

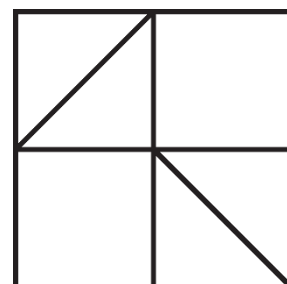
Magpie Projects

COLLECT . CONNECT . CREATE

AI and Advanced Analytics in Anti-Financial Crime Compliance

White Paper

December 2025



Executive Summary

Financial institutions operate within a complex landscape fraught with risks stemming from various illicit activities such as money laundering, fraud, and corruption. Effectively managing these risks is paramount to maintaining trust, integrity, and regulatory compliance within the financial sector.

In one notable case, Binance, one of the world's largest cryptocurrency exchanges, received a substantial fine of USD 4.3 billion by US regulatory authorities¹. This penalty was imposed due to significant shortcomings in Binance's KYC processes, which failed to prevent illicit activities such as money laundering and fraud on its platform.

Similarly, in 2019, the British bank Standard Chartered agreed to pay a combined fine of USD 1.1 billion to US and UK authorities for violating Anti-Money Laundering (AML) and economic sanction rules. Standard Chartered was accused of processing hundreds of millions of dollars in transaction and providing countries such as Cuba, Sudan or Syria access to the US financial system². These fines highlight the ongoing scrutiny by national regulatory authorities and the high stakes associated with KYC compliance for traditional banking institutions.

These recent high-profile cases underscore the severe consequences of failing to adhere to stringent Know Your Customer (KYC) protocols and the fines financial institutions can face.

Traditional Anti-Financial Crime (AFC) methods have proven to be prone to false positives (results erroneously indicating a given condition), consuming significant time and resources in the process. Only a small fraction of generated alerts are actually true positives (based on our previous experience approx. 2-5%) of which an even smaller fraction leads to a Suspicious Activity Report (SAR). Moreover, AFC professionals are frequently confronted with poor data quality, hindering their ability to accurately identify and mitigate risks. AI can help by automating routine tasks, improving efficiency, and providing predictive insights, allowing AFC professionals to concentrate on more complicated and critical cases.

Based on an analysis conducted with our clients, a gain of 40-60% in employee efficiency is possible by leveraging AI.

Moreover, AI-powered solutions can adapt and evolve in real-time, staying ahead of emerging threats and regulatory requirements. By leveraging AI, financial institutions can significantly enhance their AFC capabilities, bolstering their defenses against financial risks while optimizing operational efficiency, and resource allocation.

For example, the challenge of fragmented sources can be overcome by leveraging Large Language Model (LLM)-based agents. These agents can conduct internet research, search through open-source lists, comprehend additional information, and reason about true or false positives. For instance, if a possible match is found on a sanctioned individual but the date of birth of the searched-for individual is not available, additional information such as the year the individual graduated from university can be used to assess the likelihood of a match. This advanced reasoning capability allows for more accurate and efficient identification of potential risks, enhancing the overall effectiveness of KYC processes.

Integrating these AI-driven solutions not only strengthens an institution's ability to meet regulatory demands but also ensures a more dynamic and responsive approach to risk management in an increasingly complicated financial and geopolitical environment.

¹ (U.S. Department of Justice, 2023)

² (U.S. Department of the Treasury, 2019)

AI represents a paradigm shift in the fight against financial crime, offering unparalleled opportunities for innovation and advancement within the financial sector. Embracing AI-driven solutions is not merely an option but a necessary step for financial institutions seeking to safeguard their assets, reputation, and regulatory compliance in an increasingly complex environment.

