saifr®

WHITE PAPER

The Efficiency Trap:

The hidden danger of false negatives

Overview

In today's business landscape, efficiency is a prized objective—easy to measure, tied to cost savings, and often seen as the fastest route to operational gains. It's no surprise that many organizations chase it relentlessly.

For senior financial services leaders—from BSA officers to CEOs—efficiency is just one important aspect of **adverse media screening in KYC processes within AML programs and monitoring in e-communications surveillance programs.** Just as critical are meeting regulatory requirements, cost control, customer trust, employee engagement, and long-term organizational resilience.

In the pursuit of efficiency, there's a hidden cost to focusing primarily on false positives, instances when a system flags alerts that ultimately prove irrelevant. Risk management approaches that primarily strive to look efficient on paper can degrade the customer experience, frustrate employees, and increase organizational risk. Legacy workflows can often leave teams overwhelmed by focusing on volume, not insight. When staff are busy moving tasks instead of thinking critically, red flags get missed. Blind spots grow. And when failures occur, they're rarely due to inefficiency. Instead, they stem from inaction or delayed response. For example, a recent global bank's money laundering scandal was rooted in overlooked risk signals.

For financial firms, **the greatest risk can come from false negatives**—when critical issues go undetected and no alerts are triggered.

Traditional compliance systems are great with managing false positives, not uncovering hidden threats. Legacy systems weren't designed to process vast, unstructured datasets.

As a result, some firms trade coverage, which can make false positives unmanageable, for control.

With AI and large language models (LLMs), it's now possible to continuously monitor vast volumes of data in multiple formats, extract meaningful context, and reduce the window of time during which risk goes undetected.

The benefit: greater efficiency that many firms strive for, plus the ability to more easily surface real risk that may have been overlooked in the past.

Success today is no longer measured solely by how efficiently low-risk alerts are processed to catch bad actors. Instead, it's equally defined by how effectively high-risk anomalies are identified and addressed before they cause harm.

Striking this balance has become the cornerstone of the modern risk leader's mandate; efficiency must be paired with a focus on organizational resilience. After all, the greatest threat to a firm isn't what gets flagged, it's what slips through undetected. In the following pages, we explain why this balance matters and how to achieve it.

What's inside

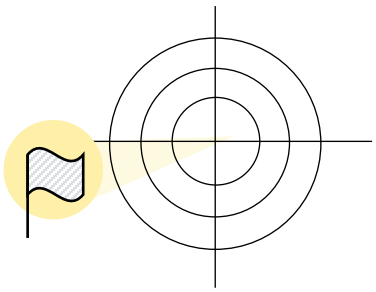
The false positive and false negative paradigm	3
Measuring your effectiveness in detecting missed risk	6
Modernizing your risk management approach	9
Defining success	10
Final thoughts	13

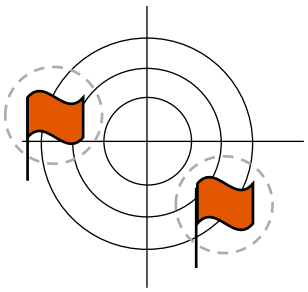
saifr

The false positive and false negative paradigm

As we all know, in the domains of adverse media screening and e-communications surveillance, two types of system errors dominate the risk conversation: false positives and false negatives.

What are they?





False positives
occur when a system flags an alert that, upon investigation, proves to be irrelevant.

False negatives
happen when a system fails to flag an issue that should have raised a red flag, allowing genuine risk to go undetected.

Why do they occur?

Understanding the world of data helps us understand why false positives and negatives occur, seeing as most KYC systems only focus on structured data:

	Insights	Updates	Ease of use	Format	Pool of data
20% **Structured data**	Incomplete if sole data source	Static, can be dated	Easy to store and manage	Predefined, fixed formats	Limited
80% **Unstructured data**	More insights when using LLMs	Active, in real time	Messy and complex	Multitude of formats	Abundant

While both error types carry consequences, they are not equally threatening. To understand their impact, let's explore how they manifest across different surveillance systems with a few examples.

False positives

The cost of over-detection

This can trace back to outdated systems that fall short in several critical areas, including:

Data limitations: Heavy reliance on structured data, struggle to process and interpret dynamic, unstructured data.

Rigid, rules-based approaches: Inflexible, often overly simplistic frameworks prevent them from capturing the full complexity of real-world scenarios.

