WCRG Global Research
WORLD COMET RESEARCH GROUP

International Journal of Innovative Research in Humanities & Technology
Vol. 2, No. 2, 2019
Doi: 10.67228/3071561X/IJIRHT-2019PII8N6Y
PP. 01-10

Original Article

Adversarially Resilient UX in AI-Powered Cybersecurity Decision Support

Dr. Rakesh Chandra

Associate Professor, University of Calcutta, India.

Received: 06-03-2019 Revised: 06-26-2019 Accepted: 07-03-2019 Published: 07-05-2019

ABSTRACT

As AI becomes integral to cybersecurity decision support systems, the attack surface expands beyond models to include user interfaces and decision workflows. This paper explores the concept of adversarially resilient user experience (UX) in AI-powered cybersecurity tools. We examine how adversaries may exploit cognitive, perceptual, and interaction-level vulnerabilities in UX to mislead human analysts or distort model outputs. Drawing on human-computer interaction (HCI), adversarial ML, and cyber threat intelligence, we propose a design framework for resilient UX in such systems. Through threat modeling, scenario analysis, and a taxonomy of attack vectors at the human-AI interface, we provide guidelines for creating decision-support environments that enhance trust, robustness, and interpretability while resisting manipulation. We validate the approach with case studies in phishing detection and network anomaly triage, and we identify key research directions for building human-AI systems that are secure not just technically, but behaviorally and perceptually as well.

KEYWORDS

Adversarial Machine Learning, Human-AI Interaction, Cybersecurity Decision Support, User Experience (UX) Design, Adversarial UX, Human Factors in Security, Trustworthy AI, Interpretability, Threat Modeling, Cognitive Security.

ISSN: 3071-561X
PP: 01-10.

1. INTRODUCTION

1.1. Motivation: Role of AI in Cybersecurity Decision Support

The growing complexity and scale of cyber threats have led to increased reliance on artificial intelligence (AI) to support cybersecurity decision-making. AI systems are now integrated into various aspects of security operations—from threat detection and incident response to anomaly detection and predictive analytics. These tools promise improved efficiency and faster reaction times, often outperforming human analysts in sifting through vast amounts of data. However, while AI enhances decision support, it also introduces new challenges related to reliability, interpretability, and user trust. The motivation for this paper stems from the need to critically examine how AI systems assist human decision-making in security contexts and whether they are truly augmenting analysts or creating new dependencies and vulnerabilities.

1.2. Growing Risk: Adversarial Threats Beyond ML Models

While the field of adversarial machine learning has primarily focused on how attackers manipulate model inputs to degrade performance or evade detection, a broader class of adversarial threats is emerging. These threats extend beyond algorithmic vulnerabilities to target the entire decision-making pipeline, including how information is presented to human operators. Adversaries are no longer just fooling models—they are also attempting to manipulate humans who rely on these models. This includes poisoning training data, crafting misleading alerts, or designing inputs that exploit cognitive biases in human analysts. As a result, the threat landscape is expanding from purely technical exploits to hybrid attacks that target both AI systems and their human users, demanding a more holistic view of adversarial risks.

1.3. UX as an Attack Surface

User experience (UX), traditionally regarded as a usability or design consideration, is now emerging as a critical attack surface in cybersecurity systems powered by AI. Interfaces that mediate human-AI interaction—such as dashboards, alert systems, and incident response tools—can be manipulated or subverted to mislead human analysts. For instance, adversaries might exploit interface inconsistencies, visual overload, or poor alert prioritization to bury malicious activity under benign noise. Such UX-level manipulations can lead to delayed or incorrect decisions, especially when compounded by time pressure and cognitive fatigue. Understanding UX as an attack vector requires rethinking the design of interfaces not only for usability but also for security and resilience against manipulative behaviors.

1.4. Contribution of the Paper

This paper makes a multidisciplinary contribution at the intersection of AI, cybersecurity, human factors, and adversarial resilience. It argues for an expanded definition of adversarial threats that includes UX-level manipulations and outlines a taxonomy of these emerging attack vectors. The paper reviews existing AI-powered cybersecurity tools and critiques their assumptions about human-AI collaboration.

It also highlights how poor interface design and cognitive overload contribute to decision-making failures. Finally, it proposes design principles and defense strategies for building AI systems that are not only technically robust but also human-aware—accounting for trust, bias, explainability, and UX vulnerabilities.

