SCHOLARLY DEBUT
by Scholarly Review

Fall 2025

AI-Driven Cybersecurity: Solving the Employee Security Risk

By Nikhil D. Bhargava

AUTHOR BIO

Nikhil Bhargava is a rising senior at the Potomac School in Mclean, Virginia. His current academic interests include business and finance, computer science, and other technologies, and on a broader scale, fields that involve some STEM. Building on his interest in STEM, he loves activities that involve strategy or problem solving, like chess and jigsaw puzzles. He plays squash and tennis for his school teams, and swims for a local swim team during the summer. He also enjoys traveling, going on several trips per year, one of his favorites being Croatia. Nikhil often volunteers and participates in various forms of community service, like coaching, farmers markets, and Habitat for Humanity. Nik overall has diverse interests and passions.

ABSTRACT

As technology advancements escalate and AI and other technological applications grow, the need for strong cybersecurity has become critical, especially as cyber threats become increasingly sophisticated. The cost of cyberattacks has become significant, with companies facing billions in losses and compromising sensitive data, as evidenced by high-profile breaches like the recent Microsoft-Crowdstrike incident, which impacted critical infrastructure worldwide [1]. Despite the deployment of multi-factor authentication and training initiatives, a significant vulnerability persists within businesses: employees. Human errors, particularly in response to phishing and spam, are primary vectors for data breaches, underscoring the need for proactive security measures [2].

This research proposes an AI-driven system that proactively identifies employees exhibiting high-risk behavior patterns, thereby enabling targeted, proactive interventions. By analyzing data on factors such as frequency of spam interactions, behavioral indicators, experience, and training attendance, the AI application can assess and flag employees who may inadvertently compromise security. An application installed on company devices gathers and evaluates real-time data, which is then reviewed to take preventative action.

This approach emphasizes early detection and mitigation of human-related cyber risks, ultimately safeguarding business continuity and reducing the likelihood of large-scale cyber incidents. Through this system, companies can better understand employee-based vulnerabilities, enabling informed decision-making to enhance critical security in a growing cyber threat world.

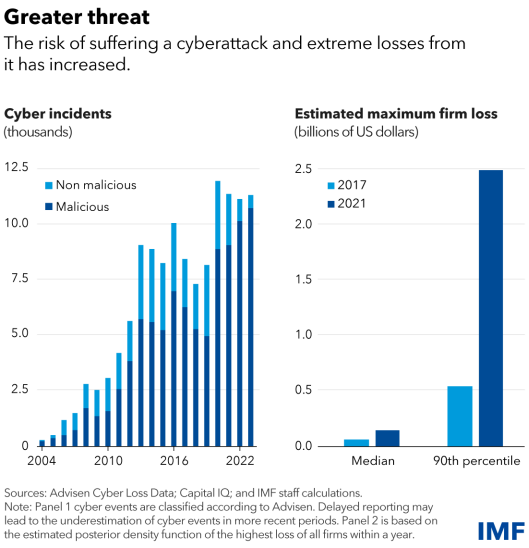
Keywords: *Cybersecurity, Cyberattack, AI-systems, Application, Employee, Spam, Security, Human errors*

SCHOLARLY DEBUT
by Scholarly Review
Fall 2025

BACKGROUND

The need for strong cybersecurity in businesses has been increasing rapidly, especially with the constant advancements in technology, and the necessity and desire to preserve and improve the economy. Companies have lost millions, even billions of dollars, sensitive data, and had other issues that caused worldwide chaos [2][3]. Still, it is only getting worse. According to the IMF (see below), cybersecurity attacks have nearly doubled since the pandemic [3].

Figure 1
Total cyber attacks over the years, and Estimated losses in billions in 2017 and 2021



IMF, Greater threat

Taking a look at the recent Microsoft CrowdStrike issue in particular, a cyber threat was able to impact way more than the security and strength of a business; the cyber threat grounded thousands of planes around the world, and put panic into the eyes of many emergency services locations [1]. Cyber attacks are becoming ever more concerning as these attacks have proven to have the potential to take down many different economies and scare big businesses (like Microsoft) that seemed to be well secured [4][5][6].