TABLE OF CONTENTS

Executive Summary	2
1 Current Challenges in Anti-Financial Crime Measures	5
2 How AI and Advanced Analytics Can Help Mitigate These Challenges	7
2.1 Data Quality	7
2.2 Varying Data Sources	8
2.3 Parameter Tuning	10
2.4 Alert Scoring	12
3 Conclusion	14

1 Current Challenges in Anti-Financial Crime Measures

Traditional Anti-Financial Crime (AFC) monitoring methods typically entail a dual approach: rules-based identification of potential risk cases combined with manual assessment. This method, however, presents several challenges for AFC professionals, such as high workloads or redundant work processes. Leveraging the capabilities of Data Analytics, particularly Artificial Intelligence (AI), offers promising solutions to address these shortcomings. In the subsequent discussion, we will delve into these weaknesses and explore how advancements in AI can enhance AFC practices.

A High Number of False Positive Alerts

Basic measures utilized by AFC professionals within financial institutions are ongoing Customer Due Diligence, the screening of customers against watchlists and continuous transaction monitoring. To increase efficiency and the quality of the work, various tools are implemented to help detect potential suspicious activities by the institution's clients. To establish an internal database for these tools it is important to screen publicly accessible watchlists and to monitor transactions. The outcomes of both measures can be protocolled in a database which must be kept up to date continuously.

One of the most common problems confronted by AFC professionals is false positives, i.e. alerts generated by detection tools and categorized as possible hits which have no criminal nature. These are especially prevalent with watchlist screenings and transaction monitoring. In the case of transaction monitoring, false positives usually occur when the respective applied weight of parameters or choice of parameters in itself triggers high amounts of false positive alerts. Financial institutions usually base these parameters on client-specific risk levels (i.e., location, occupation or business areas, PEP (Politically Exposed Persons) or SOE (State-Owned Enterprises) status etc.), information derived from behavioral information and further client observations. The screened watchlists are usually comprised of sanction lists by governments or organizations such as the EU or the UN, PEP lists, international search warrants, etc. Additionally, clients are screened for adverse media or against other internally curated lists. Watchlists are screened continuously and are subject to constant changes.

Another cause for a significant number of alerts is the generation of alerts based on types of information, such as high-volume transactions that exceed pre-determined thresholds, transactions to governmental bodies or transactions to individuals that have the same or similar names as flagged PEPs. Additionally, a suboptimal tuning of screening rules (i.e. fuzziness of possible hits) may result in a high number of alerts as well. Only a small fraction of alerts warrant further investigation and an even smaller fraction requires the filing of a SAR (Suspicious Activity Report).

Amount of Required Effort in Investigating Alerts

A common challenge faced by AFC professionals is the significant effort required to assess alerts effectively. This effort is described in the following scenarios:

- Analyzing the corporate structures of companies, especially complex holding structures, demands for extensive manual research to ascertain stakeholder information and identify ownership structures.
- Utilizing different sources and systems can result in an increased amount of time and effort spent on researching individual alerts. Additionally, sources which constantly update their information (i.e., press archives or free internet searches) must also be assessed constantly.

- Individually investigating and assessing different alerts, which result from the same suspicious transaction in the case of transaction monitoring.
- Recording the results of investigations in a comprehensible audit trail.

Data Quality and Data Sources

The data sources used for the monitoring and screening processes are often incomplete, outdated, or contain errors, which contribute to ambiguity.

A significant portion of the data managed by AFC professionals often arrives incomplete or flawed. This requires preprocessing before it can be utilized for investigative or evaluative purposes.

Checking and cleansing the data is a laborious process demanding a considerable amount of time and effort. Nevertheless, the pre-processing of data is indispensable, given incomplete or faulty data can result in false positives (or, even more alarming, in false negatives, i.e. results which wrongly indicate that a particular condition or attribute is absent) and performance issues during further analysis. Pre-processing, amongst others, reduces spelling mistakes and the number of false alerts and thereby increases the accuracy of positive hits.

In addition, external data sources used by AFC professionals may lack adequate quality control, fact-checking or governance. This is particularly evident with sources such as news articles, media content, and social media platforms. Discrepancies in information on the same client from internal or external sources may arise. These discrepancies can manifest in:

- Variations in spelling of the same client name
- Different data formats
- Duplicate entities with contradicting attributes

Other Challenges

Screening clients on watchlists can result in a multitude of potential hits. For one, this might be due to the selection of search parameters. Ubiquitous client names may also yield numerous potential hits. Given the potentially incomplete or inaccurate client data, confirming these potential hits can become exceptionally challenging, requiring a consultation on fragmented systems, internet search or looking up open-source databases.

