# RESEARCH STATEMENT

Muhammad Lutfor Rahman (https://www.cs.ucr.edu/~mrahm011)

Cybersecurity evolves with the technological advancements throughout history. The more we integrate technology into our society, the higher the chance we will become a victim of cyber attacks. According to the White House, cyber attacks cost the U.S. economy between $57 billion and $109 billion in 2016 [5]. As per research and markets data, the global cybersecurity market is expected to reach USD 267.73 billion by 2024 [6]. While tremendous efforts have been made to secure the hardware and software stack of cyber systems, all these preventive technologies might be in vain if a user falls for a phishing attack. In fact, as attackers usually target the weakest link of the security chain, we have observed a surge of phishing and social engineering attacks in the past few years; many large corporations were penetrated through targeted/spear phishing attacks.

My research aims to understand and incorporate human factors, particularly neural insights for security-relevant tasks. We know the human brain is one of the mysterious things with more than 100 billion neurons and trillions of synapses. My research explores the unconventional approach to study the treacherous world of cybersecurity and to unfold some of its mysteries by dissecting the neural underpinnings. As the first step into this promising direction, I have conducted studies for understanding humans in security tasks using cognitive load and engagement [4], demonstrated potential security and privacy issues of Brain-Computer Interface (BCI) [3], designed more reliable access control mechanisms by keeping neural signal into the loop [9], developed new analysis techniques of neural data [7], and developed applications [10] for improving human training performance. According to Google Scholar, my work has received 50 citations. My work also has a significant impact on the community, as evidenced by coverage of more than 600 high-profile media outlets (e.g., Phys, Homelandsecuritynewswire, Neurosciencenews, MIT Technology Review, International Business Times, Daily Mail, Popular Mechanics, FOSSBYTES, ZDNet) in more than twenty languages worldwide. Due to the nature of my research field, most of my projects are collaborative.

## Understanding Phishing Attacks through the Lens of BCI

Phishing is the act of impersonating a trusted third party to steal users' sensitive and private information. The stolen information can cause direct financial loss or be used to penetrate large corporations, making it one of the most severe threats to cybersecurity. To understand why users fall into phishing attacks, we conducted a study [4] that measured and characterized users' neural processes in detecting phishing websites. We used neural signals and an eye-tracker to measure users' engagement and workload during the phishing detection task in a near-realistic environment. This study showed that brain areas related to critical decision making and visual search exhibit differences in activation when participants view real versus phishing websites. We also found that users may not heed to the key areas of the website and may exhibit some differences while processing real and fake websites. Moreover, we observed that users do not spend enough time analyzing key phishing indicators and often fail to detect these attacks, although they may be mentally engaged in the task and subconsciously processing real sites differently from fake sites.

Inspired by the observation from the aforementioned study, we utilized the differences in the activation levels as features to predict whether the participants were viewing a real website or a phishing website [7, 8]. In this study, we analyzed the underlying hidden patterns in neural data using tensor-based representations of electroencephalography (EEG) data related to phishing detection tasks. Traditional feature extraction techniques, such as power spectral density, autoregressive models, and Fast Fourier Transform, can only represent data either in spatial or temporal dimension; however, our tensor modeling utilized both spatial and temporal traits in the input data. We found the level of brain areas related to the users' decision making process with respect to the real and the fake websites based on the latent factors extracted using tensor. The machine learning classifiers showed that tensor-based neural features gave the accuracy of above 94% consistently across all classifiers which actually exceed humans' decision-making accuracy of 84.82%. This pattern of results indicates that neural signatures encoded in the brain are important to correctly classifying a website as real versus phishing. A more surprising result was that the pattern did not always align with the participants' final decisions and that ML-based classifiers can actually be more accurate than participants' decisions. This phenomenon seems to suggest that phishing websites will trigger "internal" warnings in the user's brain, but will be ignored or overridden by other signals.

Based on these prior results, we know that the human brain must be processing real and fake websites differently. So, in an ongoing study, we aim to dissect the role of this neural signal, which we refer to as *skepticism*, in a systematic, hypothesis-driven manner. Specifically, we are testing the hypothesis that *visual features of real versus phishing websites will trigger skepticism, which in turn, can be captured through electroencephalography (EEG)*. To this end, we want to conduct a systematic EEG study to elicit a distinct decodable signature of skepticism when users are looking at a phishing website. Unlike previous studies where we studied the differences in neural activity when participants viewed real versus phishing websites, as a whole, in this study, we explicitly control visual feature presentation such that each feature is encoded

sequentially. This novel task design allows characterization of each feature's contributions to the skepticism signal and to the decision itself. We would like to delineate the signature of neural activities associated with different visual components (e.g., security indicator, address bar, page content). We also would like to learn how the neural activities evolve as users process these visual components. The results of this study will inform development of targeted feedback approaches. This, in turn, will make it possible to study the separable effects of novel feedback protocols (e.g., whether participants made the correct choice; if not, which visual cues they missed; etc.) on neural activity (i.e., skepticism) and on the final decision.

