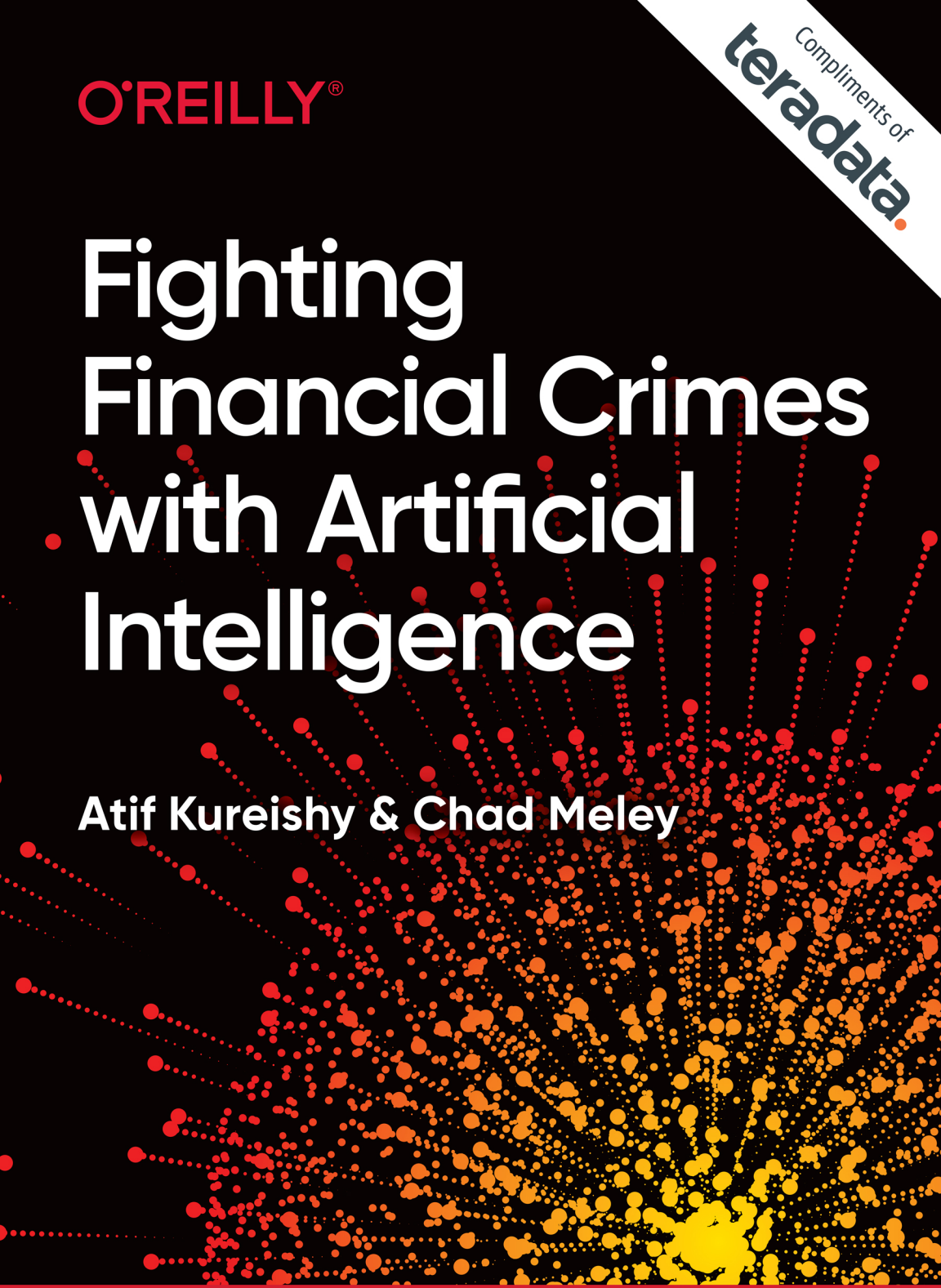
O'REILLY®

Compliments of teradata.

Fighting Financial Crimes with Artificial Intelligence

Atif Kureishy & Chad Meley



REPORT

Fighting Financial Crimes with Artificial Intelligence

Atif Kureishy and Chad Meley

Beijing • Boston • Farnham • Sebastopol • Tokyo O'REILLY®

Fighting Financial Crimes with Artificial Intelligence
by Atif Kureishy and Chad Meley

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://oreilly.com*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Acquisitions Editor: Nancy Davis
Developmental Editor: Jeff Bleiel
Production Editor: Nan Barber
Copyeditor: Octal Publishing, LLC
Proofreader: Nan Barber
Interior Designer: David Futato
Cover Designer: Karen Montgomery
Illustrator: Rebecca Demarest

April 2019: First Edition

Revision History for the First Edition
2019-04-01: First Release
2019-05-16: Second Release

978-1-492-05266-1

[LSI]

Table of Contents

Fighting Financial Crimes with Artificial Intelligence

Executive Summary

The latest research shows that an epidemic of financial crime is costing society $4.2 trillion globally. As new digital channels emerge for financial transactions, financial crime just keeps growing worse. Whole communities of fraudsters and criminals are continuously innovating new ways to steal, and they collaborate with one another and sell their tactics and techniques in a seemingly insatiable, worldwide black market. Legacy practices and traditional rules engines can't keep up. We need new approaches.

Artificial Intelligence (AI) offers a way forward. AI techniques, although new, have already been proven to thwart a variety of financial crimes. This report provides the following:

- Insights from executives who share their experiences in applying AI in the fight against financial crimes. Unlike many strategic technology areas in which methods and outcomes are closely guarded secrets to protect competitive differentiation, this report spreads ideas and best practices from real-world implementations.
- Visibility into successful outcomes, and thought leadership on challenges, countermeasures, and the problems that still need to be solved to turn the tide in the fight against financial crimes.

AI conjures different reactions ranging from skepticism to irrational exuberance, and this report aspires to leave you with a practical

understanding of the exciting current capabilities as well as current limitations of applying AI in the fight against financial crimes.

