AI Security Continuum: Concept and Challenges

Hironori Washizaki
Waseda University, Tokyo,
Japan
washizaki@waseda.jp

Nobukazu Yoshioka
Waseda University, Tokyo
Japan
nobukazu@engineerable.ai

ABSTRACT

We propose a conceptual framework, named "AI Security Continuum," consisting of dimensions to deal with challenges of the breadth of the AI security risk sustainably and systematically under the emerging context of the computing continuum as well as continuous engineering. The dimensions identified are the continuum in the AI computing environment, the continuum in technical activities for AI, the continuum in layers in the overall architecture, including AI, the level of AI automation, and the level of AI security measures. We also prospect an engineering foundation that can efficiently and effectively raise each dimension.

KEYWORDS

AI Security, Software Engineering for AI and Machine Learning, Metamodel, Security Risk Management

ACM Reference Format:
Hironori Washizaki and Nobukazu Yoshioka. 2024. AI Security Continuum: Concept and Challenges. In *Conference on AI Engineering Software Engineering for AI (CAIN 2024), April 14–15, 2024, Lisbon, Portugal.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3644815.3644983

1 INTRODUCTION

With the advancement of devices and cloud computing, the AI computing continuum has been realized from edge AI and cloud AI to federated learning and AI ecosystems. The flip side of it is the seamless provision of advanced and adaptive services based on AI and machine learning (ML), which creates a broad attack surface in devices and layers, where the non-determinism and uncertainty of AI and ML can be exploited, and attacks can continue from anywhere and at any time, and the impact can spread across layers.

This paper first describes the challenges and related work for addressing the breadth of such AI security risk. We then proposes the "AI Security Continuum" as a comprehensive framework consisting of dimensions necessary to address AI security risk. Furthermore, we prospect its engineering foundation.

2 RELATED WORK AND CHALLENGES

Although security guidelines for AI/ML systems [1–3] are useful to specify possible attacks and defenses, connections to higher levels, such as business activities and society, have not yet been addressed well. There are technical documents and toolkits to support AI risk analysis [4–8]; however, none of them systematically organized a series of security measures over layers and computing environments, making it difficult for system developers and service providers to implement comprehensive AI security measures.

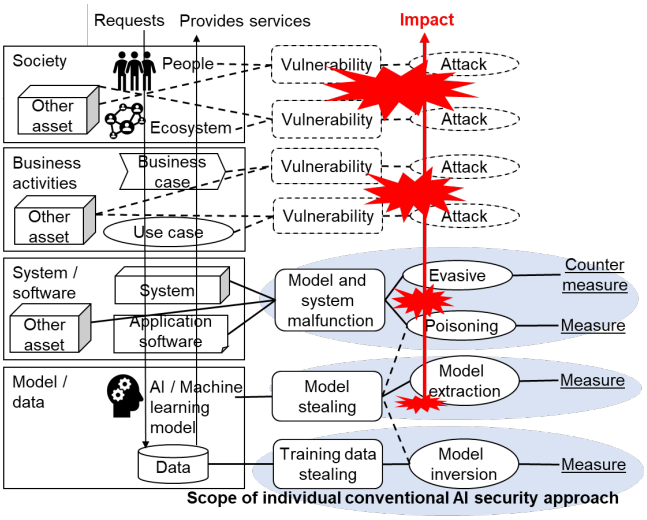

Figure 1: Layers and AI security concerns

Fig 1 shows examples of protection assets, attacks, and countermeasures at each layer. The impact of attacks is often propagated and amplified from lower to higher layers; however, no coordinated risk analysis and countermeasure efforts across layers have been obtained. Conventional efforts focus on specific risk analysis and countermeasures in individual computing environments and lack mechanisms for linking and reusing countermeasures across environments. Incremental and adaptive security engineering processes have to be supported more. Furthermore, many conventional efforts are reactive in response to security attacks or problems, with limited preventive and predictive measures. These immature supports for AI security risk management undermine the trustworthiness of AI systems, limiting the extent of AI-supported automation.

