THOMSON REUTERS

REGULATORY INTELLIGENCE

AI incidents in financial services: part two - getting ahead of the problem

Published 18-Nov-2020 by
Patrick Hall and Collin Starkweather

While technology provides organisations with immense benefits, it can also fail, suffer attacks, or be used for malicious purposes. In that respect, artificial intelligence (AI) and machine learning (ML) are no different from other technologies. Indeed, as AI/ML is adopted more widely, AI incidents are occurring more frequently.

Part one of this series focused on learning from past incidents; part two looks to the future. The authors put forward two broad types of AI/ML risks that have not been documented publicly in the financial services sector, and then discuss AI incident response as a regulatory gap and as a common sense mechanism for offsetting risk and liability stemming from AI incidents.

Novel AI/ML concerns for the financial services industry

Part one summarised two well-known AI incidents, and urged organisations to learn from the growing body of documentation concerning AI incidents to avoid them in the future. There are, however, types of incidents that have occurred in other domains or that have precedent in financial services, that have not yet been documented. The relevance of these include systemic risks or specific AI/ML security vulnerabilities.

Systemic risks

Systemic failures, such as those observed during flash crashes, could also be triggered by errant AI/ML systems. Firms should monitor AI/ML systems and regulators should monitor trading patterns for technical problems such as error propagation and feedback loops, which have the potential to cause a system to spiral out of control. Error propagation occurs when problems in the outputs of one data-driven system feed into another data-driven system's inputs, leading to degrading accuracy in each subsequent system. Feedback loops can happen whenever an AI/ML system affects reality, for example, by executing a trade, then that action appears again as a data input for the system to learn or decide about. Outside of finance, AI/ML feedback loops have already been associated with negative outcomes in predictive policing. Given that these feedback loops have been observed in other high-impact contexts, the potential for AI/ML feedback loops to produce unexpected market dynamics in financial markets appears very real.

Security and related risks

The most common kinds of AI/ML attacks discussed by researchers include adversarial manipulation of AI/ML system outcomes, insider manipulation of system outcomes, IP theft through AI/ML system endpoints, and trojans hidden in various software artifacts associated with these systems.

Some aspects of what may be legitimate algorithmic trading activity can also bear a close resemblance to AI/ML attacks. Trading strategies are by nature evolutionary, and algorithmic traders may engage in a "cat-and-mouse" game in which they attempt to gain an edge by reverse-engineering algorithmic trading models whose activities they observe in securities markets, in effect obtaining a blueprint of the algorithm.

They may deploy reverse-engineered models themselves, hoping to profit by replicating the strategy, or use their knowledge of the models' strategy to devise profitable counter-strategies by means such as adversarial manipulation of trading patterns recognised by that model. Due to the speed with which trades can be executed by automated trading systems, a formerly profitable strategy can very quickly lead to substantial losses in the face of an effective adversarial system.

To address potential cyber-security concerns, financial firms should consider internal controls and auditing specifically addressing AI/ML security problems, as is done by organisations such as Facebook, and be familiar with related regulation and guidance, such as Federal Trade Commission (FTC) reasonable security standards or breach reporting requirements. Incident response plans should also be updated to address identification and containment protocols specifically for AI/ML attacks.

Generally, financial firms with trading operations should be aware of the potential for trading strategies deployed by AI/ML systems to be compromised (possibly by other, adversarial AI/ML systems) and consider internal controls and monitoring to offset the potential for losses should an adversarial attack lead to the rapid reversal of fortunes.

AI incident response plans

Beyond media reports of AI incidents, many European nations, the UK and Singapore have already put forward specific guidance for AI/ML systems. In U.S. financial markets, regulators including the Commodity Futures Trading Commission (CFTC), the Financial Industry Regulatory Authority (FINRA), the Consumer Financial Protection Bureau (CFPB), the FTC and the Federal Deposit Insurance

Corporation (FDIC) have all signaled their interest in AI/ML oversight, but AI incident response plans are often an afterthought in government guidance or business processes.

As can be seen in Table 1, some domestic and international guidance does touch on the subject, but often leaves important details for organisations to fill in for themselves.

Table 1: A brief summary of how several domestic and international guidance documents treat AI incidents

Guidance	Addresses AI incident response plans
GDPR	In detail, but only for "traditional" data privacy incidents
UK ICO - Guidance on AI and data protection	In detail, but for "traditional" data privacy incidents
Singapore PDPC - Model AI Governance Framework	Gap acknowledged (but unaddressed)
FRB - SR 11-7	Only briefly, but with a focus on accuracy and stability
Equal Credit Opportunity Act	In detail, but only for transparency in adverse action notices
FTC - Using Artificial Intelligence and Algorithms	No
FINRA - Artificial Intelligence (AI) in the Securities Industry	Gap acknowledged (but unaddressed)
CFTC - A Primer on Artificial Intelligence in Securities Markets	Gap acknowledged (but unaddressed)

A £100 microwave oven comes with a troubleshooting document, but most of the AI/ML systems seen in the wild, often with multi-million-dollar research and development price tags and tasked with managing assets that are orders of magnitude more valuable than that, come with zero troubleshooting guidance for discrimination, privacy or security problems.

That is not to say that financial services firms are fully unprepared for AI/ML failures. As mentioned above, the discipline of model risk management on the consumer finance side is fairly mature. For traders, supervisory and compliance manuals may also address these risks, though many fail to address the unique risks associated with AI/ML. Throughout the industry, seasoned IT security professionals are familiar with building and using general computer incident response plans. As with many supervisory and compliance manuals, standard incident response manuals — such those from SANS Institute and NIST — do not specifically address AI/ML system failures or attacks. That potentially leaves firms open to the kinds of liabilities discussed in part one.

As the adage from statistician George Box — "all models are wrong, but some models are useful" — makes clear, sophisticated AI/ML systems may make bad decisions or fail outright, especially over a long period of time. As financial AI/ML systems transition from enhanced predictive models to running trading systems, chatbots, robotic process automation and beyond, they are going to fail in different and unexpected ways. The good news is that organisations such as the CFTC, law firms and consulting companies have started to recommend response plans for AI incidents. Bnh.ai, a boutique firm focusing on AI and data analytics, has released a free and open sample AI incident response checklist.

Rapidly changing landscape

Firms should take care not to believe the out-sized hype and not to get caught flat-footed, and should consider spending the relatively modest amount of time and resources required to build AI incident response plans. AI/ML is a powerful and promising technology but, like previous generations of transformational commercial technologies, it can fail, be abused, or be misused. AI incident response plans are one of the most direct and concrete steps firms can take to protect themselves from known harms as they adopt this potentially transformational technology.

Additionally, many regulators are devoting considerable attention to risks associated with AI/ML systems, and both guidance and regulation are evolving quickly. In this rapidly changing regulatory landscape, management and compliance officers need to keep up-to-date with the latest developments and be aware of any impact on best practices and compliance policies and procedures.

Patrick Hall is principal scientist at bnh.ai and Collin Starkweather is an economics and technology consultant and founder of Starkweather Economics LLC. Thanks to Andrew Burt for his research and insights on technology and regulation

Complaints Procedure

Produced by Thomson Reuters Accelus Regulatory Intelligence 18-Nov-2020