Financial Crime Continues to Increase

Recent research finds that global financial crime is massive in scale and accelerating in pace. One study put the current global cost at $4.2 trillion. Another estimated that the typical organization loses as much as 5% of revenue to fraud each year. And every $1 of crime costs businesses between $2.48 and $2.82 in total—that's about two and a half times the direct loss itself. Extra costs include regulatory fines for noncompliance, operational expenses for dealing with the aftermath of crimes, notifying and compensating the victims, and the financial fallout from reputational hits. Unfortunately, financial crime is only accelerating. In 2017, two-thirds of businesses experienced financial criminal activity firsthand—a sharp 58% increase since 2016. In a recent interview with Bank Info Security, Daniel Cohen, head of RSA's fraud and risk intelligence product suite, said this:

> If we are predicting that transaction volumes are going to grow, then, obviously, fraud cases, and I'm talking in real numbers, are also going to grow significantly. And then there's the whole operational question of, 'How do we manage and mitigate the fraud that the bank is suffering?'

Different Types of Financial Crime

Financial crime encompasses a varied and wide collection of illegal activities. For the purposes of this report, we divide them into three categories: fraud, money laundering, and cybercrime.

Fraud

Fraud schemes are sophisticated and change from day to day. Automated systems as well as human fraud analysts pour over data and documents and review suspicious activity reporting (SAR) logs to determine which transactions are genuine and which are fraudulent. Here are a few of the most common fraud practices:

Credit card fraud
: Linked to the rise in online commerce, the most common kind of credit card fraud is called *card not present* (CNP) fraud. CNP

fraud occurs when a fraudster purchases a product or service online and is not obliged to present the card or provide a PIN or signature. CNP fraud can also be a type of identity fraud in which a fraudster takes over the digital identity of a victim and opens credit card accounts or runs up bills in the victim's name. Although many experts attribute the increase in CNP fraud to the implementation of Europay, MasterCard, and Visa (EMV) chip technology, there's also the fact that online commerce is booming, and fraud prevention tools have not caught up. Financial institutions unfortunately bear responsibility for most of the money lost as a result of CNP fraud. The Nilson Report in October 2016 showed that card issuers paid a 72% share of fraudulent losses, with merchants and ATM acquirers paying the other 28%.

Synthetic identity fraud

Synthetic identity fraud is an increasingly common type of fraud. In this case, the fraudster creates and establishes credit using a false persona, which is often a sophisticated blend of faked and real personal details; for instance, using a fake name along with a real social security number, such as that of a child, to lend the persona credibility. *Puppeting* the persona, the fraudster patiently mimics legitimate financial behaviors—taking out loans, withdrawing cash advances, and running up bills, but always acting properly and paying on time to establish the credit-worthiness of the synthetic persona and grow the line of credit. Then, they perform what is called a "bust out" in which they max out the cards, max out the cash advances, default on all loans, and disappear. This type of fraud is very difficult to catch in advance because the fraudsters do nothing wrong until, all at once, during the bust out, they do everything wrong.

Account takeover

Account takeover fraud is a form of identity theft in which a third party gains access to unique details of a user's online accounts. By posing as the real customer, fraudsters change account passwords and phone numbers, buy goods and services, withdraw funds, and use the stolen information to access other accounts of the victim. Exact figures are difficult to estimate, but one study from 2017 concluded that account takeover fraud had increased year-over-year by more than 45% and was costing merchants more than $1 billion every month.

Money Laundering

Money laundering is the process by which criminals trick authorities into making it seem like the monetary proceeds of illegal activities came from legitimate economic activities. For instance, one typical money-laundering scheme involves using criminally acquired money to purchase subjectively valued items such as real estate, art, or antiques. Or a fraudster can buy chips at a casino with illegally acquired money, exchange the chips for cash, and claim that the actually ill-gotten monies were merely legitimate gambling winnings. By obscuring the identities of transaction partners, Bitcoin and other cryptocurrencies are also becoming a problem for financial institutions.

In an attempt to counter the many varieties of money laundering—old-fashioned and novel—annual global spending on anti–money laundering (AML) routinely runs to the tens of billions of dollars. Despite the enormous sums invested to prevent this type of financial crime, massive amounts of illicit financial gains still slip into the global banking system every year as a result of money laundering.

To subvert illegal activity, domestic and international governments subject financial institutions to an array of regulations to identify and keep records regarding events that could be signs of money laundering. Compliance with AML regulations is mandatory, and fines for noncompliance in 2017 reached a record $2.5 billion.

Cybercrime

Electronic fraud is becoming larger as technology innovations continue to offer new ways to do business as well as the global nature of many of these transactions. Bitcoin is a part of that, as is the ability of cybercriminals to utilize unregulated cryptocurrency exchanges to cash out the return of their criminal online activities.

Moreover, it's not just individuals that financial institutions need to be cautious about. Organized criminal organizations and state-sponsored gangs collaborate on ways to penetrate the global banking system in new ways.

One of the biggest cyberheists in history, the Bangladesh Bank robbery, occurred in February 2016, when instructions to fraudulently withdraw $1 billion from the account of Bangladesh Bank were sent to the Federal Reserve Bank of New York via the Society for World-

wide Interbank Financial Telecommunication (SWIFT) network. Five transactions worth $81 million were successfully sent to the Philippines, where they were laundered through the country's casino system. Some of the funds were recovered, but a good proportion was successfully stolen.

Fallout to Financial Services Firms of Successful Crime

As we mentioned earlier, damages done to financial institutions take a number of forms including the direct financial loss itself; fines for noncompliance with relevant regulations; the operational costs of changing systems to avoid future losses; the costs of informing and providing restitution to victims; and the reputational costs that affect costs and rates of customer acquisition, customer retention, and the cost of capital on stock exchanges.

Regulatory fines alone can take an enormous toll. The United States and Europe in total imposed $342 billion in fines on banks between 2009 and 2017 for misconduct—including almost 600 enforcement actions against financial institutions in just the United States in 2014—which is likely to grow to more than $400 billion by 2020, according to a research report by Quinlan and Associates. The firm expects to see continued increases in fines over the next few years.

In early 2018, one US bank was fined $613 million for failing to monitor suspicious transactions and other AML activities. Likewise, a German bank was fined £163 million for serious failings in its AML controls. As high-profile financial crimes prompt governments to pass ever-more stringent regulations, the risk and crime-fighting teams at financial organizations are forced to make investments to update their practices and systems to achieve and maintain compliance.

Banks are also justifiably concerned with how fraud events erode customer trust and organizational reputation. As researchers at Carnegie Mellon University recently confirmed, customers are much more likely to leave institutions where they have experienced fraud, even if they receive quick refunds of their losses. Beyond the impacts of fraud on customer relationships, financial institutions also suffer stock price drop after significant fraud events.

