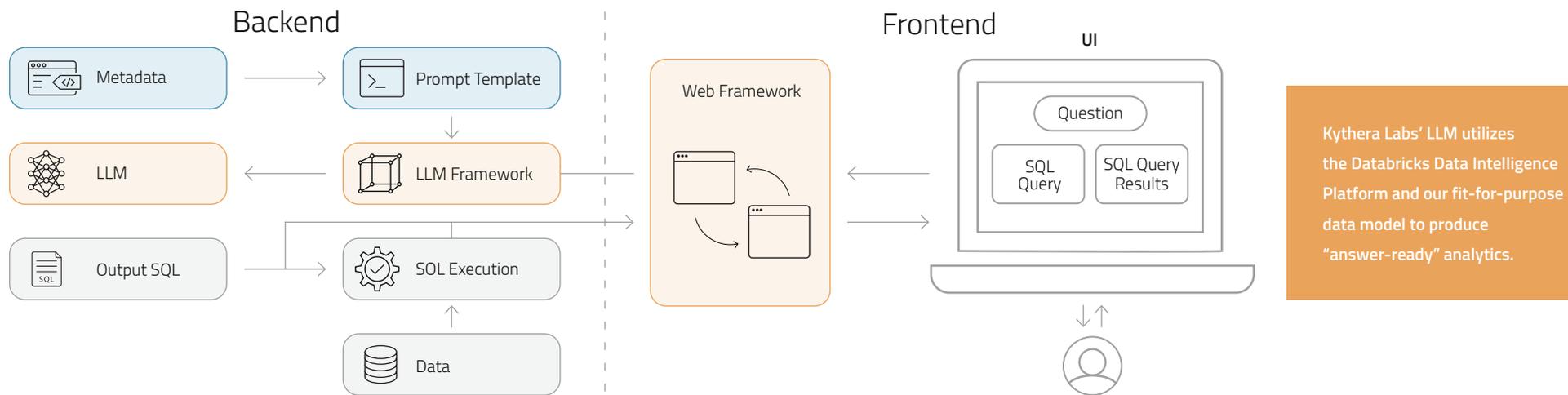


“  
Be Aware: When using third-party applications, always review terms & conditions to protect access to your data.”

# How Kythera Labs Got Started With Generative AI

To evaluate and understand the potential impact of LMMs on Healthcare data analytics, Kythera Labs built proof-of-concept LLM-powered utilities for querying healthcare datasets using our common data model and patient event datasets. We embraced this undertaking as a way to improve our internal data operations and test the hypothesis that our customers will realize better and more rapid ROI using Gen AI and LLM technology. Here's how we got started.

Rather than build our own LLM solution, Kythera partnered with [Databricks](#), leveraging their [Data Intelligence Platform](#). The Platform provides the basic architecture out of the box for hosting and tuning LLMs within an organization's own environment. However, having an appropriate architecture is just the first step. Ensuring the accuracy and interpretability of the generated results requires a tremendous amount of focus on data quality and ensuring that the data model is fit-for-purpose. Kythera Labs' history of turning petabytes of raw healthcare data into high-fidelity, use case-ready analytics provided an ideal starting point.



To produce "answer-ready" analytics, we developed a dynamic common data model that restructures and reformats data, removes incorrect information, imputes correct information, and infers the existence of missing healthcare events. The model enables us to bring together data from different sources and simplifies and organizes the data. The model works like a feedback loop that self-informs and improves the data by interacting with thousands of interrelated data points, such as demographic, genomic and directory data.

## Data Quality

For LLMs to work successfully and consistently, higher-quality data is essential. Just like the saying “garbage in, garbage out,” poor quality data can lead LLMs to generate impressive-looking but completely wrong answers. This means data needs to be accurate (reflecting reality), complete (with no missing pieces), consistent (standardized, normalized, and unified), and current (up-to-date and accessible effectively on demand).

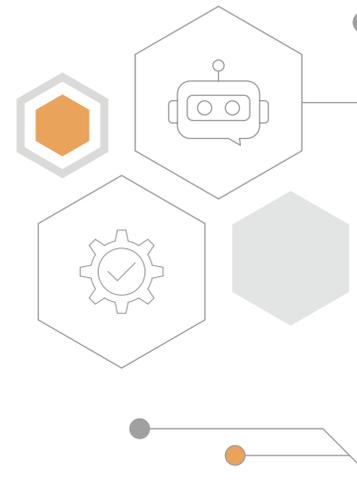
Unlocking powerful insights from data requires both high-quality and accessibility. This is crucial not just for advanced models like LLMs, but also for anyone making strategic decisions. Understanding market trends, expanding services, navigating complex systems, and staying competitive all rely on clear data analysis. By dissecting market dynamics, user needs, and overall trends, decision-makers can identify new opportunities, minimize risks, and allocate resources efficiently.

