Towards Adaptive Adversarial Robustness: A Multi-Armed Bandit Perspective

Nashrah Haque[1]
[1]Graduate School of Arts and Sciences
Fordham University
New York, NY
nhaque14@fordham.edu

Abstract

This work introduces ARBO, a conceptual framework for dynamically adapting defense strategies against adversarial attacks. ARBO balances exploration of novel mechanisms with exploitation of established approaches, using hybrid methods like Exp3 and adversarial-specific techniques based on Wasserstein distances to identify high-uncertainty regions. The framework employs a multi-objective reward model to optimize clean-data accuracy, robustness, and efficiency, while leveraging contextual bandit frameworks for dynamic defense allocation. Although this work is purely theoretical and only written for review purposes, it provides a foundation for future research and experimentation to validate and refine the proposed ideas.

1 Introduction

Machine learning systems are increasingly being deployed in high-stakes domains such as healthcare, autonomous navigation, and financial decision-making. Despite their transformative potential, these systems remain vulnerable to adversarial attacks, carefully crafted perturbations to input data that can induce incorrect outputs from even state-of-the-art models [12, 13]. Such vulnerabilities raise critical concerns about the reliability and security of machine learning systems, particularly in real-world scenarios where trust and safety are paramount.

Adversarial robustness, a field dedicated to improving the reliability of models under adversarial conditions, has seen substantial progress in recent years. Techniques such as adversarial training [16] and certified robustness [28] have emerged as prominent approaches. However, these methods often rely on static defenses tailored to specific attack strategies. Static approaches, while effective against known threats, struggle to adapt to dynamic and evolving attack patterns, leaving systems vulnerable to adversaries that continuously refine their tactics [1, 7].

The need for adaptive and dynamic defense mechanisms has become increasingly apparent. This motivates the exploration of methodologies that can respond in real-time to evolving threats. The multi-armed bandit (MAB) framework, a well-studied paradigm in sequential decision-making, offers a promising foundation for dynamic adversarial defenses. MAB algorithms are designed to balance exploration—testing new strategies—and exploitation—leveraging known effective strategies—making them inherently suited to address dynamic and adversarial challenges [14]. Despite this potential, the application of MAB frameworks in adversarial robustness remains underexplored.

This paper introduces Adaptive Robust Bandit Optimization (ARBO), a theoretical framework that applies multi-armed bandit (MAB) principles to the domain of adversarial robustness. ARBO is structured around four key components: hybrid exploration strategies, multi-objective reward modeling, contextual adaptation, and multi-agent collaboration. These components are designed to address limitations in existing defenses by dynamically adjusting strategies based on real-time observations.

The contributions of this paper are twofold:

1. **Hybrid Exploration Strategies:** ARBO combines stochastic exploration techniques, such as Exp3, with adversarial-specific methods leveraging Wasserstein distances to target high-uncertainty regions. This novel combination aims to balance the trade-off between exploration and exploitation in adversarial defense contexts.

2. **Multi-Objective Reward Modeling:** The framework introduces a reward model that simultaneously optimizes for clean-data accuracy, adversarial robustness, and computational efficiency, providing a more holistic approach to decision-making under adversarial conditions.

Although this work is purely theoretical, it serves as a conceptual foundation for future research. The proposed framework bridges the gap between deep learning, sequential decision-making, and adversarial robustness, outlining a pathway for integrating dynamic adaptation into adversarial defenses. In doing so, ARBO highlights the opportunities and challenges in leveraging MAB frameworks to enhance the security and reliability of machine learning systems.

This review is organized as follows: Section 2 and 3 provides an overview of the existing literature concerning both adversarial robustness and Multi-Armed-Bandits respectively. Section 4 reviews state-of-the-art research integrating MAB and adversarial robustness. Section 5 presents the ARBO framework, Section 6 discusses future directions and discussion. The paper then concludes with a summary of insights and recommendations for advancing the field.

