

Artificial Intelligence (AI) has made remarkable strides over recent years, with tools like **ChatGPT** and **Google Gemini** showcasing impressive capabilities in language understanding and generation. Accuracy improvements in these models have been significant, leading many to wonder: if hallucinations—the generation of incorrect or fabricated information—are getting better, why do they remain such a critical problem, especially in high-stakes environments like family offices?

The Persistent Challenge of AI Hallucinations

AI hallucinations represent one of the most subtle but insidious challenges facing the adoption of large language models (LLMs). Unlike outright errors that can be readily spotted, hallucinations often produce outputs that are plausible, sometimes highly persuasive, but factually incorrect. This blurring of fact and fiction creates considerable residual risk for users who rely on these tools without robust verification processes.

What Are AI Hallucinations?

In the context of natural language processing models such as ChatGPT and Google Gemini, hallucinations are outputs that contain untrue, misleading, or fabricated information, often generated without explicit intent. These can range from small factual inaccuracies to entirely invented claims or data points.

The reason hallucinations arise is due to how these models are built: they predict the most statistically probable next word or phrase based on their training data, without true understanding or external verification. This can produce fluent yet incorrect statements.

Why AI Hallucinations Are So Persuasive and Risky

- **Fluency Masks Error:** Outputs are coherent and well-structured, leading users to trust them implicitly.
- **Authority Illusion:** Because these tools have demonstrated expertise across countless queries, users tend to assume correctness.
- **Lack of Source Citation:** Unlike traditional research tools, many chatbots provide answers without verifiable sources, making it hard to fact-check on the spot.
- **Speed Over Scrutiny:** The convenience and fast pace at which these models deliver answers may encourage users to skip due diligence.

In high-stakes environments such as investment decision-making, this persuasive nature imposes significant risks. A single fabricated data point or misunderstood insight could lead to errors with real financial consequences.

Family Office Use Cases: Where Accuracy Matters Most

Within the family office sphere, AI tools are increasingly leveraged to aid in preliminary research, parsing voluminous reports, and generating summaries. However, the nature of these use cases makes them particularly sensitive to hallucination risk.

Preliminary Research and Due Diligence

Family offices conducting manager due diligence may prompt AI models to provide quick overviews of managers, funds, or market dynamics. While these can speed initial investigations, any hallucinated information—such as made-up performance statistics or invented track record narratives—can mislead the user and introduce bias before the final, human-led verification occurs.

Parsing Complex Reports

Investment reports, especially in private markets, often involve dense disclosures with nuanced language. AI tools can extract key points and summarize findings, saving time. Still, hallucinations here could distort critical information, such as misstating risk factors or misrepresenting investment terms.

Summarization: A Double-Edged Sword

Summaries generated by LLMs can highlight essential themes, but omissions or insertions of inaccurate details risk skewing perceptions. <https://canadianfamilyoffices.com/special-reports/technology-sr/ai-has-the-makings-of-a-powerful-tool-for-family-offices-but-frequent-and-highly-convincing-hallucinations-are-proving-troublesome/> In family office settings, where trust in reporting accuracy is paramount, such errors undermine confidence in both the technology and the data used for decision-making.



Private Markets Data Scarcity: An Amplifier of Risk

One unique challenge for family office due diligence is the scarcity of reliable data in private markets. Unlike public equities with standardized disclosures, private investments often lack transparent, frequent, or structured reporting.



In this data-thin environment, AI models have fewer concrete facts to learn from and may be more prone to fabricate plausible but incorrect details to fill gaps. This scarcity elevates the residual risk that hallucinations will sneak into outputs, making human oversight and robust data governance indispensable.

Faithfulness vs. Factual Hallucinations

In AI terminology, **faithfulness** refers to the degree that a model's output accurately reflects its input data or source. Conversely, hallucinations typically break faithfulness by adding or altering information that is not supported by the input.

Aspect	Faithful Output	Hallucinated Output	Accuracy	True and verifiable	False or fabricated	Source alignment
Consistent with input data	Yes	No	High	Yes	No	Yes
Introduces unsupported content	No	Yes	Low	No	Yes	No
Risk level	Low residual risk	High residual risk				
User trust	Maintained	Eroded if discovered				

Continuous improvements in models like Google Gemini aim to boost faithfulness by reducing factual hallucinations, but residual risk persists due to model architecture and limitations in training data.

Balancing Innovation with Risk Management

Given the residual risks from hallucinations, particularly in family office and investment contexts where high-stakes decisions are made, a cautious approach is essential:

1. **Use AI as an Aid, Not a Source of Truth:** Employ tools like ChatGPT and Google Gemini for hypothesis generation and preliminary parsing but always validate findings with primary sources.
2. **Implement Robust Data Governance:** Define clear protocols on what outputs can be accepted, how verification occurs, and when escalation is needed.
3. **Document and Audit AI Usage:** Keep logs of AI-generated insights and reason about residual risk explicitly —“What would we show an auditor?” should be a guiding question before adopting any tool.
4. **Combine AI with Domain Expertise:** Human oversight remains critical. Experienced analysts can spot implausible claims and contextualize AI outputs appropriately.

Conclusion

While accuracy improvements in AI models like ChatGPT and Google Gemini have reduced the frequency and obviousness of hallucinations, these errors remain a persistent problem due to their persuasive nature and potential impact on high-stakes decisions. This is especially true in data-challenged environments such as private markets—common in family office investing—where the temptation to fill gaps can encourage factual hallucinations.

Adopting AI tools requires balancing innovation with a strong governance framework to manage residual risk. By recognizing hallucinations as an ongoing hazard rather than a solved problem, family offices and other cautious users can harness AI's benefits responsibly without jeopardizing decision quality.

Remember: AI output is a starting point, not an endpoint. Trust but verify.