While there are many ways that companies are susceptible to cyber attacks, one ongoing vulnerability that all businesses face is the actions of their employees. Employees are often entrusted with important company components, including a company device and important information on aspects of the company. According to Tech.Co's *Impact of Technology on the Workplace* report, employees are the key weak point, with employee errors and phishing attacks being one of the top causes of data breaches in 2023 [7]. Given the increased interconnectivity and digitization, any weak point can lead to an infiltration and a significant attack. A leading cybersecurity firm, Astra, estimates that over three billion phishing emails are sent daily, and these emails cause 16% of data breaches, averaging around 5 million dollars per attack, and even at one point got close to 2 billion dollars [2]. Hackers have become increasingly sophisticated and persistent, often using various methods and finding ways to leverage familiarity, such as well-known brand names, to breach trust.

Companies recognize this vulnerability and have instituted various protection mechanisms to help, but unfortunately, they have not been enough. Multi-factor authentication has been a meaningful deterrent, and training and awareness programs have helped reduce the obvious behavioral mistakes. However, every 11 seconds, a scam email is sent, and over time, with increased sophistication. According to Astra, nearly 30% of phishing emails are still opened, leading to increased chances of downloading malicious links that contain ransomware or malware [2]. There are employees in companies who are

susceptible to accidentally revealing private company data or providing access to the system through spam links or calls, or some other misuse of a company device or email. Along with the billions of dollars lost to cyber threats, individuals have to worry about their personal business data and credentials, or even data on their health status in hospitals, being exposed.

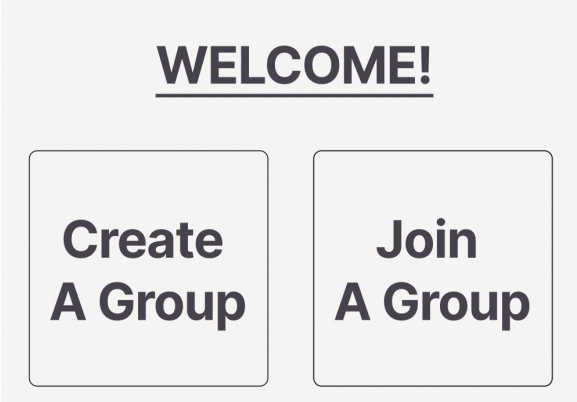
To solve this growing, critical problem, this proposed solution focused on helping businesses proactively identify the most vulnerable employees and their behaviors, and in turn, provide additional proactive efforts to help employees avoid doing the wrong thing. Every employee is human, and humans are prone to making mistakes, so it is critical that businesses find these employees ahead of time. To do this, the use of AI, and in particular machine learning and AI analytics, can detect patterns in collected data of who is the most likely to perform the wrong behavior (for example, clicking on the wrong email) before the cyber hacker gets access and anything is leaked. I identified indicator data that would be most associated with people who are most vulnerable: First, data on who receives the most spam calls and emails; Second, data on who is likely to click on ads and bad emails; Third, data on who uses their time to explore sites that are not work related; Fourth, data on who skips training; Fifth, data on who has performed the wrong behavior in the past. Then, by training a machine based on the data, the AI application will detect and identify any employees who are at higher risk of attack. By creating an application that can detect the issue before an employee hurts the company, the application will protect the business and lower the risks of major crashes.

METHOD

Initial Setup

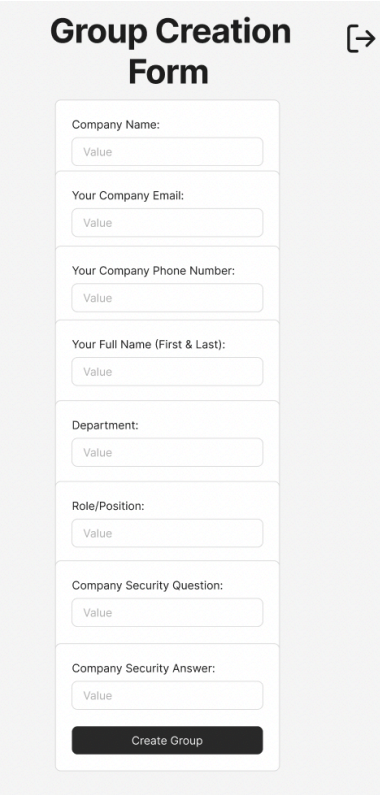
In order to solve this problem, an application is installed on every company device that tracks individual and group data on individual experience and computer-based activity regarding ads, spam calls, and spam emails. It then analyzes the data and reports that data to whoever is running the group as the manager. This application can be installed on each device by the company's IT department or can be installed manually through the device's application store. Each employee and manager will be greeted by a welcome page and the choice to create a group or join a group (Figure 2). The manager, or whoever is designated to manage the group, will start by creating a group, clicking on "create a group". The group manager will reach a login page that requires the following information: their company name, company email and/or company phone number (i.e. the number that the employee would receive business calls on, complaints, etc.), full name, department and position, and a security question to help verify the employees that sign in (Figure 3). After creating the group, any other important existing company or employee-related information that the manager keeps track of, including training attendance and whether or not an employee has misused a device in the past, can be entered into the algorithm. This input can be done in several ways, including using pre-existing spreadsheets or documents, or manually typing the data in. The manager can then inform the employees of a unique code of numbers and letters that will be used for employees to join the group. Each employee will input their personal information into a form similar to the manager's, with their name, department, email, device identification, group code, and a response to the security question. The IT department can help speed up this process if they have the required data.

