McKinsey & Company

McKinsey Direct

Generative AI in finance: Finding the way to faster, deeper insights

Generative AI technologies can automate time-consuming tasks for finance professionals, but can they be trusted to give the right answers?

This article is a collaborative effort by Steve Eklund, Lisa Kaufman, Avani Kaushik, Andrii Kurdiuk, Jan Svoboda, and Edward Woodcock, representing views from McKinsey's Operations Practice.

January 2023

Business leaders are excited about generative AI (gen AI) and its potential to increase the efficiency and effectiveness of corporate functions such as finance. A May 2023 survey of around 75 CFOs at large organizations found that almost a quarter (22 percent) were actively investigating uses for gen AI within finance, while another 4 percent were pursuing pilots of the technology.

That enthusiasm has been tempered by concerns over safety, privacy, accuracy, copyright, and social manipulation. In the case of finance, where numerical data and the accuracy of mathematical operations are fundamental, there is also concern over the possibility that gen AI systems could produce inaccurate or misleading information, a phenomenon known as "hallucination."

Developing a finance assistant using generative AI

One European consumer goods company recently developed a proof-of-concept gen AI assistant for finance professionals and business users. The new tool was built in about six weeks by a team of data scientists, engineers, and finance experts. It allows users to ask questions about financial performance in everyday language and rapidly receive answers that aid them in understanding and interpreting the data.

This proof-of-concept exercise was only the company's first step in applying gen AI in the finance function, but it offers several useful lessons for organizations seeking to capture the benefits of these technologies while managing the risks.

1. Start with high-impact, internally facing use cases.

Gen AI pilot projects should solve meaningful problems for the organization without creating new ones. That requires a high degree of control over the data that will feed into the model and a design that minimizes security challenges or the potential for data misuse. In this case, the company chose a single in-house dataset and a user group that was already familiar with it.

The proof of concept was designed to address a key pain point for the business. Finance analysts were often overloaded with requests for information from managers, and the managers were frustrated that it could take several days to get answers to relatively straightforward questions. Solving that problem helped to create interest in, and enthusiasm for, gen AI technology (Exhibit 1).

2. Assemble the right tools and capabilities for the job

A diverse array of gen AI solutions is available, ranging from on-premises options to cloud-based solutions and the choice of deployment should reflect the data sensitivity and specific requirements of the use case. Taking advantage of existing IT and digital infrastructure may shorten development time and can also ensure that gen AI solutions integrate existing workflows with minimal disruption.

After aligning on its preferred technology stack, the company assembled a highly integrated cross-functional team to drive the development process. In addition to data scientists and IT engineers, that team included senior members of the finance team, who helped to shape the design and acted as superusers during the development and testing phases. Their input was critical to ensure that the new tool met the organization's quality, reliability, and usability requirements.

3. Put humans in the loop, and keep them there

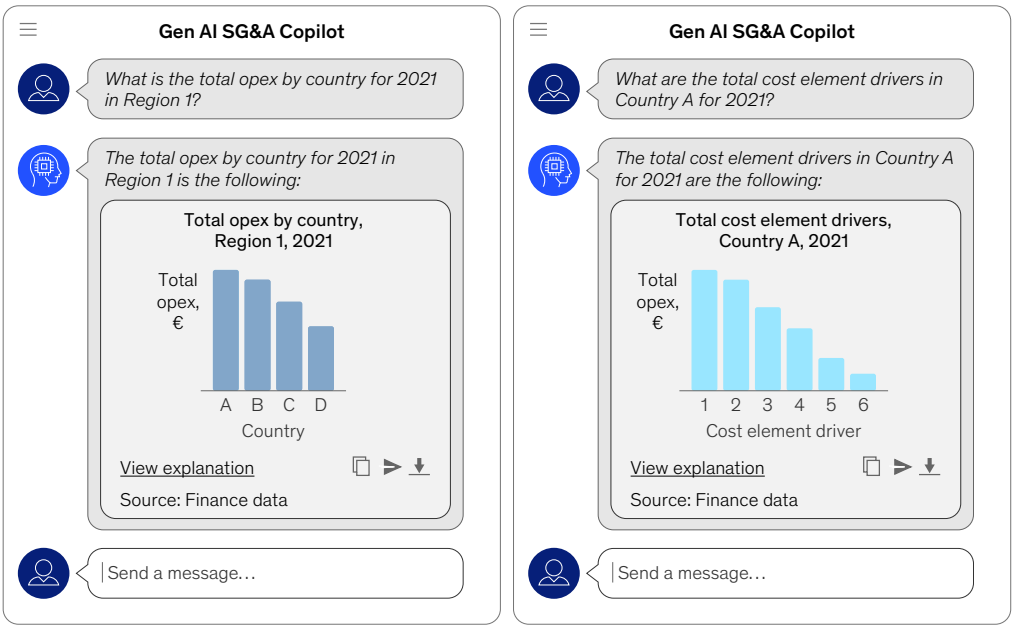
The company also paid significant attention to design of the user interface, maintaining a consistent structure and "look and feel" to provide a straightforward and streamlined user experience. One early request from users was for the ability to drill down rapidly through finance data to get to the information they needed. To facilitate this, the tool offers instant graphical visualization of results, so staff can quickly absorb and interpret its answers. All data is also made available in a tabular format that can be shared with other analytical tools.

