

AI's Creativity for Comprehensive Coverage: On Responsible Testing of Financial Technology



AI's Creativity for Comprehensive Coverage: On Responsible Testing of Financial Technology

In the fast-evolving landscape of financial technology, many industry players find themselves facing the question of whether they need to respond to the latest trends and harness the power of artificial intelligence (AI) to keep their competitive edge. In the financial services domain, data plays a crucial role, and the abundance of data makes a strong case for leveraging AI in most of the numerous use cases. Whether it is transactional data, market data, customer data, or other financial datasets, AI can extract valuable insights and boost efficiency in the associated tasks.

Risks or opportunities?

Over the past year, Large Language Models (LLMs) and Generative AI (GenAI) have come to the forefront of innovations permeating various sectors, including the financial services industry.

As this technology gains momentum, questions arise regarding its applicability and limitations. While the creativity of GenAI shows promise, concerns about its accuracy, often referred to as "hallucinations," raise doubts about its worthiness for practical use, especially in the financial sector – a domain traditionally prone to financial and reputational risks.

Even without AI-related complications and risks, financial technology is well-known for its complexity. Ensuring its reliability and robustness is a challenging task... which, quite ironically, can itself be a good case for applying Generative AI – to improve the efficiency of testing against complexities stemming from a multitude of interdependent parameters across numerous business flows, participants, protocols, asset classes, and other permutations typical for financial software.

The concept of “good” testing

What exactly can GenAI *improve* in testing? If we expect to improve something (i.e. make it better), a first step is to settle on the definition of “good”.

Some industry practitioners envision an ideal test process as possessing such characteristics as full automation, easy maintenance, speed, consistency, system-agnosticism, vendor independence, transparency, and low cost. But aiming for meeting these criteria alone carries the danger of goal misalignment – a concept that in the AI domain is associated with reward hacking, when the objective function is formally achieved without actually delivering the intended outcome. In other

words, one will always find a way to satisfy the above criteria of “ideal” testing, with the most evident one being not performing any testing at all.

The true objective function of software testing is finding defects and communicating them to the stakeholders in the most effective manner, and that’s the main purpose of testing as a complex cognitive activity, a deliberate effort. “Good” software testing is an information service, and its effectiveness is measured by the accuracy, relevance, and accessibility of the information about system behaviour. Making the case for Generative AI to improve testing, we would expect it to significantly augment the ability of the testing effort to provide such information.

The case for GenAI

Software testing and, even more broadly, software engineering are areas where Generative AI can bring substantial improvements: according to [Gartner](#), “by 2025, 30% of enterprises will have implemented an AI-augmented development and testing strategy, up from 5% in 2021.”

For testing, the power of GenAI lies in its ability to automatically generate diverse and realistic test scenarios, leading to enhanced test coverage. Just by combing through more data points, AI models can hit those rare parameter combinations that are necessary for detecting issues that would have stayed undiscovered by tests created by human testers. Leveraging GenAI’s creativity can help create more comprehensive test libraries capable of detecting more defects.

Is “good” good enough? Creativity vs responsibility

While GenAI’s unparalleled degree of creativity helps in achieving better test coverage, it does not guarantee testing efficiency. Generating an abundance of test scenarios has serious limitations, such as scarcity of the computational resources of typical test environments and the limited capability of human specialists to interpret vast volumes of test results. To test effectively, we need more possible data combinations. But to test *efficiently*, we need to differentiate between them: a highly creative Generative AI needs to be balanced out with a more restrictive method.

Software testing is a complex cognitive activity that requires a high degree of responsibility – a quality typically found in good human software testers. However, AI lacks inherent responsibility, and there’s no way to make it feel responsible. To ensure accountability in the testing process, it is crucial to introduce a discriminative peer that evaluates the creative outputs of generative AI.

Adding a reasoning mechanism to sift through gigabytes of automatically generated data and select the most meaningful entries is not a trivial task. A possible solution may be to enrich test data with even more data. By labelling test data entries and assigning weights to different test

coverage points, we can gain valuable insights into the coverage strength of each test scenario within a particular dataset: among arbitrarily many unique test scenarios, it is crucial to distinguish those that have unique and non-unique coverage. Based on this data, the model used in AI-assisted testing prioritises high-weight scenarios for future execution while filtering out those with lower weights, optimising test libraries and overall human and hardware resource utilisation. More importantly, the approach helps evaluate whether test coverage is adequate to the complexity of the system under test.

In this combination, the generative AI component allows for a greater degree of freedom in generating test scripts from code prompts, providing a solid foundation for comprehensive test coverage. Complementing its generative counterpart, the discriminative AI component, serving as an implementation of responsibility, operates within a more rule-based framework. Its primary objective is to make sure that critical errors are not overlooked, ensuring robust testing practices within the financial domain. Such an approach creates a balance between AI's generative creativity and more transparent, rule-based techniques. Integrating both types of AI ensures comprehensive coverage while maintaining accountability and traceability.

The value of responsible testing

The value of software testing can be measured across three dimensions: quality, speed, and cost. Improving software testing entails enhancing the ability to detect and interpret defects while reducing timeframes and costs. AI holds the potential to deliver value across all three dimensions, empowering clients with enhanced software reliability, faster time-to-market, and optimised resource utilisation.

Exploring the potential of Generative AI in software testing for the financial services industry reveals the need for a balanced approach that combines creativity with responsibility. By harnessing GenAI's capabilities alongside discriminative techniques, testing becomes more comprehensive and efficient. Using different AI methods as complements to each other serves the purpose of test library refinement, focusing test execution and analytical efforts on high-impact scenarios and exploring the system under test on a fundamentally different scale.

As AI continues to evolve, the path to improved software testing lies in harnessing its creative potential while upholding responsibility, enabling organisations to deliver higher quality software at a faster pace and reduced cost.