But riskier yet is the possibility of undetected hits. These false negatives might be the outcome of incomplete or faulty data or wrongly set parameters. Undetected financial crime hits can have a disastrous outcome and may potentially be passed on through further checks resulting in follow-up errors that could lead to substantial fines or penalties levied against the financial institution in addition to potential reputational damage.

2 How AI and Advanced Analytics Can Help Mitigate These Challenges

2.1 Data Quality

As mentioned previously, unprocessed data often comes with obstacles. Typical challenges include unstructured, outdated, missing, duplicated and incomplete data. After the data collection from suitable sources, standard procedures involve setting up a data structure that suits the internal database, assessing and validating the data, cleaning it, formatting it into the target format and loading the data into the internal database (or data warehouse). The summarization of these tasks is characterized as the ETL (Extract, Transform, Load) process.

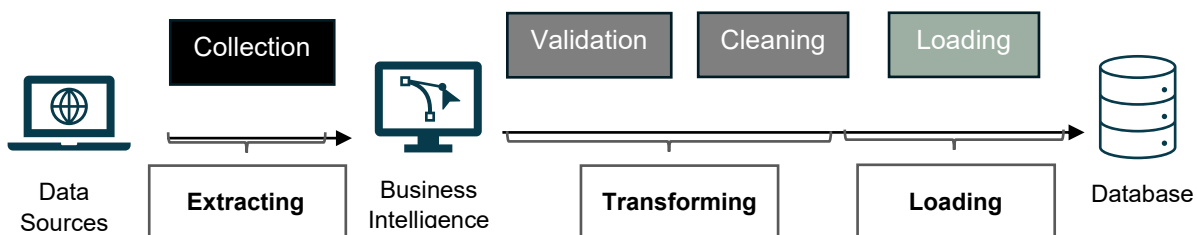


Figure 1: Schema of an ETL process.

AI can help to automate every part of the process chain and facilitate its adaptation by the human user. AI models consisting of certain selected ML (Machine Learning) algorithms can be trained to transform data regarding its quantity, initial quality, structure and the target use. While using it iteratively, the model's sensitivity can be improved, and the rate of errors decreased. Thus, the initial quality of the data is improved, and an enhanced outcome can be expected. The space of acceptable data sources gets larger, and more information can get processed with less effort.

Large language models (LLMs) utilize natural language processing (NLP) by adapting deep learning techniques to understand and generate human language. For example, LLMs can be used to apply company detection logics or to identify the gender of those individuals where the gender flag is missing. Information from unstructured "Additional Information" tags such as biography, year of graduation etc. can be extracted leveraging LLMs and placing in a structured format.

By using prompts, it is possible to interact with the LLM with a query in the form of textual input. The model will process the query and output a result to the query based on the model's degree of precision and number of iterations. Prompts can be used to give commands regarding which data should be used.

Unstructured textual data can be processed with the help of LLMs. Unnecessary text contents can be filtered out as part of the extraction task. For example, when screening watchlist entries, it is possible to condition the search process for certain types of information. By understanding the context and interpreting the informative value of textual data, LLMs can pre-process data and derive conclusions from the watchlist. While results ultimately have to be checked by a human user, AI nonetheless increases the efficiency of the ETL process and improves the reliability of automated solutions.

When comparing publicly available information from different sources, LLMs may also serve to validate the data. The user can categorize data sources and attach individual degrees of reliability to them. By applying these levels, information from higher-rated sources can simply outrate

information from lower-rated sources. This means the human user can weigh the informative value of data relying on AI-generated preparatory work.

A further issue that can be solved by AI during the data transformation process is the treatment of multiple hits on one client. LLM models can handle the ambiguity of results and assign them correctly. This reduces the effort that is demanded by employees for subsequent checks.