ISSN: 3071-561X
PP: 01-10.

2. BACKGROUND AND RELATED WORK

2.1. AI-Powered Cybersecurity Tools (e.g., SIEMs, SOAR, Threat Hunting)

AI has become integral to modern cybersecurity ecosystems, particularly in Security Information and Event Management (SIEM) systems, Security Orchestration, Automation and Response (SOAR) platforms, and advanced threat hunting tools. These systems leverage machine learning algorithms to process and correlate data from multiple sources, detect anomalies, prioritize threats, and even automate responses. For example, SIEMs use AI to reduce alert fatigue by clustering similar alerts and highlighting unusual activity patterns. SOAR platforms take this further by executing predefined playbooks in response to threats. While these tools greatly enhance operational efficiency, their effectiveness often hinges on how their outputs are interpreted and acted upon by human analysts. This reliance introduces challenges related to interpretability, trust, and cognitive load, particularly when alerts are opaque or misleading.

2.2. Overview of Adversarial Machine Learning

Adversarial machine learning (AML) is a subfield focused on how attackers can exploit vulnerabilities in ML models by providing deliberately crafted inputs. Classic examples include adversarial images that fool classification systems or malicious network traffic that bypasses intrusion detection. Techniques such as evasion attacks (where adversaries modify inputs at inference time), poisoning attacks (where training data is corrupted), and model inversion (where sensitive training data is extracted from models) illustrate the range of threats. While significant work has been done on defending ML models against such attacks, much of it remains narrowly focused on algorithmic defenses. The emerging perspective is that these attacks can also propagate downstream to affect human users, making it essential to integrate human-centered concerns into AML defense strategies.

2.3. UX in Cybersecurity Systems

The user interface (UI) and overall user experience (UX) in cybersecurity tools play a pivotal role in determining how effectively analysts can detect and respond to threats. Good UX design can help analysts quickly interpret system outputs, prioritize responses, and maintain situational awareness. Conversely, poor UX—characterized by cluttered dashboards, ambiguous alerts, or inconsistent terminology—can slow down decision-making and contribute to errors. In AI-powered systems, UX becomes even more critical, as the user must interpret complex model outputs and reconcile them with other sources of information. Emerging research suggests that adversaries can indirectly manipulate UX elements to mislead users, for example by triggering alert floods or crafting inputs that lead to misclassification but appear normal to users. This shifts UX from a usability challenge to a security-critical design component.

2.4. Human Factors and Decision Biases in Security Operations

Human analysts operating within cybersecurity environments are subject to various cognitive and psychological biases. These biases—such as confirmation bias, anchoring, or automation bias—can significantly affect how security events are interpreted and how responses are prioritized. For instance, an analyst might over-rely on AI recommendations (automation bias), even when they are incorrect, or fail to notice anomalies because they align with prior expectations (confirmation bias). Under high-stress, high-volume conditions, these biases are amplified, leading to potential security lapses. Human factors research in security has highlighted the importance of designing systems that support decision-making under uncertainty and pressure. However, the integration of AI introduces

ISSN: 3071-561X
PP: 01-10.

new cognitive dynamics that are not yet fully understood, warranting deeper investigation into how AI interfaces can mitigate, rather than exacerbate, human error.

2.5. Trust and Explainability in AI Interfaces

Trust is a fundamental prerequisite for effective human-AI collaboration in cybersecurity. If users do not trust AI systems, they may ignore valuable recommendations; if they trust them too much, they may accept incorrect outputs without sufficient scrutiny. Explainability is a key factor in building calibrated trust—it enables users to understand how and why an AI system arrived at a given conclusion. Explainable AI (XAI) techniques, such as feature attribution, counterfactual explanations, or natural language summaries, can help bridge the gap between complex models and human reasoning. However, explainability alone is not enough; it must be integrated into intuitive, actionable interfaces that align with users' mental models. Moreover, adversaries can potentially exploit explainability features to probe or reverse-engineer systems, adding yet another dimension to the adversarial landscape. This dual role of explainability—as both a trust enabler and a possible attack vector—demands careful consideration in interface design.

