Cybersecurity Management of AI Systems: Managing an Attempted Breach at E-Fortress Capital*

June 2020

Sanjana Shukla
George Wrenn
Dr. Keri Pearlson

Financial services has been a highly regulated industry, and firms have faced a myriad of regulatory, trust, and privacy concerns, including stolen intellectual property (IP), breach of personal identifiable information (defined in **Exhibit 1**), and loss of other valuable data.

Background

E-Fortress Capital (**Exhibit 2**) was an emerging hedge fund with headquarters in New York City which, in an increasingly challenging environment created by post-financial crisis market conditions, had managed to generate impressive returns posted at nearly a 13% gain in the previous year. The fund had also built a compelling senior management team. By definition, a hedge fund was an alternative investment fund focused on making investments designed to protect investment portfolios from market uncertainty, while generating positive returns (alpha) in both bull and bear markets[1]. Since its establishment, E-Fortress's assets under management (AUM) had surpassed $1Bn, and the firm had become a highly regarded and trusted hedge fund. Its clients included endowments, foundations, corporations, pension funds, and high net-worth individuals[2]. The hedge fund also used a discretionary investing approach as opposed to a purely systematic approach, and as a result its strategy relied heavily on the skill and judgement of the portfolio manager when making investment decisions[3].

The Securities and Exchange Commission (SEC) required individuals to qualify as accredited investors to use hedge funds, including E-Fortress Capital, as investment vehicles. At a minimum, such an investor was required to have an annual income of $200,000 and a net

[1] Hedge Fund Definition. *Hedge Fund Marketing Association*. Link: https://www.hedgefundmarketing.org/hedge-fund-definition/
[2] Where Hedge Funds Get Their Capital. *Investopedia*. Link: https://www.investopedia.com/ask/answers/121614/where-does-hedge-fund-get-its-money.asp
[3] Discretionary vs. Systematic: Two Contrasting Hedge Fund Approaches. *Preqin*. Link: https://docs.preqin.com/newsletters/hf/Preqin-HFSL-Jun-14-Systematic-Discretionary-Funds.pdf

worth above $1 million. E-Fortress, however, reserved its hedge fund investments for clients with at least $3 million in investable assets[4].

Jack Torrens was a high-net worth individual who had $5 million in investable assets entrusted to E-Fortress Capital. As an accredited investor, Jack was a "qualified investor", a high net-worth individual who thought he understood the unique risks associated with high speed algorithmic trading hedge funds, but not AI trading funds, and had a sophisticated understanding of personal finance, investing, and trading.

David Ambrose was a portfolio manager (PM) at E-Fortress Capital. With an undergraduate degree in statistics from Imperial College London and an MBA from MIT Sloan School of Management, he was a successful portfolio manager respected by the firm's senior leadership.

David had recently read an article in the Wall Street Journal about how financial services firms using applications with artificial intelligence had profited from better predictions, fewer errors, and greater efficiency[5] in making investment recommendations and trades as compared to firms who did not use artificial intelligence.

Fundamental analysis was a method for measuring a security's intrinsic value by examining related economic and financial factors with the goal of arriving at a number that an investor could compare with a security's current price in order to determine whether the security was undervalued or overvalued[6]. David was also familiar with online literature discussing certain advantages to supplementing investment and trading decisions, fundamental or algorithmic, with AI-based recommendations. These advantages included: emotional trading elimination, greater discipline in following strategy rules, more consistent investment behavior, and reduced losses[7]. David believed that integrating such an AI-based application would help portfolio managers at E-Fortress improve their final investment decisions and trades. He also believed that integrating AI-applications would give E-Fortress a competitive advantage to distinguish itself from other hedge funds. The success of such an application would also allow

[4] Hedge Funds in High Net Worth Portfolios. Susan B. Weiner, CFA. *Advisor Perspectives.* Link: https://www.advisorperspectives.com/pdfs/newsltr08-2-2-1.pdf

[5] Smart money: AI transitions from fad to future of institutional investing. PwC. Link: https://www.pwc.com/us/en/industries/financial-services/library/artificial-intelligence-investing.html#:~:text=Some%20firms%20are%20using%20AI,efficiency%20for%20the%20investment%20industry.

[6] Fundamental Analysis. *Investopedia.* Link: https://www.investopedia.com/terms/f/fundamentalanalysis.asp

[7] Designing Automated Trading Systems for Commodity Trading – Theoretical Aspects. Petr Tucnik. *ResearchGate.* Link: https://www.researchgate.net/publication/271642533_Designing_Automated_Trading_Systems_for_Commodity_Trading_-_Theoretical_Aspects

David to move up the pecking order at the company and become a technical communicator with senior management.

