Research Program Title: "Proactive Cyber Threat Intelligence for Scientific Cyberinfrastructure: An Artificial Intelligence (AI)-enabled Analytics Perspective"

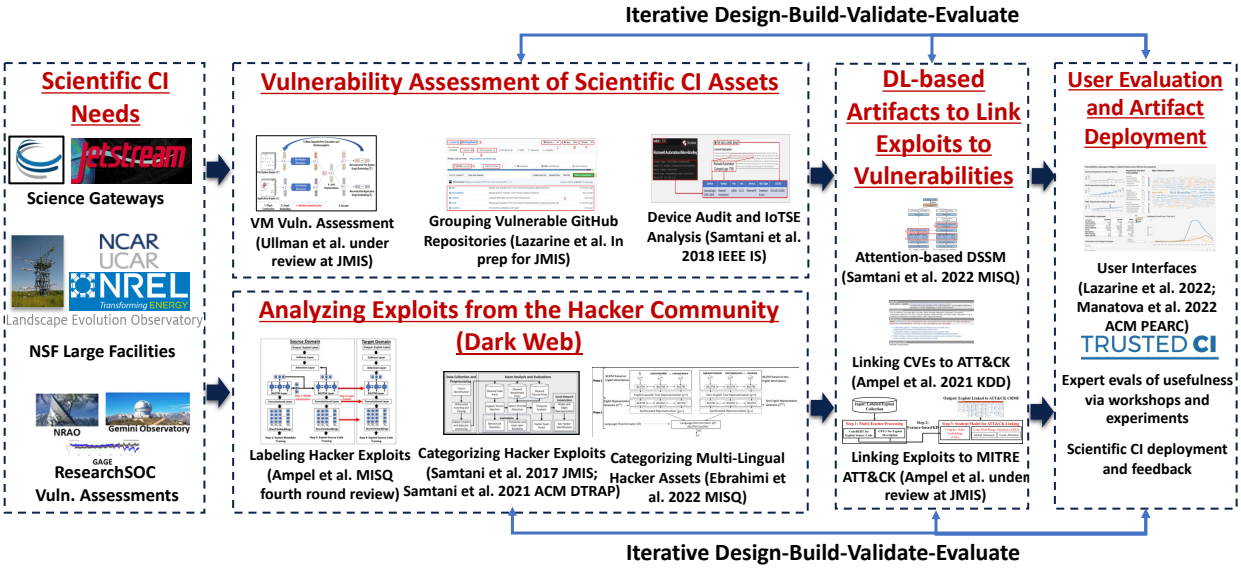
Nominees: Sagar Samtani (Assistant Professor and Grant Thornton Scholar, Indiana University) and Hsinchun Chen (Regents' Professor, University of Arizona, Director of the Artificial Intelligence Lab, AAAS/IEEE/ACM Fellow)

Project Summary (500 Words):

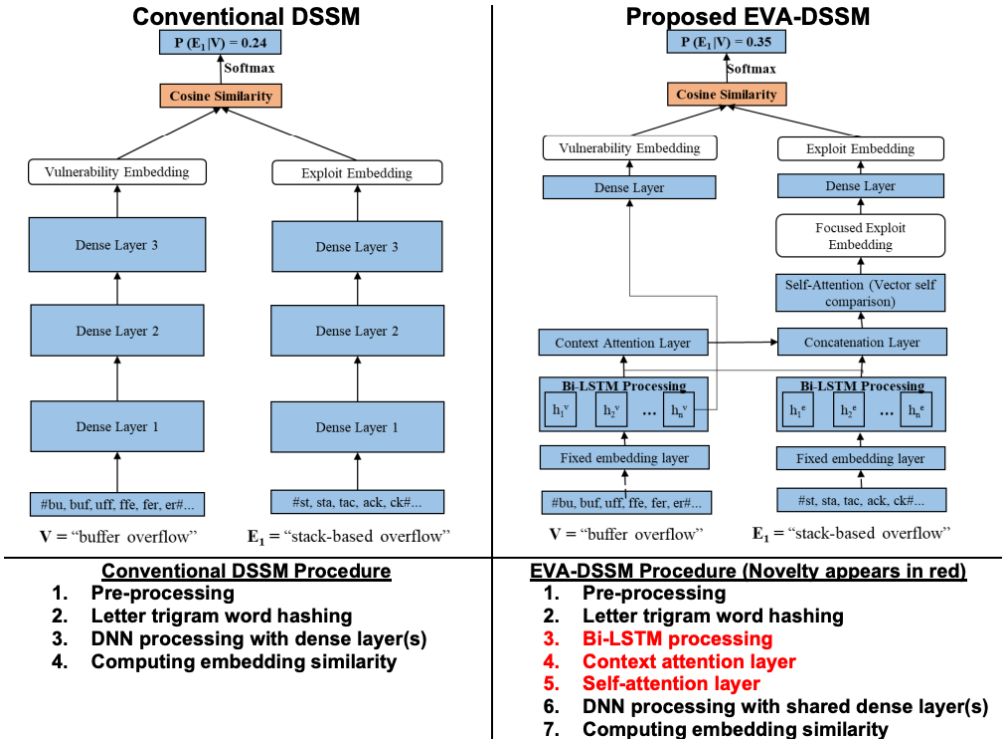
Identification and Statement of Need:

Federal funding agencies such as the National Science Foundation (NSF) have invested significant funds into developing and providing advanced scientific cyberinfrastructure (CI), such as high-performance computing, open-source coding repositories, containers, and more. Although these resources accelerate the rate of scientific discoveries, many scientists inadvertently introduce thousands of vulnerabilities into CIs. Many CIs lack security teams to group, rank, and prioritize vulnerabilities, and instead delegate responsibilities to third-party Security Operations Centers (SOCs). However, these entities are often inundated with large quantities of vulnerability and exploit data that significantly hamper their ability to produce cyber threat intelligence (CTI) for CIs.

This research effort, funded by the NSF Cybersecurity Innovation for Cyberinfrastructure program, aims to develop proactive CTI IT artifacts for scientific CIs by (1) collecting and analyzing hacker exploits from the Dark Web; (2) scanning CI assets for vulnerabilities; (3) linking exploits to vulnerabilities; and (4) and integrating algorithm results and the original exploit and scan data into dashboards to support SOC analyst CTI decision making. The figure below illustrates the major thrusts of our work.



We designed multiple deep learning (DL) algorithms to analyze exploits from the Dark Web and vulnerability data from CI assets. One essential DL-based approach we designed is the Exploit Vulnerability Assessment Deep Structured Semantic Model (EVA-DSSM; compared to conventional DSSM below) to link hacker exploits and vulnerability data.



Iterative Design-Build-Validate-Evaluate Activities:

Our artifacts are systematically evaluated against state-of-the-art DL methods and through user evaluations. For example, the usefulness of EVA-DSSM's ranked exploit-vulnerability pairs was evaluated by 45 cybersecurity analysts (see supporting paper). Dashboard designs were guided by semi-structured interviews with SOC analysts and received extensive feedback at Vulnerability Management Workshops.

Feasibility and Value of the Designed IT Systems and Artifacts:

Major CIs outsource their CTI efforts to SOCs; thus, the primary artifact users are half a dozen security analysts and engineers at ResearchSOC that develop CTI for the following CIs (each with multi-million dollar investments from NSF): CyVerse, National Radio Astronomy Observatory, Gemini Observatory, and Geodetic Facility for the Advancement of Geoscience. One measurable outcome from our program includes a 3% decrease in overall vulnerabilities for a CI since one of our dashboard's deployments (from January to March). ResearchSOC's letter further summarizes the benefits of the proposed artifacts on their operations.

Design Principles, Insights, and Lessons Learned:

- **Design Science Insights and Contributions to IS and Beyond:** This research program's artifacts are generalizable to other domains and disciplines. For example, the EVA-DSSM for linking exploits to vulnerabilities can be leveraged in short-text matching contexts (e.g., for question-answer). The dashboards designed to visualize vulnerability assessment results could be adopted into commercial vulnerability scanners and/or serve as the basis for cybersecurity HCI studies.
- **Lessons Learned:** We learned important lessons about key characteristics that CTI artifacts for scientific CIs should possess when operationalizing our IT artifacts, including: (1) computationally accounting for an exploit's age and the quantity and severity of vulnerabilities in DL algorithms and (2) appropriate layout, color, functionality, and utility of vulnerability dashboards.

Verification:
This project is principally led by university-based faculty for R&D.

Supporting Documents:

- **Journal Manuscript:**
  - Samtani, S., Chai, Y., and Chen, H. 2022. "Linking Exploits from the Dark Web to Known Vulnerabilities for Proactive Cyber Threat Intelligence: An Attention-based Deep Structured Semantic Model," *MIS Quarterly*, (46:2), pp. 909-946. (https://doi.org/10.25300/MISQ/2022/15392).
- **NSF Project Websites:**
  - **Project Abstract:** https://eller.arizona.edu/departments-research/centers-labs/artificial-intelligence/research/cici
  - **CICI Project Abstract:** https://www.nsf.gov/awardsearch/showAward?AWD_ID=1917117&HistoricalAwards=false
- **Vulnerability Management Dashboards:**
  - **Link to Dashboard for Identifying the Vulnerability Landscape of GitHub Ecosystems for CIs:** https://public.tableau.com/app/profile/ben.lazarine/viz/CICI-Dashboard-Filterable/Dashboard1
  - **Link to Dashboard for Identifying Managing the Conventional Vulnerabilities for ResearchSOC:** https://public.tableau.com/app/profile/dalyapraz/viz/VulnerabilityManagementDashboardtoEnhanceSecurityAnalystsDecisionMakingProcesses/Overview
- **Video of Vulnerability Management Dashboard (for conventional vulnerabilities):** https://iu.mediaspace.kaltura.com/media/t/1_h4zdaale
- **Community Support Letters and Testimonials:**
  - Support letter from a former NSF Cybersecurity Innovation for Cyberinfrastructure (NSF program that funded our research program) and Secure and Trustworthy Cyberspace (SaTC) Program Director, Anita Nikolich.
  - Support letter from Dr. Inna Kouper, the former Director of Research Engagement for ResearchSOC an NSF-funded entity focused on developing vulnerability management, and CTI capabilities for scientific CIs, homed in the internationally-recognized OmniSOC.

