Chapter 9
AI Hallucination in Financial Systems: Threats, Controls, and the Future of Risk Management

Ganesh Harke
https://orcid.org/0009-0005-3947-0960
Independent Researcher, USA

ABSTRACT

The adoption of Generative Artificial Intelligence (GenAI) in the global financial services sector is characterized by distinctive predictive analytics, customer service, and automation. However, it introduces a critical vulnerability: AI hallucinations, instances in which models generate outputs that are linguistically sound but factually incorrect or baseless. GenAI's capacity to fabricate data, misinterpret policies, or generate false market signals poses serious operational, reputational, and systemic risks. The chapter offers a detailed analysis of AI hallucinations in financial systems, covering definitions and a financial domain risk taxonomy. In addition, it examines the agency behind model failures. It also discusses the distinct threats that these models pose to a financial system. Additionally, the chapter introduces the Control Framework, which integrates Retrieval Augmented Generation (RAG) and other protocols. Finally, it examines the future of risk management, anticipating the development and evolution of regulations for stochastic models.

INTRODUCTION

Artificial intelligence (AI) has created new capabilities in automation, analytics, and judgments across industries. Recent developments in GenAI have further

DOI: 10.4018/979-8-3373-7534-2.ch009

advanced these capabilities, thereby giving rise to generative cognition. Recent advancements in GenAI have further developed these capabilities, thereby giving rise to generative cognition. It is a capability to produce text, code, and analysis that mimics human communication. The financial services industry (FSI) is not lagging in its adoption of GenAI. According to NVIDIA (2024), 91% of financial services companies are either assessing AI or already using it in production. These firms are using AI to drive innovation, improve operational efficiency, and enhance customer experiences (State of AI in Financial Services: 2024 Trends, 2024). This integration of GenAI into the global financial system is a technological innovation comparable to the introduction of algorithmic trading or the use of digital ledgers. From hedge funds to retail banking, international financial institutions are adopting large language models (LLMs) to automate complex tasks, including regulatory compliance analysis, credit risk assessment, and the drafting of personalized financial plans. However, this instantaneous integration of GenAI into an industry built on deterministic presuppositions poses a profound, paradoxical risk of hallucination (Erdem et al, 2025). Unlike legacy software bugs, which are logical and predictable, hallucinations are plausible but factually false outputs (Tonmoy et al., 2024). As financial systems grow more autonomous, the risk of fabricating financial data, misinterpreting legal policies, or giving incorrect financial advice increases. This threat extends beyond basic operational risk, potentially harming reputations and systemic stability (Erdem et al, 2025; Sigma, 2025). To address these challenges, this chapter examines the fundamental areas:

- **RQ1**: What are the domain-specific mechanisms driving AI hallucinations in financial high-stakes environments?
- **RQ2:** How adequate are current mitigation controls against intrinsic vs. extrinsic fabrication?
- **RQ3:** What regulatory frameworks are emerging?

What Are AI Hallucinations?

A hallucination is an instance in which a generative model produces text that is linguistically sound and syntactically flawless but remains ungrounded in either the source input or factual reality. While the term may feel like AI dreams, it technically refers to a failure mode inherent to the architecture of modern LLMs. LLMs function as next-token prediction engines. They do not possess epistemological knowledge. They compute the statistical likelihood of the following word in sequences using