After further consideration, he decided to approach the firm's Chief Economist Alisha Singh and Chief Technology Officer (CTO) Sheryl Smith and seek their guidance. Alisha told David:

> "I'm glad you brought that up. At this week's meeting, senior management discussed integrating an AI-based trade recommendation system into the fund's investment approach, but we all agreed that introducing AI applications would come with potential cybersecurity challenges. A preliminary study showed promising signs that AI-based systems outperform non-AI systems, but we were cautiously optimistic because the data supporting this finding was early at best. That being said, the study showed that yield, Sharpe ratio, maximum drawdown, and customer satisfaction did better with AI-based systems in place. So the team came to believe that adopting an AI system for making investment recommendations would benefit our clients and allow us to further differentiate ourselves from competitors, but we would need to integrate the technology in a secure manner and reduce risks associated with early AI adoption. This would come with certain guardrails that we'd need to build around the AI system, from a fund governance and data custodian standpoint. As a PM, you would have to be willing to accept certain restrictions, even if they interfere with a business opportunity."

David agreed with the sentiment, and upon Alisha's advice, decided to speak with the firm's software development team. In the meantime, Alisha informed the hedge fund's Chief Legal Officer (CLO) and Chief Compliance Officer (CCO), and they immediately engaged the Legal and Compliance departments in exploring the legal considerations and requirements for adopting an AI-based investment recommendation system. The CLO recommended a review of 23 NYCRR 500[8] as a first step. David downloaded the PDF and started reading the section of the regulation reproduced in **Exhibit 3**. The section was titled "New York State Department of Financial Services 23 NYCRR 500".

Designing the System

At the outset, the IT team was baffled. The team's primary responsibility had always been maintaining the firm's IT infrastructure and networks, and the developers had never built an AI-based application, let alone have expertise on the NYC DFS laws, to assist investment and

[8] New York State Department of Financial Services 23 NYCRR 500: Cybersecurity Requirements for Financial Services Companies. Link: https://www.dfs.ny.gov/docs/legal/regulations/adoptions/dfsrf500txt.pdf

trading decisions at E-Fortress. Also, the CLO just learned of pending legislation named the "SHIELD ACT," also known as the "Stop Hacks and Improve Electronic Data Security Act," which she believed to be directly applicable to E-Fortress as well.[9]

Constructing the AI-application for E-Fortress required designing a system at the intersection of technology, law, public policy, and finance, and the system would use data models and ancillary algorithms to inform investment recommendations.

The engineering team understood that for the system to be trusted by PMs, it needed to be cyber secure. This security requirement led to the question: What were the cybersecurity governance and management issues that the team needed to consider when designing the system?

As a smaller hedge fund, there was no dedicated Chief Information Security Officer (CISO). For this reason, Sheryl the CTO decided to hire cybersecurity consultant Jim Donoghue. Jim told Sheryl's team to consider the framework shown in **Exhibit 4** before beginning system development. In the framework, the Trading Systems were well known by the existing IT team, and the CLO completed a legal search to capture the Legal Requirements so everyone on the team would share the same goal of Compliance for the updated system. What was missing was a working framework and proof, or Lemma, to manage the new system's uncertainty and demonstrate compliance in the system.

The CLO decided this was a discussion important for senior leadership to have, and during the next management meeting, the team collectively devised the framework illustrated in **Exhibit 5**.

As part of the collective framework, the CLO proposed a set of legal considerations, the CTO proposed a few technology considerations – which she believed were enough to push the system into production, and the team together identified several technical, ethical, business impact, and consumer perception considerations.

Leadership also came up with a list of model manageability considerations, which they acknowledged would require an overwhelming majority of work because model manageability encompassed keeping the system's data up to date. As a result, the team knew it needed to hire one dedicated employee to focus solely on managing the system's data.

With respect to the business impact concerns, one key consideration was that the system would need to be attractive to the customer base. The team had spoken with some customers

[9] SHIELD Act. Link: https://legislation.nysenate.gov/pdf/bills/2019/S5575B

including Jack Torrens. Initial stage research had also shown that firms who spoke about early technologies with customers learned that their customers were interested but cautiously optimistic about using AI for money management. This tied into the team's customer perception concerns, specifically into the uncertainty of trusting an AI system over an experienced portfolio manager. The team asked itself: If results were shared with the customer, could the customers trust those results? Would customers rather trust an AI system or an experienced PM with a proven track record?