Lack of contextual awareness: Without the ability to apply context, firms can face significant challenges in accurately assessing and prioritizing true risks.

False negatives

The risk you don't see coming

This can stem from two key issues:

Lack of data access: No system captures every data source; important signals may simply not be ingested.

Poor signal recognition: Systems that over rely on name matching often miss critical risks embedded in context, location, or identity details that can provide stronger and more precise risk signals that a firm should pay attention to.

In KYC and adverse media scenarios

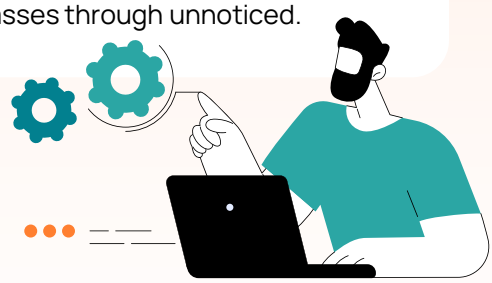
A customer named John Smith (born 1955) is flagged due to adverse media tied to another John Smith (born 1985)—same name, different person.

A long-term customer conducts large financial transactions for years without triggering any alerts. Later, under law enforcement scrutiny, suspicious patterns emerge that should have raised alarms—but the system failed to detect them.

In e-communications monitoring

Standard phrases such as "We think stock XYZ is going to tank" or "Let's keep this between us" may trigger alerts suggesting insider knowledge, concealment, or policy violations even when used innocuously.

Employees exchange material non-public information (MNPI) using coded or subtle language that the model isn't tuned to detect—and the communication passes through unnoticed.



The impact

Operational friction: Teams must spend valuable time investigating non-issues, increasing alert fatigue and slowing down meaningful reviews.

The true blind spots: They are the red flags no one saw, the misconduct that can quietly escalate, the compliance failure revealed only after damage is done.

Why false negatives are more dangerous

It's natural for organizations to focus on reducing false positives. They're highly visible, time-consuming, and resource-intensive. However, false negatives can pose the greater existential threat. They can expose firms to:

- Regulatory action for missed red flags
- Financial and reputational damage from undetected misconduct
- Strategic risk from flawed assumptions about system efficacy

	False positives	Example	False negatives
	Flagging a customer for adverse media related to a different individual		**Failing to flag a customer who later proves to be suspicious**
↓			
	Highly visible to operations teams; creates alert fatigue and operational burden	Visibility	Often invisible until after damage occurs; represents a serious blind spot
↓			
	Wasted time, resource strain, investigation overload	Impact	Undetected risks, regulatory breaches, financial/reputational harm
↓			
	Inefficient processes but generally manageable consequences	Outcome	Critical failures that can escalate into major crises

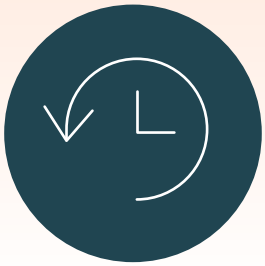
The bottom line:

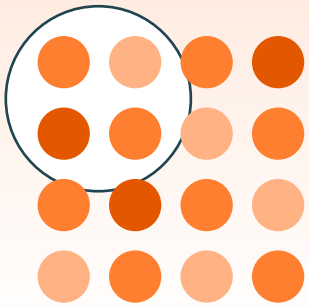
False positives are noisy, but false negatives are silent. And it's the silence that can create the greatest risk.

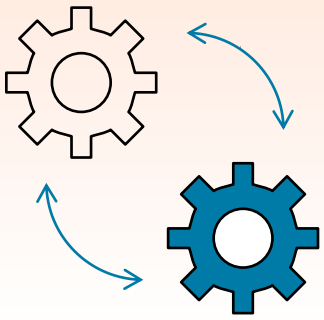
Measuring your effectiveness in detecting missed risk

Unlike incorrectly flagged risks (false positives), missed risks (false negatives) are much harder to quantify than simply tracking alerts. So how can one gain a clearer understanding of these overlooked risks, and why is it so important to reduce them?

We encourage leaders to begin by intentionally estimating false negatives. To support this, we outline three practical methods that shed light on the potential blind spots within current systems, each offering valuable insights to help strengthen your risk management approach.







1

Post-incident review

One of the most direct methods is reviewing past incidents in which risky events were missed.

After a conviction for a major crime, someone is clearly a bad actor; and you can review your firm's history of activities with the involved individual.

