Getting the GenAI balance right: Varying approaches to RAG system validation depending on scale, complexity and sensitivity of your application

Abstract

The adoption of Generative AI (GenAI) within regulated sectors such as banking, credit lending and insurance can be unnecessarily slowed due to concerns about reliability, security, and regulatory compliance. In many cases, users engage with GenAI in a limited capacity—primarily through direct interactions with a Large Language Model (LLM), such as ChatGPT, to complete specific tasks. However, this approach does not enhance or refine the foundational model, potentially leading to responses that stakeholders may find unhelpful, especially for enterprise specific tasks.

A more robust and context-aware utilisation of GenAI can be achieved through Retrieval-Augmented Generation (RAG), a technique that integrates external knowledge sources into LLM outputs to improve relevance and accuracy. While the use of LLMs and RAG introduces multiple components that increase system complexity, it does not require equally complex validation methodologies. Instead, the approach to validation can rely on a selection of principles already familiar to predictive model specialists and should be proportionate to the system's use case, balancing rigor with practicality.

This presentation explores the fundamental principles of validating GenAI systems, emphasising the extent of validation required based on the intended application. Key evaluation areas—visibility, reliability, efficiency, representativeness, security, and trustworthiness—can be examined in relation to system performance and risk mitigation.

We will also discuss the escalating validation demands associated with high-risk applications and increasing system complexity. Specific topics include data pre-processing, chunking, retrieval and prompt augmentation, as well as objective and subjective methodologies for validating prompt augmentation and LLM generated responses. By outlining structured validation frameworks, we aim to provide actionable insights into ensuring the robustness and reliability of GenAI RAG implementations across diverse use cases in the application of credit scoring.

Authors & Affiliations

Mr James Barlow[1], Mr Gordon Baggot[2], Mrs Lucy Worsley[1]
[1]4most, London, United Kingdom. [2]4most, Edinburgh, United Kingdom