July 2023

What Lying Chatbots Tell Us About Model Design in AI-Driven Investing

Call them hallucinations, fantasies or glitches — fact-challenged answers from ChatGPT and other generative AI tools demonstrate the critical function of rigorous model development and human oversight.

Chrissy Bargeron, CFA
Client Portfolio Manager

Highlights

- New human-like chatbots illustrate the tremendous potential of AI, but also serious concerns about accuracy, security and bias, which can be deal-breakers for AI investment integrations.
- These AI pitfalls can be addressed by training models using high-quality, reliable data sources and implementing multiple layers of risk management at each step of the investment process.
- Voya Machine Intelligence combines extensive knowledge in both data science and investment management, bringing more than a decade of experience testing, running and refining models in a systematic process that integrates the best of human and machine talents.

When AI goes off the rails

The incredible promise of ChatGPT, launched in late 2022, has taken the world by storm, sparking the fastest adoption of any consumer application in history by a mile.[1] The chatbot and its growing list of competitors are based on a form of generative artificial intelligence called large language models (LLMs), which train on vast amounts of data scraped from the internet and use that information to provide seemingly relevant, human-like responses to prompts.

In many instances, LLMs have demonstrated remarkable skill across a wide range of domains, beating the average human in activities such as SAT tests, wine recommendations and medical expertise.[2] They have demonstrated proficiency at summarizing large documents, writing code and generating recipes. In more creative endeavors such as poetry, results range from surprisingly good to hilariously awful to downright disturbing.

But sometimes, LLMs give answers that can be misleading or outright false.

VOYA ® INVESTMENT MANAGEMENT

In March, the still-very-much-alive data privacy expert Alexander Hanff asked ChatGPT about himself. Its response:

"...Tragically Hanff passed away in 2019 ... [His] death has been reported in several news sources, including his obituary on the website of The Guardian ... " https://www.theguardian.com/technology/2019/apr/22/alexander-hanff-obituary" [the link was entirely fabricated and never existed]
The Register, "Why ChatGPT should be considered a malevolent AI – and be destroyed," March 2, 2023.

Large language models are dependent on their training data set and aren't designed to make judgments about accuracy.

Why do such errors happen? LLMs are designed to interpret and generate language, not to determine accuracy. That makes LLMs dependent on both the phrasing of a user's prompt and the quality of their data source (typically, the public internet), forcing AIs to contend with conflicting and often false information.

These problems aren't new:

- In 2013, IBM's Watson, which famously beat Jeopardy champions Ken Jennings and Brad Rutter, teamed with the University of Texas MD Anderson Cancer Center to develop models to eradicate cancer. Watson started giving cancer patients dangerous treatment recommendations. It turns out that IBM engineers had built Watson's diagnostic models with a limited set of hypothetical data rather than real cancer patient data. In other words, the models were an example of "garbage in, garbage out."

- In 2016, Microsoft introduced Tay, a Twitter chatbot. Internet trolls corrupted Tay in less than a day, flooding it with racist tweets and other offensive posts, which rapidly altered Tay's personality. The problem was that Tay had been launched without adequate filters, leaving it vulnerable to malicious actors. These instances serve as stark reminders of the risks associated with AI when followed blindly and without appropriate risk controls.

These challenges demand careful consideration, robust safeguards and responsible regulation to ensure AI tools are used ethically and for societal benefit.

What does this have to do with investing?

The impressive information processing capabilities of ChatGPT has, unsurprisingly, put increasing pressure on investment managers to adapt. We believe it is crucial to approach these technologies with careful deliberation, informed by AI's past failures.

Watson's mistake in trying to predict cancer treatment demonstrates the importance **of training models using data sources that contain material information (not just a large quantity of data) and that are reliable and bias-free.** Building "quality in, quality out" is no small endeavor; it takes years of person-hours to curate data and identify its essential features before one can start to build AI models with potential to enhance investment processes.

Tay's vulnerabilities illustrate the need for **multiple layers of risk management throughout the investment process.** For asset managers, the deep work here lies in designing virtual tools with strong filters to provide confidence in the output. Moreover, although the role of humans in the investment process may shift, it doesn't go away. On the contrary, the use of machine learning tools puts a premium on the experience and skills of investment professionals who have a deep understanding of both investing and data science.

"AI represents a 'Blockbuster moment' for the investment industry – a paradigm shift that will create winners and losers."

Gareth Shepherd, PhD, co-head of Voya Machine Intelligence

Voya's approach to machine learning: Keep humans in the loop

At Voya, *"machine intelligence"* refers to our overall process for AI-driven investing, which integrates machine, human and quant insights for stock selection and risk management."

Voya Machine Intelligence is rooted in the principle that the dual knowledge of data science and investing is essential when applying machine learning to security selection to generate positive outcomes and mitigate unintended consequences. This has not always been the case within the industry. During the rise of quantitative finance in the late 1990s and early 2000s, we saw many asset management firms hire individuals with strong mathematical backgrounds in areas such as linear algebra, calculus and differential equations, often over those with finance-oriented educations.

When it comes to machine learning, the depth of relevant human expertise is of the utmost importance. Building accurate models for stock selection requires a solid foundation in investing, not just data science. The Voya Machine Intelligence team consists of pioneers in this field, bringing more than a decade of experience testing, running and refining models, which in our view gives Voya a strong first-mover advantage.

Practical experience surpasses academic knowledge

Not only do the humans operating the AI model need knowledge of investing, it's also important to recognize that the experience of the AI model can significantly affect its accuracy. Our machine learning models (known as **"virtual analysts"**) have been undergoing training since 2011, accumulating more than 10 years of real-world experience. This lengthy period has allowed them to learn from mistakes, fine-tune their algorithms and enhance their stock selection skills. As industry professionals understand, knowledge gained on the job far surpasses that acquired solely in academic or training settings.

Supervision by virtual traders to optimize execution

Building its models on carefully curated data, Voya incorporates multiple layers of risk management throughout the investment process. For example, Voya Machine Intelligence has built **"virtual traders"** — similar to traditional quantitative models — to recommend precise timing of trades and filter out potentially risky recommendations from the virtual analysts. Human portfolio managers thoroughly review all trades before execution, and the trades are executed by Voya's traders in NYC. We believe these measures ensure that a comprehensive risk management framework is in place with mechanisms to override poor recommendations from the ML tools.

In our view, this fusion of knowledge is indispensable, which is why the co-founders of Voya Machine Intelligence's, via their predecessor firm G Squared, possess this rare blend of skills. The integration of data science with investment acumen is paramount to ensure the models' robustness, accuracy and alignment with the investment objectives of our clients.

Symbiosis of human and machine

As the investment industry integrates AI into its decision processes, demand is making data more widely available and thereby more commoditized. Simply having access to the data itself no longer constitutes an advantage. What counts is the insight you are able to draw from the data. Voya's approach seeks to build a symbiosis among its three sources of information strength: fundamental analysis, machine learning and quantitative analysis.

"In our view, if you're able to marry the best of the human with the best of the machine – particularly in long-term investing – that's the way to invest for the future."
Gareth Shepherd

[1] UBS; 02/01/23. Months to reach 100 million active users — ChatGPT: 2, TikTok: 9, Instagram: 30. Ed. note: Meta's Twitter rival, Threads, looks likely to beat ChatGPT's record shortly.
[2] Lifearchitect.ai; 06/23: Alan D. Thompson: "Integrated AI: The Sky Is Entrancing"; benchmarking based on ChatGPT-4.

VOYA® INVESTMENT MANAGEMENT