The goal is to see what, in hindsight, should have triggered an alert.

By creating a dataset of such missed risks, you can begin to identify patterns or gaps in your system.

2

Sampling and validation

This involves sampling recent alerts. If you process, for example, 100 alerts daily and you know that approximately 20 are true positives (legitimate risks), you could randomly review the next 100 alerts to estimate how many true positives were missed.

Essentially, the first 100 you sample is your baseline to benchmark future results against. By performing this process regularly (e.g., every month or quarter), you can estimate the false negative rate over time.

While more time-consuming, it can be highly effective for detecting whether your system is improving or drifting.

3

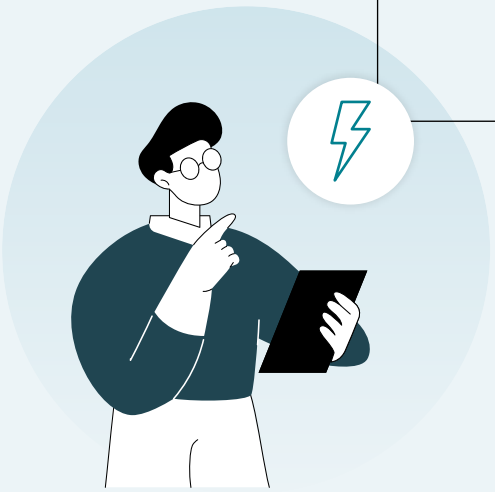
Comparing systems

A more advanced, and potentially more accurate method, involves comparing your system with a competitive technology.

A firm can run a parallel test by feeding the same data into two different systems and comparing the alerts generated.

For example, if System A flags 100 genuine issues out of 1,000 of these alerts and System B flags 200, it's clear that System A may have missed 100 critical issues, giving you a direct insight into its false negative rate.

By evaluating the quality of the alerts and how well they match the data, you can begin to understand where your system might be falling short.



Your next step: addressing false negatives

Once you measure your false negative rate, the next step is to take action. Some organizations might be content with the false positives (misleading alerts). However, a false negative (missing a critical risk event) can have much more serious consequences.

The real challenge for many firms is the trade-off between focusing on false positives versus false negatives. While false positives are time-consuming and tax your organization, false negatives can allow dangerous risks to slip through unnoticed and can result in meaningful damage to your firm.

Determining the potential marginal cost of false negatives

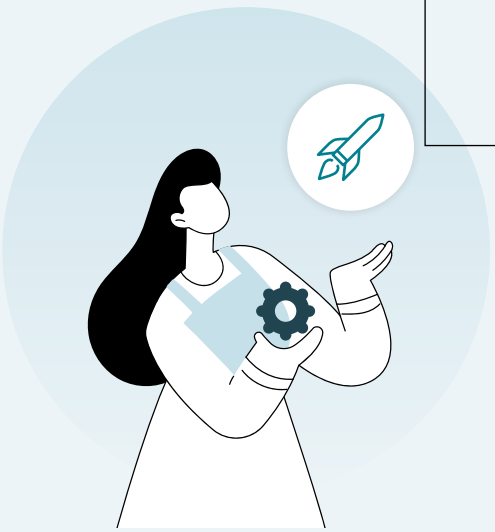
To make a compelling case for improving the detection of false negatives, firms may want to assign a dollar value to both false positives and false negatives.

For illustrative purposes, let's assume that each missed false negative costs your organization $10,000 (fines, reputational damage, or fraud losses). If your system is missing 1,000 of these false negatives every year, you're looking at a potential loss of $10 million.

Now compare this potential loss to the cost of reviewing a false positive. What if, for illustrative purposes, the cost of review for each false positive costs just $100 in terms of the analyst's time.

By doing the math in this illustration, one should feel compelled to accept the trade off 100 false positives for 1 false negative.

The bottom line is that by understanding the marginal cost of missing a false negative versus reviewing a false positive, organizations can objectively make decisions about thresholds and risk tolerance when seeking to optimize their systems and allocate resources.



The path forward

Organizations should embrace a mindset shift to actively address false negatives. The first step is acknowledging the risk. Just as you wouldn't ignore a growing fraud trend or a wave of false positives, you shouldn't ignore the possibility of missing critical alerts.

To get started, leaders should consider performing regular checks and evaluations of their systems, applying cost-benefit analysis to weigh the risks.

