EasyChair Preprint

№ 15301

Elevation of MLsec: a Security Card Game for Threat Modeling Machine Learning Systems

Elias Brattli Sørensen

October 25, 2024

Elevation of MLsec: a security card game for threat modeling machine learning systems

Elias Brattli Sørensen[1]

Kantega AS https://kantega.no elias.sorensen@kantega.no

Abstract. Machine learning based systems have gained a massive adoption the last few years. These systems bring with them inherent risks that should be mitigated with proactive security engineering. Threat modeling and risk analysis can play an important role in the efforts for securing machine learning. I propose Elevation of MLsec, which is a threat modeling card game targeting machine learning systems. This paper gives a brief overview of the the objectives, design considerations and testing experiences during the creation of the game.

Keywords: Threat modeling · Machine learning security · Serious games.

1 Introduction

Artificial intelligence (AI) is in a new spring, with machine learning (ML) through deep neural networks driving the most notable innovations in the last few years. Compared to traditional software, ML poses different security and privacy challenges due to factors like their statistical nature and the way data are represented internally. These challenges should be met with proactive security engineering, to design systems that are secure, robust and resilient [5]. Threat modeling is a well-established method for proactively reducing the security risk of software [1]. Security games like Protection Poker [3] and Elevation of Privilege (EoP) [4] have played a role in making threat modeling more accessible to developers. However, these games follow traditional risk frameworks that are oriented around software programs. Machine learning introduces a different risk landscape that necessitates frameworks like the architectural risk analysis published by Berryville Institute of Machine Learning (BIML) [5]. Elevation of MLsec is a machine learning security extension of EoP that adopts the risk framework introduced by BIML into a card game. This paper explains the objectives, design process and test results from the creation of Elevation of MLsec.

2 Related Work

Security games are security-oriented serious games, often designed to facilitate threat modeling. Introduced by Laurie Williams [3], Protection Poker has served as a spearhead for other security games. Elevation of Privilege by Shostack [4]has acted as a natural follower. Tøndel et al. have studied the reception of both

Protection Poker [7], and EoP [8], and find that they lead to discussions on security, as well as increased security awareness and knowledge in the team, though it is unclear how it affects actual security outcomes in a project.

Other threat modeling tools come in the form of cards, though they do not have an explicit gaming mechanism. LINDDUN GO [9] is a privacy-oriented threat library that promotes privacy by design. PLOT4AI[1] is a LINDDUN-inspired threat modeling library that targets machine learning systems. The cards can be played as a game, but the gaming concept is not the primary focus.

3 Elevation of MLsec

Elevation of MLsec (EoML) is a machine learning security (MLsec) extension of the threa modeling game EoP [4], suitable for 3-6 players. The contents on the cards of EoML are based on the risks in the framework published by BIML [5]. EoML is licensed under a CC BY-SA 4.0[2] and publicly available for download, as well as commercially available for purchase of printed decks. The creation of the game started as a learning project to better understand machine learning security. The purpose of EoML is to serve as a medium for awareness and learning about machine learning security, grounded in the BIML risk framework. The game offers a catalog of 48 risks that can be used to identify specific security risks in a machine learning based system, and four wildcards that target player creativity. EoML targets identification of risks, and does not offer controls.

To make the BIML-78 risk framework [5] apply for a standard 52-card deck, its components have been projected into four risk categories: Dataset risk, Input risk, Model risk and Output risk (DIMO) as shown in Figure 1. In DIMO the suits are derived from the four *objects* in BIML: 3. datasets, 6. inputs, 7. model, and 9. outputs. The ovals in BIML components 2, 4, 5 and 8 (and the polygon in 1) are processes or data pools that form *interfaces* between the objects. Usually the risks about the system as a whole can also be isolated to one component. Some risks from BIML's LLM publication [6] and OWASP top 10 for LLMs[3] are also included.

After completing the deck, a play-test of was conducted on 39 first to third year Norwegian university IT students who were invited to an afternoon workshop. After a short introduction, students were divided into five groups consisting of 7-8 students, and played EoML for an hour. After playing 26 participants gave open-ended responses to their game experience through an anonymous Mentimeter[4] poll, as shown in Table 1.

[1] https://plot4.ai/
[2] https://creativecommons.org/licenses/by-sa/4.0/
[3] https://mltop10.info/
[4] https://www.mentimeter.com/