The team shared its framework with Jim, who also pointed out that AI-based projects were unique because of the specific skills involved, so what roles did the E-Fortress development team need for the project, and what AI skills were available in-house? According to Jim, answering this would further help the team evaluate whether to build the system in-house or outsource its development. Furthermore, could the firm's existing DSML vendors work well during this transition to using AI, or did the team need to re-evaluate its current vendor partnerships? In an all-IT team meeting led by Sheryl, he further remarked:

> "Most people design an AI system before testing for cybersecurity risks. You brought me in early as the cybersecurity expert, and this will give you an advantage because we can design your system with cybersecurity in mind. Your AI system will have several components that communicate and work together to ultimately output a recommendation (**Exhibit 6**). The system will first and foremost have a data component. This will include raw data in the world that undergoes processing before it's used by the system's internal training and validation models. Your system's model component will correspond to the learning models and underlying algorithms. It will also have a communication component, which includes how the system will communicate internally between processes and externally with human stakeholders and other systems. This communication component also references a human factor component – the role of humans both in interpreting outputs by the system to inform business decisions and in choosing which output data to use as input data for the system's self-learning. We will also need to consider the overall AI system by factoring in the trust that research analysts and PMs will place on it. Overall, these components tell us that there are five high-level risk areas: data, model, communication, human factor, and overall system risks. My job will be helping you identify risks across these key areas as you begin development (**Exhibit 7**)."

Since data management was the first component Jim had highlighted, the developers decided to start there. They knew that if the system lacked correct and necessary data, it would have compromised the PM's business needs. A flourishing market had emerged for new forms of

alternative data to inform investment recommendations[10]. Examples of alternative data included GPS locations from mobile phones to understand foot traffic and point of sale data to predict same store revenues. Furthermore, data associated with the system also included PII, such as the system's log-in passwords, usernames with system access, and associated email addresses (of PMs, data vendors, developers, and other third party stakeholders) used to communicate information about the system. Sensitive information pertaining to the system also included data upon which the system trained, including any PnL (i.e. profit and loss) information, alternative data pipelines and data streams, and datasets from the fund's internal databases.

The team further devised a set of models that would be important to build in the system, including a learning algorithm. For example, one model the team built into the system was a natural language processing (NLP) enabled financial text interpreter, which parsed financial news stories on websites such as Yahoo Finance along with minute-by-minute stock price data, and then used the former to predict the latter.

Furthermore, the team considered how the system would communicate – both internally within its processes and externally between itself and other applications. One way that the system would communicate with human stakeholders was via the PM's role both in determining whether the system's output data was reliable and in interpreting the system's results to inform the PM's investment decisions. Another communication component was the feedback loop created when outputs of the system were re-entered as input data used for iterative self-learning.

While developing the system, Jim remarked:

> "There are a number of potential attack vectors that can exploit system vulnerabilities during a cyber-attack (**Exhibit 8**). When you think about data, models, people, and networks, you want to design a system that not only fits your business requirements but also is cyber secure."

To test for cyber security, Jim recommended that the team assess the trade recommendation system and its associated software and network and communication channels for vulnerabilities, and he helped the development team implement a solution in conjunction with firm's mandated legal and compliance practices[11]. Sheryl commented:

[10] What Machine Learning Will Mean for Asset Managers. Robert C. Pozen, Jonathan Ruane. *Harvard Business Review.* Link: https://hbr.org/2019/12/what-machine-learning-will-mean-for-asset-managers
[11] What is an IT Security Consultant? *CareerExplorer*. Link: https://www.careerexplorer.com/careers/it-security-consultant/

> "We simulated a series of attacks against the system and considered what vulnerabilities were present. Through discussions with Jim, we realized that there's usually a chicken-and-egg problem with doing vulnerability identification. The more black-box AI applications are treated to be, the harder it becomes to respond effectively to a breach or to detect a threat as quickly and as efficiently as we could in an AI system which we fully understood. This created a dilemma for my team: as the system developers, they had to figure out how much of a black-box to make the system while still ensuring that it was easy to use for any investment research analyst and portfolio manager. We also re-evaluated E-Fortress's existing DSML vendor partnerships. There were three vendors before, and in order to reduce the number of stakeholders interacting with the system, we reduced the number of vendors to just one."

The system underwent a 4-week pre-development process and 2-month development phase. It was deployed within David's investment team, which was the sponsoring team, as a test system shortly after, and it was continuously monitored for any flags indicating potential breaches by the humans monitoring it.