Table 1: Comparison of AI-Powered Cybersecurity Tools

Tool Type	Description	Common Features	Role of AI	UX Challenges
SIEM	Security Information and Event Management systems	Log collection, correlation, alerting	Pattern detection, alert prioritization	Alert overload, poor visualizations
SOAR	Security Orchestration, Automation, and Response tools	Automated playbooks, incident response	Automates response based on AI logic	Opaque automation, lack of trust
Threat Hunting	Proactive threat search across network and endpoints	Hypothesis testing, anomaly analysis	Anomaly detection, predictive analytics	High cognitive load, data overload
EDR/XDR	Endpoint/Extended Detection and Response platforms	Real-time threat detection and response	Behavior-based detection	Unclear alert rationale, interface clutter

Table 2: Categories of Adversarial ML Attacks

Attack Type	Description	Target	Impact	Defense Techniques
Evasion	Modify input data to avoid detection by ML models	Inference phase	Model misclassification	Robust training, input sanitization
Poisoning	Corrupt training data to introduce model biases	Training data	Long-term model degradation	Data validation, anomaly detection
Model Inversion	Reconstruct training data from model outputs	Trained model	Privacy breach, sensitive data leakage	Differential privacy, output limits
UX Manipulation	Exploit human trust via misleading alerts or visual noise	Human users	Cognitive overload, delayed response	UX hardening, alert curation

3. THREAT MODELING THE HUMAN-AI INTERFACE

As AI systems become more embedded in cybersecurity workflows, the human-AI interface itself emerges as a significant attack surface. Traditional threat modeling has largely focused on network layers, software vulnerabilities, and algorithmic robustness. However, adversaries are increasingly targeting the seams between human users and AI systems—where perception,

ISSN: 3071-561X
PP: 01-10.

interpretation, and decision-making occur. Threat modeling at this level involves identifying how interface-level vulnerabilities can be exploited to mislead or confuse human operators. This expanded threat landscape includes UX-specific adversarial tactics designed not to break the system directly, but to exploit how humans interpret and act on AI outputs.

3.1. UX-Specific Adversarial Attack Surfaces

These attack surfaces exist at the intersection of interface design and human cognition. They do not necessarily require corrupting the underlying model or data. Instead, attackers exploit how information is presented through dashboards, alerts, graphs, or decision-support tools. For example, attackers may trigger specific model outputs that appear benign but are visually emphasized or de-emphasized in misleading ways. This type of manipulation leverages trust in the interface itself rather than in the model's raw predictions. As a result, even well-performing AI systems can be compromised if their interfaces are poorly designed or manipulated.

3.2. Misleading Visualizations

Visualizations are core to how security analysts interpret complex data. However, when these visualizations are misleading either intentionally through adversarial manipulation or unintentionally due to poor design they can become vectors for misinformation. Examples include threat maps that downplay the geographic origin of attacks, bar charts that exaggerate trends, or heatmaps that obscure patterns through color overload. Adversaries may craft inputs to skew these visuals in subtle ways, causing analysts to underestimate or overlook real threats. Misleading visualizations can manipulate attention and shape false mental models, resulting in poor decision-making.

3.3. Biased or Ambiguous Explanations

AI explainability interfaces are intended to make model decisions more transparent to users. However, if these explanations are biased, vague, or overly confident, they can mislead users rather than inform them. For example, a phishing classifier might explain a decision by highlighting benign-looking keywords while omitting riskier indicators. Ambiguous or incomplete explanations can lull analysts into a false sense of security, leading them to approve risky actions. In adversarial scenarios, attackers might exploit explanation mechanisms to hide malicious activity or produce deceptive rationales that align with user expectations or biases.

3.4. Manipulable Confidence Indicators

Many AI interfaces display a "confidence score" to indicate the system's certainty in its predictions. While useful in theory, these scores can be easily manipulated—either by adversarial inputs or through model miscalibration. A high-confidence score on a misclassified threat can mislead analysts into trusting faulty outputs. Conversely, a low-confidence score may cause undue skepticism about valid alerts. If users are conditioned to associate confidence with correctness, they may become vulnerable to interface-level deception. Confidence indicators, when not anchored to a clear and trustworthy explanation, become easy tools for adversarial manipulation.