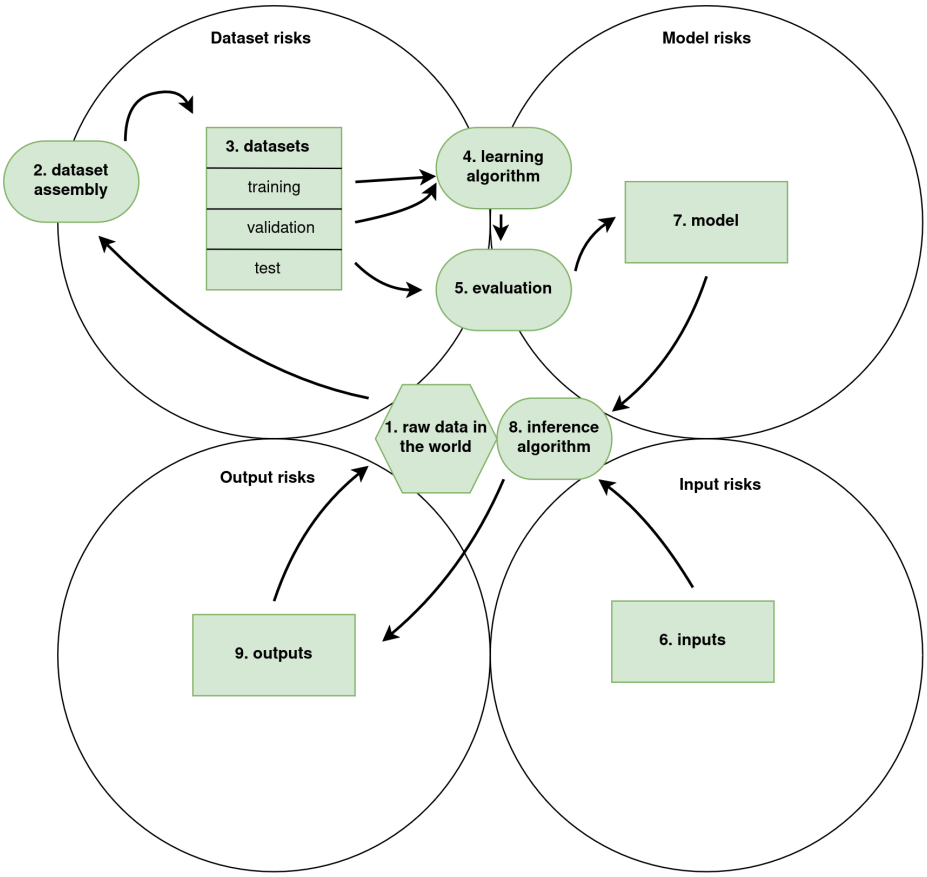

Fig. 1. The Dataset, Input, Model, Output (DIMO) risk framework illustrated as a derivative of the BIML model for a generic ML lifecycle [5]. The 9 components of BIML are mapped into the four DIMO components. Then tenth (system-wide) category is freely mapped into DIMO.

Table 1. Feedback categorized – count in parentheses for the number of responses matching the category of feedback.

Positive feedback (count)	**Constructive feedback (count)**
The game was instructive (6)	The rules were not well enough explained (6)
The game/topic was interesting (5)	The game should have a clearer competitive aspect (2)
The game was fun and /or engaging (5)	It was unclear whether the value on the cards was related to risk complexity (1)
The workshop was good (4)	Too much information in a short period of time (1)
	Hard to understand how arguing about risks affects points (1)
	The cards should be translated (1)

4 Discussion and conclusions

I propose Elevation of MLsec as a novel machine learning security adaption of the threat modeling card game EoP. The level of engagement observed during the play-testing with university students shows promise for EoML as a training tool, while the constructive feedback also shows room for improvement. Testing on university students cannot necessarily be generalized to industry practitioners or real-world projects. More play testing is warranted to get more insight into the game's performance in the real world.

Future work includes play testing on machine learning practitioners and security practitioners in real industry projects. More can be learned about the reception of EoML by repeating the study that Tøndel et al. [8] conducted on EoP with EoML as the variable. An advantage of sharing the game design with EoP is the ability to play the games together by mixing the decks. An ML system is unlikely to live alone, but will instead often be part of a bigger system where their risks are mixed. An interesting research subject is to study the use of games like EoP together with EoML to reason about a system with both ML components and other software component.

References

1. Shostack, A.: "Experiences Threat Modeling at Microsoft". In: Modeling Security Workshop In Association with MODELS '08 (2008).
2. McGraw, G.: "Software Security: Building Security In". In: Addison Wesley (2006).
3. Williams, L., Meneely, A. and Shipley, G.: "Protection Poker: The New Software Security 'Game'". In: IEEE Security & Privacy, vol. 8, no. 3, pp. 14-20, May-June (2010).
4. Shostack, A.: "Elevation of privilege: Drawing developers into threat modeling". In: USENIX Summit on Gaming, Games, and Gamification in Security Education (3GSE 14) (2014).
5. McGraw, G., Figueroa, H., Shepardson, V., & Bonett, R.: "An architectural risk analysis of machine learning systems: Toward more secure machine learning". Berryville Institute of Machine Learning (BIML), Clarke County, VA (2020).
6. McGraw, G., Figueroa, H., McMahon, K., & Bonett, R.: "An architectural risk analysis of large language models: Applied machine learning security". Berryville Institue of Machine Learning (BIML), Berryville, VA, USA, Tech. Rep (2024).
7. Tøndel, I. A., Jaatun, M. G., Cruzes, D., and Oyetoyan, T. D.: "Understanding challenges to adoption of the Protection Poker software security game". In: Computer Security: ESORICS 2018 International Workshops, CyberICPS 2018 and SECPRE 2018, Barcelona, Spain, September 6–7, 2018, Revised Selected Papers 2. Springer International Publishing (2019).
8. Tøndel, I. A., Oyetoyan, T. D., Jaatun, M. G., & Cruzes, D.: "Understanding challenges to adoption of the Microsoft Elevation of Privilege game." Proceedings of the 5th Annual Symposium and Bootcamp on Hot Topics in the Science of Security (2018).
9. Wuyts, K., Sion, L., and Joosen, W.: LINDDUN GO: A Lightweight Approach to Privacy Threat Modeling: In: IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), Genoa, Italy, pp. 302-309 (2020).