Moving Quickly to System Production

Initial beta testing had been a success, and David started trusting the test system's outputs to inform his trading decisions.

Every quarter since he had become a PM at E-Fortress, David had emailed his team of research analysts a cybersecurity awareness note. These emails generally contained descriptions of a cyber-attack that occurred in another company, reports from the web about the state of cybersecurity in various industries, or other articles he found insightful on the topic. They usually took the form of short paragraph synopsis, were addressed to his entire team, and were never daunting technical pieces of literature. David titled these emails "Cyber Awareness from Last Quarter."

Kim Li was an investment research analyst on David's team. On one particular morning, when Kim was reviewing the system's recommendations for the trading day, she noticed that the system had recommended buying a significant quantity of equity EQ shares and suggested that EQ stock was expected to benefit from a short-term gain in the market, which she thought was unusual since she knew EQ had just fired its CEO due to insider trading charges and, as a result, general market sentiment was that EQ's stock price would drop in the short-term. This was an unusual recommendation by the system; even though it had been producing both intuitive and counterintuitive outputs, all of which had been reasonable and therefore trustworthy, this output particularly stuck out to Kim. She wondered how this

recommendation could have come to be but couldn't find any indication as to why a recommended buy of EQ shares would be reasonable given the company's current market perception.

She then remembered reading an article about a former hedge fund called Gold Brook Group in one of David's quarterly cyber awareness emails. The article discussed how Gold Brook was affected by a $400 million software error a few years ago, and this error had compromised the U.S. equities trading firm at a time when the fund had a market share of nearly 15% on the NYSE and Nasdaq[12]. The loss had been due to a data poisoning attack in which a group of attackers had injected carefully crafted data samples to contaminate a Gold Brook trading system's training data in a way that eventually impaired the system's normal functions and led to a disastrous outcome[13] for the firm.

At first, Kim dismissed the idea that her own team's AI system had been tampered with. After all, the system had performed well as a beta and trusting the system's recommendations thus far had led to positive outcomes. However, for some reason she couldn't get herself to accept this system decision, and she decided that before following through with the trade recommendation, she would look further into why the system had made this recommendation. She knew that E-Fortress had a culture of, "If you see something, say something." Its senior management considered it a personal responsibility to shape the hedge fund's cyber security culture from the top-down to be a culture in which portfolio managers and analysts would not be fearful of regulatory backlash or penalization for reporting an error. This "see something, say something" culture had further permeated through David's team because it was enabled by David himself.

That's why Kim felt assured in reaching out to the system development team and inquiring about its decision-making process for the EQ stock recommendation. The development team had audited every aspect of the system and managed to efficiently review each system component. They found Kim's apprehension to be justified. What was even more surprising was that the AI had failed to detect the error, and the system had proceeded to produce an output with a high level of confidence. The CCO eventually remarked:

> "Oh my god, we realized that we had actually just missed a bullet… a $20 million bullet. When Kim reported the issue and compliance got involved, we were surprised to learn that every aspect of the system had been compliant

[12] Case Study 4: The $440 Million Software Error at Knight Capital. Henrico Dolfing. Link: https://www.henricodolfing.com/2019/06/project-failure-case-study-knight-capital.html
[13] Women in AI: IBM's Lisa Amini Takes On AI Security and Reasoning. Maribel Lopez. *Forbes*. Link: https://www.forbes.com/sites/maribellopez/2019/10/03/women-in-ai-ibms-lisa-amini-takes-on-ai-security-and-reasoning/#d97228921b24

> with firm policy and the regulations we had reviewed before starting production."

The CLO agreed:

> "After reviewing the applicable laws, we also realized that none of them had anything to do with artificial intelligence. This was all a very new legal area, and there was simply no right approach."

Alisha the Chief Economist commented on what the impact would have been had the system's trade recommendation been trusted by David's team:

> "The EQ trade that would have been conducted using a machine learning algorithm in the system was really stopped in time. I think this also taught the other research analysts to not entrust blind faith in the system by treating it like a black box whose output they could have trusted without human intuition, even though the system had passed the beta stage and was soon going to be deployed across multiple desks. It goes without saying – Kim's intuition was the reason the breach was caught; she stopped the trade when the system had failed to do so. Had the EQ trades been submitted, E-Fortress would have been in violation of NY law, and we could have even leaked out a trading strategy for our high net-worth clients like Jack Torrens."