3 AI SECURITY CONTINUUM

We have identified the following five dimensions (1)–(5) as necessary to address the AI security risk continuum sustainably and systematically. We name the comprehensive framework consisting of these dimensions the "AI Security Continuum," shown in Fig 2, as a holistic view over the aforementioned problem.

(1) AI layer continuum is the continuum of layers, ranging from data and AI/ML models to software applications and systems, business activities, and social. AI security risk management efforts need to address all layers holistically and coordinately across layers to address the propagation and amplification of threats and attacks.

(2) AI computing continuum is the continuum of AI computing environments, ranging from individual mobile and edge devices

CAIN 2024, April 14–15, 2024, Lisbon, Portugal

ACM ISBN 979-8-4007-0591-5/24/04
https://doi.org/10.1145/3644815.3644983

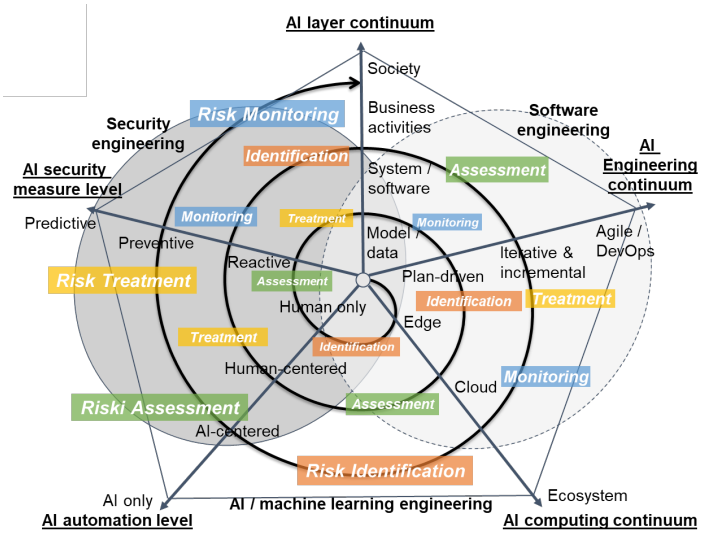

Figure 2: Dimensions of AI security continuum

to fog/cloud, distributed federated learning, and AI ecosystems. Efforts need to identify commonalities across environments through abstraction and knowledge management and provide mechanisms for linkage and reuse of countermeasures over environments.

(3) AI engineering continuum is about continuous engineering processes ranging from plan-driven processes to iterative and incremental, and exploratory and adaptive Agile/DevOps ones. Iterative and Agile/DevOps processes are desirable to continuously verify and modify hypotheses, gradually fleshing them out from vague requirements and situations due to the uncertainty of AI.

(4) AI security measure level is the level of security measures ranging from reactive to preventive, which incorporates basic countermeasure mechanisms based on assumptions of possible attacks and problems, and predictive, which identifies specific future problematic situations and applies specific countermeasures in advance. Preventive and predictive measures are desirable to keep risk to a minimum while incorporating AI uncertainties and possible changes.

(5) AI automation level is the level of AI automation, ranging from almost entirely human-controlled in behavior and decision-making to human-centered with AI support, AI-centered with human intervention, and almost wholly AI-controlled [9]. Higher is better to increase the efficiency and effectiveness of social and business activities while considering higher AI risks.

Based on this framework, the degree to manage AI security risks can be analyzed from multiple perspectives. Engineers and organizations can be guided from the lowest state (i.e., reactive security measures in a plan-driven process in the supportive use of AI on a device-by-device basis) to the highest state (i.e., predictive security measures in the continuous engineering process handling business-to-society impacts through the entire incorporation of AI in distributed systems and ecosystems).

4 EXPECTED ENGINEERING FOUNDATION

An engineering foundation is needed to efficiently and effectively raise each dimension of the AI security continuum. Expected features of the foundation can include the following (a)–(e):

(a) Handling assets, attacks, and corresponding countermeasures over layers with traceability and consistency: This can be supported by our achievements in modeling and ML workflow pipeline integration in MLOps contexts [10] based on metamodels [11, 12].