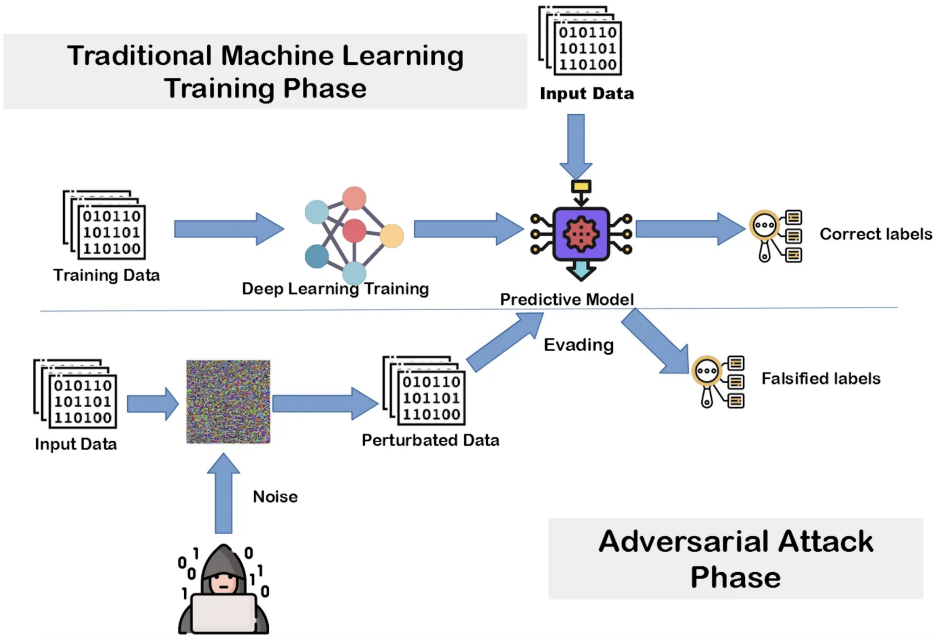

Figure 1: Illustration of adversarial attacks in machine learning. The figure compares the traditional machine learning training and prediction pipeline (top) with the adversarial attack phase (bottom).

Figure 1 provides a comparative visualization of traditional machine learning workflows and adversarial attack scenarios. In the traditional phase, the model learns from clean training data and performs accurate predictions on test inputs. However, in the adversarial attack phase, attacker-crafted noise is added to the input data, creating perturbations that lead the model to make erroneous predictions. This process underscores the need for robust defenses like Adaptive Robust Bandit Optimization (ARBO) to counter such vulnerabilities.

2 Literature Review: Adversarial Robustness

Adversarial robustness is critical for machine learning systems deployed in high-stakes domains such as healthcare, autonomous navigation, and cybersecurity. Adversarial examples, small, imperceptible perturbations of input data, can significantly compromise the reliability of these systems [24, 12]. Adversarial robustness focuses on developing defenses to ensure that machine learning models remain reliable under adversarial conditions. Research has shown that adversarial examples exploit the inherent linear nature of neural networks[12], leading to vulnerabilities that can be efficiently exploited using methods such as gradient-based attacks [16]. This section reviews the challenges of adversarial robustness and provides real-world examples to illustrate their impact.

2.1 Key Challenges

The primary challenges in achieving adversarial robustness are:

1. **Generalization to Unseen Attacks:** Static defenses often fail to counteract adaptive attacks, which evolve based on observed defenses [1].
2. **Robustness-Accuracy Trade-off:** Models optimized for robustness often experience reduced accuracy on clean data.
3. **Scalability and Computational Costs:** Many robust training methods, such as adversarial training, require substantial computational resources to generate adversarial examples during training [20].

