

Generative AI

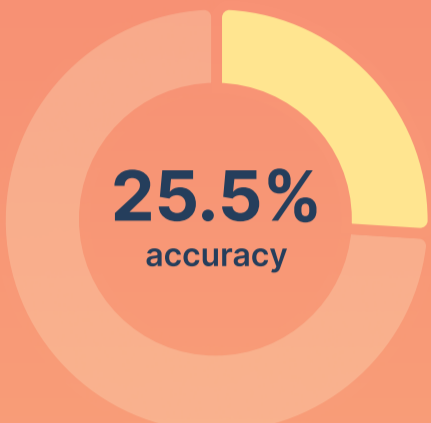
Accuracy in the Enterprise

New benchmark* examines Large Language Models (LLMs) accuracy in the enterprise and the improvements offered by a Knowledge Graph.

16.7%
average accuracy

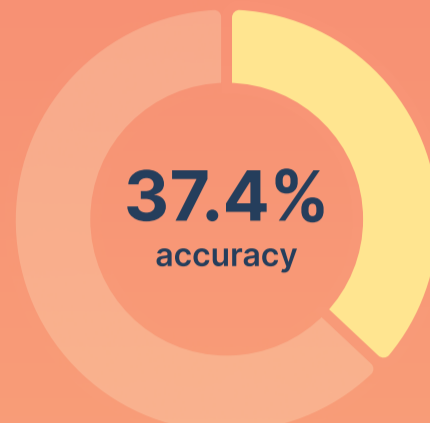
LLMs **struggle** to produce accurate responses to business questions

LLM accuracy is low for day-to-day analytics & operational questions



Day-to-day Analytics

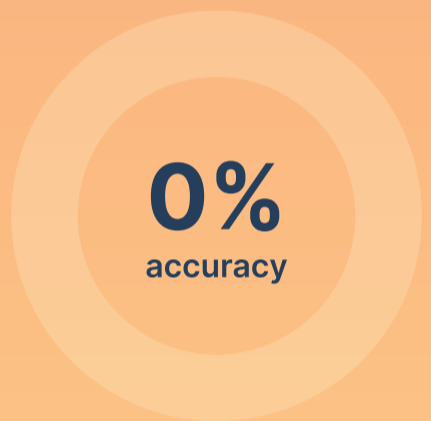
These questions are common in the day-to-day work of an analyst or business user, like "Return all the claims we have by claim number, open date, and close date."



Operational Analytics

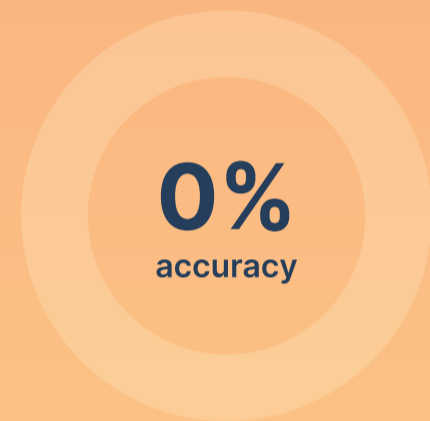
These questions are fundamental to running and optimizing the operations of the business, like "What is the average time to settle a claim policy?"

For more complex business questions, LLMs failed to return accurate answers



Metrics & KPIs

These questions are important for measuring the health of the business and success of certain initiatives, like "What are the loss payment, loss reserve, expense reserve amount by claim number?"



Strategic Planning

These questions are important for setting the course for the business, maximizing opportunities, and reducing risk, like "What is the total loss of each policy where loss is the sum of loss payment, loss revenue, expense payment, expense reserve amount?"

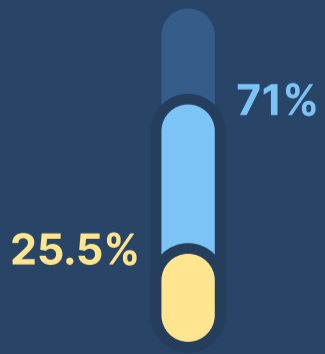
The Knowledge Graph Difference

3x
improvement

A **Knowledge Graph** improved the accuracy of LLM responses by an average of 3x.

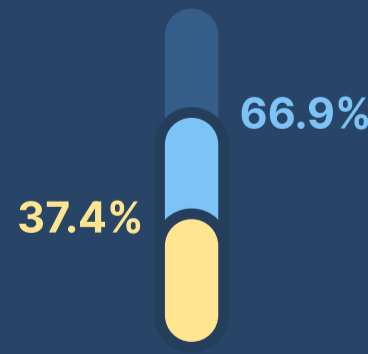
45.6%
improvement

Day-to-day Analytics



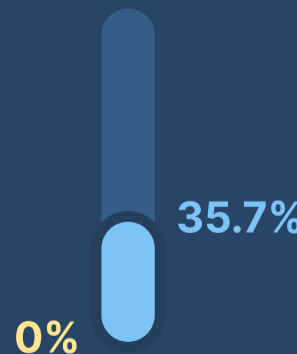
29.5%
improvement

Operational Analytics



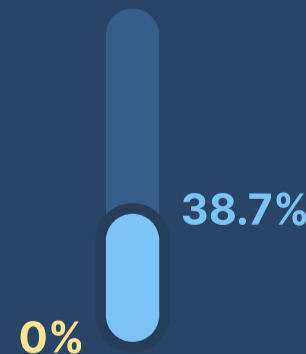
35.7%
improvement

Metrics & KPIs



38.5%
improvement

Strategic Planning



● LLM Accuracy Alone ● LLM Accuracy with a Knowledge Graph



To learn more about the AI benchmark, see the full study [here](#). And read [this whitepaper](#) to find out how a data catalog built on a Knowledge Graph can create the foundation for scalable AI.

*The benchmark uses the enterprise SQL schema from the [OMG Property and Casualty Data Model](#) in the insurance domain. Accuracy was measured using the metric of Execution Accuracy (EA) from the [Yale Spider benchmark](#). Using this schema and accuracy metric, the benchmark compared the accuracy of responses to 43 questions of varying complexity, ranging from simple operational reporting to key performance indicators (KPIs).