HAWK

Whitepaper

Using AI to Detect Financial Crime

Contents

Introduction

Artificial Intelligence (AI) has become essential for detecting money laundering risk.

However, questions about how to apply this technology for anti-money laundering (AML) purposes remain unanswered for many financial institutions (FIs).

How does AI improve AML processes and operations? How is AI different from legacy rules-based technology? How should FIs implement AI? What pitfalls should they avoid?

In this white paper, we explain how AI can optimize AML operations by reducing false positives and detecting additional risk. We will outline the differences between AI and rules-based technology. We will also offer practical advice for implementing AI for AML purposes, including how to steer clear of common AI stumbling blocks.

Defining AI in an AML Context

AI refers to any model or algorithm that imitates human intelligence to accomplish specific tasks. In an AML context, AI technology enhances the detection of suspicious money laundering activity.

AI delivers efficiency gains to AML teams via false positive reduction and alert prioritization, as well as effectiveness gains via anomaly detection and pattern recognition. AI empowers AML teams to identify more suspicious behavior and focus investigative resources where they can make the greatest impact.

> It's critical for FIs to remember that AI does not replace a compliance professional's judgement for decision-making.
>
> It enhances their capacity to make more quality decisions using fewer resources.

Machine Learning Models

Machine learning models can be either supervised or unsupervised. These models are trained on historic data and used to predict future outcomes. Machine learning models are typically used to characterize a portfolio and identify anomalies and outliers.

Deep Learning Models

Deep Learning and Large Language Models (LLMs) leverage multi-layer neural networks (NNs), often chaining multiple NNs together, to make predictions. LLMs learn patterns to interpret and generate text.

These models require vast datasets and significant computational resources to train; they are often employed by institutions as pre-trained models (e.g. ChatGPT). FIs can leverage LLMs to accelerate and automate manual processes, improve the quality of documents and inputs, and gain insights into unstructured data (complaints, issues, breaches, risk assessment questionnaires, regulation library, and policies).

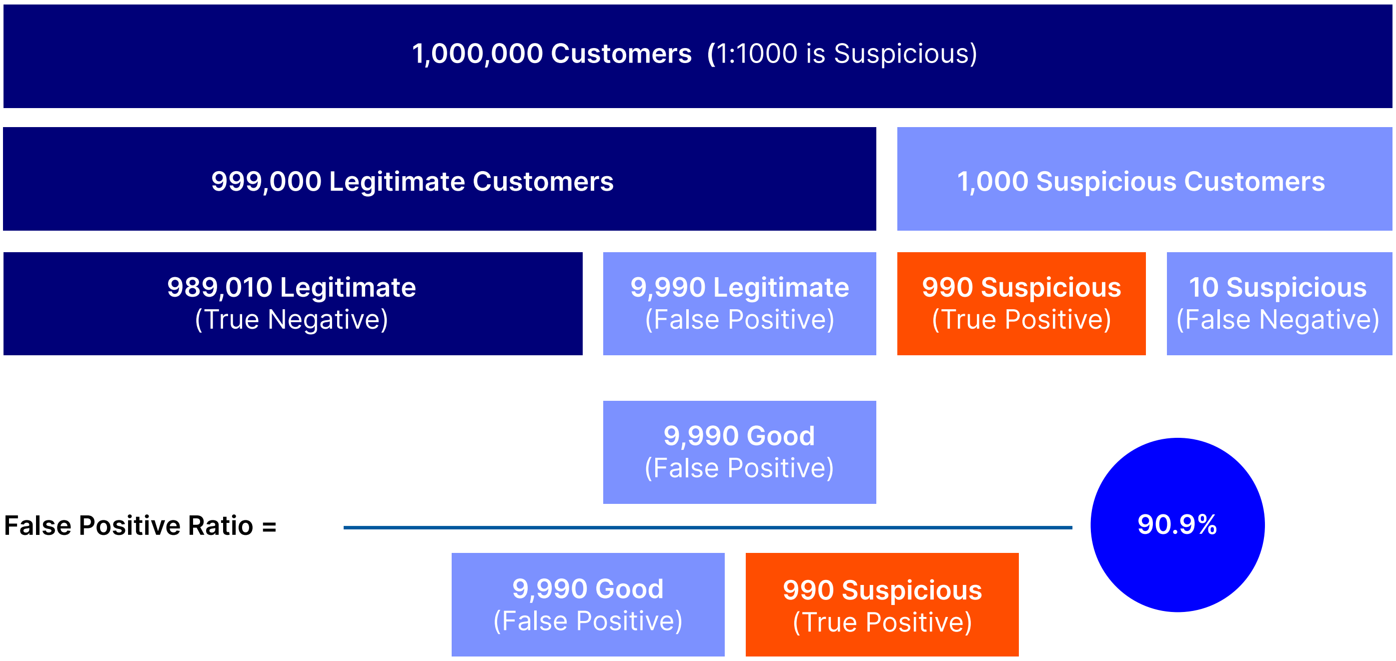
Facing Rising Tides of False Positive Alerts

