Balancing Innovation and Risk: The Transformative Potential of GenAI for KYC Onboarding in the Banking Sector

Dr. Henry Balani*

June 25, 2024

Abstract

This paper aims to highlight the impact of Generative Artificial Intelligence (GenAI) on the Know Your Customer (KYC) onboarding process, providing banks with a deeper understanding of the potential advantages and risks of adopting this new tool. The different roles of 'traditional' Artificial Intelligence and Gen AI are discussed, along with the complementary roles these technologies play in addressing banking regulatory requirements while also balancing the benefits of cost efficiencies through innovation with the risks of new technology introductions. The paper also describes use cases that leverage GenAI specifically for KYC onboarding and associated considerations when implementing in a bank's KYC process.

Keywords: GenAI, AI, KYC, AML, Financial Crime, Banking, Regulations

*Corresponding Author: Henry Balani, Global Head of Industry & Regulatory Affairs, Encompass Corporation. Email: henrybalani@encompasscorporation.com

1. Introduction

Artificial Intelligence (AI) is driving significant transformations across various business functions within the banking sector, redefining traditional approaches to regulatory processes and promising unprecedented efficiencies. Generative Artificial Intelligence (GenAI) tools, including Open AI ChatGPT, Google Gemini, Anthropic Claude AI are at the forefront of this transformation. GenAI has the potential to revolutionize customer interactions, turning what was once mundane and routine processes into a sophisticated system of workflow processes and intelligent chatbots. Chatbots can, among other capabilities, provide personalized and efficient service through online banking platforms along with improved regulatory filing requirements and financial crime detection.

Moreover, GenAI has improved risk assessment processes in a range of financial services, including credit scoring and investment advice, setting new standards and paving the way for significant innovation in banking. However, the adoption of AI and GenAI is not without challenges. Concerns about security, data privacy, lack of transparency, and bias in model development and application all need to be considered. Therefore, it is crucial to weigh the benefits against the risks to determine the true value of generative AI technology.

In the context of KYC onboarding, GenAI offers a streamlined approach to data and risk management. It enhances the user interface and customer experience while automating mundane tasks, thus reducing errors and increasing customer satisfaction. This article will explore how GenAI can be utilized to develop transparent and comprehensive customer profiles based on behavioral patterns, potentially revolutionizing KYC and Anti-Money Laundering (AML) risk management processes.

2. Understanding GenAI's potential in KYC Onboarding

To comprehensively understand the potential of GenAI in the KYC onboarding process, it is essential to first describe current methodologies employed. Regulated entities, including banks, are mandated to assess the risk profiles of potential customers—both individuals and corporate entities—to determine their suitability and process for onboarding. This requirement is driven by regulations that require a customer identification process and screening potential customers against sanctions and other government-issued lists. Once a risk profile is established, the appropriate due diligence process is followed, with more comprehensive examination and validation of higher risk customers.

For instance, in the United States, banks must first undertake a customer identification process and screened against the Office of Foreign Assets Control (OFAC) Specially Designated

Nationals (SDN) list[1]. This list includes individuals and companies owned or controlled by, or acting on behalf of, targeted countries. It also enumerates individuals, groups, and entities such as terrorists and narcotics traffickers designated under non-country-specific programs. According to these regulations, banks are required to block the assets of entities identified on these lists, with prohibitions against U.S. persons dealing with them, under the threat of criminal enforcement.

Similar sanctions lists are published across various jurisdictions, including the UK Sanctions List[2], the EU Sanctions List[3], and the UN Security Council Consolidated List[4]. It is important to note that KYC onboarding is a component of a bank's broader financial stability risk program, designed to assess a customer's risk beyond mere regulatory compliance, thus categorizing their overall risk exposure.

The initial phase of the KYC onboarding process involves the collection of relevant documents and proof of identity. Banks then verify these credentials and documents for authenticity and relevance. Based on the assessed risk level, the bank will perform either Customer Due Diligence (CDD) or Enhanced Due Diligence (EDD) before an account is opened. Furthermore, banks conduct ongoing reviews of customer accounts, the frequency of which depends on the assigned risk level and the additional services required by the customer. Subsequent customer transactions that potentially suspicious will require filing of suspicious activity reports to regulatory authorities or risk penalties.

3. Use of AI in KYC Onboarding

While GenAI is a relatively new entrant to the KYC onboarding landscape, the utilization of AI to streamline and enhance this process is not unprecedented. AI has been employed in KYC processes for several years now, and thus, concerns about safety, reliability, and the potential displacement of existing compliance staff are not exclusive to GenAI. These issues have been thoroughly explored within the broader context of AI. Both 'traditional' AI and GenAI should be perceived as augmentative tools that enhance the roles of Compliance Analysts and Money Laundering Officers tasked with identifying financial crime and regulatory reporting.