Figure 2
Welcome page to create or join a group



What each manager and employee will see when opening up the application

Figure 3
Image of the form for managers to create a group



Data Gathering and Analysis

Once employees are in, the application will use various technologies to gather, verify, store, and analyze data that will ultimately detect threats in a certain department or from a specific employee and report it to the group manager and whoever the threats apply to. To do this, every day, the application will trace the origin of data and information to first verify whether or not an email or call is spam or fake, and how many ads appear on the device, and securely store the information. Then, the application will use a combination of AI and a teachable machine to detect important information and patterns, and will continuously improve the accuracy of its results based on the patterns found from the new daily information. Through machine learning, the AI will specifically be trained to notice patterns, including when a device scores above

average in the amount of ads and/or spam, with the average determined by running averages. The benchmark averages will include the average of the group, the average across relevant groups, and the average across the company. The application can also group other important, specific information, as seen below. Since company data, as seen below, is likely either not tracked or released to the public, Chat GPT 4.0 was used to create mock data, which represents a company that is keeping track of 30 employees, what the data may look like, and how it may be organized. Both examples divide the employee data by department and give data on employees in that department. Along with employee names in each department, example one presents the average daily ads, spam emails, and spam calls. While example two gives an employee ID and employee information on misused devices and meeting attendance, alongside the name and department division.

Example 1

Average Daily Ads, Spam emails/calls for HR

Human Resources Department

Full Name	Avg. Daily Ads	Avg. Daily Spam Emails	Avg. Daily Spam Calls
Nicole Harris	11	8	2
Brian Scott	10	7	2
Laura King	12	9	3
Steven Wright	11	8	2
Karen Young	13	10	3

Average Daily Ads, Spam emails/calls for Finance Dept.

Finance Department

Full Name	Avg. Daily Ads	Avg. Daily Spam Emails	Avg. Daily Spam Calls
Thomas Baker	9	6	1
Elizabeth Green	10	7	2
William Adams	8	5	1
Patricia Evans	11	8	2
James Collins	9	6	1
Mary Nelson	10	7	2

Average Daily Ads, Spam emails/calls for Sales Dept.

Sales Department

Full Name	Avg. Daily Ads	Avg. Daily Spam Emails	Avg. Daily Spam Calls
John Smith	12	8	3
Emily Johnson	15	10	4
Michael Brown	11	7	2
Sarah Davis	14	9	3
Robert Wilson	13	8	4
Jennifer Lee	16	11	5
David Martinez	12	9	3

Average Daily Ads, Spam emails/calls for IT Dept.

IT Department

Full Name	Avg. Daily Ads	Avg. Daily Spam Emails	Avg. Daily Spam Calls
Jessica Thompson	8	5	1
Andrew Garcia	7	4	1
Michelle Rodriguez	9	6	2
Ryan Martinez	8	5	1
Stephanie Lewis	10	7	2
Matthew Turner	7	4	1

Avg. daily Ads, Spam emails/calls for Marketing Dept.

Marketing Department

Full Name	Avg. Daily Ads	Avg. Daily Spam Emails	Avg. Daily Spam Calls
Lisa Anderson	18	12	2
Kevin Taylor	20	15	3
Rachel Moore	17	11	2
Daniel White	19	13	3
Amanda Clark	21	14	4
Christopher Hall	18	12	2

Example 2:

Employees attendance to cybersecurity meeting and history on misused devices per department

IT Department

Employee ID	Name	Attended Cybersecurity Meeting	Misused Devices
001	John Smith	Yes	No
006	Laura Taylor	Yes	No
011	Amanda Clark	Yes	No
016	Daniel Walker	Yes	Yes
021	Olivia Roberts	Yes	No
026	Aiden Murphy	Yes	No