Modernizing your risk management approach with AI and LLMs can change the operational calculus as these systems provide a way for firms to track more potential risk with the same level of effort.

Modernizing your risk management approach

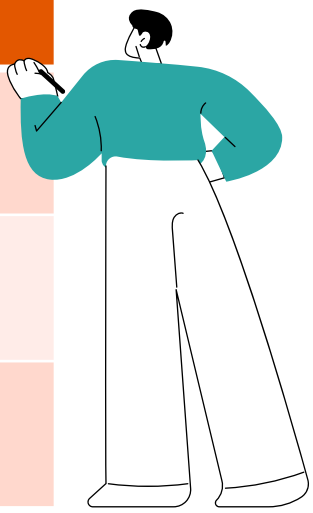
For operational teams, false positives are an obvious pain point since they are time-consuming, repetitive, and frustrating. But for Chief Risk Officers and executive leaders, false negatives represent the true existential threat. These are the *unknown unknowns*: the red flags missed, the misconduct undiscovered, the reputational crises that surface only after the damage is done.

The industry narrative, largely shaped by legacy vendors, has long promised comprehensive coverage. In practice, it has proved harder than it appears. Again and again, our side-by-side comparisons reveal that AI-based systems that rely on LLMs can analyze vast, unstructured data sets, extract meaningful context, and surface true risks **without overwhelming operations with noise.**

What's more, these models are highly customizable for your organization's needs and can be trained to understand your specific organization's language. As a result, they can detect and alert upon identifiable types of risk within established parameters. Their ability to tap into real-time data **takes risk management to new heights of speed and precision** that traditional algorithms can't begin to compete with.

The result: broader, deeper coverage with less manual burden.

	Traditional system	AI-based LLM solution
Searches and analyzes dynamic, unstructured, real-time data	✗	✓
Understands context of results which can lead to more precise and accurate search results	✗	✓
Ongoing improved results over time, thanks to natural fine tuning of system, like machine learning	✗	✓



With the flow of data in multiple formats reaching unprecedented levels, it's clear that organizations relying primarily on traditional screening for KYC and e-communications may be falling behind. A recent McKinsey study reports that the world generates five quintillion bytes of data every 48 hours.[1]

Given these challenges, it's time for risk leaders focused on adverse media and e-communications surveillance to proactively manage the invisible danger of false negatives.

Because the greatest risk may not be what your system flags, it may be what it misses.

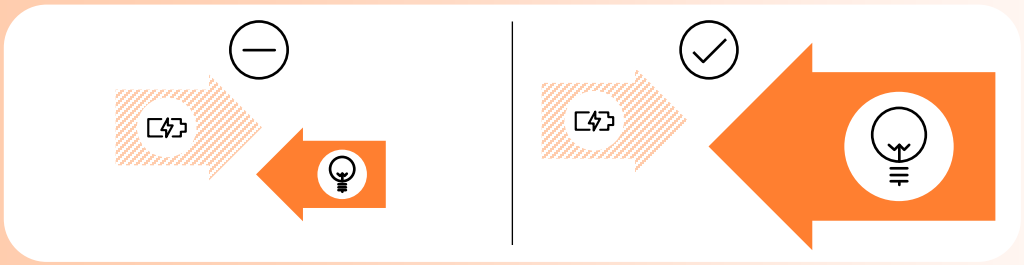
1 https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-data-driven-life.

Defining success

When a leader adopts an advanced, LLM-reliant AI system, success isn't just about detecting more potential threats or reducing false positives. It's about navigating the delicate balance between risk exposure and operational effort. Here's a closer look at what true success entails and some key steps for achieving it.

1 Increase risk detection without increasing effort

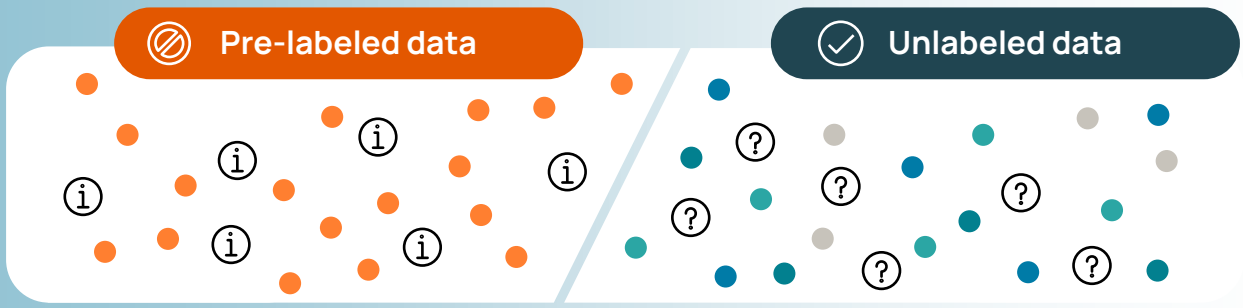
At a minimum, a successful system should uncover more potential risk with the same level of effort. This is the most immediate and measurable win. If a new solution identifies a greater volume of risk indicators without requiring more input or resources, it's a strong sign of effectiveness.