However, the path from raw data to actionable insights presents numerous challenges. Fragmented data, compatibility issues, data governance, and privacy concerns can all hinder the smooth flow and usability of this valuable information. Rigorous data quality checks, data governance frameworks, and advanced data integration and cleansing techniques can help overcome these barriers.

Investing in higher-quality data sources and data technology is critical for LLM success. Clean, accurate data powers these models to generate reliable and trustworthy answers, ultimately driving more informed decision-making across various domains faster than ever before.



A chatbot cannot turn bad data into good answers.”



## Fit-For-Purpose

Fit-for-purpose data is essential for informed decision-making, enabling evidence-based insights, and achieving desired outcomes. For LLMs, that means using data that is suitable, relevant, and appropriate. A challenge faced when developing LLMs for writing data analytics queries is that they are not yet good enough for complex reasoning or solving problems that require multi-step logic to produce a correct answer. Our findings indicate that the closer the data elements align with the concepts relevant to the analyst's query, the better the models perform in generating queries that produce accurate results. Here are some examples of relevant concepts for analyst queries:

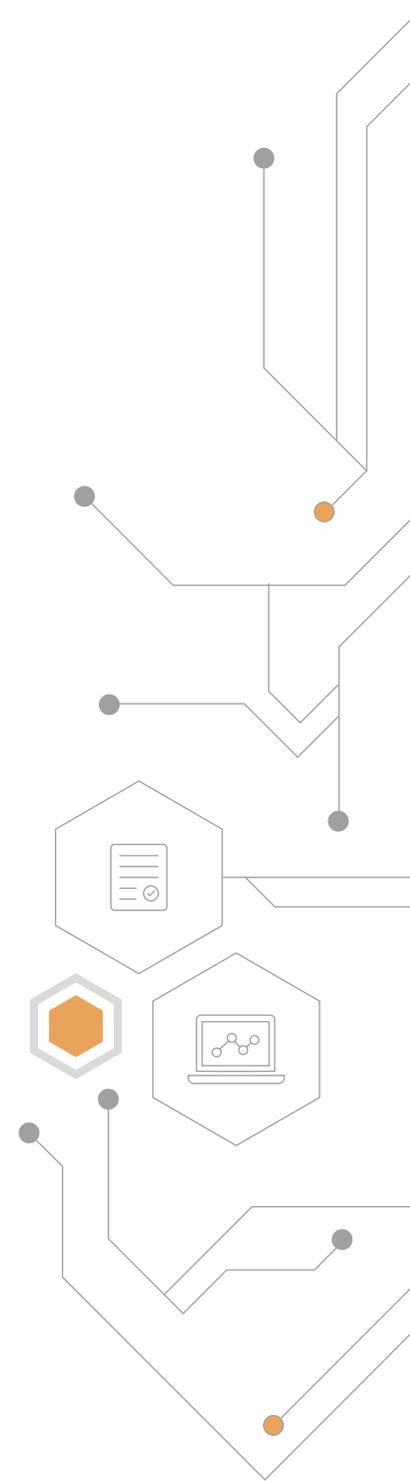
Erosion of surgical volumes from inpatient to outpatient facilities.

Referral alignment among physician practices.

Common treatment patterns for patients with chronic disease.

Data specifically structured to answer these types of questions performs far better than raw source claims data which is not structured into patient events or patient journeys. Kythera Labs' longitudinal patient event dataset curated and optimized for these types of analyses, provided an excellent starting point to train our models.

If you are considering getting into AI, you need to fully understand the data being used. First, ask yourself if you can evaluate the effectiveness and quality of the data that you will be using in your model. Also consider if the data is "fit-for-purpose." A fit-for-purpose dataset to train the model will help ensure better model success.