## Neural Signals in the Loop

The "neural signals in the loop" is defined as a system that requires feedback from neural signals. Here, neural signals involved in a feedback loop to train, tune, and test the system. It makes the system more accurate and more confident. We have used neural signals in the loop both in the hardware [9] space for improving security in the devices and in the software [10] space for improving human training performance.

**Neural feedback for access control.** Access control is the core security mechanism of an operating system (OS). Ideally, the access control system should enforce context integrity, i.e., an application can only access security and privacy-sensitive resources expected by users. Unfortunately, existing access control systems, including the permission systems in the modern OS, such as iOS and Android, fail to enforce context integrity; thus, these systems allow apps to abuse their permissions. In our Intent-driven Access Control (IAC) study [9], we explored the feasibility of a novel approach to enforce the context integrity – by inferring what task users wanted to do under the given context from the Event-related potentials (ERPs) of their neural signals. ERPs are small but measurable (with an EEG sensor) voltage changes generated by the brain in response to a stimulus event. In our experiments, performing a given task with mobile apps is the stimulus event. During normal operations, the OS will continuously monitor neural signals through the BCI device as well as the user's interaction with the system to create and cache the most recent ERPs. ERPs are bound to the app to which the input event is delivered (e.g., the most foreground app at that moment) and will expire after a context switch. This prevents one app from "stealing" another app's ERP. Upon an application's request to access a protected resource (e.g., camera), the access control system will retrieve the most recent ERP. The ERP will then be fed into the trained classifier to infer whether the user intended to perform a task that requires access to that resource. If so, permission is automatically granted to that request; if the intended task does not require the permission or the confidence of the classification result is not high enough, IAC will fall back to prompt the user to make the decision. The ground truth collected from the prompt window is then to update the machine learning (ML) model. As demonstrated in the previous work [11] and our experiment, this feedback is important for fine-tuning the ML model to improve the precision of the prediction. The idea of using brain electrical signals to help OSes make dynamic access control decisions is interesting and especially relevant if BCI will be popular in the future, as the idea could be smoothly integrated into BCI supported systems.

**Neural feedback for adaptive training.** Training is a systematic approach to acquiring skills that improve performance in a task of interest. Adaptive training is one kind of training where the difficulty of the training is varied in an attempt to keep it within the optimal ranges for the trainee. An adaptive training, based on behavioral metrics, outperformed non-adaptive training in several cognitive tasks [2]. In our study [10], we added neurophysiological traits into a behavioral-based adaptive training system as a close correlation has been found between the performance of cognitive tasks and the neural signals. In this method, a task is presented to a trainee and measures the trainee's theta-alpha-ratio (TAR) from the neural signals using the EEG sensor. The task difficulty changes based on the TAR and performance score. From past research, it is revealed that alpha (8-12 Hz) power increases, focus and attention decrease, the opposite direction observe for theta power (3-7 Hz). We combined these two neural features as a ratio since they demonstrate negative and positive relationships with the behavior of interest (focus). The neural loop-based adaptive training system improved 10% over the non-adaptive training system and 17% improvement over the behavioral adaptive training system in the subsequent transfer task. The benefits of our novel approach are broad within the context of interactive training (e.g., game-based cognitive training) and could allow for improved transfer performance, reduced training times, and reduced training costs.

## Privacy of BCI

**Side-channel attacks on BCI devices.** Consumer-grade brain-computer interface devices are becoming mainstream for gaming, education, and entertainment. Considering current efforts by big companies [1] towards the BCI domain, it is inevitable that BCI devices will be part of our daily lives. The increase in the usage of the BCI devices introduces a new form of attack. In our study [3], we have shown that a malicious third-party app developer or EEG device would be able to guess PINs and passwords by monitoring a person's brain waves. BCI devices are used by gamers to play games controlled by their minds. The main idea of this kind of new attack is that a person who paused a video game and logged into a bank account while wearing an EEG headset is at risk for having sensitive data stolen by a malicious third-party app. With

this study, we raised the awareness of potential security and privacy risks associated with this emerging technology and developed viable solutions to malicious attacks.