Marketing Department

Employee ID	Name	Attended Cybersecurity Meeting	Misused Devices
002	Sarah Johnson	No	Yes
008	Linda Martinez	Yes	No
012	Brian Lopez	Yes	Yes
018	Christopher Hall	No	Yes
022	Ethan Green	Yes	No
028	Lucas Carter	Yes	Yes

Sales Department

Employee ID	Name	Attended Cybersecurity Meeting	Misused Devices
003	Michael Brown	Yes	No
007	James Anderson	No	Yes
013	Jessica Thompson	No	No
017	Rachel Wright	Yes	No
023	Sophia Adams	No	Yes
027	Isabella Ross	Yes	No

HR Department

Employee ID	Name	Attended Cybersecurity Meeting	Misused Devices
004	Emily Davis	Yes	No
009	Robert Moore	Yes	No
014	Kevin Harris	Yes	No
019	Ashley Young	Yes	No
024	Jackson Baker	Yes	No
029	Grace Edwards	Yes	No

Finance Department

Employee ID	Name	Attended Cybersecurity Meeting	Misused Devices
005	David Wilson	Yes	Yes
010	Patricia Lee	No	No
015	Megan King	Yes	No
020	Anthony Scott	Yes	Yes
025	Abigail Perez	No	No
030	Mason Hughes	No	No

Results and Scenarios

With these organized data sets, the group manager can view specific data about the employee or department. As discussed, the AI will go further and detect patterns or anomalies in the data, and can combine those anomalies with other information. One key use would be tying high-risk behaviors, such as an employee missing training sessions, to employee function to elevated risk levels for the device. To further understand how the AI will detect patterns and make its reports, mock scenarios were created and input into Chat GPT 4.0 to get an AI response, indicated by the quotations, that would not only indicate the likelihood of risk, but also recommended solutions. Three examples are provided below, with separate scenarios, a description of the mock AI (Chat GPT 4.0) response, and the actual AI response to the scenario.

Scenario 1: Employee X attends a small number of the training meetings, receives noticeably above average ads and spam every day, and has been caught misusing a device within the past couple of years. Also, Employee X is in Sales, which requires the employee to reply to emails and calls, leading them to often face more spam.

AI response: The AI will pick up the common theme of the unusual amount of ads and spam daily compared to other employees in the company and the department, then it will pull in other important information, like training, misuse of devices, and department data. The AI will then send a report to both the employee and the manager stating a message along the lines of: "Employee x has a higher likelihood to release data. Employee X may benefit from a combination of better training, stronger spam filtering, clear guidelines for device use, and better support for handling their high communication load in sales. Addressing these areas can help reduce distractions, improve performance, and minimize potential security risks. Additionally, ongoing conversations about expectations and performance could guide Employee X toward improving their work habits and better aligning with company goals."

Scenario 2: Employee Y sometimes attends training meetings, receives more spam emails than average, but just an average amount of ads and spam calls, and has never misused a device before. Employee Y is also new to the company and has not been able to build a lot of trust. The employee is young, and they work in the Finance department.

AI response: The AI will bring out the issues from the data, such as the increase in ads and not being connected fully to the rest of the company, and analyze them. The AI will report back to the manager and employee that while the employee has been attending training sessions, and has received average ads and calls, the employee can still be at high risk because of the high level of emails, low familiarity with company policy, and inexperience in the workforce. As a result,

the manager can spend more time with the employee, provide additional training, and set up a mentor for the employee to provide additional guidance.

"Employee Y appears to be an employee who is conscientious, but is still adapting to their new role and the company culture. Some key areas for improvement include more consistent training attendance, reducing the spam overload, and integrating more fully into the company's social and professional environment to build trust. With proper guidance, support, and a little patience, Employee Y can gain confidence and develop into a reliable and trusted member of the Finance team. By offering practical solutions to these challenges, you can help Employee Y become more productive, secure, and well-integrated into the organization."

Scenario 3: Employee Z has attended the training sessions diligently, receives just below average on both ads and spam, but is older and not familiar with the typical types of tricks hackers have used in recent years.

AI response: The AI will report that employee Z has a lower risk, but the employee needs to continue to be thoroughly informed and taught about the dangers and how they could easily detect the threats. Since cyber attacks have been continuously changing and finding new ways to attack companies, keeping a well-informed team will provide a safer environment.