It turned out that one of the classification models used by the system had been altered. The tampered model was responsible for classifying whether an equity was likely to increase in price above a threshold, decrease below a threshold, or remain between the thresholds. The team hypothesized that the cause of the tampering was a targeted misclassification attack, in which some external attacker or group of attackers had anticipated system behavior and misclassified a specific label. They hypothesized that attack had perturbed the system until the model misclassified what would have been a "sell" classification for EQ to a "buy" classification. The development team was surprised that they had failed to back test for such an attack, and arrived at the conclusion that it was probably because the system was pretty much immediately put to use in David's team as part of the testing phase after a quite rapid pre-development and development cycle.

One of the developers commented:

> "Going forward, one valuable takeaway from this incident is to improve production monitoring. We had automated processes in place but having a

> human eye on outputs such as transaction volume recommendations would be helpful[14]."

Aftermath and Next Steps

E-Fortress brought in a team of forensics experts to further uncover what had happened. Further investigation discovered that a renegade group of hackers in Russia had been attempting cyber-attacks against the AI system at E-Fortress Capital. These hackers were led by an insidious former insider who had been ousted from the fund's founding team due to ethical concerns. The insider was motivated by a personal vendetta against E-Fortress. As hypothesized by the system's developers, the hackers had broken in and changed the system's classification algorithm. They had wanted to use E-Fortress's trading strategy in a way that would benefit them by mucking up oil futures, and David's desk had been a perfect target since it covered the energy space. Furthermore, no one realized that the hackers' leader had remembered one of the authentication passwords which had been unchanged from his time at the firm. Authorities were now following up on this former insider because he had engaged in illegally conspiring against E-Fortress.

Upon uncovering the motivation behind the incident, Jim provided further advice to Alisha and Sheryl on next steps:

> "Security training will become increasingly important as the AI system is adopted across all teams, not just David's. Through publicity campaigns and awareness, you can motivate people to both constantly think about cybersecurity and to become more diligent. Going forward, I suggest that you run anti-cyber-attack campaigns for your research analysts and PMs. The more you can keep the issue top of mind, the more you are likely to have your people thinking about cybersecurity as they transition to and continue using AI. For example, you can send employees fake erroneous outputs and monitor whether they try to submit those erroneous outputs as actual trade requests. When this happens, you can the employee an email letting them know it was an anti-cyber-attack campaign and to be more vigilant about the system's outputs next time. You can do so repeatedly and determine if the percentage of analysts and PMs that incorporate these fake recommendations decreases. Ensure that you do this for people at all levels of the company, including executives. You also can't forget about the basics: unchanged passwords were one of the attack surfaces used by the hackers. Teams adopting AI cannot forget about basic cybersecurity precautions."

[14] Software Testing Lessons Learned From Knight Capital Fiasco. Matthew Heusser. *CIO*. Link: https://www.cio.com/article/2393212/software-testing-lessons-learned-from-knight-capital-fiasco.html

E-Fortress's senior leadership was very determined to ensure that no such breach occurred in the future. The incident highlighted that even the newest of regulations lacked AI-based system development concerns. It also highlighted the role of human stakeholders when designing and using the system. Going forward, E-Fortress's leadership decided to evaluate its AI system design model and devise management response plans for analysts and PMs who used AI-based systems.

The biggest realization for the team was that once there had been a hack and the compromised AI system began outputting results that were trusted by its users, it was difficult to tell the difference between whether the system had been tampered with and whether the system was simply outputting a trustworthy yet counterintuitive result based on its self-learning. Leadership realized that E-Fortress was essentially betting itself on this emerging technology. If there were ethical issues, and the team had already devised examples of ethical issues that existed, then E-Fortress was effectively risking losing its trading license if its AI system was ever comprised.

Exhibit 1: Glossary

Acronym	*Term*	*Description*
AI	Artificial intelligence	AI refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The ideal characteristic of artificial intelligence is its ability to rationalize and take actions that have the best chance of achieving a specific goal[15]. E-Fortress implements an AI system, the goal of which is to provide investment recommendations to its human users.
ATS	Autonomous trading system	Also known as an algorithmic trading system, an ATS allows traders to establish specific rules for both trade entries and exits that, once programmed, can be automatically executed via a computer[16]. Note that the AI system developed by E-Fortress used by David's team is not an ATS, since David is a discretionary investor, but ATS systems are known to employ AI in decision making.
GPT	General purpose technology	GPT is a term coined to describe a new method of producing and inventing that is important enough to have a protracted aggregate impact[17] on the entire economy. E-Fortress's senior leadership classify the AI system to be an example of a GPT in their collectively framework of considerations.
ML	Machine learning	ML is a sub-field of AI that employs algorithms to identify patterns and relationships in data that allow the machine to make predictions about data it has not seen before[18]. E-Fortress's AI system uses algorithms with underlying ML methods to make predictions about a company's stock price and output an investment recommendation about that stock based on these predictions.
PII	Personal identifiable information	PII is defined as any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means[19]. As a trusted hedge fund, E-Fortress has PII on its clients, and this PII is a potential vulnerability for hackers with malicious intent to exploit.

