ECONOMIC VIEWPOINT

Human Against Machine: Are Professional Economists or AI Better at Forecasting Inflation?

By Jimmy Jean, Vice-President, Chief Economist and Strategist, Randall Bartlett, Senior Director of Canadian Economics, and Samuel Turcotte, Master's Student in Applied Economics, HEC Montreal

> ▶ As interest grows around the potential of artificial intelligence (AI) to forecast economic indicators, the US Federal Reserve has demonstrated some success in using it to forecast inflation.
>
> ▶ Similar results have been obtained by broadening the Fed's research to include other large language models (LLMs), countries and time horizons. LLMs have generally been at least as good as professional forecasters at projecting inflation since 2018, with 2023 being the only year that US and Canadian forecasters outperformed AI platforms in this area.
>
> ▶ But AI isn't without its limitations. In examining historical forecasting performance, it's very difficult to prevent LLMs from "cheating" by taking into account information that wasn't available when professional forecasters issued their projections. Further, some of the AI platforms' technical characteristics can cause them to just naively repeat prior forecasts.
>
> ▶ All told, AI should be considered a useful tool in the economic forecaster's toolbox, especially since it incorporates different information than traditional tools. However, until an analysis is undertaken which properly controls for all of AI's current limitations, it isn't yet time to replace the human with the machine.

There has been a lot of excitement around the capabilities of generative artificial intelligence (AI) since its use became commonplace a couple of years ago (see our recent note on AI and its applications). In economics, as in many other domains, it has simplified tasks like writing, data analysis and computer programming. But there is also an emerging field of research on how large language models (LLMs) can be leveraged for economic forecasting.

How Does AI Stack Up Against the Pros?

In the very near term, economic forecasts using high-frequency data (known as "nowcasting") are being complemented by information derived from LLMs. AI is also being used to forecast economic indicators at longer-term time horizons, such as four or more quarters into the future.

United States

The Federal Reserve Bank of St. Louis (St. Louis Fed) published a research paper earlier this year exploring the idea of forecasting economic indicators at longer time horizons using AI. In it, the reserve bank compared US CPI inflation forecasts generated by AI to those from the Survey of Professional Forecasters (SPF). The authors concluded that "LLM forecasts generate lower mean-squared errors overall in most years, and at almost all horizons. LLM forecasts exhibit slower reversion to the 2% inflation anchor."

In a recent research note, Turcotte (2024) explored this result by applying the St. Louis Fed's approach to the Bloomberg forecast survey for different countries using a wider variety of AI packages. His findings for the US were broadly similar to St. Louis Fed's conclusions. Two of the three LLMs used provided forecasts that were closer to the actual data from Q1 2019 to Q1 2024 than the average of the contributors to the Bloomberg survey of
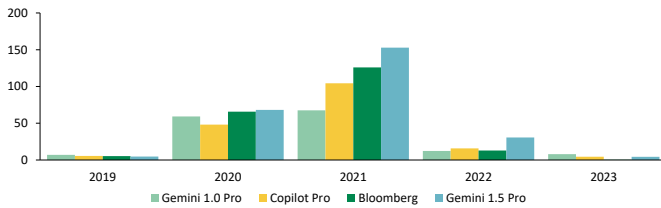
Graph 1
US Forecasters Had Larger Inflation Forecast Errors than 2 of 3 LLMs

Root mean square errors of US inflation forecasts

LLM: Large Language Model
Bloomberg, Bureau of Labor Statistics, Turcotte (2024) and Desjardins Economic Studies

forecasters (graph 1). And much like in St. Louis Fed's analysis, private-sector forecasters outperformed during periods when inflation was around or falling toward 2%.

Canada

Turcotte's findings for Canada were different, however. His analysis found that Canadian forecasters who contributed to the Bloomberg survey had smaller average forecast errors overall than two out of three LLMs (graph 2). Canadian economic forecasters also tended to perform better versus AI than their US peers during periods of high and rising inflation. But even then, at least one LLM beat the forecasters' average in four out of five years in Canada, a similar outcome to that observed for the US.