Challenges to Keeping on Top of Financial Crime

When the attacks are constant and ever-evolving and the fallout is costly in so many ways, it is imperative to find effective responses to the threats of financial crime. But financial institutions face diverse challenges when it comes to successfully combating fraud, money laundering, and cybercrime. Here are just some of them:

Regulatory divergence
: Regulations continue to multiply. What makes the regulatory situation worse from a compliance management perspective is that these regulations are not unified but instead are being imposed at regional, country, and global levels. The volume, variety, and often conflicting nature of these mandates adds complexity to an already difficult job.

Growing popularity of mobile channels
: Today, consumers use their phones and other mobile devices to shop, check their bank balances, and even open new credit card or banking accounts. Criminals are likewise shifting their focus to mobile. One report found that 54% of businesses have seen a rise in mobile fraud, including *device cloning*, in which a fraudster creates a software duplicate of a victim's smartphone. According to the latest Kaspersky Cybersecurity Index, 35% of consumers use their smartphones for online banking, and 29% for online payment systems; up from 22% and 19%, respectively, in the previous year. The arrival of mobile adds to the proliferation of channels that financial institutions must monitor and protect.

Demand for real-time transactions and payments
: Increasing demand for fast, easy transactions—made anywhere, including international payments—means that financial institutions need to make ever-more-rapid decisions about the integrity of transactions. Financial firms are under pressure to increase both the accuracy of their transaction monitoring systems and the speed with which transaction approval takes place. It is a massive challenge to respond effectively to fast-evolving, increasingly sophisticated threats, and at the same time to meet ever-rising customer expectations for speed and ease in transac-

tions. With these conflicting pressures, it's perhaps unsurprising that many costly mistakes are made.

Cryptocurrencies as targets
: As more financial institutions begin to explore the use of cryptocurrencies like Bitcoin, cybercriminals are homing in on cryptocurrencies as good vectors for attacks, including for implanting malware. Mining malware is increasing year by year, and this is expected to continue.

Pressure to innovate
: Financial service firms are being forced to innovate to maintain competitive advantage not only against traditional competitors but also against increasing numbers of innovative upstarts and tech giants. For example, Square has introduced a Bitcoin exchange and Facebook began obtaining banking licenses. Together with new financial product or service offerings come untested and unknown vulnerabilities, making these new offerings relatively easier targets for criminals.

Criminals Are Innovating Relentlessly

Financial criminals move very quickly to respond to shifting vulnerabilities and opportunities, and their schemes change constantly. It can be difficult or impossible for a bank or a financial institution to keep up. For example, when EMV was mandated, the change in attack tactics from the criminal community was dramatic and immediate. Point-of-sale fraud decreased, as was intended with the EMV mandates, but criminals simply shifted their activities online, resulting in an explosion in the aforementioned CNP fraud. Other CNP scenarios, like the phishing of financial services' call center employees, also increased, with such fraud attempts rising six-fold the year EMV was implemented.

Whereas criminals have shown a reliable ability to change strategies rapidly, financial institutions have lagged. Money laundering strategies can change literally overnight, but, by comparison, financial institutions' AML measures adapt at what can seem a glacial pace. Indeed, a recent study found that 71% of financial institutions consider their inability to keep up with criminal innovation to be their biggest challenge.

A major factor holding financial services firms back from reacting swiftly to criminals' ever-changing tactics is *technical debt*: the large

investment made in legacy systems in terms of dollars, technology, and person-hours. "The fact is that today's investigation tools are simply not good enough. They date back 10 to 15 years, and were created for scenarios and behaviors that bank fraud specialists identified in the late 1990s—20 or more years ago," says Will Griffith, industry consulting practice lead at Teradata Consulting. The surveillance systems that are trying to identify and interdict some of these behaviors are not sophisticated enough to catch criminal attacks that evolved in the 2000s, he says, "let alone the 2015-to-2018 time period, or what still lies ahead in the 2020s and beyond."

Current State of Anticrime Measures in Financial Institutions

To counter the onslaught of diverse financial crime vectors, banks and other financial institutions are investing heavily in anticrime technologies. According to a PwC study, 42% of companies spent more on fighting crime in 2018 than they did in 2017, and 44% plan to increase spending over the next two years—specifically on better technology tools. But also according to this PwC survey, only 54% of global organizations said they performed a general fraud or crime risk assessment in the past two years. So, a surprising number of financial institutions are investing a significant and growing amount of money without always knowing the ways they are vulnerable to attack.

Today's Transaction Monitoring Systems

Most financial institutions' crime-fighting initiatives center on rules-based transaction monitoring systems (TMSs). They have invested substantial sums into these systems over the past several decades, and even today more than half (54%) of financial institutions still allocate the majority of their anticrime outlays to maintaining and improving their TMSs.

Why transaction monitoring? Well, a large percentage of activity in the financial services sector involves a transaction: buying things, paying for things, borrowing things, making payments, signing up for subscriptions, filling out applications, and so on. A TMS monitors transactions in as close to real time as possible. It compares the activity against a set of known behaviors or a set of known risks—

these are the "rules"—and either approves the transaction, flags it for further investigation, or prevents it from taking place.

To level-set with a basic example, consider a wire transfer of $10,000 that occurs between AT&T and 3M every third month. It might be that AT&T buys Post-Its in bulk every quarter. In this case, the transaction is predictably recurring and between two known entities; two features of the transactions that indicate they are unlikely to be fraudulent. By contrast, a wire transfer of $100,000 between a beauty salon in Brooklyn and an import-export shop in Bangladesh will trigger rules cautioning that this kind of transaction carries a lot of risk, and most institutions will use the rules to stop the transaction pending additional, manual review.

Although TMSs have earned a central role in crime defenses, they suffer from serious limitations. These systems require financial institutions—or regulatory bodies, customers, or law enforcers—to first identify criminal behavior. They then must create a specific rule to recognize and flag that behavior whenever it occurs again in the future. For example, an automated rule could be that any deposit coming from the country of Colombia that is more than $1 million is flagged for a manual review. Crime detection teams can use their understanding of crime tactics to specify rules manually or use behavioral and comparative algorithms to identify anomalies, or use a combination of both types of rule. As more and more rules are added to a TMS, it can become a difficult-to-manage tangle of code. The rules must then be tested to ensure that they stop or flag only criminal transactions, and let legitimate transactions pass. Rules engines, even when well designed and tuned, are prone to produce false-positives; that is, transactions that are legitimate but which the TMS either stops outright or flags as needing further analysis.

The Burden of False-Positive Alerts in a TMS

Ask any fraud analyst, and they'll tell you that the volume of false-positive alerts are the bane of their existence. Because of poorly defined rules in TMSs, large teams of security analysts are routinely forced to wade through thousands of false alarms to find a handful of truly fraudulent transactions for which they are looking.