---

[1] Specially Designated Nationals And Blocked Persons List (SDN) Human Readable Lists | Office of Foreign Assets Control (treasury.gov), available at: https://ofac.treasury.gov/specially-designated-nationals-and-blocked-persons-list-sdn-human-readable-lists (accessed June 24 2024)

[2] The UK Sanctions List - GOV.UK (www.gov.uk), available at: https://www.gov.uk/government/publications/the-uk-sanctions-list

[3] European Union sanctions | EEAS (europa.eu), available at: https://www.eeas.europa.eu/eeas/european-union-sanctions_en

[4] United Nations Security Council Consolidated List | United Nations Security Council, available at: https://www.un.org/securitycouncil/content/un-sc-consolidated-list

These technologies can be used to address manual, repetitive tasks such as identification, screening and matching, thereby allowing analysts to focus on higher value-added activities, including interpreting, analyzing, and reporting on complex money laundering cases.

Furthermore, the efficacy of any technological tool is contingent upon the completeness and accuracy of the data employed in business processes. In the context of KYC onboarding, there can be gaps in both customer data and external reference data. Both AI and GenAI are susceptible to data challenges, which are particularly critical when accurate information is essential for determining customer risk during the onboarding process. Consequently, while AI and GenAI strive to address these data challenges, banks must ensure their databases are as accurate as possible using tools beyond the AI domain.

4. GenAI as a subset of AI

'Traditional' AI is a broad term encompassing technologies that exhibit human-like intelligence, including a range of tools and methodologies from rule-based systems to machine learning and deep learning models[5] [6]. 'Traditional' AI is utilized extensively in the banking industry, particularly in data analytics to profile customers for additional revenue opportunities. Typically a rules based approach that incorporates predefined rules and logic to analyze past banking and purchasing behaviors to reveal cross-selling opportunities for additional financial products as an example. Data analysis can also assess a customer's creditworthiness to better predict transaction risks. Risk management is a key decision point for investment banks aiming to market their products to both institutional and private investors. Banks leverage data analytics and machine learning to discover patterns of unnecessary discounts offered to customers, leading to refined product offerings and increased revenue for a bank. Understanding customer preferences by analyzing large datasets, including social media, product usage, and demographic data, is highly valued by marketing groups.

GenAI, by contrast, focuses on generating new content based on predetermined prompts. It can create text, images, music, and more, leveraging deep learning architectures trained on extensive text-based datasets. GenAI often leverages unsupervised or semi-supervised learning, where models learn patterns from unstructured data. GenAI is considered more broad, where a

---

[5] Traditional AI vs. Generative AI: A Breakdown | CO- by US Chamber of Commerce

[6] Goldman Sachs | Intelligence, available at: https://www.goldmansachs.com/intelligence/artificial-intelligence/index.html?chl=ps&plt=bi&cid=638280346&agp=1316117710182195&kid=artificial%20intelligence%20business&mtype=p&msclkid=a50913cea9cc16b82ef37696425273ba&gclid=a50913cea9cc16b82ef37696425273ba&gclsrc=3p.ds&utm_source=bing&utm_medium=cpc&utm_campaign=CPB_CBA_IntelligenceSourceOfChoice_UNB_NA_AMRS_USA_SEM_BING_AWR_O-44S2Z_253_2023&utm_term=artificial%20intelligence%20business&utm_content=CPB_CBA_UNB_Intelligence_AMRS_USA_RSA_AI

larger range of tasks can be handled, adapting to difference contexts including creative writing, art creation and chatbots. Outputs can vary, even with the same input which can lead to 'hallucinations', creating unique or occasional false results. Accuracy of output is heavily reliant on the quality and quantity of data used in its training models. The revolutionary aspect of GenAI is its human language interface, which facilitates natural interaction, requiring minimal upfront training for Compliance Analysts to become productive.

5. Comparative advantages of GenAI in KYC Onboarding

Traditional AI has been instrumental in automating KYC processes, significantly enhancing customer identification efficiency. This automation has streamlined verification processes, reducing the time required for data retrieval and interpretation of causal network relationships. Additionally, AI has improved risk profiling, facilitating the assessment of client risk levels for banks and online businesses[7]. Optical character recognition (OCR) tools, powered by AI, have further optimized onboarding by enabling automated extraction of relevant information from snapshots or scanned documents, thus expediting the KYC process.

