Al-Balqa' Applied University
Faculty of Engineering Technology

Conversational Firewall Management: Chatbot-Guided Firewall Operations Enhanced by Machine Learning

Aktham Alarabi
Network Systems and Security Department
Amman, Jordan
aktham.alarabi@gmail.com

Osamah Alananzeh
Computer Engineering Department
Amman, Jordan
ananzehosamah99@gmail.com

Bashar Alabdallat
Network Systems and Security Department
Amman, Jordan
almothanaaltaweel@gmail.com

Alaa Alnajaar
Computer Engineering Department
Amman, Jordan
Alaa_najjar2004@hotmail.com

Almothana Altaweel
Computer Engineering Department
Amman, Jordan
almothanaaltaweel@gmail.com

Ashraf Alsharah
Computer Engineering Department
Amman, Jordan
aalsharah@bau.edu.jo

Chapter 1: Introduction

1.1 Overview

The contemporary digital environment is characterized by an unceasing escalation in the volume, sophistication, and diversity of cyber threats. Malicious actors continually devise new attack vectors, ranging from large-scale Distributed Denial of Service (DDoS) campaigns and elusive zero-day exploits to stealthy Advanced Persistent Threats (APTs) that can remain undetected for extended periods. This dynamic and aggressive threat landscape persistently challenges the efficacy of conventional security measures, rendering traditional, signature-based defenses increasingly ineffective against novel and polymorphic attacks [1]. Consequently, there is a pressing and critical need for advanced cybersecurity solutions that are not only more intelligent and adaptive in their threat detection and response capabilities but also more intuitive and efficient for security personnel to manage and operate.

This project directly addresses this imperative by proposing the design, development, and rigorous evaluation of a next-generation, AI-driven firewall system. This system is engineered to provide robust, intelligent, and agile network defense. Its architecture (illustrated in Figure 1.1) is founded upon the synergistic integration of three core technological pillars, each designed to address specific shortcomings of current security paradigms:

1. **AI-Driven Anomaly Detection:** Employing sophisticated machine learning (ML) and deep learning (DL) techniques to meticulously analyze network traffic patterns and behaviors, identifying anomalous activities and subtle deviations indicative of sophisticated, and potentially unknown, attacks.
2. **Automated Attack Blocking:** Enabling immediate, autonomous mitigation actions against credibly identified threats. This rapid response capability is crucial for minimizing the window of opportunity for attackers and reducing the potential impact of security incidents.
3. **Natural Language Chatbot Interface:** Providing an intuitive, conversational medium for administrators to perform direct firewall control, configuration, monitoring, and operational oversight. This approach aims to simplify complex tasks and enhance the agility of security management.

The overarching goal of this research and development endeavor is to create a cohesive and potent security solution that significantly enhances threat detection accuracy and response timeliness. Simultaneously, it aims to simplify complex firewall administration tasks through intelligent automation and an advanced, user-centric human-computer interface, thereby empowering security operators and improving overall network resilience.

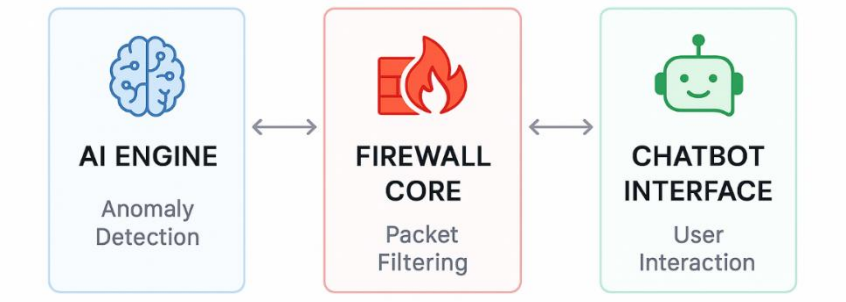

Figure 1.1: High-Level System Architecture

1.2 Problem Statement

The impetus for this project stems from several critical limitations inherent in traditional cybersecurity approaches and the significant operational challenges faced by security personnel in today's complex network environments:

Inadequacy of Traditional Firewalls against Modern Threats: Conventional firewalls, predominantly reliant on static, signature-based detection mechanisms, exhibit inherent weaknesses in countering novel, polymorphic, or zero-day attacks. Their reactive nature means they often fail to identify or are easily bypassed by attackers employing previously unseen techniques or obfuscation methods, leaving networks vulnerable [1]. This necessitates a shift towards proactive, behavior-based detection.