Many FIs that use rules-based AML technology suffer from an overwhelming volume of false positive alerts. These alerts are not truly suspicious, but AML investigators must examine them and clear them out of the case queue anyway. These alerts waste valuable time and resources, hampering FI's efforts to achieve AML compliance and risk management goals.

In the example below, an FI uses rules-based technology to detect suspicious behavior.

Out of one million customers, 1,000 are truly suspicious. The machine catches 99% of these. However, the system also mistakenly flags 9,990 good customers as suspicious, leading to a false positive ratio of 90.9%.

It will take the FI an exorbitant amount of time and resources to investigate these cases, when the FI could divert these efforts to investigate the 1,000 truly suspicious cases. These numbers are hypothetical, but the principle should ring true for FIs everywhere.



1,000,000 Customers (1:1000 is Suspicious)

999,000 Legitimate Customers

1,000 Suspicious Customers

989,010 Legitimate (True Negative)

9,990 Legitimate (False Positive)

990 Suspicious (True Positive)

10 Suspicious (False Negative)

False Positive Ratio = **9,990 Good** (False Positive) / **9,990 Good** (False Positive) **990 Suspicious** (True Positive)

90.9%

> "What the industry has been struggling with for a long time is that even if you build a really good mousetrap, a really good way of detecting financial crime, you still end up with this huge amount of false positives."
>
> **Michael Shearer** - Chief Solution Officer, Hawk

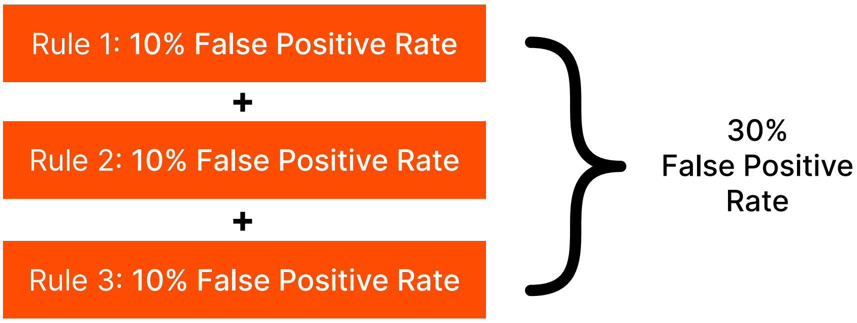
Reducing False Positives with Contextual Information

AI applies more fine-grained rules than rules-based AML technology can. We can think of AI as traditional rules “on steroids.” AI dynamically generates a network of interrelated rules for different segments of an FI’s customer portfolio. Because AI can apply more of these contextual filters simultaneously, it weeds out more false positive alerts.

The Rules-Based Approach

Consider three rules, each having a 10% false positive rate, deployed to detect suspicious behavior. When we apply these rules, the total false positive rate is 30%. The rate is high because all the rules act together, and they don’t choose any specific type of behavior to look at. The same rule applies to every single customer.

