

The research problem & the research questions

Prepared by

ALI MMADI, PhD student - Midocean University



Article: Li, Li (2024).

Comprehensive Survey on Adversarial Examples in Cybersecurity:
Impacts, Challenges, and Mitigation Strategies. *arXiv, December 16, 2024.*

INTRODUCTION

In this video, I will present Assignment 3 of our critical analysis module of a scientific article. The objective of this assignment is to evaluate the research problem as well as the research questions in a scientific article. I have chosen a recent article titled: Comprehensive Survey on Adversarial Examples in Cybersecurity: Impacts, Challenges, and Mitigation Strategies, published in 2024 by Li Li on the scientific platform arXiv.



Research problem

Problem Statement Context:

The article highlights that artificial intelligence (AI), particularly deep learning, is increasingly used to detect threats in cybersecurity (malware, intrusions, secure authentication, etc.).

Identified Problem: These deep learning models are vulnerable to adversarial examples, which can deceive them and cause false detections (either positive or negative). **Importance:** This calls into question the reliability and security of DL systems in real-world environments.



✓ **Positive points:**

The problem is well identified: the vulnerability of DL models to adversarial attacks jeopardizes cybersecurity. It clearly addresses a current issue, essential for the reliability of automated systems.

✗ **Limitations:**

The article could have been more explicit by formulating a central question such as: To what extent do adversarial examples impact the detection of threats in cybersecurity? The problem remains formulated in a general way, without targeting a specific sub-domain (e.g. network intrusion, IoT, malware).

□ **Suggestion:** Propose a more structured version: How do adversarial examples reduce the effectiveness of DL systems in detecting malware, network intrusions, or AI-assisted attacks, and what solutions can strengthen their robustness?

Research questions

Implicits

The article explores several important aspects:

Who carries out these adversarial attacks?

In what forms (malware, botnet, intrusion, encryption, authentication...)?

What impacts (false alerts, bypassing systems)?

What defense methodologies (adversarial training, gradient masking, detection, etc.) are effective?

✓ **Positive points:**

The article covers many applicable areas (six chapters dedicated to different types of attacks).It presents techniques and methods, which enriches the analysis.

✗ **Limitations:**

No research question is clearly formulated (in question form) in the introduction or at the end of the introduction.The absence of a precise list limits the readability of the main thread.

□ **Suggestion:**

Explicitly include questions such as:

- 1) What are the most common types of adversarial attacks in cybersecurity?
- 2) How do these attacks concretely affect DL performance (accuracy, false negatives)?
- 3) What defense strategies improve the robustness of DL models?

CONCLUSION

Li Li's article (2024) identifies a crucial and current issue: the threat of adversarial examples on deep learning systems in cybersecurity. It explores various applications and proposes defense strategies. However, to enhance its scientific value: A clear formulation of the problem in the form of a question is missing. The research questions are implicit, lacking explicit structuring. A more rigorous presentation, with a clear problem statement and a set of precise questions, would improve the coherence and academic impact of the article. This work demonstrates the importance of a well-defined research framework to guide readers and strengthen the rigor of the study.

Main reference of the analyzed article and link

- 1) Li, L. (2024).
Comprehensive Survey on Adversarial Examples in Cybersecurity: Impacts, Challenges, and Mitigation Strategies.
arXiv preprint arXiv:2412.12217.
- 2) Link
<https://arxiv.org/abs/2412.12217>

Other references cited in the article (important excerpts)

- 1) **Goodfellow, I., Shlens, J., & Szegedy, C. (2015).**
Explaining and harnessing adversarial examples.
In *International Conference on Learning Representations (ICLR)*.
- 2) **Papernot, N., McDaniel, P., Wu, X., Jha, S., & Swami, A. (2016).**
Distillation as a defense to adversarial perturbations against deep neural networks.
In *IEEE Symposium on Security and Privacy (SP)*.
- 3) **Kurakin, A., Goodfellow, I., & Bengio, S. (2017).**
Adversarial examples in the physical world.
arXiv preprint arXiv:1607.02533.
- 4) **Biggio, B., & Roli, F. (2018).**
Wild patterns: Ten years after the rise of adversarial machine learning.
Pattern Recognition, 84, 317–331.



مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY



Thank You

