FireEye: AI + Human for Cybersecurity

CIS 8080

October 7, 2022

Kaustuv Dutta, Zaki Khan, Swati M. Radia, Jada N. Simpson, Jeremiah Warren

Table of Contents

[Table of Contents 2](#_Toc495731798)

[Executive Summary 3](#_Toc654757298)

[Case Synopsis 4](#_Toc472729562)

[FireEye Considerations 7](#_Toc149232749)

[Risk Analysis 9](#_Toc1004757520)

[Solution to the Core Problem 10](#_Toc1300745214)

[What Has Changed between then and 2022? Does it Impact your solution? 12](#_Toc569690882)

[References 15](#_Toc1553880018)

[Exhibits and Appendices 18](#_Toc2068288917)

Executive Summary

In 2020, FireEye developed a new artificial intelligence tool called Atomicity to assist the company’s elite threat intelligence team with threat attribution. In a collaborative effort using Agile methodologies, data scientists, cyber threat analysis experts, and IT personnel designed and helped train a machine learning model that compared unclassified threat clusters (UNCs) to other threat groups within FireEye’s database. The creation and adoption of the Atomicity tool represented a strategic shift in how FireEye paired the knowledge of human analysts with the pattern identification capabilities of artificial intelligence to accomplish complex tasks like threat attribution.

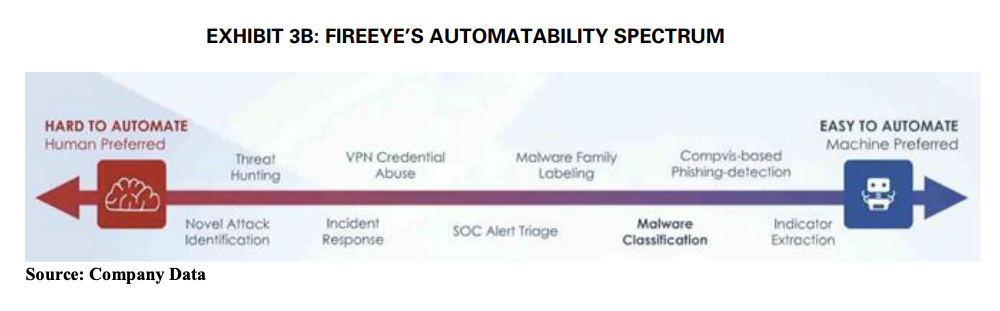
This report highlights key aspects of the development of the Atomicity tool, as well as explores FireEye's rationale for the creation of the tool. Additionally, the report will emphasize the problems facing the organization before the tool’s development, provide a detailed risk analysis of FireEye’s Human+AI approach to threat intelligence, and offers an assessment of FireEye’s Human+AI approach. It concludes with a reflection on the changes experienced by FireEye, the cybersecurity industry since 2020, and the best approach for companies to use for cybersecurity.

Case Synopsis

The FireEye case is a great example of the Human+AI approach done right. CTO, Steve Ledzian knew his company already had industry human experts on his team but wondered if integrating the help of AI tools was the correct approach for FireEye. This case describes how human ingenuity and teamwork paired with AI/ML allows FireEye to continue to be successful and be on the cutting edge of cybersecurity innovation.

FireEye is a public cybersecurity company, was established in 2004, and has a large portfolio of companies, many of which are on the Forbes international 100. They are by no means strangers in the world of cybersecurity. FireEye “provides its clients with a single platform that blends innovative security technology with threat intelligence” (Miller, Bhattacharya, 2021). FireEye won a competition sponsored by the US Navy in 2020 in which their AI-based product, MalwareGuard, was declared the winner. The team at FireEye is diverse in nature and has some impressive metrics which include, “deploying more than 17 million virtual machine sensors globally and blocking between 50,000 to 70,000 confirmed malicious events per hour” (Miller et al., 2021). The company has also constantly monitored advanced persistent threat (APT) groups, financial (FIN) threat groups as well as tracking 2,000+ uncharacterized threat groups or clusters. (Miller et al., 2021)

AI/ML solutions are implemented within many of FireEye’s solutions, and when it comes to deciding whether to lean on the human element or machine element, FireEye utilizes what they call their “automatability spectrum,” which is shown in the below figure.