Coarse-Grain Manual Rules



Rule 1: 10% False Positive Rate

+

Rule 2: 10% False Positive Rate

+

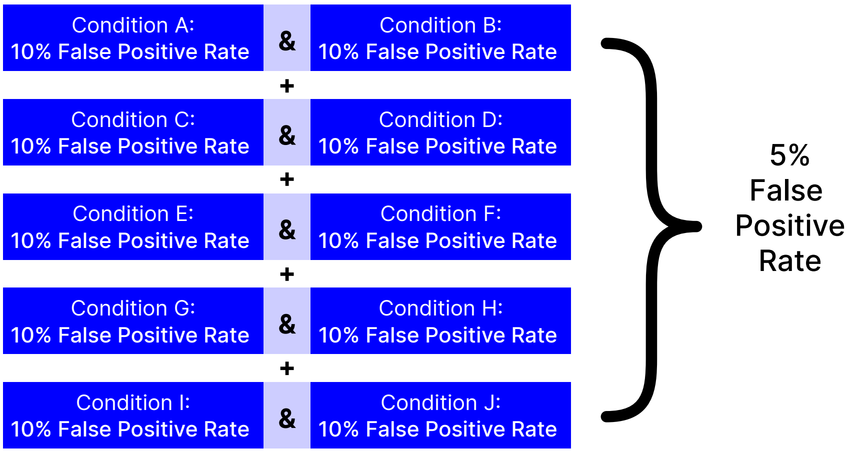
Rule 3: 10% False Positive Rate

30%
False Positive
Rate

The AI Approach

AI generates a large set of fine-grain rules that look for specific combinations of behavior. AML investigators can tailor AI much more precisely to what a particular customer does. For example, with AI we can apply five rules, each with two conditions having a false positive rate of 10%.

Fine-Grain AI-Generated “Rules”



Condition A: 10% False Positive Rate & Condition B: 10% False Positive Rate

+

Condition C: 10% False Positive Rate & Condition D: 10% False Positive Rate

+

Condition E: 10% False Positive Rate & Condition F: 10% False Positive Rate

+

Condition G: 10% False Positive Rate & Condition H: 10% False Positive Rate

+

Condition I: 10% False Positive Rate & Condition J: 10% False Positive Rate

5%
False
Positive
Rate

The rules only “fire” if both conditions are met, resulting in a false positive rate of 1% per rule. When we combine the five rules, we get an overall false positive rate of 5%. In this scenario, we’ve applied more rules and still reduced the false positive rate significantly. Imagine the resulting gains in efficiency and effectiveness when AI is applied at scale.

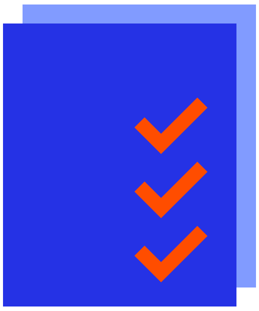
Mitigating AML Risk Earlier

Manual rule-tuning is difficult, and it takes time. The process looks something like this:

- Find a new behavior in individual casework
- Collectively identify the new typology
- Codify the typology into quantifiable behaviors
- Set thresholds, segments, test, and tuneDeploy the rule to production

AML risk accumulates throughout this time-consuming endeavor. On the other hand, the process to retrain an AI model is much faster:

- Find a new behavior in casework
- Retrain model periodically
- Deploy to production



> With AI, you don't have to go through the entire cycle manually. The machine automatically encodes emerging behavior and investigator expertise into a robust set of rules.
>
> That means you find AML risk earlier – and you can mitigate that risk earlier.

Taking a New Approach to AML Risk Detection

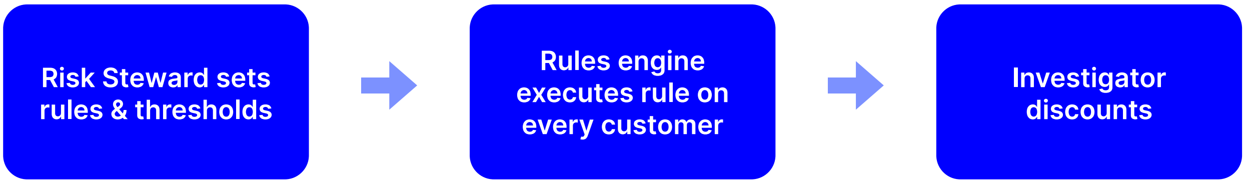
The Rules-Based Approach

The rules-based approach to detecting financial crime has traditionally been led by a risk steward and has been informed by data. With this method, an external authority (risk steward) dictates sets of rules and thresholds. The risk steward says that if these thresholds are reached, or if these events occur, you must raise a case. You program your machine to do that. The machine examines the data and generates cases, and then investigators look at those cases. This approach should feel familiar to most AML professionals.

The AI Approach

With AI, the traditional approach is inverted; it's led by data and informed by the Risk Steward. First, the AI learns from what the investigators do. It watches what cases the investigators label as suspicious. Then, it looks at what behavior caused investigators to label customers as suspicious. In this process, AI allows AML professionals to reverse-engineer desired outcomes from transaction and customer data. The AI learns from the investigator rather than applying a set of rules that were dictated to it upfront. The AML risk steward's role evolves from setting rules and thresholds to deciding what the AI needs to know.

