VIEW POINT

EXPLAINABLE AI AND INTERPRETABILITY: BUILDING TRUST AND REDUCING FALSE POSITIVES IN FINANCIAL GENAI MODELS

Infosys®
Navigate your next

The Black Box Conundrum – Unraveling the Mysteries of Financial GenAI

Artificial Intelligence (AI), particularly Generative AI (GenAI), is reshaping the financial services landscape through revolutionary products, personalized customer experiences, and streamlined operations. Yet, this promise is over-shadowed by significant challenges:

The Black Box Conundrum:

Financial institutions are wary of AI's "black box" nature. *How can they trust models whose decision-making processes are opaque?* This lack of transparency hinders regulatory compliance and erodes stakeholder trust.

The False Positive Concern:

The financial sector grapples with the high costs of false positives. Erroneous fraud alerts, inaccurate risk assessments, and unwarranted compliance flags lead to high cost, operational inefficiency and damaged customer relationships.

This paper explores how Explainable AI (XAI) offers a solution. By making AI's decision-making processes transparent and understandable, XAI can build the trust necessary for widespread GenAI adoption in finance. Furthermore, it can significantly enhance model accuracy, reducing the incidence of costly false positives.

By delving into XAI's potential, this paper aims to guide financial institutions towards a future where they can confidently harness the power of GenAI while maintaining the transparency and accountability crucial to their industry.

Deep Dive – The High Cost of Black Box AI in Finance

The opaque nature of GenAI models refers to the difficulty in understanding how these models arrive at their decisions or predictions. Unlike traditional statistical models, which often rely on transparent mathematical formulas, GenAI models, particularly deep neural networks, operate on complex, non-linear patterns within vast datasets. This complexity poses significant risks in the financial sector, where trust, accountability, and clear explanations are paramount.

These risks can be categorized into operational, reputational, and regulatory risks.

Fig 1: Questions in the mind of the business leaders while thinking of the opaque nature of GenAI models

Operational Risks

False Positives:

A primary concern is the generation of false positives. In **fraud detection**, for instance, a false positive can lead to the wrongful highlighting of legitimate transactions, increasing operational costs in resolving and eventually approving the transactions, but with a potential delay.

Model Drift:

GenAI models are susceptible to model drift, where the model's performance deteriorates over time due to changes in data distribution. Without explainability, it becomes difficult to identify the root causes of performance degradation, leading to inaccurate predictions and poor decision-making.

Reputational Risks

Loss of Customer Trust:

Customers are increasingly demanding transparency and accountability from financial institutions. Black box AI models can erode trust by making it difficult for customers to understand the rationale behind decisions that impact them, such as loan approvals or insurance premiums.

Brand Damage:

High-profile failures of AI systems can have severe consequences for a financial institution's brand reputation. A lack of transparency in these situations can exacerbate the damage by creating an impression of irresponsibility and lack of control.

Increased regulatory scrutiny:

Financial institutions operating in a complex regulatory environment and using GenAI may face heightened scrutiny. The inability to explain AI-driven decisions can lead to regulatory penalties and reputational damage.

Regulatory Risks

Compliance risks:

Financial institutions must adhere to a complex web of regulations, such as anti-money laundering (AML) and Know Your Customer (KYC) requirements. Explainable AI is crucial for demonstrating compliance and providing evidence in case of regulatory inquiries.

Model Governance risks:

Without explainability, it is difficult to assess the fairness, bias, and ethical implications of AI models. This can lead to regulatory non-compliance and reputational risks.

Further, there is an overall increased Audit and Monitoring Costs – The lack of transparency in AI models necessitates increased audit and monitoring efforts to identify and mitigate potential risks. This translates to higher operational costs and resource allocation.

Explainable AI – Shedding Light on the Black Box

Explainable AI (XAI) is a set of techniques that enable humans to comprehend and trust the results and output created by machine learning algorithms. With XAI, it is possible to understand the 'why' behind AI's decisions, providing crucial insights into the factors that influence model outcomes.

Explainable AI (XAI) is therefore, key to unlocking the potential of Generative AI while mitigating its risks. By providing transparency into the decision-making process, XAI transforms the black box into a glass box.