(Miller et al., 2021)

“To guide its internal thinking on identifying scenarios in which a machine or a human expert would be the most effective approach to solve cybersecurity challenges, the firm has conceptualized an ‘automatability spectrum’ which took multiple factors into consideration to determine the degree of automatability of a task” (Miller et al., 2021). Based on how right or left on the spectrum the team gets would identify if the respective task should be given to a human expert or automated. A huge aspect of this approach allows AI to handle tedious automatable tasks so human experts can focus on larger/impactful tasks.

Steven Stone described the work of his team, Adversarial Pursuit (AP), as “Hunting Big Evil” (Miller et al., 2021). “In 2018, the AP team had deployed an ML tool called Atomicity to support its threat attribution work. The development of the Atomicity tool had been triggered by an operational need – processing large amounts of data to sift through information and decipher meaning from data” (Miller et al., 2021). FireEye had a plethora of data available from previous attacks/threat techniques and wanted to find automated ways to leverage their data—this is where ML techniques were able to come in to help assist the team. The team was able to create an ML solution to help identify/create threat clusters which in turn helped offload repetitive/tedious tasks. This, however, did not come without its challenges.

The team had to ensure that data inputted into the ML model was in the “right” state as well as “cleaned & sanitized.” This would require some human intervention as well as planning/coordination. The team had to be on the same page and “a team was formed consisting of experts from each area. Agile methodologies were implemented to develop the application as a collaborative effort” (Miller et al., 2021). With a project as large as this, the whole team had to be operating cross-functionally on the same wavelength, so Agile was implemented as a control to ensure constant iteration and feedback. This resulted in great success, as Stone’s team had success with working with the Atomicity tool and were even able to utilize it to identify APT33 attacks. “APT33 was an identified group of malware attackers with suspected attribution to Iran, which targeted the aerospace and energy industries” (Miller et al., 2021).

AI/ML has its capabilities as well as limitations. Due to these limitations, there will be a continued need for human touch. FireEye demonstrated efficient ways to integrate AI/ML and ensured all stakeholders were involved from the start using an Agile approach. FireEye understands that AI can be hugely beneficial for cybersecurity protection and future threats. “The FireEye application of AI to support the AP team was an example of a symbiotic process of human knowledge initializing the ML platform solution verification, and then humans learning and sharpening their judgment based on the ML model insights” (Miller et al., 2021). There is a constant need to improve in the realm of cybersecurity, hackers are “skilling up,” so it’s important for cybersecurity companies to stay ahead of the curve. FireEye sets the standard with implementing AI/ML + Human approach, and in the future, many other companies will follow suit if they have not already.

“Roles may evolve but we believe the human element will always continue to be present” (Miller et al., 2021).

# FireEye Considerations

The threats to organizations are constantly changing. This is increasingly true as attackers are becoming more innovative, forcing organizations to do the same. One such way that hackers are using AI is to test the success of their malware. Through machine learning, hackers can train their models on the existing defense methods and how they can avoid it. Another method used by attackers is to saturate the AI models used by organizations with irrelevant data. Since the models are data dependent, attackers will feed the model with useless information in the hopes that they will be able to retrain what it sees as an attack, and thus bypass the defenses. An additional way that AI is used maliciously is by mapping the AI models used by organizations. This tactic has allowed hackers to again retrain the models and be able to gain an advantage through evasion of detection from companies like FireEye. With the advancements in the threat environment, it is important for FireEye to explore all possible options in cybersecurity.

Though FireEye has no true issues outlined in the case, there are constant threats to them and the organizations that hire them. This coupled with the desire to be the best pushed them to develop the Atomicity tool, enabling FireEye to process data much faster and much more efficiently. The focus of this tool was to collect data that could be used to assess the similarity of the attacker's patterns, malware information, and general techniques of the attackers. AI/ML was used by the company to alleviate human analysts from the more mundane tasks required for identifying Uncharacterized Threat Clusters (UNCs). AI/ML has numerous capabilities and functionalities for threat attribution. Namely, artificial intelligence can be used to analyze large amounts of data that would be time consuming for an individual. Through trained models, AI can be used to consistently monitor threats, observe patterns and help overall with the classification of threat clusters. More specifically, models could be trained to detect anomalies, identify and even remediate when certain attacks more common attacks occurred, and in some cases predict. Developing a tool such as this allowed FireEye to be extremely marketable to its clientele.