[15] Artificial Intelligence (AI). *Investopedia*. Link: https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp
[16] Automated Trading Systems: The Pros and Cons. *Investopedia*. Link: https://www.investopedia.com/articles/trading/11/automated-trading-systems.asp
[17] General Purpose Technologies. Boyan Jovanovic, Peter L. Rousseau. *New York University*. Link: https://www.nyu.edu/econ/user/jovanovi/JovRousseauGPT.pdf
[18] Artificial Intelligence: A Primer. Sam Kwok. *Garage Technology Ventures*. Link: https://www.garage.com/artificial-intelligence/
[19] Guidance on the Protection of Personal Identifiable Information. *U.S. Department of Labor*. Link: https://www.dol.gov/general/ppii

Exhibit 2: E-Fortress Capital Management Structure

Name	*Position*	*Responsibility*
Alisha Singh	Chief Economist	Responsible for managing an organization's investment portfolios, the Chief Economist oversees a team which manages and monitors investment activity, maintains investor relations, and develops short-term and long-term investment policies at the firm.
Sheryl Smith	Chief Technology Officer	Responsible for managing the technological requirements of the firm.
Karen Aaronson	Chief Legal Officer	Responsible for the legal affairs of the entire firm, the CLO helps the firm minimize its legal risks by advising the company's other officers and board members on any major legal and regulatory issues the firm confronts.
Makeda Bankole	Chief Compliance Officer	Responsible for policy and procedure management (defining, communicating, training, and attesting to corporate policies and procedures), compliance monitoring (evaluating and measuring the state of compliance across the organization), and managing investigations into any wrongdoing in violation of regulatory or legal requirements.
David Ambrose	Portfolio Manager	Each portfolio manager focuses on covering a specific industry (e.g. retail, media technology, biotechnology, etc.) and is responsible for the investment research analysts that are a part of the PM's team (a PM's "desk"). PMs manage their own PnL (profit and loss).
Kim Li	Investment Research Analyst	Investment RAs inform the investment strategy of their portfolio manager by conducting due diligence, investment research, and providing recommendations to their PM.
Jim Donoghue	Cyber Security Consultant	Cyber security consultants play both the attacker and the defender in computer systems, networks, and software programs. They identify system strengths and weaknesses to prevent exploitation.

Exhibit 3: New York State Department of Financial Services 23 NYCRR 500

Section 500.00 Introduction. The New York State Department of Financial Services ("DFS") has been closely monitoring the ever growing threat posed to information and financial systems by nation-states, terrorist organizations and independent criminal actors. Recently, cybercriminals have sought to exploit technological vulnerabilities to gain access to sensitive electronic data.

Cybercriminals can cause significant financial losses for DFS regulated entities as well as for New York consumers whose private information may be revealed and/or stolen for illicit purposes. The financial services industry is a significant target of cybersecurity threats. DFS appreciates that many firms have proactively increased their cybersecurity programs with great success.

Given the seriousness of the issue and the risk to all regulated entities, certain regulatory minimum standards are warranted, while not being overly prescriptive so that cybersecurity programs can match the relevant risks and keep pace with technological advances. Accordingly, this regulation is designed to promote the protection of customer information as well as the information technology systems of regulated entities.

This regulation requires each company to assess its specific risk profile and design a program that addresses its risks in a robust fashion. Senior management must take this issue seriously and be responsible for the organization's cybersecurity program and file an annual certification confirming compliance with these regulations. A regulated entity's cybersecurity program must ensure the safety and soundness of the institution and protect its customers.

It is critical for all regulated institutions that have not yet done so to move swiftly and urgently to adopt a cybersecurity program and for all regulated entities to be subject to minimum standards with respect to their programs. The number of cyber events has been steadily increasing and estimates of potential risk to our financial services industry are stark. Adoption of the program outlined in these regulations is a priority for New York State.