Business Benefits of an AI-based data approach

- A holistic and automated ETL process. It can increase the speed and quality of AFC checks, reducing the demand for oversight by employees while simultaneously reducing costs significantly.
- A decrease in the error rate. By using a suitable AI model and choosing optimized parameters, false positives and negatives can be diminished. This not only minimizes unnecessary follow-up work and escalations but also increases the reliability of alerts and case assessments.
- AI models improve continuously. Therefore, the learning process of AI models does not end with the initial training process. The feedback received during the application with real-life data generates further optimization in precision, prediction accuracy and case-handling confidence.
- Tailored end-to-end solution. AI unifies the path from data collection to loading into the internal database. This makes the ETL process easier to follow and comprehend. Fragmentation and micro-management become less necessary, and the AI analysis provides a holistic view of the partial activities.

Challenges and Risks of an AI-based data approach

- AI models are often referred to as “black boxes” for several reasons. Contemporary models often consist of millions of parameters. This complexity makes it difficult to understand how individual input features are transformed into output predictions. For the acceptance by employees and supervisory authorities, AI applications should provide the determinants and decisive factors of their predictions to increase transparency and explainability.
- A model that is not sufficiently trained has a higher exposure to failure in each of the ETL steps. This can lead to overconfident filtering or insufficient cleansing of the data. It may also derive wrong interpretations and conclusions from watchlist entries which would lead to false positives and false negatives.
- Despite high accuracy, AI outcomes still require periodic human review, as some cases demand contextual judgment.
- AI models must be selected and tuned carefully; inappropriate training data or hyperparameter choices can slow learning, reduce stability, or cause overfitting.

2.2 Varying Data Sources

Since lists, e.g. compliance databases, the offshore leaks database of the ICIJ (International Consortium of Investigative Journalists), or PEP lists are the main source used for watchlist screenings, the internal handling of updates of these lists is of great importance. While data providers handle some cases, the regular update of internal lists is often a fragmented and manual process.

