AI incidents in financial services: part one - learning from the past

Published 17-Nov-2020

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 forms of technology. Indeed, as AI/ML is adopted more widely, AI incidents are occurring more frequently. The Partnership on AI incident database now contains more than 1,000 incident reports, and these are just the publicly acknowledged incidents. The actual numbers are doubtless substantially higher.

Despite these incidents, many financial firms have a head start in some operational areas on AI/ML risk mitigation. Their efforts to comply with anti-discrimination, data privacy and security and model risk management guidance — coupled with mature IT security practices — represent some of the most widely used best practices for controlling AI/ML risks. These practices may, however, be applied unevenly, especially across large and diverse institutions, and the financial services sector still has some gaps when it comes to managing the myriad risks of AI/ML.

In part one of this series, the authors examine two of the most high-profile AI incidents from the broader financial industry, with an eye to the prevention of future incidents.

1. Alleged discrimination in consumer finance

In these politically charged times, lenders are unlikely to want viral allegations of discrimination to be associated with the rollout of a new, high-end consumer credit product. This is exactly what happened last year, however, when Goldman Sachs and Apple released their new credit card, and regulators are taking the allegations seriously. While regulatory damages have not yet been discussed publicly, Consumer Financial Protection Bureau-reported penalties for recent consumer finance discrimination matters have ranged from $95 million to $225 million.

What happened?

A number of instances were reported in which women received substantially lower credit limits than their male spouses, including a highly-publicised tweet from Steve Wozniak, co-founder of Apple, in which he highlighted differences between the credit limits extended to him, and those offered to his wife. The couples took to social media to vent their frustration, resulting in reputational blowback and regulatory investigations.

Why this is a problem

"Any algorithm, that intentionally or not results in discriminatory treatment of women or any other protected class of people violates New York law," said Linda Lacewell, spokeswoman for the New York Department of Financial Services. Moreover, litigation and regulatory fines are not the only potential negative outcomes here. An unknown number of consumers may have been subject to discriminatory credit decisions. Both Apple and Goldman Sachs will have incurred legal and public relations expenses associated with the incident, and the opportunity costs of wary consumers choosing to apply for different cards could also be substantial.

How to prevent similar incidents

Anything related to consumer credit has the potential to draw a lot of public, media, or regulatory scrutiny. There are many AI/ML risks to consider.

Algorithmic discrimination

All technical aspects of a lending decision should be examined for discrimination in terms of output decisions (referred to as "disparate treatment", or "disparate impact" depending on the nature of the discrimination), and likely output accuracy too (sometimes referred to as "differential validity"), across protected groups and subgroups. Technology alone will not solve algorithmic discrimination problems, however. Organisations should also update their culture and processes as necessary to empower a diverse group of personnel to audit AI/ML systems and respond to potential incidents quickly, from technical, reputational and legal standpoints.

Federal anti-discrimination regulations

The U.S. Equal Credit Opportunity Act (ECOA) prohibits varying "the terms of credit offered, including the amount, interest rate, duration, or type of loan" based on protected class, including race, colour, religion, national origin, sex, marital status, age, or source of income. New York State fair lending law extends the ECOA to additional protected classes, such as military status and sexual orientation. Beyond traditional fair lending regulations such as ECOA, the Federal Trade Commission (FTC) has also telegraphed its interest in algorithm-driven consumer credit decisions, down to the level of data furnishers. Organisations must ensure that novel AI/

ML lending systems are in compliance with traditional fair lending regulations, and should watch the horizon for new guidance from the FTC, Congress, or other U.S. federal agencies.

Local anti-discrimination laws

Local laws may also provide a means to address potential discrimination. In the absence of general federal regulations on AI/ML, there are numerous state and other local anti-discrimination laws that could potentially be applied to AI/ML systems, for example, Washington D.C.'s Human Rights Act. To account for this additional legal complexity, data scientists should work with attorneys or compliance personnel to consider all applicable anti-discrimination laws in the design stages of AI/ML systems.

The potential risks must not be underestimated. If AI/ML systems can go wrong for two of the world's best-resourced organisations, it can happen at any organisation.

2. Suing an AI-managed hedge fund

In a case that could reverberate around the financial and AI/ML worlds, a billionaire investor is suing an AI/ML-based quantitative hedge fund for $23 million in losses. This case is important because there are few precedents from which to gauge the appropriate allocations of risk and liability when "black box" AI/ML systems make autonomous decisions that may result in losses, and the disclosures and representations that are appropriate for reliance on these systems.

What happened?

Beginning in late 2017, Samathur Li Kin-Kan, son of Hong Kong billionaire Samuel Tak Lee, provided Tyndaris Investments with assets through British Virgin Islands investment vehicle MMWWVWM Ltd (VWM) to manage using an AI/ML trading system. At its peak, Tyndaris was managing an investment by VWM of roughly $2.5 billion.

The investment followed a demonstration of the system's performance during simulations by the chief executive officer. In live trading, however, the AI/ML-managed fund's performance failed to live up to expectations. The fund experienced losses in December 2017 and January 2018, and on one particularly bad day in February 2018, the system allegedly suffered $20 million in losses. Li subsequently withdrew his funds and sued Tyndaris for $23 million in losses. Tyndaris is counter-suing Li for $3 million in unpaid fees.

Why this is a problem

The plaintiff has alleged that Tyndaris misrepresented the capabilities of the AI/ML system. The litigation continues, and the legal issues may yet remain undecided should the parties settle. Many general questions of AI/ML liability and negligence remain untested. Aside from a small number of U.S. cases, such as Rodgers v Christie (795 Fed Appx 878 (2020)), and recent EU policies, the assignment of liability should an AI/ML system behave unexpectedly is something of a legal and regulatory grey area. This potentially leaves operators, consumers and investors in financial and legal limbo when AI/ML systems go wrong.

How to prevent similar incidents

As with any investment opportunity, investors must do their due diligence with AI/ML based trading systems. From a technical standpoint, those who rely on AI/ML systems, whether for algorithmic trading or in other applications, should not trust them blindly. AI/ML systems can make bad decisions millions of times faster than humans, which in a trading context can accelerate the accumulation of losses and limit the capability of managers to perform timely interventions. So, supervisory policies and procedures for any high-stakes deployment of AI/ML should be structured to anticipate potential weaknesses and quickly identify issues that may arise and supervisors should likewise have a plan in place to act quickly when they misbehave.

Learning lessons

The AI incidents discussed above are two very public examples from the news media. Several additional lists of public AI incidents are now available, and the Partnership on AI is compiling a large, interactive database of incidents. Just as people continue to study and learn from transportation incidents, organisations should learn from the examples set out in this article, and those tracked elsewhere, to reduce the risk of future incidents.

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

Complaints Procedure

Produced by Thomson Reuters Accelus Regulatory Intelligence 17-Nov-2020