(b) Security knowledge base to reuse the results and processes of past and ongoing good design, maintenance, and evolution of AI systems over environments: This can be supported by our achievements in ML software engineering patterns [13], security patterns [14], and model-driven security engineering methods [15].

(c) Security verification and repair for AI models and code to identify vulnerabilities and fix them while incorporating the knowledge base: This can be supported by our achievements in code repair and refactoring approaches [16, 17].

(d) Analyzing security and ethical risks for organizations and societies in which AI systems operate and continuously reflecting them in the requirements, design, and revision: This can be supported by our achievements in AI and business alignment approaches [18].

(e) Integration and continuous evolution of all of these activities and artifacts in a consistent and change-responsive manner

5 CONCLUSION AND FUTURE WORK

We proposed the AI security continuum consisting of dimensions necessary to address the broad AI security risk sustainably and systematically and envisioned a desired engineering foundation to manage them. We plan to establish the foundation via multidisciplinary approaches over software, AI/ML, and security engineering and evaluate it through real-world case studies.

ACKNOWLEDGEMENT

This work was supported by JST-Mirai JPMJMI20B8, JSPS KAKENHI 21KK0179 and 23K18470.

REFERENCES

[1] AIST, “Machine Learning Quality Management Guideline”, https://www.digiarc.aist.go.jp/en/publication/aiqm/, 2023.
[2] QA4AI Consortium, “QA4AI Guideline”, https://www.qa4ai.jp/, 2022.
[3] MLSE, “Machine Learning System Security Guidelines”, https://github.com/mlse-jssst/security-guideline, 2023.
[4] MITRE, “ATLAS”, https://atlas.mitre.org/.
[5] The European Union Agency for Cybersecurity, “Artificial Intelligence Cybersecurity Challenges”, https://www.enisa.europa.eu/publications/artificial-intelligence-cybersecurity-challenges.
[6] Microsoft, “Threat Modeling AI/ML Systems and Dependencies”, https://learn.microsoft.com/en-us/security/engineering/threat-modeling-aiml.
[7] ICO, “AI and data protection risk mitigation and management toolkit”, https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ai-and-data-protection-risk-mitigation-and-management-toolkit/.
[8] NIST, “NIST IR8269: A Taxonomy and Terminology of Adversarial Machine Learning”, https://csrc.nist.gov/publications/detail/nistir/8269/draft.
[9] R. Feldt, et al., “Ways of applying artificial intelligence in software engineering,” RAISE 2018
[10] J. Runpakprakun, et al., “Towards Integrated Model-Based Machine Learning Experimentation Framework,” DSA 2023
[11] J. Husen, et al., “Metamodel-Based Multi-View Modeling Framework for Machine Learning Systems,” MODELSWARD 2023
[12] T. Xia, et al, “Cloud Security and Privacy Metamodel: Metamodel for Security and Privacy Knowledge in Cloud Services,” MODELSWARD 2018
[13] H. Washizaki, et al., “Software Engineering Design Patterns for Machine Learning Applications,” IEEE Computer 55(3) 2022
[14] E. Fernandez, et al., “Abstract security patterns and the design of secure systems,” Cybersecurity 5(7) 2022
[15] T. Kobashi, et al., “Validating Security Design Pattern Applications by Testing Design Models,” IJSSE 5(4) 2014
[16] R. Ishizue, et al., “Improvement in Program Repair Methods using Refactoring with GPT Models,” ACM SIGCSE 2024
[17] H. Washizaki, et al., “A Technique for Automatic Component Extraction from Object-Oriented Programs by Refactoring,” SCP 56(1-2) 2005
[18] H. Takeuchi, et al., “Enterprise Architecture-based Metamodel for a Holistic Business - IT Alignment View on Machine Learning Projects,” ICEBE 2023