Operational Complexity, Alert Fatigue, and Human Error: Modern network environments generate vast quantities of security data, logs, and alerts. Security operators are frequently overwhelmed by this deluge, leading to "alert fatigue," which can result in delayed incident response and an increased likelihood of overlooking critical threats. Furthermore, the manual configuration and ongoing management of complex firewall policies are intricate processes, prone to human error. Such misconfigurations can inadvertently create security vulnerabilities, undermining the intended security posture.

Lack of Agile and Intuitive Control Mechanisms: Managing sophisticated security appliances often requires specialized expertise and proficiency in navigating complex command-line interfaces (CLIs) or dense graphical user interfaces (GUIs). This can hinder rapid response during security incidents and make dynamic policy adjustments cumbersome and time-consuming, especially when under pressure. There is a distinct need for more intuitive, accessible, and agile mechanisms for interacting with and controlling network security infrastructure.

Gap in Integrated, Explainable AI-Driven Systems: While Artificial Intelligence is increasingly being applied to threat detection, a significant gap remains in the seamless integration of AI-driven insights into fully automated response mechanisms. Moreover, many AI systems operate as "black boxes," lacking transparency. There is a critical need for systems that not only automate actions but also provide clear,

interpretable explanations of AI decisions and system actions, aligning with the principles of eXplainable AI (XAI) [9], to foster trust and enable effective human oversight.

This project seeks to comprehensively address these deficiencies by developing an integrated system that not only intelligently detects and autonomously responds to advanced threats but also empowers administrators with direct, understandable, and efficient control over their network's security posture.

1.3 Aims and Objectives

The primary aim of this project is to design, develop, and systematically evaluate an intelligent, AI-driven firewall system. This system will innovatively integrate advanced anomaly detection capabilities, automated threat mitigation functionalities, and a natural language chatbot interface to provide comprehensive, agile, and user-friendly firewall control and management.

To achieve this overarching aim, the following specific, measurable, achievable, relevant, and time-bound (SMART) objectives have been defined:

1. **To Develop a High-Fidelity AI-Driven Anomaly Detection Module:**

   Investigate, select, and implement appropriate machine learning (ML) and deep learning (DL) algorithms (e.g., exploring architectures such as Recurrent Neural Networks for temporal traffic analysis, Autoencoders for unsupervised anomaly detection, or ensemble methods, drawing inspiration from seminal works [1, 2, 6]) capable of identifying sophisticated network threats, including zero-day exploits, polymorphic malware indicators, and anomalous traffic patterns, with a target of high accuracy and low false positive rates.

   Design the module to learn from evolving network behavior and adapt to new and emerging threat vectors, potentially incorporating mechanisms for continuous learning or periodic retraining.

2. **To Implement a Responsive Automated Attack Blocking Module:**

   Design and integrate a robust mechanism that automatically executes pre-defined or dynamically determined mitigation actions (e.g., dropping malicious packets, blocking offending IP addresses, isolating potentially compromised devices, or rate-limiting suspicious connections) in real-time upon credible threat identification by the AI-Driven Anomaly Detection module, informed by research in automated response [4, 9].

   Ensure that all automated responses are meticulously logged and auditable, providing a clear record of actions taken for forensic analysis and compliance purposes.

3. **To Create a Robust and Intuitive Natural Language Chatbot Interface:**

   Develop a conversational AI interface that allows administrators to directly configure firewall rules, manage security policies, query system status, and execute control commands using intuitive natural language (e.g., "Block traffic from IP address X.X.X.X to internal server Y," "Modify rule ID 10 to allow

port Z for host A," "Show all active policies related to the DMZ segment," "What are the latest critical alerts?"), referencing advancements in NLU for cybersecurity [8, 15, 16].

Enable the chatbot to provide users with real-time system statistics, comprehensive threat reports, and, critically, clear, interpretable explanations for AI-detected anomalies and automated system actions [5, 11, 13]. This functionality will be designed in alignment with XAI principles [9] to enhance transparency and user trust.

4. **To Integrate, Test, and Evaluate the Complete System Holistically:**

Successfully integrate the AI-Driven Anomaly Detection, Automated Attack Blocking, and Natural Language Chatbot Interface modules into a cohesive, functional firewall system prototype.