Traditional Approach: Forward Filter

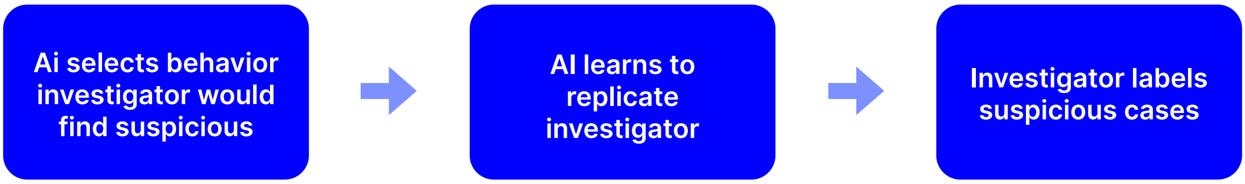


Risk Steward sets rules & thresholds

Rules engine executes rule on every customer

Investigator discounts

Supervised AI Approach: Reverse Filter



Ai selects behavior investigator would find suspicious

AI learns to replicate investigator

Investigator labels suspicious cases

> AI is really good at mirroring the behavior of your best investigator. If the investigator is seeing things that they think are of concern, then the machine can copy that. That applies to any type of behavior that the investigator sees where the machine also sees the same type of data.
>
> **Michael Shearer** - Chief Solution Officer, Hawk

Training AI with Quality Transaction and Customer Data

To successfully implement AI for AML risk detection, you must understand that its needs are different from those of a rules-based system. Most of these requirements boil down to training AI with quality transaction and customer data.

Here are a few ways AI detects AML risk differently from rules-based technology:

AI learns just like we do…

By Example: AI needs sample cases of suspicious behavior. The more cases an AI model sees, the better it learns how to respond to different situations. By Pattern: An AI model needs to review lots of normal behavior to spot the abnormal.

AI needs to see the bigger picture via…

More Attributes: AI needs a fuller view of the customer. AI can't ask questions to fill in gaps, so it only works with the attributes you give it. Stability Over Time: AI needs a stable period of historical behavior to establish what's normal behavior and what's unusual or anomalous.

AI needs precision and correlation via…

Data Quality: AI needs consistent, precise, and complete data. In other words, the data must be well organized. AI needs clear lines between Category A and Category B.

Cause & Effect: AI needs to see causal relationships between input data and case outcomes ("If X, then Y"). If investigators are making different decisions based on data that the machine doesn't have, it can't learn that the behavior is bad. There must be a correlation between the behavior that you teach it and the outcomes you show it.

> Importance of quality data

Helping AI learn to detect AML risk depends on using quality data to train AI models.

If your data quality is poor, you can clean it up. Start by selecting the data that you do trust. Train your AI on this data instead of throwing everything you have at it.

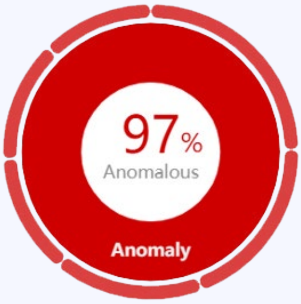
It takes some skill and time to identify the good data and discard the bad data, but it pays dividends. When you do this, the machine can work with the good data, and you can move forward as you clean the lower-quality data.

Explaining the AML Risks Detected by AI

As part of their FI’s risk-based approach, AML investigators need an AI model to explain why it flagged behavior as suspicious.