Although the direct costs of hiring and maintaining a team of analysts to perform manual transaction review are substantial, there are also costs to having a system that routinely wrongly flags legitimate

transactions. All those false-positives are transactions that your good customers are counting on you to enable. Instead, by raising false alerts, you frustrate those good customers from accomplishing the activities they want to do and that you want them to do—things like buying and borrowing. Frustrating customers is no way to build customer loyalty, and, as noted above, churn rates are higher among customers whose transactions have been interrupted with false alerts.

Part of the massive challenge of false-positives stems from the premise on which TMSs are based. TMSs typically use behavior detection algorithms—frequently constrained to comply with regulations—to identify customers who behave in unexpected or suspicious ways—ways that differ from how other customers behave.

But, of course, criminals go to great lengths to blend in with legitimate customers, to not look like suspicious outliers. "Unless they are very unintelligent, they will not behave in ways that are different from everyone else," says Simon Moss, vice president of industry consulting at Teradata. "They aim to get lost in the crowd." And very often the criminals succeed at blending in: TMSs can generate tens of thousands of alerts every month, and yet the TMS rules will not find clear anomalies in behavior to distinguish the criminal from legitimate, because the criminals are so savvy, says Moss.

So, your TMS software can't just look for the proverbial needle in a haystack, argues Griffith. "It needs to look for a needle in a stack of needles, which is just an incredibly difficult problem."

The Emergence of AI-Based Crime-Fighting Systems

The good news is that new tools are coming. Dynamic, agile tools that can catch up with criminals and fraudsters, adapt as the market changes, yet make sure that decades of huge investments in current crime-fighting systems and processes, especially TMSs, are utilized and improved, not wasted.

Specifically, AI technology has arrived. At the 2017 Money20/20 conference, the analyst firm Business Insider Intelligence noted the incredible scale and speed with which AI is emerging in financial services. Big data and analytics tools, risk and fraud prevention platforms, and regulatory technology offerings are among the many

areas in financial services where AI is already driving significant returns.

AI is not a single technology or approach; it is actually a category of many technologies and techniques that allow machines to emulate and in some cases improve upon the thinking, reasoning, and judgment capacity of humans. The three technologies most relevant to the application of AI for crime mitigation in financial services are *machine learning*, *neural networks*, and *deep learning*. Let's take a look at each of these:

Machine learning
: Machine learning is a set of computational algorithms, or *models*, that predicts outcomes based on data, and which automatically learns and improves its predictions based on new data, without being explicitly programmed.

Neural networks
: Neural networks are a family of nonlinear machine-learning algorithms loosely modeled on the human brain. Training neural networks typically relies on special-purpose hardware called graphical processing units (GPUs). Unlike current common practices among security analysts that involve manually defining financial crime detection rules, neural networks automatically achieve results that previously required financial crime subject matter expertise or expert data scientists' skills, or both.

Deep learning
: Deep learning is a subset of neural networks and machine learning that applies many layered neural networks to prediction problems. All machine learning uses mathematical and statistical techniques to better understand and train models to make predictions. But deep learning has proven especially effective at automatically squeezing predictive accuracy from large amounts of data across a variety of tasks.

For a long time, AI has had the reputation of being mostly hype. But that is changing: "We're seeing people establish clear objectives of what they want to do with AI—say, reduce false-positives in an AML system—and then achieve those objectives," says Nick Switanek, marketing director for artificial intelligence at Teradata. "These successes are real. And that's the prophylactic against AI hype. It's not hype if it works." Although there are many opportunities to apply AI

in financial services, AI has provided early adopters significant advantages by augmenting traditional rules-based TMS.

Benefits of AI-Based Models to Fight Financial Crime

Financial services firms are finding that AI-based models deliver measurable value, even at this fairly early stage of deployment. Here are some of those benefits:

AI automates rules creation
: In a traditional rules-based TMS, a human must identify rules that will cleanly distinguish criminal from legitimate transactions and then program the rules into the system. By contrast, an AI-powered system can be thought of as an automated rules-making machine. Instead of requiring a human to look at, analyze, and identify criminal patterns, rules emerge directly from the data through the process of training the AI model.

AI can catch crimes traditional rules can't
: Machine-made rules, derived from massive quantities of complex, nonlinear, and time-delayed data, can be subtler and more accurate than those that a human can discover or articulate. AI makes "hiding in the noise" much more difficult.

AI decreases false positives
: Not only do AI models improve the accuracy with which crimes are detected, but AI approaches can simultaneously reduce the flood of false-positives that are so costly to internal security operations and to the institution's reputation. Traditional machine learning (ML) techniques are almost universally plagued by a trade-off between false-positives and false-negatives: With AI, it's now possible to use data to improve both problems at the same time.

AI is agile and adaptive
: AI's agility and its ability to help you stay current with ever-changing threats is one of the biggest advantages of AI. That's because AI can adapt on the fly to changing threat behaviors. And when an AI system sees something it hasn't seen before, it can be programmed to issue an alert and to say, in effect, "this is something I've not experienced previously, therefore I need to have a human review it." The more the system is used, the more it learns, and the better it performs—unlike a rules-based sys-

tem, which simply reapplies the same rules over and over again. All results can then be fed back into the TMS based upon evidence-based investigation rather than guesswork.

AI accelerates human teams
: AI-based models can automate many of the currently manual steps involved in a transaction investigation so that the complete vetting of even complex transactions can be done in minutes rather than the half hour to an hour that it can take using traditional rules-based systems.

AI is objective
: AI models make "adaptive workflows" consistent and disciplined, and—very important—clearly document them. You can make investigations specific and tailored to each alert, case type, business unit, asset class, or behavior type, while at the same time reducing the amount of analyst subjectivity and increasing the reliability and auditability of the investigation.

AI increases analyst value add
: With detection accuracy up and false-positives down, analysts no longer need to process high numbers of mostly false-positive alerts. Instead, they can follow only those leads more likely to actually be fraudulent or otherwise illegitimate and perform only the higher-level manual analyses that at this point are still beyond the capabilities of AI.

AI cuts costs without risking your existing investments
: You will be able to reduce the size of your investigative units, cutting costs by as much as 50%—but also realizing 50% greater efficiencies with the TMS technologies and processes you already have in place. AI doesn't replace existing systems; rather, it integrates with them: AI supplements the strengths of the transaction monitoring and case-management systems and makes up for their weaknesses.

