The Alchemy of AI and Cybersecurity:
Who Shapes, Frames, and Evaluates AI for Cybersecurity?

Christian Gustavsson
PhD Student, Linköping University
christian.gustavsson@liu.se

Abstract

The use of artificial intelligence (AI) in cybersecurity is growing rapidly, both as a threat and as a defense mechanism. AI enables powerful new capabilities, from intrusion detection and malware analysis to adversarial simulation and automated penetration testing. Yet many cybersecurity professionals remain cautious, wary of tools that offer little in the way of assurance, explainability, or formal guarantees.

This study investigates the intersection between AI and cybersecurity by systematically analyzing how research in this domain is framed, conducted, and published. We review peer-reviewed publications with respect to AI models and methods used, cybersecurity tasks addressed, evaluation strategies employed, and signs of interdisciplinary collaboration. Based on this analysis, the aim to develop a framework that maps the current landscape of AI-for-cybersecurity research, identifying dominant trends, methodological blind spots, and underexplored areas. This framework aims to support future research and foster stronger integration between the AI and cybersecurity communities.

We conduct a mixed-methods study of publications from ten major research venues - five from AI (AAAI, CVPR, ICLR, ICML, NeurIPS) and five from cybersecurity (ACM CCS , IEEE S&P, NDSS, RAID, USENIX Security) - between 2022 and 2025. From a corpus of over 49,000 accepted publications, we identify those relevant to both fields.

The focus for a talk will be on presenting method and dataset, inviting feedback from the security community. Hopefully, the extensive dataset could also spark ideas for future cooperation with other researchers.