Source Link: https://www.dfs.ny.gov/docs/legal/regulations/adoptions/dfsrf500txt.pdf

Exhibit 4: Jim's Proposed Framework to Sheryl's Team

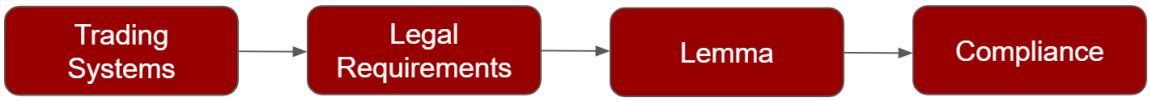
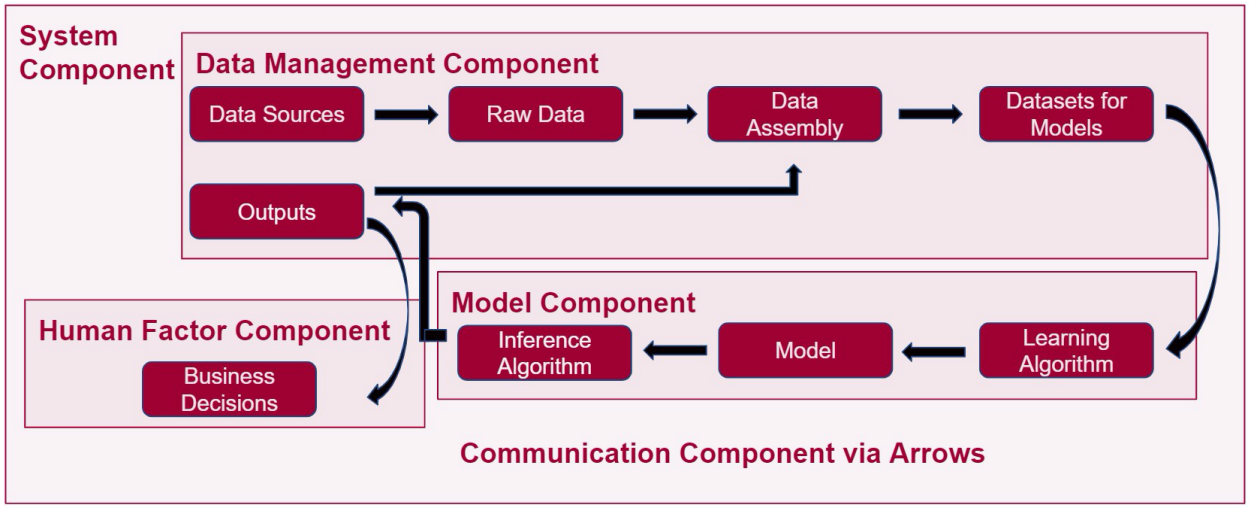


Exhibit 5: Senior Leadership's Collectively Proposed Framework

Type	*Considerations*
Legal	• NYC Financial Services Laws • GDPR fines • SHIELD Act
Technology	• Early AI technology • General purpose technology (GPT)
Technical	• Cybersecurity
Ethical	• Traceability • Explainability • Integrity • Privacy
Model Manageability	• Resources needed • Expertise • Data source considerations • Data quality considerations
Business Impact	• Higher transaction rate • Value-added differentiator • Attractive to customer base
Customer Perception	• Uncertainty of using an AI instead of an experienced PM • Trust in the results • Fear of computer decision making • Loss of control

Exhibit 6: General AI System Model



Inspired by Gary McGraw, Ph.D., Harold Figueroa, Ph.D., Victor Shepardson, Richie Bonett
Berryville Institute of Machine Learning (BIML)

Exhibit 7: Cybersecurity Risks in AI

Source: Sanjana Shukla, George Wrenn, Keri Pearlson, DBA
Cybersecurity at MIT Sloan (CAMS)

Exhibit 8: Attack Vectors and Definitions

Attack Vector	*Definition*
Data poisoning attack	The attacker injects carefully crafted data samples to contaminate the AI system's training data in a way that eventually impairs the system's normal functions.
Neural net reprogramming attack	Through a hack that reprograms the neural network, it is possible for an attacker to make the system's model "see" things that are not actually present.
Evasion attack	A type of training time attack, evasion attacks tamper with a system by adding an invisible (to humans) layer of data noise onto an image, leading a neural network (i.e. an underlying model in the system) to report back with high confidence that the image is something other than what it originally was.
Availability attack	Availability attacks are another kind of attack vector used against an AI system's models, specifically against natural language processing (NLP). An availability attack against an NLP model in a trading system, for example, affects a system that relies on sentiment analysis. Over an extended period of time, an attacker could publish and promote a series of adversarial social media messages designed to trick sentiment analysis classifiers used by system's learning algorithms. One or more high-profile trading algorithms trade incorrectly over the course of the attack, leading to losses for the parties involved.
Misclassification attack	Misclassification attacks can take the form of targeted and untargeted misclassification attacks. In an untargeted attack, the attacker seeks to degrade model performance by causing misclassification on any label, or final output, in the model. In a targeted attack, the attacker anticipates the behavior of the victim and seeks to misclassify a specific label.
Model stealing attack	These attacks can compromise PII used by the system along with other proprietary information including trading strategies or IP.
Voice command attack	An attack on voice commands seeks to compromise this voice recognition ability or circumvent it entirely. An attack against an AI system's voice recognition models can jeopardize the ability of voice recognition software to protect against approval of unauthorized trades by tucking into broadcasts garbled voice commands that can control smartphones and other communication equipment without the users ever noticing.
Social engineering attack	This is a type of attack that coaxes someone into giving up sensitive information and is one of the most common types of attackers used against vulnerable employees.