Though AI/ML allowed FireEye to focus its efforts on more difficult tasks, AI is not without limitations. A few of the limitations include the inability to make decisions in cases of anomalies. This is to say that in the case of novel threats or newly emerged threat clusters, the models will not have enough data to adequately categorize the attack type or perform any remedial actions. For instances such as these, it is obvious that human capability and intervention is necessary. Additionally, there is the need to constantly train the model, and ensure that the integrity of the data being fed into the model stays intact. An additional limitation of the Atomicity that we identified is the inability to prioritize incident response. Though the model could be trained to rank the severity and impact of an attack. The issue of the need to be trained again occurs. In the event of a novel threat, the model is essentially useless beyond the identification stage.

The approach that FireEye took to threat attribution allowed them to use human and AI capability in a complimentary manner, analysts doing what they are best suited for and the model vice versa. The team at FireEye aimed to innovate while managing costs, as constant adaptations to the model could prove expensive. Through an iterative process and human expertise, the model was able to be consistently trained. Though it seems that FireEye had developed a model that performed as intended, innovations in machine and deep learning illude to a fully automated cybersecurity future. Many begin to fear that this could mean an end to cybersecurity roles, however, the roles will likely change as the capabilities of technology change. Humans are needed to monitor alerts provided by AI and determine genuine attacks, patch known or detected errors and vulnerabilities, and understand implicit and explicit concepts.

FireEye works with large amounts of sensitive information. Thus, it is important for them to identify the risks associated with any changes in operational processes at every step of implementation.

# Risk Analysis

Based on the information provided in the case study we performed a Risk Analysis. Considering the Human + AI approach we outlined the assets, threats, vulnerabilities, and controls (See Exhibit 1). Our objective was to determine the various threats and vulnerabilities that this approach could have. Our team identified four threats that could affect the Human + AI approach (See Exhibit 2). In building the Atomicity ML model, data that was stored in the centralized database and the remote systems that sent telemetry data was very important. So, anything that affected data would result in a threat that’s characterized as high risk. We identified that any threat that would affect integrity and availability of data would be of real concern to the organization.

We determined that the controls that FireEye has in place are good. Specifically, the case mentions how the company had good infrastructure and data management in place. Thus, we assessed the likelihood of any of these threats happening to be medium. However, if either of these threats occurs then the business impact would be significant. If the ML model is trained using the data that is modified or infiltrated, then the model would not provide desired results and would be rendered useless for the intended purpose. The availability of data is also very crucial as that helps in training the model to be better. FireEye had a vast amount of data at their disposal, but it was important for them to appropriately categorize the data. Improper categorization of data would again result in skewed results from the model.

The other threat we thought was going be of high impact was if there was a Novel attack to happen. It’s clear from the description of building the Atomicity ML model that it gave results iteratively. In case of Novel attacks, there is not enough data/ knowledge around it. Hence, it would be difficult to train the model to recognize these threats and find similarities between UNC’s (uncharacterized threat clusters).

Lastly, we cannot ignore the fact that there is always a threat from the humans involved. It could be either erroneous categorization of UNC’s or the fact that an employee is involved with an adversary. Both will result in a high risk to the AI approach. Based on this analysis we designed our qualitative and quantitative risk analysis (See Exhibit 2 and 3).

The Adversarial Pursuit team had good control measures in place to handle the issues related to the AI approach. They used agile methodologies to work on the ML tool. Involved people were from various teams across the company and worked with the goal of test-learn-improve to train the model to get better results. These methods helped them to build the model rather quickly and any issues would get highlighted and fixed much faster.

To mitigate the occurrence of the threats, FireEye should adopt stronger authorization and authentication techniques. To avoid the risk of human error good access control should be enforced. And having better security with regards to the data centers and having multiple backups would help.

Solution to the Core Problem

For the foreseeable future, there will be a continued use of both human and AI capability, not only in cybersecurity, but across many industries. There is an increase in investment efforts to implement AI across all industries. Due to the limitations of AI and ML, human analysts must continue to play a critical role. We believe that FireEye comprehended these limitations well and crafted a company strategy that reflects this understanding.

The rapid growth of new technologies, especially IoT, provides attackers with more attack vectors. Hackers are now using AI tools to find loopholes and circumvent defenses that companies have in place. FireEye and other cybersecurity companies must continue to enhance their capabilities to protect themselves from the evolving threat landscape, and we believe the Human+AI approach provides FireEye with the best possible strategic foundation.