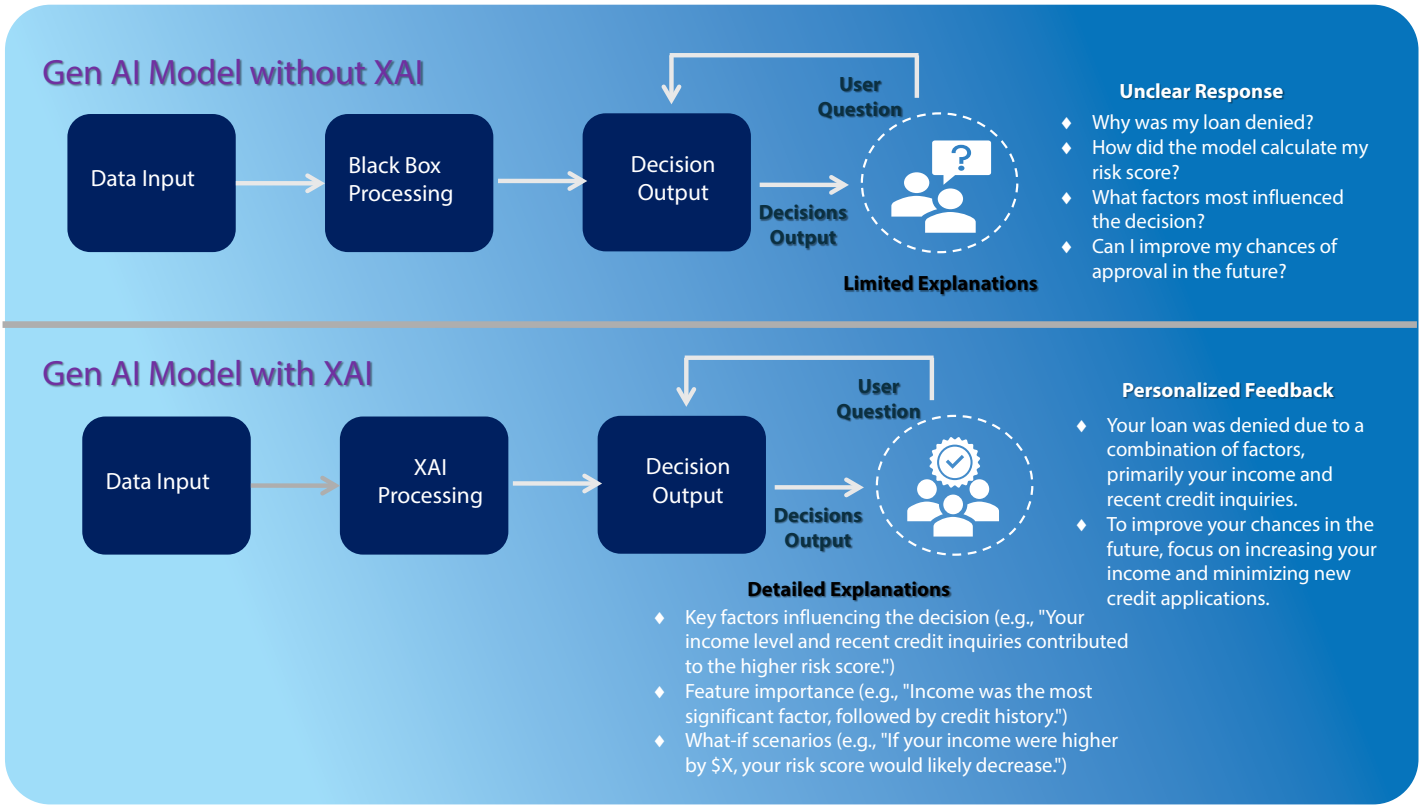

Fig 2: Comparison of outputs through GenAI to assess loan applications

Operational Benefits:

Improved Decision Making:

XAI offers insights into why a model makes specific recommendations, allowing businesses to identify patterns, trends, and anomalies. This leads to more informed decisions, reducing errors and optimizing processes.

Enhanced Model Understandability:

By understanding how a model arrives at its conclusions, organizations can identify and address biases, errors, or outdated information, leading to more accurate and reliable models.

Increased Accuracy:

XAI can streamline operations by automating tasks like root cause analysis and error detection, freeing up valuable resources.

Reputational Benefits:

Building Trust:

Transparency is the cornerstone of trust. By explaining the logic behind AI-driven decisions, financial institutions can build customer confidence and loyalty.

Mitigating Negative Publicity:

In case of errors or failures, XAI can help explain the situation, demonstrating transparency and accountability.

Strengthening Brand Reputation:

A commitment to XAI positions a financial institution as a leader in responsible and ethical AI usage.

Regulatory Benefits:

Compliance Adherence:

XAI provides a clear audit trail, making it easier to demonstrate compliance with regulations and respond to inquiries.

Risk Mitigation:

By understanding the factors influencing model decisions, organizations can identify and address potential regulatory risks.

Reduced Audit Burden:

XAI can streamline the audit process by providing clear explanations of model behavior, reducing the time and resources required.