Conduct comprehensive testing and evaluation of the integrated system's performance. This will involve assessing threat detection accuracy using benchmark datasets (e.g., CICIDS2017, NSL-KDD) and simulated attack scenarios; measuring response timeliness of the automated blocking module; and evaluating the usability, effectiveness, and user satisfaction of the chatbot interface for firewall control and information retrieval through structured user studies.

1.4 Contribution of the Project

This project aims to make a significant contribution to the field of network security by:

Developing an Integrated Tri-Pillar Solution: Offering a novel integration of AI-driven anomaly detection, automated blocking, and a natural language chatbot interface within a single firewall system, addressing limitations of siloed approaches.

Enhancing Operational Efficiency and Agility: Simplifying complex firewall management tasks and enabling faster response times through intuitive conversational control and automation, thereby reducing operator burden.

Promoting Explainable AI in Security Operations: Providing a practical implementation of XAI principles via the chatbot, making AI-driven security decisions more transparent and understandable to human operators.

Advancing Adaptive Defense Mechanisms: Contributing to the development of more intelligent and adaptive firewall systems capable of better contending with the evolving tactics of cyber adversaries.

To fully elaborate on these contributions and detail the research undertaken, the remainder of this document is structured as follows: Chapter 2 provides a comprehensive review of existing literature relevant to AI-driven anomaly detection, automated threat mitigation, and the application of conversational AI in cybersecurity contexts. Chapter 3 will detail the proposed system architecture, outlining the design choices, algorithms, and methodologies for each core module and their

interconnections. Chapter 4 will describe the implementation process, including the specific software tools, programming languages, hardware platforms, and datasets utilized in the development of the prototype. Chapter 5 will present the experimental setup, rigorous evaluation metrics, and a detailed analysis of the system's performance against the defined objectives. Finally, Chapter 6 will conclude the project, summarizing the key findings, discussing the limitations encountered, and suggesting potential avenues for future research and development in this domain.

Chapter 2: Related Works

This chapter surveys the current research landscape supporting the development of an AI-driven firewall. The review focuses on prior work in three critical areas: AI-powered network anomaly detection, automated mechanisms for attack blocking, and the application of natural language chatbots to cybersecurity operations.

Modern intrusion detection systems (IDS) face challenges in detecting unknown threats due to reliance on static rulesets. Rozendaal et al. (2023) propose neural network (NN)-assisted IDS to improve adaptability, using architectures like Feedforward Neural Networks (FFNN) for attack classification and Backpropagation Neural Networks (BPNN) for iterative learning. Their study highlights benefit such as real-time threat detection but notes drawbacks like computational overhead and difficulty detecting stealthy attacks (e.g., R2L/U2R) [1].

Advanced AI-driven cybersecurity solutions are transforming threat detection and response by leveraging machine learning (ML), deep learning (DL), and natural language processing (NLP) to address evolving cyber risks. Tanikonda et al. (2022) highlight the effectiveness of ensemble and hybrid AI models in detecting zero-day exploits, polymorphic malware, and advanced persistent threats (APTs). Their study emphasizes AI’s role in Security Orchestration, Automation, and Response (SOAR) platforms, enabling real-time threat mitigation with minimal human intervention. Challenges such as adversarial AI attacks, data privacy concerns, and computational overhead are noted, alongside solutions like federated learning and explainable AI (XAI) [2].

Lee et al. (2024) propose an AI-based approach to refine firewall rules in high-performance computing (HPC) networks, addressing human errors in policy management. By analyzing four years of firewall logs (289,320 error instances), they applied machine learning (e.g., KNN, SVM) and deep learning (NN with cross-entropy loss) models, achieving 98% accuracy in classifying misconfigured rules. Key innovations include parallel log parsing in HPC environments and feature engineering to enhance detection. While the

study demonstrates AI's potential to automate firewall optimization, challenges like computational overhead and real-time scalability persist [3].

Sharma (2023) presents an AI-driven anomaly detection system for advanced threat detection, leveraging machine learning (ML) and deep learning (DL) to identify zero-day attacks, APTs, and insider threats. The proposed system integrates unsupervised learning (e.g., autoencoders, GANs) with behavioral analytics to reduce false positives and enable real-time responses like isolating compromised systems. Challenges such as adversarial attacks and model interpretability are noted, with Explainable AI (XAI) suggested to enhance transparency. Survey results highlight high user satisfaction (avg. score: 4.5/5) in threat detection effectiveness and automated response [4].