GenAI however, brings substantial advancements to the KYC onboarding process, particularly in investigation and screening. With its text-based approach, GenAI offers a more intuitive and interactive user experience[8]. A GenAI-powered chat-based KYC workflow can serve as an interactive and transparent tool for investigating and assessing risks associated with individuals or entities. The chatbot can collect relevant information from prospective customers, ensure forms are fully completed and risk rate the application for appropriate follow up by human analysts as required. By integrating data from various trusted sources, GenAI can efficiently construct and clarify risks based on behavior patterns throughout the KYC/AML process[9]. Relevant follow up actions can include creation of suspicious activity reports for high risk applications for automatic regulatory filing. This approach enhances the customer experience by simulating human behavior and workflow processes, automatically requesting additional information, leveraging face and image recognition for validation, and providing guided instructions to improve the onboarding process without human intervention. This represents a significant improvement over traditional AI, which often involves manual, repetitive

[7] Tatsiana Isakava (2022), "AI for KYC automation: revamping risk profiling" available at https://indatalabs.com/blog/ai-for-kyc-automation

[8] Towards interactive smart screening with generative AI in KYC workflows (moodys.com) available at: https://www.moodys.com/web/en/us/kyc/resources/insights/towards-interactive-smart-screening-with-generative-ai-in-kyc-workflows.html

[9] Maksym Biellai (2022), "How AI Improves Customer Onboarding in Fintech" available at: https://www.fintechweekly.com/magazine/articles/how-ai-can-improve-the-customer-onboarding-process-in-fintech

interactions across multiple parties using disparate, non-integrated tools, typically slowing down the onboarding process and affecting the bank's competitive market positioning.

A key benefit of GenAI is its ability to comprehend and execute complex tasks using natural human language. Unlike traditional AI, which frequently struggles with natural language understanding and contextual interpretation, GenAI can use chat bots to follow entire conversations, maintain context, and respond appropriately[10]. Users can utilize pronouns, refer to previous segments of the conversation, request summaries, or update outputs based on predefined parameters, and GenAI will respond accordingly. This capability enhances the overall workflow efficiency and reduces the need for manual interpretation by analysts.

GenAI also excels in automating data entry and verification, thereby minimizing errors and accelerating the onboarding process. Enhanced accuracy in document identification and categorization is another notable benefit. While traditional AI uses OCR to identify specific data elements on documents for further processing, GenAI extends this capability by identifying relevant data required for the onboarding process. For instance, when onboarding a new corporation, GenAI can extract necessary data from various document types such as articles of incorporation, annual returns or trust documents to include the entity's name, address, nature of operations, owners, managers, directors, and incorporation date as examples. Explanations of the extracted data elements that provide justification for audit purposes can be included. The tool would generate the relevant JSON schema ingested into a structured database for review and prioritization. If required data is missing, GenAI can automatically source other documents, either internally or externally as relevant, extract the relevant data elements and attempt to complete the customer profile prior to further KYC processing. Manual input would be required only when the necessary data elements cannot be obtained.

Gen AI tools and other AI services can even be used in sequence to independently validate the accuracy and relevance of the extracted data. GenAI can flag missing or incomplete data and review additional documents automatically, reducing the need for human intervention and expediting the overall KYC process. Essentially, GenAI facilitates the process of intelligent document processing, automating manual, repetitive steps to improve the efficiency and accuracy of the KYC onboarding process.

6. Considerations when integrating GenAI in KYC Onboarding

The integration of GenAI into the KYC onboarding process, like any new technological tool, is not without its downsides and presents several challenges and risks. The inherent

[10] Aarti Nair (2022), "How Conversational AI in KYC Improves Verification Process" available at: https://verloop.io/blog/ai-in-kyc/

complexity, unpredictability, opaqueness, and potential for data breaches in these models requires robust risk management strategies. For instance, when relying on GenAI to review documents such as an 'articles of incorporation,' we assume that the OCR process accurately scans the document and that GenAI correctly categorizes the data elements to assess the risk of the corporation being onboarded. This accuracy is contingent upon the model used to train the GenAI tool and the effectiveness of the OCR engine. However, the algorithms employed, the data collected, and the basis for data categorization are often opaque and difficult to explain. As KYC onboarding is a regulatory requirement for banks, the implementation of GenAI necessitates careful review and constant adaptation of internal operations and processes, posing significant challenges for organizations to maintain accuracy.