References:

Ampel, B., Samtani, S., Ullman, S., and Chen, H. 2021. "Linking Common Vulnerabilities and Exposures to the MITRE ATT&CK Framework: A Self-Distillation Approach," ACM KDD Workshop on AI-enabled Cybersecurity Analytics, pp. 1-5. (https://arxiv.org/pdf/2108.01696.pdf).

Ebrahimi, M., Chai, Y., Samtani, S., and Chen, H. 2022. "Cross-Lingual Cybersecurity Analytics in the International Dark Web with Adversarial Deep Representation Learning," *MIS Quarterly*, (46:2), pp. 1209-1226. (https://doi.org/10.25300/MISQ/2022/16618).

Lazarine, B., Manatova, D., Samtani, S. and Zhu, H. 2022. "Visualizing the Vulnerability Landscape of Major Scientific Cyberinfrastructure GitHub Ecosystems" *ACM Practice and Experience in Advanced Research Computing Interact!,* pp. 1-1.

Manatova, D., Kouper, I., and Samtani, S. 2022. "Designing a Vulnerability Management Dashboard to Enhance Security Analysts' Decision Making Processes" *ACM Practice and Experience in Advanced Research Computing, PEARC,* pp. 1-4. (https://doi.org/10.1145/3491418.3535176)

Samtani, S., Chinn, R., Chen, H., and Nunamaker, J. F. 2017. "Exploring Emerging Hacker Assets and Key Hackers for Proactive Cyber Threat Intelligence," *Journal of Management Information Systems* (34:4), pp. 1023–1053. (https://doi.org/10.1080/07421222.2017.1394049).

Samtani, S., Yu, S., Zhu, H., Patton, M., Matherly, J., and Chen, H. 2017. "Identifying SCADA Systems and their Vulnerabilities on the Internet of Things: A Text Mining Approach" *IEEE Intelligent Systems*, (33:2), pp. 63-73. (https://doi:10.1109/MIS.2018.111145022).

Samtani, S., Li, W., Benjamin, and Chen, H. 2021. "Informing Cyber Threat Intelligence through Dark Web Situational Awareness: The AZSecure Hacker Assets Portal" *ACM Digital Threats: Research and Practice*, (2:4), pp. 1-10. (https://doi.org/10.1145/3450972).

Samtani, S., Chai, Y., and Chen, H. 2022. "Linking Exploits from the Dark Web to Known Vulnerabilities for Proactive Cyber Threat Intelligence: An Attention-based Deep Structured Semantic Model," *MIS Quarterly*, (46:2), pp. 909-946. (https://doi.org/10.25300/MISQ/2022/15392).

Ullman, S., Samtani, S., Lazarine, B., Zhu, H., Ampel, B., Patton, M., and Chen, H. 2020. "Smart Vulnerability Assessment for Scientific Cyberinfrastructure: An Unsupervised Graph Embedding Approach," *Proceedings - 2020 IEEE International Conference on Intelligence and Security Informatics*, *ISI*, pp. 1–6. (https://doi.org/10.1109/ISI49825.2020.9280545).

Selected manuscripts from the research program that are in preparation or under review:

Ampel, B., Samtani, S., Zhu, H., and Chen, H. "Labeling Hacker Exploits for Proactive Cyber Threat Intelligence: A Deep Transfer Learning Approach" In Preparation for Fourth Round Review (Minor Revision) at *MIS Quarterly*.

Ampel, B., Samtani, S., Zhu, H., and Chen, H. "Linking Hacker Exploits to a Cybersecurity Risk Management Framework: A Knowledge Distillation-based Approach" Under First Round Review at *Journal of Management Information Systems*.

Lazarine, B., Samtani, S., Patton, M., Zhu, H., Nunamaker, J.F., and Chen, H. "Identifying Vulnerable GitHub Repositories and Users for Scientific Cyberinfrastructure: An Unsupervised Graph Embedding Approach" Targeted at *Journal of Management Information Systems*.

Sachdeva, A., Lazarine, B., Zhu, H., Samtani, S. "User Profiling and Vulnerability Introduction Prediction in Social Coding Repositories" Targeted at *Management Science*.

Ullman, S., Samtani, S., Zhu, H., Lazarine, B., and Chen, H. "Smart Vulnerability Assessment for Scientific Cyberinfrastructure: A Multi-View Representation Learning Approach" Under First Round Review at *Journal of Management Information Systems*.

Ullman, S., Zhu, H., Samtani, S., and Chen, H. "Linking Vulnerabilities in Cyberinfrastructure with their Remediations: A Contrastive Representation Learning Approach" Targeted at *Information Systems Research*.