Modern AI-driven chatbots are increasingly being leveraged for intrusion detection in edge networks, addressing gaps in traditional cybersecurity approaches. Asif et al. (2024) propose an ethical AI chatbot architecture that integrates machine learning (ML) models (e.g., Decision Tree, Random Forest) with Raspberry Pi-based edge networks for real-time threat detection. Their system emphasizes user consent via OTP authentication and transparent monitoring policies, achieving high accuracy (86.84% for Decision Tree) on the NSL-KDD dataset. While the study demonstrates the chatbot's efficacy in balancing security and ethical considerations, challenges such as computational overhead and scalability in large-scale deployments remain [5].

Recent advancements in deep learning have significantly enhanced the capabilities of intrusion detection systems (IDS), particularly for real-time web traffic analysis. Kim et al. (2020) propose AI-IDS, a scalable deep learning framework combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to detect obfuscated and zero-day web attacks. Their system leverages normalized UTF-8 encoding for efficient spatial feature extraction from raw HTTP payloads, achieving 98.07% accuracy on real-world datasets (CSIC-2010, CICIDS2017). Deployed via Docker for high-performance computing, AI-IDS addresses limitations of traditional signature-based IDS by automating Snort rule updates and reducing false positives. However, challenges such as computational overhead and dependency on large-scale training data persist [6].

AI is transforming cybersecurity by integrating advanced technologies like machine learning (ML), natural language processing (NLP), and generative AI to enhance threat detection and user experience (UX). Shete (2023) explores AI-driven solutions such as Text-to-Visualization for real-time data analytics, Breach Prediction for proactive threat mitigation, and Multi-Modal Data Protection for comprehensive security. The study emphasizes the role of user-centric UX design in cybersecurity tools, advocating for interfaces that balance automation (e.g., system-generated dashboards) with user control (e.g., NLP-based

querying). While AI improves efficiency and adaptability, ethical challenges like bias and privacy persist [7].

Recent research highlights the potential and limitations of Large Language Models (LLMs) in generating cybersecurity rules, such as firewall and Intrusion Detection System (IDS) configurations. Louro et al. (2024) evaluated the capability of widely available chatbots, including ChatGPT and Mistral 7B, to produce accurate iptables and Snort rules. Their study revealed that pre-trained models achieved limited success (e.g., ChatGPT 3.5: 42% accuracy), often generating syntactically correct but technically flawed rules. However, fine-tuning approaches significantly improved performance, with Mistral 7B reaching 89% accuracy when trained on structured prompts. The study underscores the necessity of human oversight and domain-specific fine-tuning to ensure rule correctness and efficiency [8].

The integration of artificial intelligence (AI) into cybersecurity is revolutionizing threat detection, response, and mitigation by automating repetitive tasks and enhancing accuracy. Kaur et al. (2023) conducted a systematic literature review (SLR) of 236 primary studies, categorizing AI applications in cybersecurity using the NIST framework (Identify, Protect, Detect, Respond, Recover). Their findings highlight AI's effectiveness in automating vulnerability assessments, intrusion detection, and predictive threat intelligence, with machine learning (ML) and deep learning (DL) being the dominant techniques. Key benefits include real-time anomaly detection and adaptive security policies, while challenges such as adversarial AI, data heterogeneity, and computational costs are noted. The study underscores the need for explainable AI (XAI) and multi-source data fusion to improve transparency and robustness [9].

The translation of natural language queries into structured database queries has been extensively studied. Roturier et al. (2019) address the gap for document-oriented databases like Elasticsearch, commonly used in Security Information and Event Management (SIEM) systems, by proposing a hybrid translation approach. Their system combines translation memory, information extraction, and text classification to generate Elasticsearch queries from natural language inputs, supporting cyber-threat hunters by bootstrapping queries from limited training data. Challenges include handling flexible schemas and ambiguous entities, mitigated through rule-based components and interactive user interfaces [10].

Intrusion Detection Systems (IDS) are critical for network security but often pose challenges for non-experts in interpreting alerts and responding effectively. Jüttner et al. (2023) propose ChatIDS, a generative AI approach leveraging large language models (LLMs) like ChatGPT to translate technical IDS alerts (e.g., from Snort, Suricata) into intuitive explanations for home users. Their study demonstrates feasibility, though limitations include occasional non-intuitive terminology and incomplete

countermeasure descriptions. The authors also identify interdisciplinary challenges, including privacy risks, legal liabilities, and ethical concerns [11].