However, even when a modest increase in risk detection requires slightly more effort, the trade-off may still be worthwhile—especially if the newly detected risk indicators are more meaningful or costly.

2 Move beyond easy outcomes: avoid indexing on "what you know"

A common pitfall in evaluating new systems is focusing too heavily on how well the technology identifies known risks. For example, giving a system a pre-labeled data set of 10,000 entities, 100 of which are known bad actors, and evaluating its performance solely on how many of the 100 bad actors it identifies can give a misleading picture because you're stacking the deck in favor of the current system.



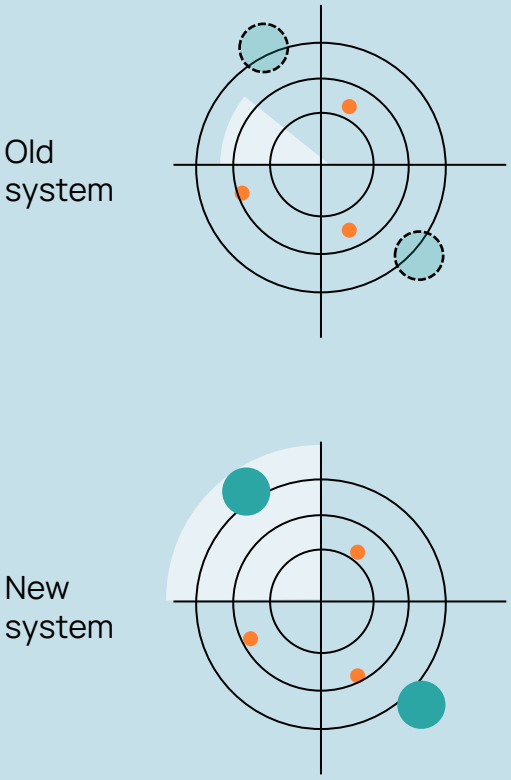
Instead, success lies in how the system performs in an environment full of threats not yet surfaced. Real-world bad actors are rare events within large, complex data sets. Therefore, evaluating how a system handles fresh, real-world data is critical. Blind tests, where **two systems are tested in parallel** on unknown data, are much more effective for evaluating true performance.

Using parallel testing to compare real-world impact

One effective approach to measure the success of a new system is parallel testing with fresh data. This method involves running your incumbent system alongside the new technology using fresh data.

For instance, the system that previously helped detect 100 bad actors in 1,000 alerts is compared with a new system, which might detect a total of 130 issues. But the key to success lies in understanding the full picture—just because a system alerts to help find 130 genuine issues doesn't mean it's perfect. Often, these systems might identify a number of false positives, or worse, miss critical risks.

What can set a system apart is how many new risks it can help identify beyond what the old system finds. Ideally, you want to know that the new system isn't just alerting to a portion of the same risks, but expanding your visibility by uncovering more hidden potential threats. This broader scope of detection, paired with the ability to manage false positives efficiently, is where AI-based technologies shine.

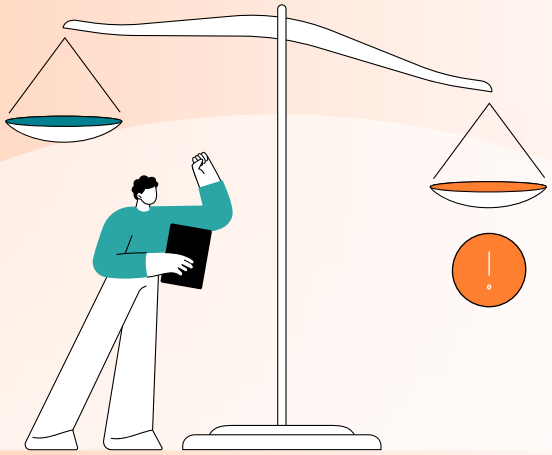


3 Balance risk and reward: weighing up false negatives vs. false positives

A crucial part of success is managing the trade-off between false positives and false negatives. This comes down to assigning a dollar value to the costs associated with each type of mistake. Using the same example as described earlier in this white paper, let's assume for illustrative purposes the following cost comparison between alerts:

False positive

$100
in time and resources to investigate



False negative

$10K
in legal penalties, plus potential reputational damage

Leaders should evaluate whether spending an extra $100 per alert is worth avoiding a $10,000+ oversight. Modern AI systems help shift this balance by reducing false positives and allowing teams to focus on more important alerts that identify true risks.

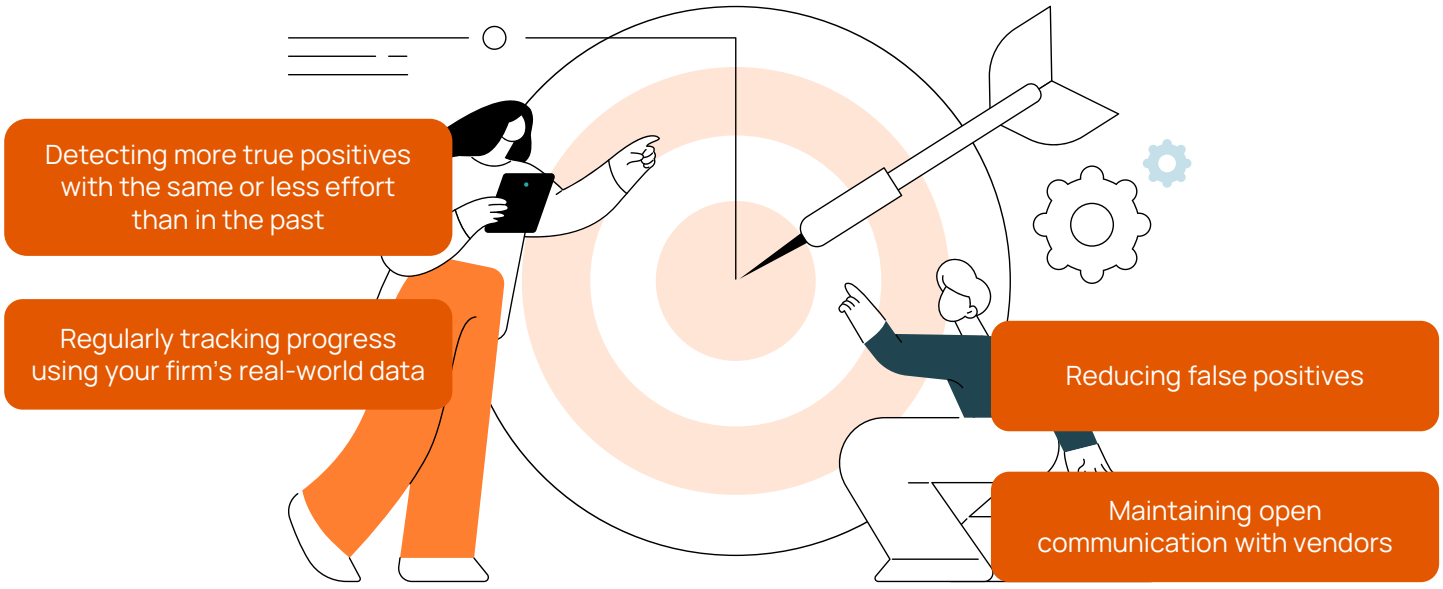
4 Collaborate with vendors to improve over time

If the new system initially helps to find fewer known risks than the incumbent, it doesn't necessarily indicate failure. It may still be surfacing previously unknown risks. For instance, if the old system helps find 30 out of 100 known threats and the new system alerts to 20, this could still be a step forward—provided the new system is helping discover different, valuable risks the old one missed.

Ongoing dialogue with the vendor is essential. If the new system helps to find 130 issues where the old found 100, the question becomes: can it eventually detect 200? Clear expectations and feedback loops help vendors refine their models and better meet your evolving needs.