## Future Directions

**Better understanding of users and neural signals.** Previous studies have shown that well-designed phishing websites have very high success rates in fooling users, and many users do not have enough understanding of how to avoid phishing attacks. Therefore, a better understanding of how our brains classify a phishing website from the cognitive neuroscience perspective will provide theoretical explanations for these empirical observations. Importantly, it will also lay the foundation for better user interface (UI) design and better anti-phishing training processes. For example, the analog skepticism signal may be a better way to evaluate whether a new UI design or training protocol is effective. Next, I want to study the applications of BCI devices in improving the user-experience in security tasks (e.g., security training).

**Applications of neural signals in the loop.** In our recent work [10], we observed the performance improvement in a stimulus recognition transfer task in a neural-based training system compared to both the control and behavioral-based system. The general approach of incorporating the neural feature into adaptive training systems could be applied to existing adaptive training of perceptual or cognitive tasks in which blocks of practice are subject to an adaptive difficulty modification (e.g., tutoring, exercise, simulated training). The incorporation of brain signals into Virtual Reality (VR) headsets opens the door for improving human training. We are anticipating better learning with the addition of neural measures into VR systems than the conventional VR system.

**Privacy of BCI with the help of AI.** We foresee a next-generation computing platform that will be based on BCI. The attacks like hacking, identity theft, fraud, phishing, spam, malware, ransomware, and data breaches will be common in this emerging platform. It is essential to analyze the potential unforeseen security and privacy risks associated with this emerging technology and protect the human thoughts from security breaches. In the long-term, I am planning to develop solutions against any unforeseen security breach in these new computing platforms with the help of AI.

## Multidisciplinary Collaborative Research

Due to the interdisciplinary nature of my research, most of my projects are collaborative. My understanding of the phishing attacks studies [4, 7, 8] is mostly a collaborative effort of psychologists, neuroscientists, and security experts. I have conducted studies to learn representations of brain signals measured via BCI [7] with machine learning and data science experts. I am eager to collaborate with researchers from academia and industry from the above mentioned domains. I believe that my past collaborative experiences have prepared me for future collaborative works.

# References


[1] A. Brian. Neuralink, facebook, and kernel compete on direct brain-computer interface. https://www.linkedin.com/pulse/direct-brain-interface-brian-ahier, 2017.

[2] J. Holmes, S. E. Gathercole, and D. L. Dunning. Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental science*, 12(4):F9–F15, 2009.

[3] A. Neupane, **M.L. Rahman**, and N. Saxena. Peep: Passively eavesdropping private input via brainwave signals. In *International Conference on Financial Cryptography and Data Security*, pages 227–246. Springer, 2017.

[4] A. Neupane, **M.L. Rahman**, N. Saxena, and L. Hirshfield. A multi-modal neuro-physiological study of phishing detection and malware warnings. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 479–491. ACM, 2015.

[5] C. Report. https://www.whitehouse.gov/articles/cea-report-cost-malicious-cyber-activity-u-s-economy.

[6] L. Shanhong. Size of the cybersecurity market worldwide, from 2017 to 2023. https://www.statista.com/statistics/595182/worldwide-security-as-a-service-market-size, Oct. 2019.

[7] **M.L. Rahman**, S. Bardhan, A. Neupane, E. Papalexakis, and C. Song. Learning tensor-based representations from brain-computer interface data for cybersecurity. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 389–404. Springer, 2018.

[8] **M.L. Rahman**, A. Neupane, S. Bardhan, E. Papalexakis, and C. Song. Phishing detection based on neural signals: An eeg study. Under Submission.

[9] **M.L. Rahman**, A. Neupane, and C. Song. Iac: On the feasibility of utilizing neural signals for access control. In *Proceedings of the 34th Annual Computer Security Applications Conference*, pages 641–652. ACM, 2018.

[10] **M.L. Rahman**, B. T Files, A. D. Passaro, P. Khooshabeh, A. H. Oiknine, K. Pollard, and C. Song. Cat beats bat: Combining behavioral performance and neural signals to improve adaptive training. Under Review.

[11] P. Wijesekera, A. Baokar, L. Tsai, J. Reardon, S. Egelman, D. Wagner, and K. Beznosov. The feasibility of dynamically granted permissions: Aligning mobile privacy with user preferences. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 1077–1093. IEEE, 2017.