AI anticipates future regulation
: In late 2018, the US Treasury Department's AML unit and federal banking regulators issued a statement encouraging financial institutions to use AI as part of their approaches to fighting financial crime. Although neither providing safe harbor to innovative firms that use AI nor describing new compliance protocols, the move by US regulators strongly suggests that they

are moving toward a world that welcomes if not mandates sophisticated new anticrime approaches that make use of AI.

Challenges of Deploying AI Models when Fighting Financial Crime

Yes, there are many advantages to using AI to fight financial crime, but there are also challenges. AI is not a panacea against financial crime. There are some significant challenges—challenges that most financial firms need expert help to overcome—to deploying AI successfully.

AI Models Can't Just Be Handed Over to IT

Historically, the data science team created a machine learning model or analytics algorithm. The model would then be handed over to an IT team, which would implement it and integrate it in production systems. Because of this handoff, there would typically be a six-month gap between developing, testing, and tuning the analytical model on archival data and actually deploying it so that the model could make operational decisions on live data. Two totally different organizations would be involved.

"This doesn't work. The pace at what you want to deploy new models is much too fast for that," says Ben MacKenzie, director of artificial intelligence engineering at Teradata. Recall that financial criminals and their attack tactics evolve constantly and quickly. "You need a suitable governance process in place before deploying new AI models, of course, but if you wait six months, the models you deploy may well be obsolete."

What's the answer? "Be clear that you can't turn every anticrime initiative into an IT project," advised MacKenzie. "You require push-button deployment and that requires automating the deployment of that model into production. Otherwise, IT will take it apart. They will reprogram it. They will build new software to house the model. It will become a custom IT project to get that particular model deployed. But this is neither a swift nor a sustainable option," says MacKenzie.

Model "Explainability" Can Be Difficult

Another challenge is *explainability*. Many regulations require that your decision-making processes be transparent. "You may build a sophisticated, advanced machine learning way of detecting money launderers, but if you can't explain it to a regulator, and if you can't ensure that it aligns with key aspects of a regulation, you're not going to get very far with it," says Atif Kureishy, Teradata's global vice president for artificial intelligence and deep learning.

It's not good enough to just tell regulators that the machine and deep learning models flagged this transaction as fraudulent. You need to explain all of the attributes, the probabilities, and why and how much a particular attribute contributes to the suspicion of fraud. "And then if the transaction, let's say, involves a bank that operates in the European Union, you got to make sure that you're in compliance with GDPR," says Kureishy. "Or, in the United States, are you compliant with Fair Credit Lending and the Bank Secrecy Act? If you turn someone down for a loan, you have to explain why."

But this, unfortunately, is currently one of the disadvantages of machine learning, especially neural nets and deep learning: AI models are often quite opaque. Many AI researchers are actively trying to solve the problem that many AI models are proverbial black boxes, generating predictions on the basis of inputs without providing human-interpretable transparency about how the model used the input data to generate the prediction. The current, black-box nature of many AI models is also one of the reasons why rules continue to have a place. In the credit card market, for example, some very basic rules top the crime-fighting hierarchy of actions. Is this credit card number one of ours? Is this account still open? Does it have an available credit line? Is the customer current on his or her payments? These and similar questions are binary rules, and until recently they made up the bulk of the detection logic within TMSs. Answers to these types of questions have the advantage of being easy to explain to regulators.

Fragmentation of Teams—and Therefore Siloed Data

In a panel discussion at the 2017 Money20/20 conference, Apple cofounder Steve Wozniak noted that the key to industry success will be investing in AI and building "centralized teams" focused on deploying it in ways that augment rather than replace humans.

This is an important insight with many implications, because financial institutions frequently maintain distinct teams to handle different types of crime. Such organizational fragmentation leads to fragmentation in the data that banks need to sort through to determine legitimate versus illegitimate transactions. How does this fragmentation happen? To take a current example, as banks digitize, the number of ways that they interact with customers grows. With each new touchpoint (point of contact) comes the very appropriate demand to monitor and secure the touchpoint against attacks. Each touchpoint typically comes with its own software applications and data streams, and thus tends to result in separate financial crime-fighting suborganizations within the larger financial services organization. It's all too common, for instance, for an institution to have a financial fraud unit for online banking, another unit for CNP transactions, and still another for call-center operations.

Many financial institutions also possess dated ecosystems that are not well-integrated. For example, many credit card banks store transaction data in legacy warehouses. This is typically good, detailed, and well-structured data. Along with the details of each transaction, the system has captured customer profile information, account status data, and perhaps even recent credit scores. But, until now, behavioral data is missing from these data warehouses. What has been the browsing behavior before the consummation of an application for a line of credit? Has this individual, or device (through the use of cookies), attempted this before and abandoned it? Have they attempted, under multiple different identities, to access an account? Measures of much of the behavioral context is typically stored in a separate database.

As a result, different teams are myopically focused on serving a particular customer in a particular context or through a particular touchpoint. Rather than seamlessly sharing data and analytics assets between teams, significant gaps and frictions exist, both organizational and technical. These gaps create vulnerabilities that are readily exploitable by criminals.

Even within an organization with a culture of high diligence, the current customer due diligence (CDD) and enhanced due diligence (EDD) processes often remain manual and highly fragmented across different providers and functions. Indeed, most firms invest large sums hiring third parties to perform EDD because their own CDD data access and management processes are so rudimentary and

cumbersome. But integrating and achieving consistency across CDD and EDD data sources has the potential to enable "Know Your Customer" (KYC) decision-making and onboarding processes that are strongly secure against criminal attacks.

The technical issues of siloed data are frequently driven by the particular infrastructure investments that have been made by subgroups operating within financial institutions. Different subgroups use different kinds of storage systems operating on different time latencies, with different consistency guarantees, using different data schema and indexes. Without coordinated infrastructure investments and a future-proof architecture, however, financial institutions make it near impossible to achieve "single source of truth" confidence in their data and analytics. Institutions with infrastructure fragmentation frequently struggle with even the most basic of analytic tasks, finding they have surprisingly little confidence that a customer appearing in one dataset on one platform is actually the same as one in another platform.

The technical barriers erected when there is fragmentation of data and analytics infrastructure is then compounded by the distinctive makeup of individual teams. Very different types of analysts, or operators, or investigators, work on different types of anti-crime efforts, introducing yet another hurdle to effective sharing of data and analytics. Even the data scientists from group to group might not agree on best practices. It's not uncommon for different data science and analyst groups to prefer using distinct analytic frameworks, languages, and statistical tools. These technical and cultural differences create still-further barriers to pooling data and analytics to present a united defense against financial criminals.