Graph 2
Canadian Forecasters Had Smaller Forecast Errors than 2 of 3 LLMs

Root mean square errors of Canadian inflation forecasts
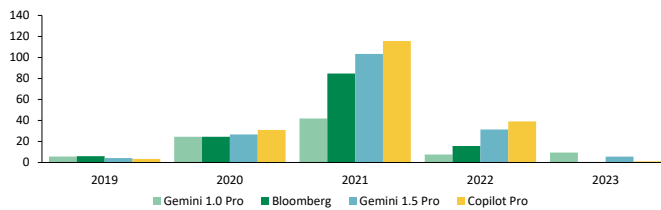
LLM: Large Language Model
Bloomberg, Statistics Canada, Turcotte (2024) and Desjardins Economic Studies

In the same study, Turcotte also analyzed the model performance rankings for each forecast year, as this approach provided a better understanding of the overall performance and placed less emphasis on the 2021 outcome. With this approach, Canadian and US private-sector forecasters only outperformed the LLM forecasts in 2023 (table 1). The overall analysis of the annual performance reveals that Bloomberg consensus forecasts perform best in the US case. Meanwhile, Gemini 1.0 Pro's greater accuracy for the years 2020–2022 is what allowed the LLM to slightly outperform Canadian forecasters (table 2).

Table 1
US Professional Inflation Forecasts Bested All LLMs in 2023 Only

YEAR	ANNUAL PERFORMANCE RANKING OF US INFLATION FORECASTS			
	FIRST	SECOND	THIRD	FOURTH
2019	Gemini 1.5 Pro	Bloomberg	Copilot Pro	Gemini 1.0 Pro
2020	Copilot Pro	Gemini 1.0 Pro	Bloomberg	Gemini 1.5 Pro
2021	Gemini 1.0 Pro	Copilot Pro	Bloomberg	Gemini 1.5 Pro
2022	Gemini 1.0 Pro	Bloomberg	Copilot Pro	Gemini 1.5 Pro
2023	Bloomberg	Gemini 1.5 Pro	Copilot Pro	Gemini 1.0 Pro
Overall ranking	**Gemini 1.0 Pro**	**Bloomberg**	**Copilot Pro**	**Gemini 1.5 Pro**

LLM: Large Language Model
Bloomberg, Turcotte (2024) and Desjardins Economic Studies

Table 2
Canadian Professional Forecasters Ranked Second to Gemini 1.0 Pro

YEAR	ANNUAL PERFORMANCE RANKING OF CANADIAN INFLATION FORECASTS			
	FIRST	SECOND	THIRD	FOURTH
2019	Copilot Pro	Gemini 1.5 Pro	Gemini 1.0 Pro	Bloomberg
2020	Gemini 1.0 Pro	Bloomberg	Gemini 1.5 Pro	Copilot Pro
2021	Gemini 1.0 Pro	Bloomberg	Gemini 1.5 Pro	Copilot Pro
2022	Gemini 1.0 Pro	Bloomberg	Gemini 1.5 Pro	Copilot Pro
2023	Bloomberg	Copilot Pro	Gemini 1.5 Pro	Gemini 1.0 Pro
Overall ranking	**Gemini 1.0 Pro**	**Bloomberg**	**Gemini 1.5 Pro**	**Copilot Pro**

LLM: Large Language Model
Bloomberg, Turcotte (2024) and Desjardins Economic Studies

What Explains the LLMs' Forecasting Performance?

Due to limitations in how far back in time private-sector inflation forecasts can be compared with AI-generated forecasts, it's difficult to produce a meaningful long-term comparison. Indeed, it hardly seems fair to isolate the evaluation period to a time when both the US and Canada experienced the highest inflation in over 40 years. But there are still some helpful insights to be gleaned from the results.

The first consideration is around inflation expectations. If there is any group for which inflation expectations are well anchored to central banks' 2% targets, it's likely among economists. Given the success of inflation-targeting regimes in anchoring inflation expectations among consumers and investors alike, it makes sense that forecasters' projections would ultimately converge to 2%. In contrast, apart from Gemini 1.5 Pro, which seems to have unwavering confidence in what central bank actions can do, there is little evidence that LLMs intuitively assume that inflation converges to 2% over time.

Second, and related to the prior insight, LLM inflation forecasts seem to generally be more momentum-driven than private-sector inflation forecasts. For instance, LLMs may not account for idiosyncratic behaviour in some subcategories of inflation in the face of very specific shocks, such as expected tax hikes

(e.g., value-added tax on alcohol or tobacco) or subsidies to government programs (e.g., childcare). That said, these models display impressive capabilities for inferring the impact of global events like the China–US trade war, OPEC quotas or major geopolitical conflicts.