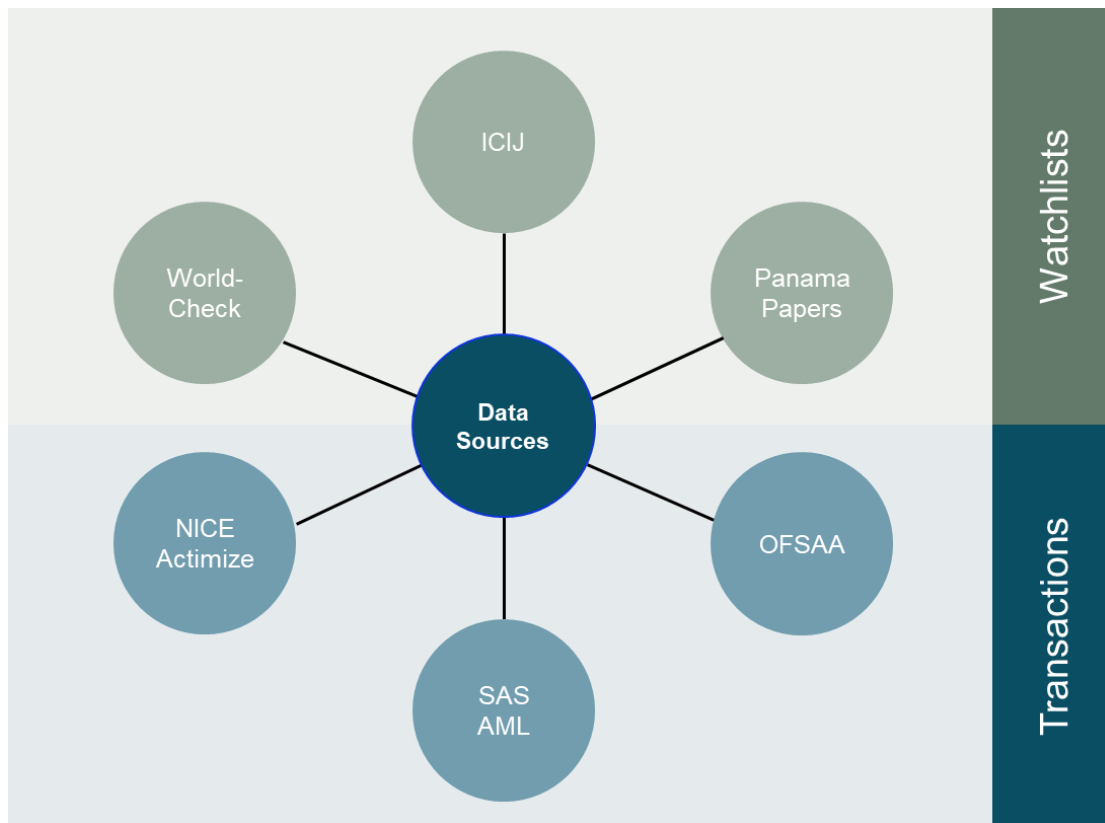


Figure 2: Exemplary watchlist and transaction monitoring sources.

Extensive data enrichment requires additional screening of adverse media. Typical sources are publicly accessible media, like news articles or legal reports. This includes both historic and current published documents.

As stated before, alternating sources can result in differences in spelling, format and other textual attributes. The ability to understand texts makes LLMs usable for linguistic similarity checks which can help to solve these issues. LLMs are also suitable for the comparison of different format types (e.g. European and US-American data formats). The following list states some of the steps LLMs can assist with, to ensure the quality of the input from varying data sources.

- **Spell Checks:** LLMs can serve as an advanced spell checker. They can identify misspelled words in texts by comparing them to a vast database of correctly spelled words. This helps to catch typos and spelling errors.
- **Contextual Understanding (reasoning):** Unlike traditional spell checkers, LLMs understand context. They can identify words that are spelled correctly but might be the wrong word in the context. For example, "there" versus "their", or "to" versus "too."
- **Handling Variations:** LLMs can recognize variations in spelling, including different regional spellings or alternative spellings of words. They can help normalize these variations to a standard form, depending on the specific requirements of the task. For example, when comparing data from the SEC (the US Securities and Exchange Commission) with data from the British counterpart, FCA, deviations in spelling can be neutralized by choosing one of the spelling types as default by the user. Further use cases are date formats, country codes and country-specific classification standards.
- **Customization:** Some LLMs can be fine-tuned or customized for specific domains or applications. This allows them to learn domain-specific terminology and spelling variations, making them even more effective in screening data within those domains.

Additionally, long text-chains can be trimmed to the data of interest by feeding the model with examples.

- Text formatting: Raw results can be evaluated and brought into the shape required for the input to the internal database. The shape may be adopted from the existing format or alternated by the user. For example, normalizing country names to ISO standards where the names are stored in different languages and formats.
- LLMs can evaluate text results immediately and rate them according to the parameters of interest. For example, when observing PEPs for suspicious activity, it is possible to predefine a scale of relevance for certain events. By using this scale as a parameter, the model can assess recent events accordingly.

Furthermore, AI can help with the screening of sources. By predefining the type of sources of information, the user can limit the universe of sources without listing every valid source by name. For example, by limiting the source set to 'US governmental', the LLM will process the limits and screen only hits that are relevant to the restriction, such as the FBI Public Records, US Department of State and The Office of Foreign Assets Control (OFAC).

Business Benefits of using AI for data sourcing

- AI can simplify the research process of data sources. It can search through the internet for suitable data sources within set boundaries and present them to human analysts for further review. This saves employee capacities and time.
- AI can be used for the automatized documentation of data sources that were used. For example, publicly accessible websites can go offline or change their URLs. In such a case, this would be recognized and documented by the model.
- In the case of pretrained language models, fewer examples are necessary (i.e. Few-Shot Learning).
- LLMs automate textual checks and lower error rates. This makes the results less vulnerable to human errors.
- Reduction of the number of false positives by filtering hits that triggered an alert because of variations in spelling or typos.