References

[1] Hedge Fund Definition, *Hedge Fund Marketing Association*. Link: https://www.hedgefundmarketing.org/hedge-fund-definition/

[2] Where Hedge Funds Get Their Capital, *Investopedia*. Link: https://www.investopedia.com/ask/answers/121614/where-does-hedge-fund-get-its-money.asp

[3] Discretionary vs. Systematic: Two Contrasting Hedge Fund Approaches, *Preqin*. Link: https://docs.preqin.com/newsletters/hf/Preqin-HFSL-Jun-14-Systematic-Discretionary-Funds.pdf

[4] Hedge Funds in High Net Worth Portfolios by Susan B. Weiner, CFA. *Advisor Perspectives.* Link: https://www.advisorperspectives.com/pdfs/newsltr08-2-2-1.pdf

[5] Smart money: AI transitions from fad to future of institutional investing. PwC. Link: https://www.pwc.com/us/en/industries/financial-services/library/artificial-intelligence-investing.html#:~:text=Some%20firms%20are%20using%20AI,efficiency%20for%20the%20investment%20industry.

[6] Fundamental Analysis. *Investopedia.* Link: https://www.investopedia.com/terms/f/fundamentalanalysis.asp

[7] Designing Automated Trading Systems for Commodity Trading – Theoretical Aspects. Petr Tucnik. *ResearchGate.* Link: https://www.researchgate.net/publication/271642533_Designing_Automated_Trading_Systems_for_Commodity_Trading_-_Theoretical_Aspects

[8] New York State Department of Financial Services 23 NYCRR 500: Cybersecurity Requirements for Financial Services Companies. Link: https://www.dfs.ny.gov/docs/legal/regulations/adoptions/dfsrf500txt.pdf

[9] SHIELD Act. Link: https://legislation.nysenate.gov/pdf/bills/2019/S5575B

[10] What Machine Learning Will Mean for Asset Managers. Robert C. Pozen, Jonathan Ruane. *Harvard Business Review.* Link: https://hbr.org/2019/12/what-machine-learning-will-mean-for-asset-managers

[11] What is an IT Security Consultant? *CareerExplorer*. Link: https://www.careerexplorer.com/careers/it-security-consultant/

[12] Case Study 4: The $440 Million Software Error at Knight Capital. Henrico Dolfing. Link: https://www.henricodolfing.com/2019/06/project-failure-case-study-knight-capital.html

[13] Women in AI: IBM's Lisa Amini Takes On AI Security and Reasoning. Maribel Lopez. *Forbes*. Link: https://www.forbes.com/sites/maribellopez/2019/10/03/women-in-ai-ibms-lisa-amini-takes-on-ai-security-and-reasoning/#d97228921b24

[14] Software Testing Lessons Learned From Knight Capital Fiasco. Matthew Heusser. *CIO*. Link: https://www.cio.com/article/2393212/software-testing-lessons-learned-from-knight-capital-fiasco.html

[15] Artificial Intelligence (AI). Investopedia. Link: https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp

[16] Automated Trading Systems: The Pros and Cons. *Investopedia*. Link: https://www.investopedia.com/articles/trading/11/automated-trading-systems.asp

[17] General Purpose Technologies. Boyan Jovanovic, Peter L. Rousseau. *New York University*. Link: https://www.nyu.edu/econ/user/jovanovi/JovRousseauGPT.pdf

[18] Artificial Intelligence: A Primer. Sam Kwok. *Garage Technology Ventures*. Link: https://www.garage.com/artificial-intelligence/

[19] Guidance on the Protection of Personal Identifiable Information. *U.S. Department of Labor*. Link: https://www.dol.gov/general/ppii