3.5. Cognitive Attack Vectors (e.g., Anchoring Bias, Alert Fatigue)

Human cognitive biases present another layer of vulnerability. Anchoring bias, for instance, may cause an analyst to rely heavily on the first piece of information presented—such as an initial alert classification—regardless of subsequent evidence. Alert fatigue, a well-documented issue in SOCs (Security Operations Centers), results from the overwhelming number of alerts, leading analysts to ignore or superficially triage notifications. Adversaries can exploit these weaknesses by

ISSN: 3071-561X
PP: 01-10.

creating conditions where real threats are buried under benign-looking noise or where misleading indicators are placed early in the analysis process. The goal is to subtly manipulate the analyst's decision path using interface-level cues and psychological pressure.

3.6. Social Engineering through Interfaces

Beyond technical and cognitive manipulation, interfaces can also become conduits for social engineering. Attackers may exploit interface patterns that users trust—such as known branding, default behaviors, or auto-suggested actions—to inject malicious actions into a workflow. For instance, an attacker might craft an input that triggers a familiar-looking popup or alert that prompts the user to take risky actions. In AI-assisted environments, where users expect automation and assistance, this type of manipulation is particularly potent. Interfaces that mimic legitimate system behavior or exploit trust in AI decision-making can be weaponized to carry out social engineering attacks at scale.

3.7. Real-World Examples and Case Studies

Numerous incidents illustrate these vulnerabilities. For example, some phishing attacks have exploited spam filters and mail triage systems that misclassify emails due to adversarial crafting of message metadata or embedded code. In another case, attackers used log injection techniques to confuse visualization tools in a SIEM dashboard, masking their activities by polluting the interface. In both cases, the AI model functioned as expected, but the human interface was the exploited vector. These examples demonstrate the necessity of incorporating UX-focused threat modeling into cybersecurity practices, especially as AI systems become the gatekeepers of human attention and decision-making.

4. DESIGN PRINCIPLES FOR ADVERSARIALLY RESILIENT UX

To address the challenges outlined above, it is essential to define design principles that guide the creation of AI interfaces resilient to adversarial manipulation. The goal is not only to prevent misuse but to actively support robust, informed, and trustworthy decision-making by human users. These principles are grounded in interdisciplinary research from security, HCI (Human-Computer Interaction), psychology, and AI ethics.

4.1. Resilience Goals: Robustness, Transparency, User Empowerment

A resilient interface should first be robust capable of functioning under uncertainty or adversarial stress without misleading the user. Second, it must be transparent, clearly communicating model logic, system limitations, and data integrity. Finally, it should promote user empowerment, enabling users to question, verify, and override AI decisions when necessary. These goals set the foundation for the design principles discussed below and serve to align technical reliability with human-centered usability.

4.2. Redundancy and Cross-Verification

To guard against single points of manipulation, interfaces should support redundancy—multiple views of the same data or decision, ideally from independent sources. Cross-verification mechanisms allow users to compare AI outputs with raw logs, alternative models, or peer-reviewed analysis. This approach helps catch inconsistencies or subtle manipulations that may not be apparent in a single view. For instance, displaying both model confidence and historical trends side by side can help users detect anomalies even if one element is compromised.

ISSN: 3071-561X
PP: 01-10.

4.3. Explainability without Oversimplification

While explainability is essential, overly simplified or "one-size-fits-all" explanations can be misleading or even dangerous. Interfaces should present layered explanations, allowing users to drill down from high-level summaries to granular technical details. The system should also make it clear when it is uncertain or when multiple interpretations exist. This helps avoid the trap of false clarity, where users trust explanations because they seem straightforward, not because they are accurate or complete.

4.4. Resisting Spoofed or Manipulated Data

Interfaces must be able to detect and flag potentially manipulated inputs before they are visualized or acted upon. This includes checking for adversarial perturbations, log injection attempts, or anomalous metadata. Systems should alert users when the data source is untrusted or when the integrity of input data is in doubt. Moreover, visual indicators (e.g., warning badges or degraded trust ratings) can help ensure users are not blindly influenced by spoofed artifacts.

4.5. User Attention Guidance (Mitigating Bias)

Interface design can be used proactively to steer users away from known cognitive pitfalls. For instance, rather than emphasizing the first alert in a list, interfaces can randomize display order or highlight statistically unusual items. To counter alert fatigue, smart prioritization and summarization techniques can be used, such as collapsing repetitive alerts or offering digestible threat summaries. The design should nudge users to consider alternative hypotheses, promoting more balanced and less biased decisions.