During development, the company recruited a group of superusers from its finance function as product testers. They were encouraged to apply

Exhibit 1

The gen AI finance assistant helps users interrogate financial data with speed and precision.

Illustrative examples



the tool in real-world situations and to assess and provide feedback on its reliability and usability.

Everyday users also play a central role in the model's quality assurance and risk management. Up front, they are given clear guidance about the types of questions they can ask and how to formulate them. Additionally, to make it easier to check and validate outputs, the model is designed to generate a plain-language explanation of the calculations underlying each set of results it produces.

4. Play to the strengths of gen AI

While gen AI is a powerful solution for many tasks, other technologies may be a better fit for some elements of the overall solution. By designing solutions that move functions that can be handled independently of gen AI to other technologies, organizations can enhance the stability and performance of the tools they build.

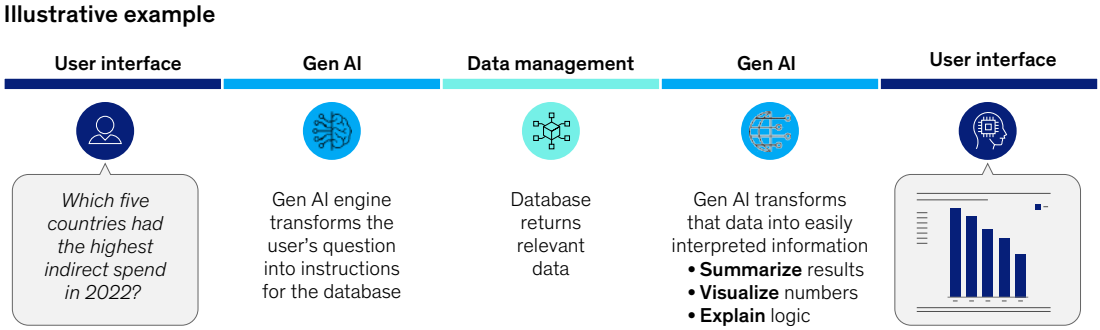
One limitation of the large language models (LLMs) that underpin many modern gen AI systems is innumeracy. These systems were designed to operate on natural language, not to perform precise calculations, and have been observed to struggle with mathematical computations. Engineering teams are working hard to address this limitation. One approach is to build a hybrid model, using the LLM in ways that, build on its strengths and other data tools to perform mathematical manipulations. For its proof of concept, the company used the gen AI system to translate user requests into database queries, then gave the calculation work to a dedicated data analytics platform (Exhibit 2). The output from the analytics system is passed back to the AI platform, which uses that data to produce several end products, including a text precis of the results, a summary of the calculation methodology, and the computer code used to create the graphical visualization.

5. Optimize the prompts

The team behind the AI assistant also took great care to minimize the opportunity for error and hallucination by engineering the prompts used to trigger the gen AI system. Shorter, simpler prompts encourage large language models to produce more predictable results. They are more efficient, too, since the computational workload of a gen AI system is closely related to the number of language "tokens" it handles in each interaction. For the finance assistant, the team divided the work into three distinct interactions, each handled separately by the gen AI model:

Exhibit 2

The gen AI platform provides the interface to financial data but does not perform underlying calculations.



translate the user's question into a data query, generate a summary and explanation, and create a visualization .

The company developed examples of effective prompts for interacting with the system plus guidance and training materials to help users generate their own. Those training materials were another chance to emphasize the importance of the human-in-the-loop, reminding users of their responsibility to check the assumptions in the language of their queries and behind the results of every interaction.

6. Build a robust test environment

Fine tuning a gen AI model is a highly iterative process. The ability to test and validate rapidly is key to successful development. To achieve it, the team built a comprehensive automated test suite, which operated using a list of high-priority queries provided by senior finance analysts. This test was run after every update of the model during development, with the results used for further refinement and operation.

7. Put effective governance systems in place

As companies move from pilot projects to mainstream adoption of gen AI tools, they will need appropriate governance frameworks to maintain quality and manage risks at scale. These might include ongoing monitoring and auditing mechanisms to assess AI system behavior, ensuring it aligns with established ethical guidelines. Cross-functional teams of AI experts, ethicists, and legal advisors should evaluate AI models and applications for potential biases or ethical concerns. After the success of its finance assistant proof of concept, the consumer goods company is now exploring appropriate governance structures that will support the wider deployment of gen AI tools.

Generative AI is already changing the way professionals do their work in the finance function and beyond. To capture the benefits of these exciting new technologies while controlling the risks, companies must invest in their software development and data science capabilities. And they will need to build robust frameworks to manage data quality and model engineering, human–machine interaction, and ethics. However, as these case examples show, these technologies can accelerate and enable access to critical business information, giving human decision makers the information to make thoughtful and timely choices.

Steven Eklund is a partner in McKinsey's Stamford office, where **Edward Woodcock** is a senior expert; **Lisa Kaufman** is a specialist in the London office; **Avani Kaushik** is an expert in the Southern California office; and **Andrii Kurdiuk** is a senior data scientist in the London office, where **Jan Svoboda** is an associate partner.