What Has Changed between then and 2022? Does it Impact your solution?

Since FireEye adopted a human and AI approach to threat intelligence in 2020, there have been tremendous changes in the way FireEye, the cybersecurity industry, and the world utilize artificial intelligence. FireEye, now known as Trellix after being acquired by STG (Symphony Technology Group) and combined with McAfee Enterprise, continues to invest in artificial intelligence solutions to enhance their capabilities and better protect their clients. New machine learning-based, threat intelligence products like Trellix Insights anticipates threats and designs remedies before their clients are attacked. On the other end of the automatability spectrum, Trellix continues to develop and refine their security analytics tools for automated threat detection. Last month, Trellix created the Trellix Advanced Research Center to centralize their threat intelligence capabilities. This new center, coupled with the Extended Detection Response (XDR) platform with integrated AI tools, underlines the importance Trellix places on utilizing their human analysts and AI tools to better protect their clients.

Trellix is not the only cybersecurity firm investing in artificial intelligence. Competitors like Crowdstrike and Palo Alto Networks now boast strong AI-assisted security offerings. This follows the industry trend, as experts project the market for AI within cybersecurity to reach 32.7 billion USD by 2026. Additionally, clients now expect AI solutions, and cybersecurity companies are innovating to meet this demand.

Also driving this investment is the adoption of AI tools by malicious actors to better target and attack their victims. While artificial intelligence provides invaluable analytics tools for businesses, it also allows malicious attackers to discover patterns and design stronger attacks. Attackers utilizing machine learning to craft spear-phishing attacks and to create deep fakes is particularly concerning. These types of attacks require a level of sophistication and appear more convincing than normal attacks, which raises the probability of success and therefore the impact of these attacks. While experts think attackers are only beginning to incorporate AI tools in their attacks, it will be imperative for cybersecurity firms to get out head of these threats and design tools to mitigate the effectiveness of AI-assisted attacks.

Cybersecurity is far from the only industry investing in artificial intelligence. A recent report claimed that 91% of leading businesses have ongoing investments in artificial intelligence. Artificial intelligence is spreading into all sectors of the economy, including unexpected industries like art and medicine. Incorporating artificial intelligence into business practices is inevitable, and it will be up to organizations to decide how much they rely on artificial intelligence for core business functions. The power of artificial intelligence has grown exponentially through advanced methodologies such as deep learning neural nets coupled with increased computing power. As this usage and effectiveness of artificial intelligence grows, it will be important to monitor how reliant organizations are on these technologies for decision-making. It will be critical to evaluate the fairness, accountability, and transparency of machine learning in this decision-making process.

Recent world events offer a disturbing insight into the future of warfare. With cyber warfare increasingly being used to augment conventional warfare, it stands to reason that artificial intelligence will become another weapon of war. Autonomous weapons like drones are only the beginning of this shift in warfare. Going forward, it will be important to follow how artificial intelligence is incorporated into warfare and what form of regulation the global community puts on their usage.

After surveying the cybersecurity industry, general business trends, and human history since 2020, FireEye’s AI approach appears to be remarkably prescient. FireEye’s approach understood both the limits and capabilities of artificial intelligence and the importance of human analysts in the decision-making process of threat attribution. Understanding the limits of AI, FireEye’s approach ensured that they did not overly rely on their AI tool. However, FireEye did understand how the Atomicity tool could be used to improve the capabilities of their human analysts. This approach highlights the usefulness of artificial intelligence in the business process, especially when companies leverage the knowledge of human experts. As a wise professor once said, “All models are wrong some are useful.”

Conclusion

The FireEye case study shows us why the Human+AI approach is here to stay along with why other companies will begin/continue to adopt this approach. Hackers are already utilizing similar measures and attacks are only going to grow more sophisticated in nature. It is essential for cyber defense companies to ensure they are continually skilling up. FireEye was able to get internal alignment on integrating AI/ML and did so successfully using the Agile approach. They were able to maximize efficiency but more importantly ensured the team was on the same page while doing so. There is no doubt AI/ML investment will only grow and the Human+AI approach will become even more prevalent within both the cyber industry and within all industries.

Is the future Human+AI? It appears so.