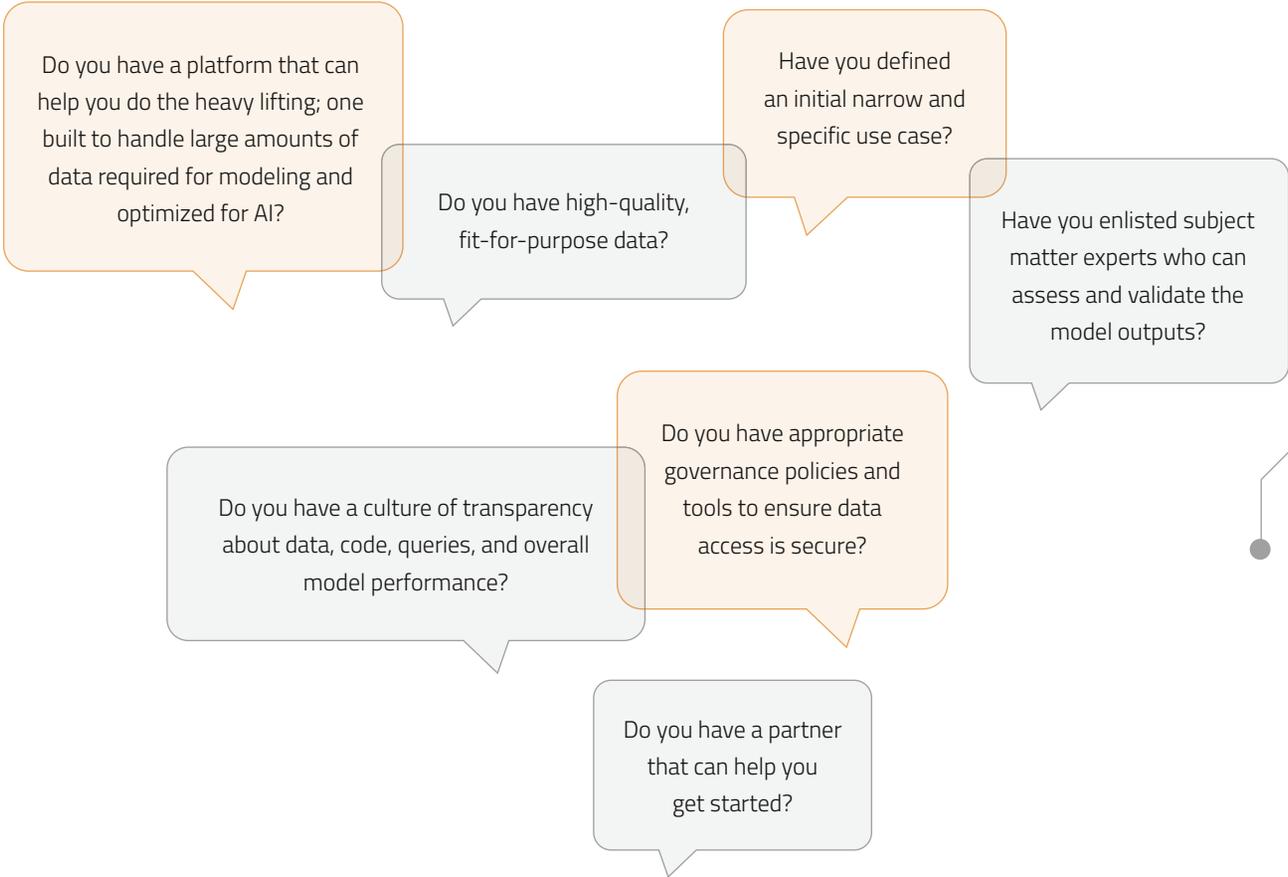
## Pick The Right Use Case

The use cases for LLMs and Gen AI are expanding, so it is imperative to keep a few key things in mind. Rather than starting with a model or specific technology, it is advisable to begin by identifying a specific problem or discrete use case. In keeping the use case narrow, you can measure model effectiveness and then understand if the model is scalable. Gen AI aids your decision-making by analyzing huge datasets; however, it still requires human expertise, intervention, and supervision. If you start too wide, you may end up with questionable outputs without knowing how the model produced them.

As Kythera Labs began developing and tuning our model, we used a small number of use cases typical of healthcare strategy and commonly faced by market growth professionals. As we monitored the models performance, we began to fine-tune and identify best practices. Our prompt engineering grew more refined based on the outputs we were seeing. Transparency about the data used, the code used to generate the answers, and how the model functions is critical for gaining acceptance and having confidence in the outputs. As we fine-tuned, we cataloged instruction sets for each use case to reduce future time to insights.

# Conclusion

The Gen AI landscape is rapidly evolving, and there are several things that you should consider right now.



The benefits of faster time to insights and more efficient use of resources for technical and non-technical staff can be achieved through the application of Gen AI for Healthcare analytics. There’s no doubt that Gen AI is already having a profound impact on the future of healthcare (and we are still in the early stages). At Kythera Labs, we are inspired by the opportunities and eager to tackle the challenges that lie ahead.



**About Kythera Labs** Kythera Labs is a data and technology company that enables Healthcare and Life Sciences organizations to rapidly integrate, access, and analyze healthcare data with scale and speed using the Wayfinder data science platform and pre-configured data pipelines, data science toolkits, and remastered datasets for competitive advantages across markets. Learn more at [www.kytheralabs.com](http://www.kytheralabs.com) or follow Kythera Labs on [LinkedIn](#).



346 Main Street  
Franklin, TN 37064

[www.kytheralabs.com](http://www.kytheralabs.com)

Copyright© 2024 Kythera Labs Inc. All rights reserved.