Challenges and Risks of using AI for data sourcing

- AI cannot rate sources without any input by a human user. The AI model must be trained on a set of sources based on conditions in order to integrate comparability and scalability to the contained information.
- The choice of qualified and non-qualified data sources must be reviewed by human analysts. In some cases, modifications must be made manually.
- The filtering of data sources can lead to the deletion of important information.

2.3 Parameter Tuning

Parameter tuning is a critical process in transaction monitoring to improve the detection of suspicious activities while minimizing false positives. In order to identify alerts, a rule set must be implemented that applies scoring mechanisms and sets thresholds. In transaction monitoring, behavioral and static flags are two types of criteria used to detect suspicious activities that could indicate fraud, money laundering, or other forms of financial crime.

Behavioral flags are designed to identify unusual behavior by comparing current transactions against an established baseline of historical behavior. These flags often involve statistical models or algorithms that track transaction patterns over time. Behavioral rules, however, are not adaptive

and consider how a client's transaction amounts, frequencies, or destinations are based on a predefined parameter set. For instance, if a customer who typically makes small, local transactions suddenly initiates large, international transfers, this change might trigger an alert only based on previously derived patterns. With time these patterns might change, and this requires a continuous tuning of the parameters. Some examples of the behavioral rules are:

- Transactions of the same entity or individual as an originator with same or very similar amounts above a certain threshold.
- Transactions of the same entity or individual as an originator or beneficiary with round amounts below a certain threshold.
- Transactions with similar amounts containing the same or a very similar sounding originator to beneficiary instructions.

Static rules are set within the context of look-up lists, countries of origins, non-resident status or other such predefined attributes. These are static and are not based on the behavior of the entities. Some static rules or flags could be:

- The domicile of a customer or a transaction counterparty in a predefined high-risk country
- The counterparty is a high-risk entity (such as payment service providers, virtual asset service providers, merchant service providers)
- The entity or individual in question has counterparties that appear in known shell company lists, or are named on open-source lists

Depending on the category, the parameters in question vary. Identifying these parameters and deriving them based on historical data requires a lot of manual effort. Often the parameter derivation based on empirical evidence is challenging due to the lack of data. In these cases, human users give input based on expert opinion. However, these approaches have limitations which can be solved with the help of AI.

Behavioral rule parameters

Some of these parameters include:

- Transaction amounts thresholds
- Aggregated transaction amounts thresholds across a certain period
- Transaction counts related thresholds

The challenge for derivation is the identification of what is to be considered normal behavior, to avoid false positives. Conventionally, this is done by analyzing historical data, which means that it remains static and requires continuous manual update. However, we know that human behavior evolves over time. An algorithmic approach for detecting anomalies in past transactions can identify changes which would not have been registered with traditional rules-based models. This can be achieved by implementing anomaly detection algorithms. Setting relative parameters can also enhance the anomaly detection logic. For example, rather than looking at plain thresholds, relative thresholds can be used on an entity or segment level. Finally, due to lack of proper historical data, parameters can be given by expert input.

Static rule parameters

Some of these parameters might include

- Name similarity thresholds (e.g. Levenshtein Distance, Cosine Similarity, Jaccard Distance etc.)
- Name lists for customer segmentation

- Address similarity or matching logics

Setting up such parameters is challenging as fuzzy matching mechanisms are used for identifying entities in certain look-up lists. If the thresholds for similarity are too low then these result in an excessive number of false positives, whereas a threshold set too high can lead to a coverage gap and false negatives. AI, especially language models, can be helpful in solving this task. LLMs can be used for generating labelled data for which threshold tuning can be applied. Additionally, LLMs can be leveraged to segment known clients into different groups and avoid fuzzy matching altogether.