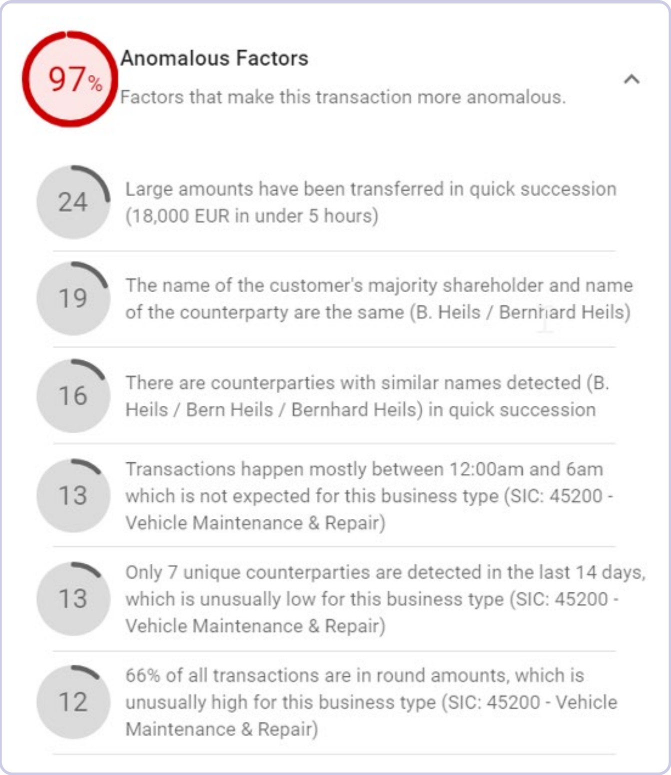
With manual rule configuration, this has been relatively straightforward. You would employ a relatively small number of rules, which were readable by a human, used a limited number of data points, and operated on a simple on/off binary. In contrast, AI will employ a large number of rules, utilize rules that are not easily readable by a human, use many features, and derive insights via statistics.

This would normally make the AI model less accessible to both AML professionals and regulators. However, we have made great strides in Explainable AI technology. Now, AI models can deliver explanations, both in natural language and visually, of why they flagged a given behavior for investigation.



In the example on the right, we see that the AI model has assigned a risk score of 97% to a particular case of suspicious activity.

We also see the individual risk factors that contributed to this risk score, such as “large amounts have been transferred in quick succession” and “transactions happen mostly between 12am and 6am”.



97%

Anomalous Factors

Factors that make this transaction more anomalous.

24 Large amounts have been transferred in quick succession (18,000 EUR in under 5 hours)

19 The name of the customer's majority shareholder and name of the counterparty are the same (B. Heils / Bernhard Heils)

16 There are counterparties with similar names detected (B. Heils / Bern Heils / Bernhard Heils) in quick succession

13 Transactions happen mostly between 12:00am and 6am which is not expected for this business type (SIC: 45200 - Vehicle Maintenance & Repair)

13 Only 7 unique counterparties are detected in the last 14 days, which is unusually low for this business type (SIC: 45200 - Vehicle Maintenance & Repair)

12 66% of all transactions are in round amounts, which is unusually high for this business type (SIC: 45200 - Vehicle Maintenance & Repair)

Making Tradeoffs
AML Risk Coverage vs. Alert Efficiency

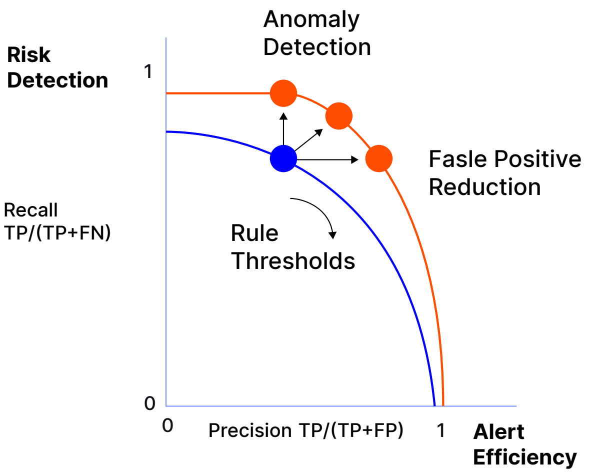
AI is not a silver bullet; put simply, it just generates a better set of rules than what you've used before. The same tradeoff between AML risk coverage and alert efficiency inherent to rules-based technology also applies to AI.

High Thresholds, High Alert Efficiency

If we set high thresholds on our rules, we look for very egregious behavior before we will alert. The result? All our alerts are truly suspicious. A customer must demonstrate extremely suspicious behavior to get above the threshold. The side effect, however, is that risk coverage suffers. We've set the threshold so high that we're missing behaviors that are truly suspicious, but not extreme. We therefore open ourselves to regulatory and reputational risk. In the image below, this would put us at the bottom right quadrant of the graph.

Low Thresholds, High Risk Coverage

Alternatively, we can significantly reduce thresholds. This improves risk coverage because we catch pretty much every customer who behaves suspiciously. The problem with this method is that many low-risk customer behaviors also sit above the lower threshold. The result? We get poor alert efficiency, i.e. large volumes of false positive alerts. This would put us in the top left quadrant of the graph below.