"Employee Z is an employee who demonstrates a strong commitment to training and learning, which is excellent. However, they may face challenges in adapting to modern cybersecurity threats due to their unfamiliarity with recent hacking tactics. Offering targeted training on current cybersecurity risks, fostering mentorship opportunities, and providing additional resources to keep them updated on modern threats would be beneficial. The goal should be to build on their existing knowledge while helping them adapt to new risks and technologies, ultimately ensuring that Employee Z feels both confident and secure in their role. With continued support, they can continue to be a valuable asset to the company, contributing both their experience and growing understanding of current best practices."

CONCLUSION

While the final stages of this application's development are ongoing, the simplicity and effectiveness of the concept holds the potential for substantial impact on businesses worldwide. The process of creating a solution that is both user-friendly and robust against cyber threats presented unique challenges, but ultimately highlighted the critical role employees play in a company's cybersecurity strength. This research phase focused on understanding and evaluating employees' vulnerability to cyber threats, exploring how attackers exploit human error, and analyzing the potential impacts of these attacks on businesses. Human actions can be unpredictable, and mistakes happen; therefore, employing technologies like machine learning and other forms of AI offers an invaluable approach to identifying and mitigating these risks proactively.

As mentioned before, this development process is ongoing, but with the creation of how the application has demonstrated promising results in monitoring and identifying high-risk behaviors among employees, allowing managers to implement timely interventions. Certain aspects, such as refining the AI's accuracy in detecting patterns of risk, worked well, while other areas, like ensuring privacy and user compliance, required iterative improvements. Future development will focus on enhancing the adaptability of the AI to recognize emerging phishing tactics and continuously updating risk

metrics to remain aligned with evolving cyber threats.

Businesses could leverage these findings to bolster their cybersecurity strategies, applying insights and data from this application to target employee vulnerabilities directly and preemptively. As discussed, cybersecurity involves multiple layers, from technical defenses to human-based vulnerabilities. This application is designed to bridge that gap, complementing existing security measures with a focus on the human element.

Looking ahead, the future of cybersecurity is likely to become even more complex and dynamic. In the next year to five years, we may see businesses adopting increasingly sophisticated AI-driven tools to stay ahead of the evolving tactics of cybercriminals. Within a decade, technologies like machine learning could become integral components of all cybersecurity frameworks, with continuous advancements in detecting nuanced behavioral risks. In the longer term, say over a century, as technology continues to evolve and our dependence on digital infrastructure grows, cybersecurity will likely require a transformative approach, potentially integrating AI and machine learning into every facet of employee interactions.

In a landscape where cyber threats will only grow more sophisticated, businesses must adapt, leveraging advanced technologies to limit data leaks and financial losses. This application, with continuous refinement and adaptation to future cyber threats, exemplifies how companies can harness current technologies to protect their operations and ensure resilience against an ever-evolving cyber threat landscape.

REFERENCES

[1] Suliman, A., Rosenzweig-Ziff, D., Dau, E., & Paúl, M. L. (2024, July 19). What caused the Windows outages affecting flights, companies around the world. *Washington Post*. https://www.washingtonpost.com/world/2024/07/19/crowdstrike-microsoft-outage-cause-windows-explainer/

[2] Palate, N. J. (2024, June 5). *81 Phishing Attack Statistics 2024: The Ultimate Insight*. Astra. https://www.getastra.com/blog/security-audit/phishing-attack-statistics/#Conclusion

[3] Natalucci, F., Qureshi, M. S., & Suntheim, F. (2024, April 9). Rising Cyber Threats Pose Serious Concerns for Financial Stability. *IMF*. https://www.imf.org/en/Blogs/Articles/2024/04/09/rising-cyber-threats-pose-serious-concerns-for-financial-stability

[4] Center For Strategic And International Studies. (n.d.). *Significant Cyber Incidents*. CSIS. Retrieved February 11, 2025, from https://www.csis.org/programs/strategic-technologies-program/significant-cyber-incidents

[5] Kerner, S. M. (2025, January 15). *Treasury Department hacked: Explaining how it happened*. TechTarget. Retrieved February 11, 2025, from https://www.techtarget.com/whatis/feature/Treasury-Department-hacked-Explaining-how-it-happened

[6] Lydon, M. (2025). 11 Companies Who Have Recently Faced a Cyberattack. *Growbo*. https://www.growbo.com/recent-cyber-attacks-on-companies/

Drapkin, A. (2024, February 26). *8 Worrying Cybersecurity Statistics You Need to Know in 2024*.tech.co. Retrieved December 23, 2024, from

https://tech.co/news/cybersecurity-statistics-2024