Chen et al. (2024) propose ControlNet, a firewall mechanism for Retrieval-Augmented Generation (RAG)-based Large Language Model (LLM) systems to prevent toxic or malicious outputs during inference. By integrating a multi-stage classifier with reinforcement learning (RL) feedback, ControlNet evaluates retrieved documents and candidate generations to filter unsafe content. The system enhances safety without retraining the underlying LLM, achieving high detection accuracy and minimal latency overhead, demonstrating a modular AI firewall that filters harmful inputs and outputs [12].

Deussom et al. (2023) propose a chatbot integrated with the ELK Stack SIEM tool to enhance Security Operations Center (SOC) capabilities in identifying, analyzing, and remediating cybersecurity incidents. The system processes voice commands for real-time event supervision, leverages natural language processing (NLP), and uses Python-based AI modules on a Raspberry Pi for speech recognition and interaction. Benefits include centralized threat visibility, faster incident response, and cost reduction, showcasing AI's role in augmenting human oversight in security monitoring [13].

Lin and Yao (2022) introduce a firewall anomaly detection model using an asymmetric double decision tree to identify and manage rule conflicts in large-scale policies. Their approach constructs an equivalent decision tree to separate valid and anomalous rule spaces, enabling faster incremental detection when new rules are added. Key innovations include converting policies into simplified black- or whitelist models and generating detailed anomaly reports. Experiments show up to 90%-time savings for incremental detection compared to full rule analysis, validating this model's efficiency [14].

Alzantot et al. (2024) introduce GOLLUM, an LLM-augmented system designed to assist in firewall configuration by translating high-level security intentions into accurate low-level rules. Leveraging large language models (LLMs) such as GPT, GOLLUM improves policy specification accuracy, reduces human error, and enhances usability through natural language interfaces. The study evaluates GOLLUM on real-world firewall datasets and benchmark tasks, showing improved correctness and user satisfaction, though challenges in interpreting ambiguous intents and ensuring rule consistency remain [15].

Ramakrishnan et al. (2022) propose an IoT-integrated chatbot system using natural language processing (NLP) to support real-time aquaculture monitoring. The system uses sensors (e.g., for pH, temperature, dissolved oxygen) connected via NodeMCU to collect environmental data, which is processed and visualized through a cloud platform like ThingSpeak. A Dialogflow-based chatbot allows users to access this data through natural language queries, improving usability and reducing technical barriers for farmers. Benefits include real-time alerts, reduced manual oversight, and improved decision-making.

While not cybersecurity-focused, this work informs our project's chatbot interface by demonstrating the effectiveness of NLP for intuitive, real-time interaction with technical systems [16].

The reviewed literature collectively validates the core tenets of our proposed AI-driven firewall system. The body of work in AI-driven anomaly detection [1], [2], [4], [6] clearly supports the use of advanced ML and DL techniques for identifying sophisticated network threats, with the successes of various NN architectures like FFNNs, CNN-LSTMs, autoencoders, and GANs providing a strong foundation for our project's AI-Driven Anomaly Detection module and its potential for high accuracy. However, persistent challenges such as computational overhead [1], [2] and detecting stealthy attacks [1] inform our design considerations for optimized models and robust feature engineering. This foundation in detection is complemented by research underscoring a clear trend towards leveraging AI for intelligent automation in cybersecurity, spanning both proactive policy management [3], [14] and reactive threat response. The emphasis on automated response capabilities in SOAR platforms [2], adaptive security policies [9], automated rule generation [6], and system isolation [4] strongly supports the rationale for our project's Automated Attack Blocking feature, which aims to integrate AI-driven detection directly with immediate mitigation actions, though challenges like real-time scalability [3] and reliable translation of AI detections into correct responses are key implementation considerations. Finally, the literature strongly indicates a growing adoption of conversational AI and advanced NLP techniques [5], [11], [13] to improve human interaction with cybersecurity systems, directly validating our Chatbot Interface for security monitoring, alert explanation, and system interaction. Our chatbot aims to fulfill these roles by providing summaries, reports, and explanations of actions from the other modules, with research by [8], [15], and [16] highlighting the critical need for fine-tuned models or validation layers for translating natural language commands into safe firewall actions. Data querying features [7], [10] and LLM output safety [12] further inform its design, as the complexity of AI-detected anomalies and automated actions necessitates a clear, interpretable interface, aligning with the broader call for eXplainable AI (XAI) [4], [9].