Legend

TP: True Positive
FP: False Positive

FN: False Negative
TN: True Negative

● Rules
● Rules + AI

> The more you de-risk your AML program, the more difficult detection becomes. This remains true when you use AI.
>
> What is also true, however, is that AI detects more financial crime while generating fewer false positive alerts, wherever you draw the line between risk coverage and alert efficiency.
>
> It's up to every FI to determine an appropriate balance based on its unique customer portfolio.

Avoiding Commmon AI Implementation Errors

Even when you know that AI is not a silver bullet, it can be easy to jump to conclusions about the technology's effectiveness.

Avoiding these common errors will help you get the most out of your AI technology and have an accurate view of whether it's working or not:

Don't rush the process

Getting data takes time. It's not just "wrangling" the data; you need to approve and develop it as well. Getting your data ducks in a row before implementation will help smooth the rest of the process.

Don't judge too soon

AI usually creates a sizeable uplift at implementation, but the measurable rate of improvement can plateau or even dip. However, even after a dip, AI is still much more effective and efficient than using legacy rules-based technology. Make sure to control for any leveling off in your analysis of AI effectiveness.

Don't train too soon

If you train an AI model on cases you haven't worked to conclusion, you may find that the results get skewed. When you train on unclosed cases, the machine has only seen half the story. Waiting until you have a robust set of completed cases before training should prevent this issue.

Don't train on insufficient casework

If a case has been detected for one reason, and then gets escalated or filed for another reason, don't try to train an AI model on that data. If the reason for the escalation is something the machine doesn't have a data point for, then it will flag incorrectly. Weed these cases out and only train on the cases where the machine has all the data that it needs to learn.

These common errors all stem from not training your AI models on quality data. When you do train your AI systems with quality data, you can rest assured that your AI technology will work as intended. You'll also be more likely to avoid headaches at implementation and beyond.

HAWK

Employing Practical AI Tips and Tricks

- Use AI for any task you do repeatedly and based on the same inputs. AI excels at tasks of this nature.
- Before you predict with AI, make sure you don't already have the data you're looking for. Data can often get buried in an organization, so doing a search can prevent you from wasting time trying to predict something you already have.
- Invest in data first and case outcome labelling second. What did your investigators find? Did they find suspicion? What sort of suspicion? That information is gold for a machine.
- Differentiate rule parameters and AI features. You're likely used to rules having certain parameters. AI models have features, which are similar, but there's a crucial difference: an AI algorithm may ignore a feature because it doesn't correlate with the outcome it's tryin to predict.
- Beware of time travel. It's possible to train a machine learning algorithm with data that spans multiple time periods, meaning it can see into the future. If you do this, your machine looks very clever on your test data. However, it will perform poorly in production because it won't have the future view it had in training.
- Demonstrate risk coverage equivalence at the aggregate level, rather than rule-by-rule. A side-by-side comparison of rules and AI can lead to a dead end. Instead, evaluate your risk coverage equivalent. Are you catching better numbers of financial crime at the global level?
- Be clear about material change. AI is still risk-based detection. Your model will get stuck if you don't retrain it regularly. On the other hand, you don't want to lose control of what your model is doing. You need a level of model governance. It's about finding the sweet spot between control and innovation.
- Evaluate performance fairly. Human investigators make mistakes, and machines do too. Take care not to compare your AI's performance against an unachievable ideal.
- Use rules. Don't throw the baby out with the bathwater. There is still a place (albeit a smaller place) for rules. You may want to alert on a particular type of behavior, regardless of customer identity, context, or any other factor. A rule is still the best way to do this. Be mindful of this approach, as it can cause an increase in false positives.

Conclusion

By integrating built-for-purpose AI into AML operations, FIs can enhance their ability to detect and manage money laundering risk.

Despite the challenges associated with data management and implementation, the benefits of AI are clear, offering more accurate risk assessments, reduced false positives, and streamlined compliance procedures.

As financial crime proliferates online, adopting AI-powered AML technology is not just an option; it's a necessity.

By integrating built-for-purpose AI into AML operations, FIs can enhance their ability to detect and manage money laundering risk.

HAWK

Hawk AI GmbH
Friedenstrasse 22B/i3
81671 Munich
Germany

Hawk AI USA Inc
230 Park Ave, Floors 3 & 4
New York, NY 10169
U.S.A.

Hawk AI APAC Pte Ltd
160 Robinson Road, #14-04
Singapore Business Federation Center
068914 Singapore

Follow us on LinkedIn >

✉ info@hawk.ai >

www.hawk.ai >