Banks operate within a heavily regulated industry, and given their impact on national economies, the models used require scrutiny. Continuous examination is necessary to ensure that appropriate data sources are used for their intended purposes. Sam Altman, co-founder of OpenAI and the firm responsible for launching ChatGPT, has described it as "incredibly limited, but good enough at some things to create a misleading impression of greatness," and he advises against using the tool for essential tasks[11]. GenAI models often rely on third-party data sources for training, which can include confidential, anecdotal, or biased information, potentially leading to inaccurate decisions. For example, in sanctions screening, a significant proportion of names on the OFAC SDN list are of Hispanic or Islamic origin, especially Mohammad[12]. Consequently, GenAI models trained on Anglo-centric name screening techniques may produce poor quality matches that fail to account for the nuances of non-Anglo names, resulting in biased or inaccurate screening outcomes. These biases can be further reinforced as GenAI models learn from past customer identification processes, leading to magnified errors.

The adoption of GenAI also raises significant legal and regulatory risks. GenAI models can legitimately be considered "black boxes" that make decisions based on algorithms and rules, which KYC analysts subsequently rely on. Regulatory processes require high levels of transparency and predictability for audit purposes, and the lack of transparency in GenAI models means Regulators may be uncertain about the outcomes and consistency of risk decisions for onboarded customers. Furthermore, GenAI's natural language inquiry capabilities and flexibility can lead to variability in how follow-up inquiries are framed for risk decision-making. Maliciously crafted inputs (prompts) can bypass carefully developed compliance procedures, leading to unsatisfactory KYC onboarding decisions, adversely impacting the bank's KYC policies, and risking regulatory penalties. Therefore, a balance is required to allow analysts the

[11] Terence Jackson (2023), "Exploring the Security Risks of Generative AI" available at: https://www.forbes.com/sites/forbestechcouncil/2023/04/19/exploring-the-security-risks-of-generative-ai/?sh=490dd7873594

[12] Why Is Mohammad The Most Sanctioned Name? — Castellum.AI

flexibility to use GenAI while adhering to the required risk decision framework as per the bank's KYC policies.

Privacy laws present another significant concern from a legal and regulatory perspective[13]. The use of GenAI to request additional customer information for compliance purposes during the KYC onboarding process can result in breaches of privacy laws such as GDPR[14] in Europe and HIPAA in the USA[15]. GenAI models often train on initial source data and may subsequently access confidential sources, raising the risk of data being shared with third parties. While companies like OpenAI assert that they do not access confidential data sources for model training, there remains a security risk that these tools could access specific confidential sources as part of their operations. This risk is exacerbated when banks are unaware of the nature of outreach activities, potentially leading to fines and negative court judgments for privacy and confidentiality violations. Additionally, negative public perceptions of individuals or corporate entities used to train models can lead to false positives, requiring manual investigation and creating inefficiencies.

From an infrastructure perspective, GenAI can exacerbate data and privacy risks due to its reliance on large datasets and the generation of new data for training other models. This interdependence can introduce biases, poor-quality data, and unauthorized access to confidential and sensitive information. Bias is particularly critical given the datasets GenAI tools are trained on. While this issue has been previously identified, further research is needed to understand how different datasets can ensure less potential bias and reduce "hallucination" in GenAI decisions.

7. Conclusion

GenAI has the potential to transform the KYC onboarding process by offering numerous benefits, such as an improved user experience, efficient data management, enhanced accuracy, and automation. GenAI offers regulatory benefits by improving accuracy and identification. When used along with traditional AI, GenAI can provide comprehensive KYC solutions that leverage the strengths of both technologies, potentially reducing the overall costs of regulatory compliance.

However, it is crucial to acknowledge the challenges accompanying GenAI adoption. These include the complexity and opaqueness of the models, their unpredictability, the potential for data breaches, and the necessity for robust risk management strategies. While this article has

[13] Aarti Nair (2022), "How Conversational AI in KYC Improves Verification Process" available at: https://verloop.io/blog/ai-in-kyc/
[14] What is GDPR, the EU's new data protection law? - GDPR.eu available at: https://gdpr.eu/what-is-gdpr/
[15] Health Insurance Portability and Accountability Act of 1996 (HIPAA) | CDC available at: https://www.cdc.gov/phlp/publications/topic/hipaa.html

concentrated on the KYC onboarding process, similar benefits and challenges apply to other banking processes, both regulated and unregulated, impacting privacy, confidentiality, and various business issues.

Therefore, while GenAI holds great promise for enhancing KYC onboarding, banks must be cognizant of these risks and implement appropriate measures to mitigate them. Achieving a balance between leveraging the benefits of GenAI and managing its risks is essential for the successful implementation of GenAI in KYC onboarding.

---

*Corresponding Author

Dr. Henry **Balani** | Global Head of Industry & Regulatory Affairs

t +1 332-245-4398 | +1 773-771-0234
e henrybalani@encompasscorporation.com
w encompasscorporation.com