Real-Time AI Deployment: Current Realities and Constraints

One reason that many rules-based systems are still in place is because they're fast. Some types of transaction monitoring demand subsecond decisions as to whether to let a potential transaction proceed. A collection of logical rules can execute very quickly. By contrast, some AI models can be relatively slow, involving potentially lengthy sequences of intertwined operations to produce a prediction and associated probability about a transaction's legitimacy.

So, it is often a greater challenge to operationalize AI models than to develop the models. First and foremost, there are latency issues involving data movement and specialized compute. It takes time to pull a suspect transaction from the authorization system, to run it through a model to generate a score for the transaction, and then to reintegrate the score with the rules engine. In some scenarios, unless the AI model's score for a transaction can be generated and reintegrated quickly, the scoring process involving the AI model can be too slow to be useful. That, in and of itself, can be a major barrier, depending on the use case.

Real-time models are needed to identify credit cards that have been stolen or identities that have been compromised by account takeover as well as for criminal activities such as front running and insider trading. But for other financial crimes such as money laundering and terrorist financing, investigations unfold over years, not milliseconds: These are schemes on the scale of sometimes billions of dollars that involve monitoring suspects living what appear to be perfectly innocuous lives. They could eventually be found to be fronting major narcotics businesses and terrorist organizations, but there is no way to discover that in real time.

Real-time deployment is therefore very much a moving target, and crucially depends on the context and demands of each use case. People talk about "push-button deployment" in which a model is built, and you simply click a button and the model and its requirements are seamlessly and immediately deployed to operations. In financial crime detection settings, push-button deployment is not yet reality. Although some of the frictions that cost time and money can be removed or smoothed to accelerate deployment, but other frictions might appropriately remain.

To make AI systems for financial crime detection real time, there are a variety of dependencies, all of which also need to take place in real time: real-time integration with operational data and authorization systems, real-time data processing and model scoring, and real-time integration with the rules engine. This is not possible—yet. For this reason, "anyone who tells you that deploying AI engines that operate in real time is easy probably doesn't know what they're talking about. What's more, they probably don't even know what they're looking for," says Griffith.

Still, even without attaining "real-time" deployment, it is both possible and imperative to accelerate the time-to-value of AI crime-fighting systems. Today, many financial firms take as long as nine months to move a model into production. So even though a financial firm has developed a model that can defend them against a particular criminal attack, they will still suffer losses for another full nine months. What is worse is that the people perpetrating fraud and other crimes move much more quickly: hours, days, or weeks at maximum. When the financial firm finally blocks one channel of attack, the criminals move readily to another channel, confident that they can mine that channel for months before it will be blocked.

The frequently months-long lag between model creation and deployment creates real and persistent vulnerabilities. For example, there have been a number of recent attacks on ATMs globally. Criminals get hold of compromised information and adjust credit limits for cards that they have counterfeited or have had issued to themselves using synthetic or hijacked identities. Then, within hours, across the globe, they'll have hundreds of "mules," or accomplices, go to ATMs and withdraw the maximum amounts. And because they adjusted the maximum of $200 to $2,000, ATMs simply start spitting out cash.

In situations like these, real-time deployment will help, and we will indeed evolve to be able to respond rapidly to such attacks: financial firms with the right partners will be first to be able to integrate new but proven approaches as the data engineering continues to advance and the ecosystem of tools, accelerators, and products matures.

Managing AI Models in Production

In addition to the challenges of getting AI models into production and delivering results at speeds fast enough to be useful, there's the issue of monitoring and maintaining the machine learning models in production. The right tools do not yet exist. Numerous model governance issues are not yet adequately addressed by currently available software. If a model that once performed well starts behaving badly, possibly because the characteristics of new data no longer resemble the data the model was trained to handle, how do you quickly replace the model with one that will perform better? Are you monitoring the models in production, as well as a pool of candidate replacement models? Do you have back-up plans for redirecting data for scoring based on the observed performance of the produc-

tion and challenger models? Are you able to redeploy models easily? All of these types of operational challenges hold across all models and are more significant than the actual development of any particular model.

Managing Alerts from AI-Based and AI-Enhanced Monitoring Systems

Although AI models dramatically reduce the occurrence of false-positives, AI models still generate a large number of alerts requiring manual review. You need to consider how to most efficiently investigate those alerts. For example, if you receive an alert about a client engaging in suspicious behavior, you must thoroughly investigate that client, search for news about that individual, the client's financial history, spousal and family relationships, and more. You're gathering a lot of data, and then you need to make sense of it. This is where an emerging AI technology called robotic process automation (RPA) can help. When paired with AI models, RPA can accelerate the search for and analysis of the large amounts of data relevant to alert investigation, relieving the burden from your human staff.

Other Operational Challenges

A number of other challenges with operationalizing machine learning models exist. Models can become "entangled" when one model begins to "consume" another model. Model entanglement occurs when you have undocumented dependencies between models. This can result in unexpected cascading effects that in turn produce performance issues that are difficult to recognize and resolve.

Other potential problems occur when training AI models. A typical way to train a model is to take extracts of relevant data from various databases or other operational data stores and data streams, join them into a flattened table, and store the table as a CSV file on a distributed file system or local server, and use that to train your model. But by doing that, you're creating data silos. You end up with models that were trained on different data sets. When one works better than the others, you don't know whether it's because one model is better or because the data is different.

That's why it's recommended that data for training should always come from a well-designed and managed data warehouse. If you cache it somewhere for performance reasons, that's okay, but by all

means avoid a proliferation of extracts from your data warehouse, which inevitably leads to costly analytic confusion.

Training—and retraining—models is also difficult for other reasons. Accumulating sufficient amounts of labeled data can be a major challenge. It's not like you can go out and immediately stock up on volumes of images of people's faces and start training your models, points out Peter MacKenzie, Americas director of artificial intelligence and deep learning at Teradata.

"In the case of financial crime, where new attack techniques evolve constantly, you constantly need new labels. Meaning, you need to have a high volume of new transactions that are known to be either fraudulent or nonfraudulent so you can train your models to tell the difference between them," says MacKenzie. "And that takes time." In many cases, the only way to find out whether a transaction is fraudulent or nonfraudulent is to wait until a customer or vendor complains. In many instances, financial institutions watch for any complaints within an "aging period" of 60 days, and if no complaint is received in that time window, the transaction is assumed to be nonfraudulent: this delay means that there can be a wait of 60 days before necessary labeled data is available. Because the data used for training relies on human reporting by the customer or also human confirmation by an investigator, the labeling process is a human-in-the-loop system. Making that system work smoothly and quickly is difficult, but when done right enables a positive feedback loop that drives AI model performance to better, faster.