Third, due to short-term memory issues in conversations with AI platforms, some LLMs produce inflation forecasts that are just reproductions of the previous period. This also speaks to the limited transparency in AI model data, training and forecasting processes.

Fourth, the fact that Canadian forecasters had a smaller margin of error than their US counterparts may be because Canadian data are more limited. Fewer economists and market participants could also translate into a less rich information environment for AI. This could give an informational advantage to forecasters who are more familiar with local conditions and statistics in smaller jurisdictions. However, it also implies that as more information becomes available, this informational advantage could diminish over time.

Finally, the historical forecast performance of LLMs versus professional forecasters may not be comparable at all. Turcotte determined that many LLM models "cheat" in developing their in-sample forecasts, in that they don't exclude all future information despite being told to. As a result, LLM-based forecasts may include information that wasn't available to professional forecasters when they issued their projections. Therefore, the only way to truly determine whether AI forecasts outperform those of professional analysts is to undertake rolling forecasts through time and compare their performances to actual data as it is published.

What Do LLMs Tell Us About Future Inflation?

All this begs the question: if LLMs were used to forecast inflation today, what would they tell us?

For the US, the Bloomberg consensus of professional economic forecasters from August 2024 called for headline CPI inflation to gradually decelerate from 3.2% y/y in Q2 2024 to 2.3% by Q4 2025 (graph 3). When we look at how this compares to AI-generated inflation forecasts, most LLM-based forecasts expect inflation to fall lower than that over the next two years. While some models had higher inflation projections for 2024, all of the platforms examined by Turcotte forecasted inflation at or below 2% in the US by the end of 2025.

In contrast, for Canada, two out of three LLM-based forecasts track the August 2024 Bloomberg consensus of private-sector forecasts very closely (graph 4). Additionally, the Bloomberg forecast average and the LLM-based forecasts all end 2025 at around 2%, except for Gemini 1.0 Pro. For some, this

Graph 3
All LLM-Based Forecasts Expect US Inflation at or Below 2% by the End of 2025
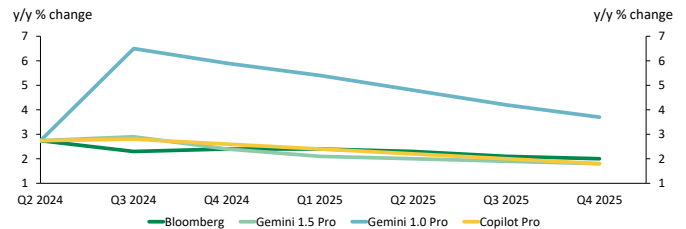US total CPI inflation forecasts
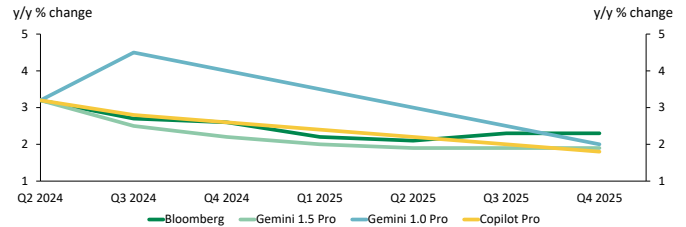
LLM: Large Language Model
Bloomberg, Bureau of Labor Statistics, Turcotte (2024) and Desjardins Economic Studies

Graph 4
Most LLM-Based Forecasts for Canadian Inflation End 2025 Near 2%

Canada total CPI inflation forecasts

LLM: Large Language Model
Bloomberg, Statistics Canada, Turcotte (2024) and Desjardins Economic Studies

may strengthen the belief that the Bloomberg consensus of private-sector forecasters is more believable. However, it should be noted that the AI platform calling for higher inflation in Canada is also the one that had the smallest forecast errors during the pandemic. The same is true in the US. Only time will tell which forecast proves correct.

What Role Should LLMs Play in Economic Forecasting?

Economists use a wide variety of statistical tools to forecast economic indicators. They also apply judgment to those forecasts, drawing on knowledge and experience that a traditional forecasting model can't incorporate. Given the strength of LLM-based forecasts, they are clearly an important tool in a forecaster's toolbox, especially since these models draw from a less conventional data set. But AI isn't ready to replace professional analysts yet, particularly in smaller, less information-rich economies like Canada.