2.4 Alert Scoring

The process of transaction monitoring involves both standard rules provided by common screening engines and a set of custom rules as per the institution's specific control needs. The screening engine takes the role of the detector of the alerts. It provides the necessary instruments for screening sources of interest, collecting data and rate the hits by a binary flag (positive or negative) or by a numerical scale of severity, based on the pre-given ruleset.

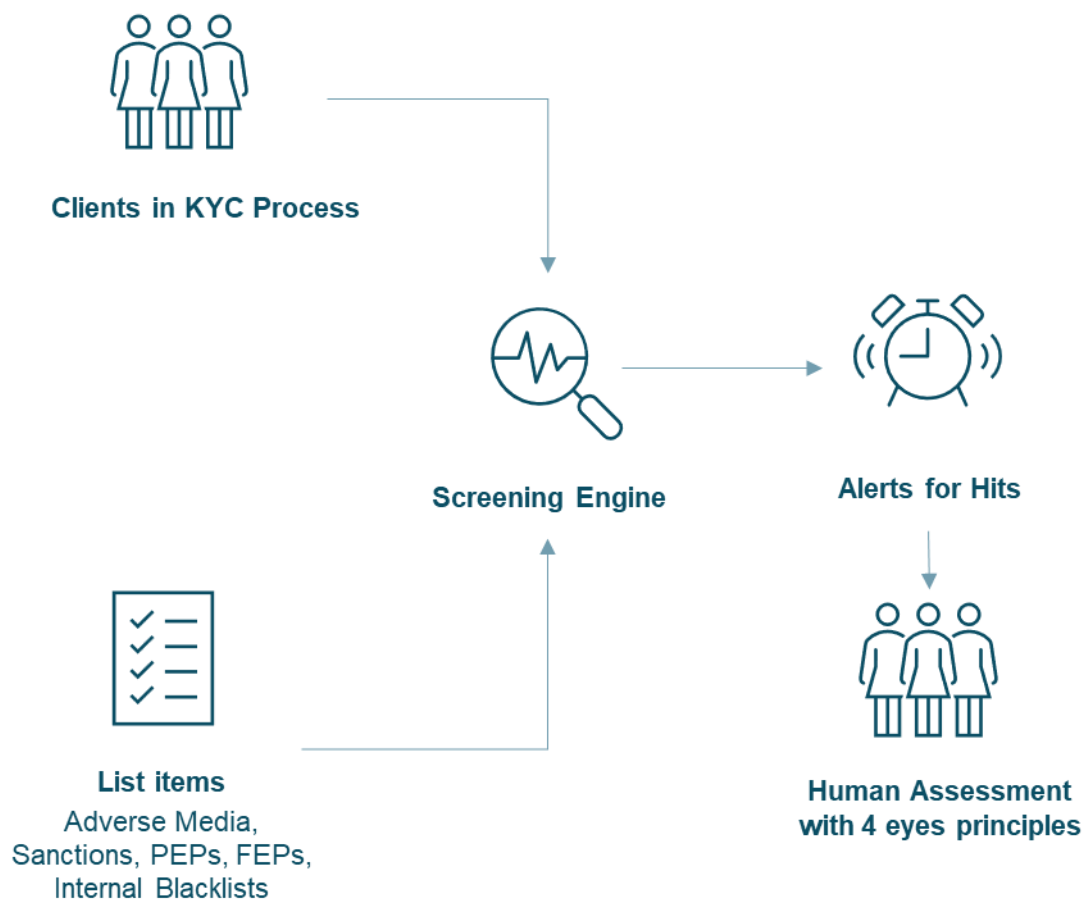


Figure 3: Typical Alert Handling Process.

Alerts are generated based on a certain scoring mechanism that weighs the generated flags and calculates a score for the scenario in question. For example, to identify suspicious activities in money laundry, flags can be generated on entities that have counterparties that are known shell entities or have repeated transactions in the same or similar amounts. The weighted scores in these scenarios are then added together and if it is above the threshold an alert gets generated. There are two approaches in this context: deterministic weights and model-based weights.

Deterministic weights would be based on expert input and stay static in the process. Model based methods can derive the weights by optimizing against previously identified SARs, and thus can capture transactions as they evolve.