Then come all the other processes required to get a model into production. You need to test it and prove that it will work; you need to get the right business approval; package it up so it is maintainable; and then, finally, you can deploy it. Given all the necessary hurdles, it's no wonder that simply getting a model deployed can take a lot of time. Automating and making all those supporting processes work quickly and robustly is a major challenge.

Then there are organizational, business-process integration challenges. It's one thing to deploy a model into an infrastructure to generate scores. The next thing is to be able to integrate it into whatever your business processes are. To the extent that your business processes include people, you're going to need to change the way they work—whether you're augmenting the information they currently have so that they can make better decisions or you're replacing some

of the decisions that they currently make with decisions made by machines. You need a broad range of people with different roles collaborating. That would include the business stakeholder, the data custodians, the AI specialists, the data engineers, and the delivery leads. Having all these people working together closely is essential.

And you need to keep in mind that a model is only as good as the input that you provide to it. If you give it the wrong input, or—more importantly—if you don't monitor and recalibrate it on a continual basis, you will not succeed. These models need to be constantly managed, calibrated, and realigned. "In effect, they are not self-learning, they are taught," says Mackenzie. "Or you can call it supervised learning. But you can't just deploy your model and go away."

Accelerating AI Model Deployment with AI-based AnalyticsOps

A framework of interdependent software tools and organizational practices that helps operationalize models more quickly, more reliably, and with more traceability is sometimes referred to as *Analytics Operations*, or *AnalyticsOps*.

As its name implies, AnalyticsOps brings together two things: analytics and operations. As discussed earlier, analytics innovation often takes months to be deployed, which is common both in financial firms and in other sectors, as well. AnalyticsOps streamlines and accelerates the handoff from development to operations by making sure there is a robust and repeatable process for getting the AI models into production.

AnalyticsOps is modeled on the DevOps framework used in software development. DevOps is a collection of Agile software development best practices and automation tools supporting continuous innovation and deployment in software. With the DevOps approach in software, application bugs and defects can be continuously removed, and new, better features are continuously added. Like DevOps, AnalyticsOps enables the operationalization of AI innovations nearly as soon as they're proven to be of value, as opposed to forcing them through a series of bureaucratic and technical bottlenecks before deployment in punctuated releases.

AnalyticsOps works best if you begin with a vision of what the results of deployed analytics will look like and then put the framework in place to support that vision.

This could mean having a model management framework and a set of agreed-upon metrics such as reducing false-positives, decreasing investigation times, or accelerating the deployment of models or rules to capture newly discovered criminal tactics. Or your vision might be an amalgam of these and other metrics.

When you have goals in mind, you and your trusted partners can look at the advanced analytics toolkit and ask what other data you might want to include to make better decisions about financial crime, and what AI tools are there for transforming that data into potential predictors, prediction probabilities, or classification probabilities. Then, because you already have explicit agreed-upon metrics that will matter to your organization, you can take the actions most likely to help you past those particular goalposts.

Don't Replace Your TMSs—Supercharge Them

For all these reasons and more, don't think about replacing your TMSs. Rules engines aren't going away. They have their place. They provide a very robust legacy capability, in that they're operational, high-volume, reliable, and redundant.

Additionally, fighting financial crime is a make-or-break strategic business activity. General prudence would dictate that you don't aggressively seek to turn off working TMSs, despite their limitations. You need to gain experience with your AI models in production over time, and make sure that you have fully mastered all of the operational challenges of maintaining the model in production before you even think of retiring elements of your existing systems.

Also, keep in mind that rules engines can be updated much more swiftly in emergency situations. Imagine that a law-enforcement agency notifies you that a particular merchant has been stealing credit cards. You can immediately put a rule into your TMS that explicitly excludes transactions from that merchant. Given time and enough labeled data, your AI model would have eventually learned to exclude this merchant, but it would have taken time—and might have required external data beyond what you have in your own dataset. Indeed, rules engines and AI-based systems complement each other.

"All too often, the approach to solving financial crime-fighting problems is to replace your current TMS with another, supposedly better TMS," says Moss. "This is like replacing a steam engine with another steam engine. You're missing out on the new technology that's now available."

So, keep the TMS you have, advises Moss. "Give it some tender loving care, and focus your budget on supercharging your AML team with a machine learning or rapid analytics layer that integrates with the TMS." This layer will take advantage of the very real value the TMS has in terms of customer history and enterprise transaction data as well as regulator-approved TMS scenarios. But it adds a complementary ability to make new and accurate inferences of criminality, connect the breadcrumbs more effectively, and identify complex behaviors that current TMS approaches find challenging.

Real-World Case Study: Danske Bank

Like many financial institutions, Danske Bank had a legacy fraud-detection system in place for many years. But it was having little success actually detecting credit card fraud—identifying just 40% of fraudulent transactions, and generating a flood of fraud alerts that were 99.5% false-positives. The bank made a strategic decision to partner with Teradata Consulting to overhaul its system with new analytics techniques—including AI—to improve its existing rules-based system.

Success came fast. Within nine months, Danske Bank slashed the false-positive rate by 60% and raised the true-positive detection rate by 50%. The first phase of the project deployed machine learning, and the bank will soon improve these numbers further with other kinds of AI models.

Overcoming Challenges

A top priority when jump-starting the initiative was making sure the team had access to a sufficient amount of clean, accurate data to train the machine learning models. But the bank had only a very limited set of accurate data. More—and higher quality—data was urgently needed.

To assemble, clean, and label the needed training data, the team undertook the tedious job of identifying and extracting historical

fraud cases from unstructured Excel spreadsheets. It had to reconstruct all historical transactions from the previous three years, following each transaction through a variety of intermediate accounts depending on type and origin. Additionally, billions of rows of relevant data existed outside of the normal business logic of the bank's real-time transactions systems, and this needed to be merged.

Next, the team needed to label the billions of transactions as either fraudulent or not to ensure that it could train on an accurate set of data. Without the ability to train the model on accurate data, the model would not be able to learn what characteristics associated with which transactions were predictive of criminal activity.

After the extensive data preparation steps, the team was able to rapidly develop and train effective machine learning models—using a portfolio of boosted decision-tree and logistic-regression models.