References

[1] K. Rozendaal, T. Dissanayaka, and A. Mallewa, "Neural Network Assisted IDS/IPS: An Overview of Implementations, Benefits, and Drawbacks," 2023.
[2] A. Tanikonda, S. R. Peddinti, B. K. Pandey, and S. R. Katragadda, "Advanced AI-Driven Cybersecurity Solutions for Proactive Threat Detection and Response in Complex Ecosystems," 2022.
[3] J.-K. Lee, T. Hong, and G. Lee, "AI-Based Approach to Firewall Rule Refinement on High-Performance Computing Service Network," 2024.

[4] S. Sharma, "AI-Driven Anomaly Detection for Advanced Threat Detection," J.P. Morgan Chase Inc., Jersey City, NJ, USA, Tech. Rep./Internal Publication (presumed), 2023.
[5] M. Asif, A. Manan, A. M. ur Rehman, M. N. Asghar, and M. Umair, "AI-Driven Chatbot for Intrusion Detection in Edge Networks: Enhancing Cybersecurity with Ethical User Consent," *Technologies*, vol. 12, no. 1, Art. no. 6, Jan. 2024. doi: 10.3390/technologies12010006.
[6] A. Kim, M. Park, and D. H. Lee, "AI-IDS: Application of Deep Learning to Real-Time Web Intrusion Detection," *IEEE Access*, vol. 8, pp. 70245–70261, 2020. doi: 10.1109/ACCESS.2020.2986747.
[7] S. Shete, "AI in Cybersecurity and User Interface Design beyond Chatbots," *J. Artif. Intell. Cloud Comput.*, vol. 3, no. 1, pp. 18–23, 2023.
[8] B. Louro, R. Abreu, J. C. Costa, J. B. F. Sequeiros, and P. R. M. Inácio, "Analysis of the Capability and Training of Chat Bots in the Generation of Rules for Firewall or Intrusion Detection Systems," in *Proc. Int. Conf. Availability, Reliability and Security (ARES)*, 2024. doi: 10.1145/nnnnnnn.nnnnnnn.
[9] R. Kaur, D. Gabrijelčič, and T. Klobučar, "Artificial Intelligence for Cybersecurity: Literature Review and Future Research Directions," *Inf. Fusion*, vol. 97, Art. no. 101804, Sep. 2023. doi: 10.1016/j.inffus.2023.101804.
[10] J. Roturier, B. Schlatter, and D. Silva, "Bootstrapping a Natural Language Interface to a Cyber Security Event Collection System using a Hybrid Translation Approach," in *Proc. 17th Mach. Transl. Summit (MT Summit)*, Dublin, Ireland, Aug. 2019, pp. 257–258.
[11] V. Jüttner, M. Grimmer, and E. Buchmann, "ChatIDS: Explainable Cybersecurity Using Generative AI," in *Proc. 17th Int. Conf. Syst. Netw. Commun. (SECURWARE)*, Porto, Portugal, Sep. 2023.
[12] Y. Chen, M. Yasunaga, Y. Mao, and J. Leskovec, "ControlNet: A Firewall for RAG-based LLM Systems," Feb. 2024. [Online]. Available: arXiv:2402.12743.
[13] E. M. Deussom Djomadji, F. A. B. Mfoum, E. Tonye, and A. Binele Abana, "Design and Implementation of a Chatbot for the Supervision of Security Events (SIEM)," *Int. J. Comput. Appl.*, vol. 185, no. 14, pp. 20–28, Aug. 2023. doi: 10.5120/ijca2023915825.
[14] Z. Lin and Z. Yao, "Firewall Anomaly Detection Based on Double Decision Tree," *Symmetry*, vol. 14, no. 12, Art. no. 2668, Dec. 2022. doi: 10.3390/sym14122668.
[15] R. Lorusso, A. Maci, and A. Coscia, "GOLLUM: Guiding cOnfiguration of firewaLL Through aUgmented Large Language Models," May 2023. [Online]. Available: arXiv:2305.12719.
[16] S. Lawal, X. Zhao, A. Rios, R. Krishnan, and D. Ferraiolo, "Translating Natural Language Specifications into Access Control Policies by Leveraging Large Language Models," in *Proc. IEEE 6th Int. Conf. Trust, Privacy Security Intell. Syst. Appl. (TPS-ISA)*, Atlanta, GA, USA, Dec. 2024, to be published 2025. doi: 10.1109/TPS-ISA62245.2024.00048.