Building confidence through transparent metrics

Ultimately, success is about objective, data-driven evaluation. Leaders should focus on:



With the right system in place, compliance teams and risk managers can **make smarter decisions, act faster, and reduce exposure to hidden threats.**

saifr

Conclusion

Risk management is about managing uncertainty, not eliminating it. But by focusing too much on operational efficiency and false positives, organizations can expose themselves to significant, unmanaged risk.

The future of risk management demands a shift toward minimizing false negatives, continuous monitoring, and early intervention.

By leveraging advanced AI technologies and adopting a mindset focused on effectiveness, organizations can significantly enhance their resilience, protect their reputations, and create safer, stronger environments.

Don't just improve efficiency, enhance effectiveness. Protect your organization by finding hidden risks before they find you.

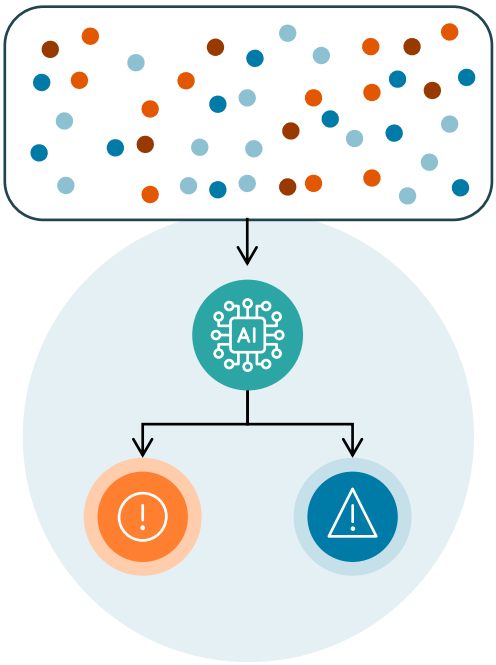
saifr

About SaifrScreen℠

SaifrScreen enables firms to more accurately and efficiently identify potential risks in full customer and vendor populations for further investigation. It leverages the latest in machine learning (ML) technology and natural language processing (NLP), including large language models (LLMs).

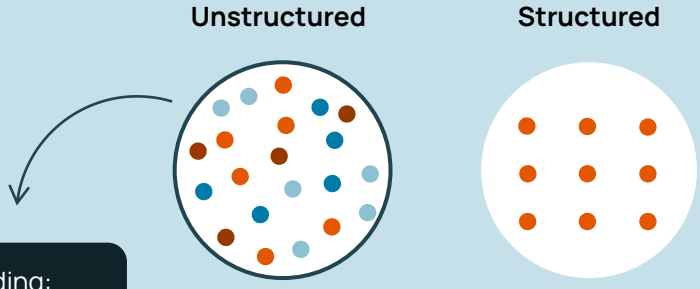
SaifrScreen continuously reviews large populations against publicly available information to **identify more potential indications of financial or reputational risk sooner.**

SaifrScreen uses behavioral science to understand context and can distinguish media that describes fraud versus murder, for example. Additionally, SaifrScreen crawls and indexes internet data 24/7 to provide ongoing review and monitoring with early warning notifications. These potential risks can feed into firms' processes for further investigation.



Most traditional AML and KYC screening and monitoring methods focus solely on structured data (e.g., sanctions, wanted, and watch lists), which **only represent ~20% of internet data** and can be **slow to be updated.**

Subject to change



SaifrScreen extends its reach to unstructured data, including:

230K	190	160	23B	millions
Online sources	countries	languages	webpages	of webpages added daily

SaifrScreen's continuously growing dataset includes sources such as news media, government sources, arrest and court record aggregators, and more. Searching this remaining ~80% of internet data can reveal valuable details and enables firms to zero in on and further investigate threats, such as fraud, as soon as they become known.

Up to **7x**
as many potential bad actors identified

Compliance officers using SaifrScreen are empowered to address more cases without sacrificing hours chasing dead ends via menial, manual methods.

About Saifr

Saifr redefines how compliance operates with advanced AI technology, the right data, and deep industry expertise. Built within Fidelity Investments' innovation incubator, Fidelity Labs, Saifr harnesses the power of AI agents to help address the limitations and inefficiencies within traditional compliance frameworks, helping safeguard organizations from regulatory and reputational risks. Saifr helps clients save time, reduce costs, and improve accuracy while protecting their firms. Our AI-powered risk prevention and management solutions include capabilities for marketing compliance review, adverse media monitoring, and electronic communications surveillance. Learn more at https://saifr.ai.

1205644.1.0