4.6. Secure Interaction Patterns

Finally, interactions with the AI system—such as approving actions, submitting feedback, or flagging alerts—should be secure by design. This means enforcing authentication, auditing actions, and preventing UI spoofing or manipulation. Any user input that can influence the model or workflow should be validated and monitored to prevent exploitation. Security controls should be baked into the interface layer itself, not just into the underlying systems.

5. TAXONOMY OF UX-LEVEL ATTACKS IN AI-SUPPORTED CYBERSECURITY

To systematically understand and defend against UX-level threats, it is useful to classify them into a taxonomy. This taxonomy outlines the nature, goals, and impacts of these attacks, providing a framework for risk analysis and mitigation.

5.1. Classification: Visual, Interactive, Cognitive, Hybrid

- Visual attacks involve manipulation of charts, colors, icons, and spatial arrangement to deceive users such as hiding alerts in visual clutter or using misleading color schemes.
- Interactive attacks exploit user input paths, default actions, and clickable elements e.g., manipulating interface behavior to encourage risky decisions.
- Cognitive attacks target psychological weaknesses using alert flooding to induce fatigue or explanations designed to confirm user biases.
- Hybrid attacks combine two or more of the above, orchestrating complex campaigns that mislead both the AI and the human in tandem.

ISSN: 3071-561X
PP: 01-10.

5.2. Adversary Models and Objectives

Adversaries in this context may range from casual hackers to advanced persistent threats (APTs). Their objectives include avoiding detection, eroding trust in AI systems, misleading analysts into false decisions, or triggering inappropriate automated responses. Some attackers aim to exploit the trust users place in AI systems and interfaces by injecting deceptive inputs that produce convincing but incorrect outputs. Others seek to exhaust analysts, increasing the probability of oversight or error.

5.3. Impact Severity and Exploitability Assessment

Attacks can be assessed based on their severity (how much damage they can cause) and exploitability (how easy they are to carry out). Visual manipulations may have low exploitability but high severity if they lead to missed intrusions. Cognitive attacks, like exploiting fatigue or anchoring, may be highly exploitable in fast-paced environments. Hybrid attacks though complex to design can be particularly dangerous due to their multi-layered impact. This assessment helps prioritize which UX vulnerabilities to address first.

6. CASE STUDIES

6.1. Case 1: Phishing Email Triage System

In this case study, we analyze an AI-assisted email triage system used to flag phishing attempts. Attackers exploit the system by crafting emails that appear ambiguous enough to bypass detection but also trigger confident “benign” classifications. The interface shows a high-confidence label along with an oversimplified explanation—e.g., “No threat detected: known sender.” However, upon closer inspection, the email metadata is forged, and the content contains a cleverly obfuscated malicious link. The UX fails to present enough contextual detail for the analyst to question the AI output. A resilient redesign includes explainability layers showing sender history anomalies, header mismatches, and visual cues indicating spoofing risks. It also adds a "verify before action" workflow that prompts user review for borderline classifications.

6.2. Case 2: Network Anomaly Detection Dashboard

This case focuses on a dashboard that visualizes traffic anomalies across a corporate network. Over time, a subtle drift in the underlying model leads to a shift in baseline behavior, causing both false positives and missed detections. However, the interface continues to present visualizations and confidence scores as if nothing changed. Analysts begin ignoring certain alerts due to repeated false positives. The trust in the AI output erodes, but without clear indicators of the model’s uncertainty or performance drift, analysts lack insight to recalibrate their decision-making. A resilient redesign incorporates model drift indicators, confidence trend graphs, and interactive alert grouping that allows analysts to flag and annotate recurring issues. This supports better human-AI trust calibration and more accurate response behavior.

7. EVALUATION METHODOLOGIES

Evaluating the adversarial resilience of UX in AI-supported cybersecurity systems demands multifaceted methodologies that test both the technical robustness and the human interpretability of the interface. One core strategy is simulated red teaming of UX, where security professionals or automated agents emulate attackers attempting to exploit weaknesses in the human-AI interface—manipulating visual elements, explanation logic, or alert presentation to mislead users. This allows researchers to observe how well the UX design withstands intentional deception. Human-in-the-loop adversarial testing complements this by embedding real users into attack simulations to study their

ISSN: 3071-561X
PP: 01-10.

decisions under manipulated interfaces, thus capturing how biases and confusion are induced in practice. Another important method is conducting user studies for resilience validation, which may include controlled experiments, A/B testing, or qualitative usability sessions to assess how design changes impact user behavior under adversarial stress. These studies help identify which UX designs foster more resilient decision-making. Evaluation also requires quantifiable metrics for adversarial resilience in UX, such as task accuracy under attack, time to detect deception, confidence calibration, and user trust variation. These metrics provide concrete evidence for the effectiveness of proposed UX defenses and can guide future design improvements.