Then the team encountered a roadblock as it moved to deploy the models. Danske Bank was founded almost 150 years ago. It possessed decades of transactions that had been processed by a mainframe server. It was going to be difficult to achieve the performance the bank required when deploying the models on the bank's legacy infrastructure.

The bank needed an architecture that would allow the models to run across millions of daily transactions in near real time. To achieve this, the team designed and implemented a new analytics platform. Leaving the existing environment intact and integrating seamlessly with it, the analytics platform made it possible to deploy the machine learning models.

After a shadow production phase when the results of the combination of machine learning models with rules engine was shown to outperform the rules engine alone, the machine learning model was expanded to run on multiple datacenters. The team did continual performance monitoring. Finally, nine months after the start of the project, with Danske Bank reassured by extensive testing and validation, the machine learning model was made live. Immediately, the bank saw a significant improvement over the former rules-based system: The rate of false-positives plummeted 50%, which cut the workload of investigators in half.

For the next phase of the project, Danske Bank will integrate deep learning software with GPU appliances to use AI models to capture

the remaining cases of fraud and reduce the false-positive rate even further.

From Machine Learning to Deep Learning

As the team shifted its attention from more traditional machine learning to the development of modern AI models, it was able to use the analytics platform it had built during the machine learning phase to test and validate different kinds of deep learning, neural network architectures. Reimagining deep learning architectures originally designed for visual detection and object recognition as tools for making predictions with sequences of transactions, the team found substantial improvements in model performance. Computer vision models are some of the most advanced of AI models, having surpassed even human performance at identifying and labeling objects, and doing it much more quickly and reliably: Danske Bank benefited from having an AI data science team able to translate the strengths of AI models designed for computer vision into a domain characterized not by images, but by sequences of transactions and transaction-related attributes.

The net result of applying the first iteration of AI models to the fraud detection use case, including computer vision and sequence models such as long short-term memory (LSTM), was a further 20% reduction in the false-positive rate—a significant improvement over traditional machine learning models.

A Platform for the Future

Through its partnership with Teradata Consulting, Danske Bank was able to build a fraud detection system that made autonomous, accurate decisions, integrated with existing business processes and systems, and met the bank's requirements with regard to security, availability, and time latency.

For Danske Bank, building and deploying a custom analytic solution that met its specific needs and utilized its data sources delivered vastly more value than an off-the-shelf fraud detection product because the custom models outperformed competing models and because they established the foundation of an agile platform for future analytics development and deployment.

With its enhanced capabilities, the solution is now ready to be used across other business areas of the bank to deliver additional value,

and the bank is well poised to continue using its data in innovative ways to deliver value to its customers.

Conclusion

In closing, AI provides a new way forward to mitigate financial crime—a new way that goes well beyond the myopic concerns that have motivated anticrime efforts of the past.

Pleasing regulators is not enough.

Regulators can say you're doing a good job complying with regulations, but regulatory compliance is not a good measure of business success. Even with the late 2018 guidance from US regulators that companies should look to incorporate AI into their AML efforts, regulatory compliance isn't enough. Of course the advantage of a focus on compliance is that compliance is clear—indeed, AI and machine learning can help to automate business processes, including communications, to ensure that they systematically comply with regulation. It is appropriate that there are investments in operations to comply with regulations. However, in spite of compliance with regulation, financial crime still costs businesses $4.2 trillion annually: it's clear that compliance with regulation is not synonymous with crime prevention. Instead of only responding to regulator guidance on what practices to implement, financial institutions need to discover, implement, and own the practices that go beyond compliance to prevent crime. They need to be proactive, to step up and define what good AML and fraud detection and other crime-fighting processes look like, to protect themselves against the crime that threatens to compromise the integrity of their services and brand.

An effective defense against financial crime is an effective defense of your brand.

There's a misconception that the financial-crimes team does not contribute to the success of the corporate brand. Anticrime measures do not only protect against immediate financial losses. As soon as a financial institution's reputation for safety and security comes into question, shareholder and brand value take a significant hit. Every anticrime advantage you have defends you against lost revenue, lost customers, and lost reputation. AI-based approaches have demonstrated step-change improve-

ments in the fight against financial crime. Defend your organization against crime like your brand depends on it, because it does.

Fighting financial crime is the responsibility of the entire organization. In the past it was thought that IT or dedicated risk and compliance employees alone are responsible for the fight against financial crime. Given that financial crime represents an existential challenge to the operating model of the firm, everyone from the board of directors on down needs to treat it with the seriousness and common purpose that it deserves. The reality of fast-evolving attacks by financial criminals demands that institutions have methods shared across the organization to respond to attacks with three A's of accuracy, acceleration, and automation: AI is the best-performing approach to power a unified and ever-improving platform to deliver what the entire organization needs in the fight against financial crime.

Virtually all financial firms at this point are either deploying AI or planning to experiment with AI. The smartest executives realize that the swift and effective adoption of tailored AI will be one of the techniques that shape the future of financial services. AI will cut compliance risk, scrutiny from regulators, and, ultimately, cost. You won't get your name dragged through the mud. You won't get fined huge sums. And you won't have an auditor sitting at the next desk, watching your every move because you're under a consent decree.

The benefits of deploying AI to fight financial crime are broad and deep. The winners will be defined not by necessarily who builds the best models, but by who has the best data foundations and who has built analytics, engineering, and operational excellence into the very fabric of their companies to bring wave after wave of AI advances to the fight against financial crime.

About the Authors

Atif Kureishy is vice president of Global Emerging Practices and AI/Deep Learning for Teradata Consulting. His teams are trusted advisors to the world's most innovative companies to develop next-generation capabilities for strategic, data-driven outcomes in areas of artificial intelligence, deep learning, and data science. You can connect with Atif on LinkedIn and on Twitter @AtifKureishy.

Chad Meley is vice president of Solutions Marketing at Teradata, responsible for Teradata's Artificial Intelligence, IoT, and Analytical Ecosystem solutions. Chad understands trends in machine and deep learning, and leads a team of technology specialists who interpret the needs and expectations of customers.

Professional awards include Best Practice Award for Driving Business Results in Data Warehousing from The Data Warehouse Institute and the Marketing Excellence Award from the Direct Marketing Association. Chad is coauthor of *Achieving Real Business Outcomes from Artificial Intelligence* (O'Reilly), and is a regular speaker at conferences, including O'Reilly's AI and Strata Conferences, DataWorks, and Analytics Universe. You can reach Chad on Twitter at @chad_meley.