Deterministic approach

In transaction monitoring, usually, human users set up rating scales that attach scores to triggered alerts. For example, the transaction of USD 1,000,000 from an American PEP to a sanctioned politician from Cyprus indicates multiple hits. Let us assume that static rules are set on a 1-10 scale, the occurrence of transactions to tax havens, such as Cyprus, generates a rating of 8 and the transaction to a sanctioned public person generates a rating of 10. Additionally, a binary flag is set for round transaction amounts (e.g. 100 or 1,000,000). If triggered, it can be treated as an alert score of $10+8+1=19$. Depending on the chosen threshold, an alert might be created or not, whereas a model-based approach would derive the weights and scores based on empirical evidence.

Model-based/Probabilistic approach

The model-based or probabilistic approach entails taking the calculated flags as binary flags in the past and comparing them against the SAR filings or historical true positives. With sufficient historical data, the model weights for each indicator can be derived where the outcome of the alert is the target and the coefficients derived by the model are the weights to be used. This gives the benefit of facilitating more flexible and smart scoring mechanisms.

Business Benefits of using AI-based Alert Scoring

- The precision of the AI-based alert generation can improve with continuous application, reducing the ratio of false positives significantly. Furthermore, the higher accuracy of AI models reduces the number of false negatives and hence lowers the risk of missing actual compliance breaches.
- The AI processing of transactions comes with a faster pace and more efficient rate. This saves time, capacities and costs.

Challenges and Risks of using AI-based Alert Scoring

- The use of AI can lead to false confidence. The 4-eyes principle does not get replaced by the AI usage. Every alert still must be investigated by human analysts.
- The reduction of false positive alerts (as intended) due to an accurate AI analysis can lead to a lack of sufficient case examples of false positives. Regarding the training of the AI model, it can become very difficult to train the model to detect false positives.

3 Conclusion

Overall, AI and advanced analytics present promising solutions for AFC professionals by automating routine tasks, improving data quality, and providing predictive insights. By improving data quality, managing varying data sources, fine-tuning parameters, and enhancing alert scoring mechanisms, AI offers substantial efficiency gains and accuracy improvements. As financial crime becomes increasingly sophisticated, the adoption of AI-driven solutions can provide an essential edge for financial institutions to maintain and protect their integrity and compliance with regulatory requirements.

Moving forward, the integration of AI in AFC is not just a technological upgrade but a strategic imperative. The continuous evolution of AI will further bolster the defenses against illicit activities, ensuring a robust and proactive approach to risk management in the financial sector. Financial institutions must embrace AI and advanced analytics to stay ahead in the battle against fraud and to safeguard their assets, reputation, and regulatory compliance in an increasingly intricate financial landscape.

In conclusion, the adoption of AI and advanced analytics in AFC represents a paradigm shift, offering unparalleled opportunities for innovation and enhancement of compliance practices. Financial institutions that leverage these technologies will be better equipped to navigate the challenges of financial crime, ensuring a secure and compliant operational framework.

Authors



Johanna Hintermayr

Senior Consultant

Telefon +49 160 98218451

johanna.hintermayr@magpieprojects.com

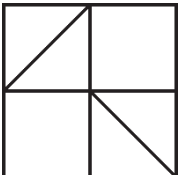


Michael Partin

Senior Consultant

Telefon +49 151 72442146

michael.partin@magpieprojects.com



Emerging from LPA Consulting, we bring over two decades of capital markets expertise but now sharpen our focus on forward-looking strategic transformation.

We act on early insights to help clients stay ahead, transforming intelligence into actionable strategies and guiding clients from traditional capital markets toward the more integrated financial markets, where public and private capital flows intersect.

We stay at the forefront of change, helping clients navigate shifts—whether in policy, technology, or changing structures. We ground ambition in experience, enabling faster decisions and measurable outcomes. As financial markets evolve, so must the tools. We embed AI and use your data to build smarter, scalable systems, always applying technology with purpose.