8. DISCUSSION

In reflecting on this research, it is essential to acknowledge the limitations and ethical concerns involved. Simulating adversarial attacks on users raises ethical considerations around deception, informed consent, and potential cognitive harm. Moreover, the proposed methods may not fully replicate real-world threat complexity. Integrating these adversarially resilient UX principles into existing cybersecurity design pipelines also presents challenges, as current development lifecycles rarely accommodate iterative human factors testing at the interface level. Additionally, trade-offs between usability and security must be navigated carefully: interfaces that are too secure may frustrate users or slow down response time, while overly streamlined UIs may hide complexity and lead to overtrust. Finding a balance between these competing needs is an ongoing design tension that requires careful consideration.

9. FUTURE WORK

Future research should prioritize formalizing adversarial UX threat models—structured frameworks that describe how attackers can target human-AI interactions, including their goals, strategies, and likely outcomes. Such formal models would enable more consistent testing and defense development. Another promising direction is the creation of AI-augmented UX design tools that can automatically identify or suggest interface changes to improve security, such as adaptive alert prioritization, trust calibration features, or explanation visualizations optimized for human cognition. Finally, this work has the potential for broader application to other high-stakes AI systems, such as those used in healthcare, criminal justice, or autonomous systems, where the interface plays a critical role in life-altering or ethically sensitive decisions.

10. CONCLUSION

This paper has examined a largely overlooked dimension of AI security: the vulnerability of user experience (UX) to adversarial manipulation in human-AI cybersecurity systems. While much attention has been given to protecting machine learning models and securing data pipelines, the interface that bridges human decision-makers and automated outputs remains a critical yet underexplored attack surface. Through a comprehensive survey of adversarial UX threats—including misleading visualizations, biased explanations, manipulable confidence indicators, and cognitive exploits like alert fatigue—this work demonstrates how attackers can indirectly undermine system integrity by targeting human perception and trust. The paper further presents a structured threat model, evaluation methodologies, and design principles for building interfaces that are resilient not only at the algorithmic level but also at the cognitive and behavioral levels. Techniques such as red teaming UX, human-in-the-loop testing, and adversarial resilience metrics enable rigorous testing of these principles in practice. This exploration also acknowledges inherent trade-offs between security and usability and outlines future paths toward AI-assisted secure UX design. Ultimately, the findings underscore a critical paradigm shift: UX should be treated not just as a matter of usability but as a

ISSN: 3071-561X
PP: 01-10.

first-class security concern. As AI systems continue to mediate complex decisions in cybersecurity and beyond, ensuring that interfaces empower users, resist deception, and foster well-calibrated trust is paramount. The paper closes with a call to action for designers, developers, and security professionals alike to expand their threat models and testing frameworks to include the human-AI interface as a potential target. By integrating adversarial resilience into UX from the ground up, we can move toward AI systems that are not only technically robust but also secure in how they guide, inform, and support human decision-making under uncertainty and pressure.

REFERENCES

[1] Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. In Proceedings of the International Conference on Learning Representations (ICLR).

[2] Kurakin, A., Goodfellow, I., & Bengio, S. (2017). Adversarial machine learning at scale. Proceedings of the International Conference on Learning Representations (ICLR).

[3] Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition, 84, 317–331.

[4] Grosse, K., Manoharan, P., Papernot, N., Backes, M., & McDaniel, P. (2017). On the (statistical) detection of adversarial examples. arXiv preprint arXiv:1702.06280.

[5] Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. Proceedings of the IEEE Symposium on Security and Privacy, 305–316.

[6] Sadeh, N., Hong, J., Cranor, L., Fette, I., Kelley, P., Prabaker, M., & Rao, J. (2009). Understanding and capturing people's privacy policies in a mobile social networking application. Personal and Ubiquitous Computing, 13(6), 401–412.

[7] Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 37(1), 32–64.

[8] Egelman, S., Cranor, L. F., & Hong, J. (2008). You've been warned: An empirical study of the effectiveness of web browser phishing warnings. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1065–1074.