References

<https://www.crn.com/news/security/stg-closes-1-2b-fireeye-buy-merges-it-with-mcafee-enterprise>

<https://www.trellix.com/en-us/assets/solution-briefs/Trellix-Security-Operations-Solution-Brief.pdf>

<https://www.trellix.com/en-us/products/security-analytics-products.html>

<https://www.businesswire.com/news/home/52921713/en>

<https://www.trellix.com/en-us/platform/xdr.html>

<https://cybermagazine.com/top10/10-cybersecurity-companies-powered-ai>

<https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-security-market-220634996.html>

<https://www.csoonline.com/article/3250144/6-ways-hackers-will-use-machine-learning-to-launch-attacks.html>

<https://connect.comptia.org/blog/artificial-intelligence-statistics-facts>

<https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html>

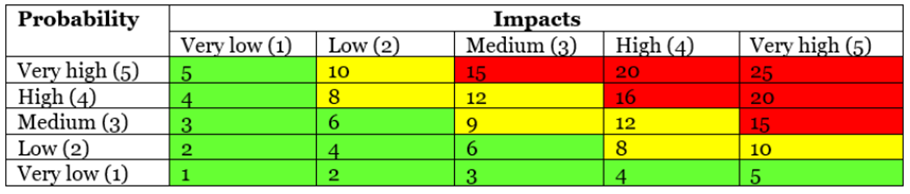
<https://news.harvard.edu/gazette/story/2020/11/risks-and-benefits-of-an-ai-revolution-in-medicine/>

<https://www.theatlantic.com/technology/archive/2021/09/i-weapons-are-third-revolution-warfare/620013/>

<https://www.fatml.org/><https://www.mpug.com/understanding-qualitative-risk-analysis/> [ENISA](http://chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https:/gastate.view.usg.edu/content/enforced2/2635688-CO.090.CIS.CIS8080.XLS.CG.20232/ENISA_RM-Deliverable2-Final-Version-v1.0-2006-03-30.pdf?_&d2lSessionVal=yEEF63x9iEVHONRGjMrte8Aqm&ou=2635688)

<https://hbsp.harvard.edu/>

Exhibits and Appendices

Reference for P.I Scoring

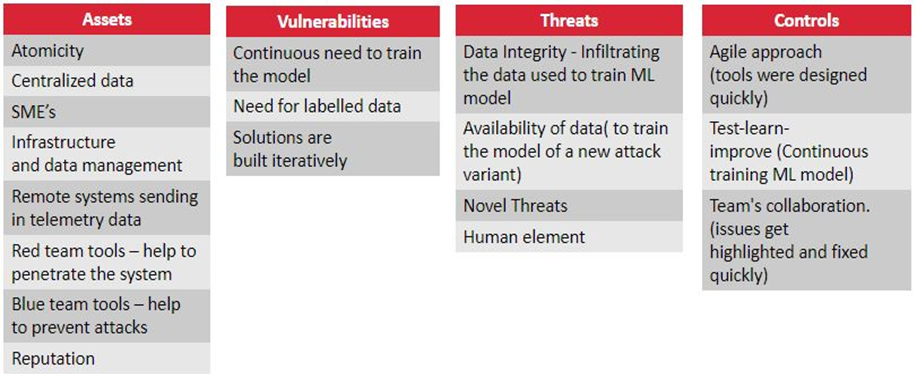


Exhibit 1

**Qualitative Risk Analysis**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Threats** | **Probability** | **Impact** | **P.I Score** | **Risk Ranking** | **Decision** |
| **Data Integrity** | 3 | 5 | 15 | 2 | Risk Mitigation |
| **Data Availability** | 3 | 5 | 15 | 3 | Risk Mitigation |
| **Novel Threats** | 2 | 5 | 10 | 1 | Risk Mitigation/  Risk Acceptance |
| **Human** | 2 | 5 | 10 | 4 | Risk Acceptance/  Risk Mitigation |

Exhibit 2

**Quantitative Risk Analysis**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Threats** | **Probability** | **Impact** | **P.I Score** | **Risk Ranking** | **Decision** |
| **Data Integrity** | 3 | 5 | 15 | 2 | Risk Mitigation |
| **Data Availability** | 3 | 5 | 15 | 3 | Risk Mitigation |
| **Novel Threats** | 2 | 5 | 10 | 1 | Risk Mitigation/ Risk Acceptance |
| **Human** | 2 | 5 | 10 | 4 | Risk Acceptance/ Risk Mitigation |

Exhibit 3