In essence, XAI is not just a technical solution; it's a strategic imperative. By embracing explainable AI, financial institutions can unlock the full potential of generative AI while protecting their bottom line, reputation, and regulatory compliance.

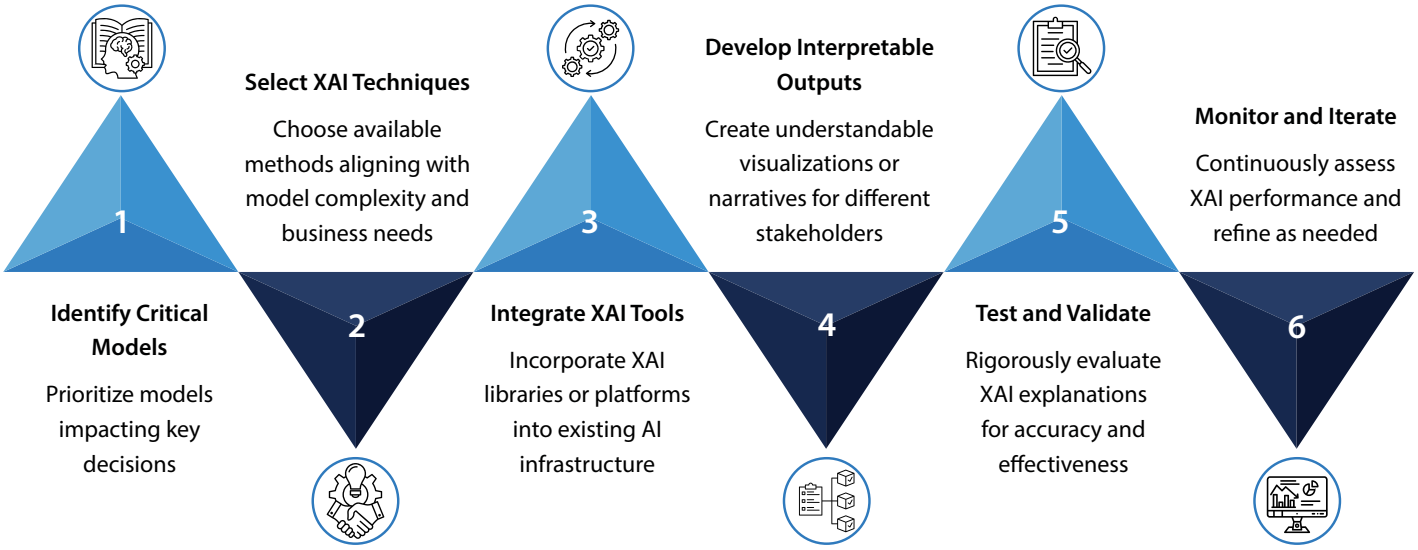

Fig 3: Six Steps to Deploy Explainable AI(XAI)

The Path Forward

XAI is not merely a technological advancement; it's a strategic imperative for financial institutions seeking to harness the power of generative AI while mitigating risks. By shedding light on the black box, XAI fosters trust, improves decision-making, and ensures regulatory compliance.

The journey to a fully AI ecosystem is fraught with risks, and therefore explainability is important. Developing effective XAI techniques for complex models remains an ongoing research area. Moreover, communicating complex XAI insights to non-technical stakeholders can be challenging.

Despite these hurdles, the potential benefits of XAI are undeniable. By investing in research, development, and implementation, financial institutions can position themselves as leaders in responsible and ethical AI, building trust and gaining a competitive edge.

Ultimately, the successful adoption of XAI requires a collaborative effort involving business, operations, and technology leaders. By working together, the financial industry can unlock the full potential of AI while safeguarding its future.

About the Authors

Sumeet Verma

Sumeet Verma is a Senior Consultant with over 15 years of progressive experience in financial services, Sumeet is a seasoned professional specializing in AI-first adoption strategies and financial crime detection and mitigation. He has a proven track record of helping clients navigate complex business challenges, delivering impactful solutions, and fostering strong client relationships across North America, Europe and APAC. His expertise in AI-driven technologies enables him to provide innovative insights and strategies that drive operational efficiency and enhance risk management for financial institutions. He can be reached at sumeet.verma@infosys.com.

Advait Rege

Advait Rege is a Senior Principal has more than two decades of experience in FSI Operations and Consulting. At IC, he is responsible for clients, pursuits, and programs in capital markets, retirement services and the wider financial services space. He can be reached at advait_rege@infosys.com.

For more information, contact askus@infosys.com

Infosys®
Navigate your next