2.2 Defense Strategies

Several defense strategies have been proposed to mitigate adversarial attacks. These strategies can be broadly categorized into the following:

2.2.1 Adversarial Training

Adversarial training is the process of augmenting the training dataset with adversarial examples. This approach, first introduced in [12], has been refined over the years to include stronger attacks, such as Projected Gradient Descent (PGD) [16]. While effective against specific attacks, adversarial training suffers from scalability issues and reduced accuracy on clean data.

2.2.2 Gradient Masking

Gradient masking refers to techniques that obscure the gradients of the model to prevent adversaries from crafting effective attacks. However, many gradient masking methods are vulnerable to stronger attacks or adaptive adversaries [1].

2.2.3 Randomization-Based Defenses

Randomization introduces stochasticity into the model or its predictions, making it harder for adversaries to optimize their attacks. While promising in some scenarios, randomization often fails against ensemble or expectation-based attacks.

2.2.4 Certified Robustness

Certified defenses provide mathematical guarantees about a model's robustness within a specified perturbation bound [28]. Techniques such as interval bound propagation and randomized smoothing have demonstrated promising results but remain computationally expensive [10].

2.2.5 Adversarial Detection

Detection-based methods aim to identify adversarial examples before they can affect model predictions. These methods leverage techniques such as statistical anomaly detection or auxiliary networks to flag suspicious inputs [17]. However, robust detection remains an open challenge due to high false-positive rates and adaptive attacks [8].

2.3 Future Challenges

While substantial progress has been made in developing defenses, several open questions remain:

1. **Adaptive Adversaries:** How can defenses evolve dynamically to counteract adaptive attacks in real time?
2. **Scalability:** Can defenses be designed to scale effectively with large models and datasets without compromising performance?
3. **Understanding Robustness:** What are the fundamental limits of adversarial robustness, and how do they vary across tasks and domains?

This overview sets the stage for exploring dynamic defense frameworks, such as multi-armed bandits, as a potential solution to the evolving challenges of adversarial robustness.

3 Literature Review: Multi-Armed Bandit

The multi-armed bandit (MAB) framework is a cornerstone of sequential decision-making and online learning, originating in the field of operations research. The term "multi-armed bandit" refers to a gambler faced with multiple slot machines (or "arms"), each with an unknown probability distribution of rewards. The gambler must decide which arms to pull in a sequence of trials to maximize cumulative rewards over time. The MAB problem epitomizes the exploration-exploitation trade-off: the tension between exploring new arms to learn their rewards and exploiting arms that have yielded high rewards in the past [14].

3.1 Theoretical Foundations of MAB

The MAB problem was first formally studied by Robbins in 1952 [19], with the goal of creating adaptive allocation rules that are optimal over time. Later work by Lai and Robbins [14] provided the first rigorous mathematical formulation of the problem, introducing the concept of regret as a key performance metric.

3.1.1 Regret and Its Variants

Regret measures the performance loss due to not always selecting the optimal arm. It is defined as the difference between the expected reward of an optimal strategy and the cumulative reward of the bandit algorithm. Formally, if the optimal arm has an expected reward μ^* and an algorithm selects arm a_t at time t, the cumulative regret after T rounds is given by:

$$R(T) = T\mu^* - \sum_{t=1}^{T} \mathbb{E}[r_{a_t}],$$

where r_{a_t} is the reward from arm a_t. The goal of a bandit algorithm is to minimize regret as T grows large.

There are two major variants of regret: 1. Cumulative Regret: Standard regret as defined above. 2. Bayesian Regret: Expected regret when the reward distributions are drawn from a known prior.

3.1.2 Optimality and Asymptotic Bounds

Lai and Robbins proved that any algorithm for solving the MAB problem must incur at least logarithmic regret, which is asymptotically lower-bounded by:

$$R(T) \geq \sum_{i:\mu_i<\mu^*} \frac{\log T}{\Delta_i},$$

where $\Delta_i = \mu^* - \mu_i$ is the gap between the mean reward of the optimal arm and suboptimal arm i.

3.2 Key Algorithms in MAB

Numerous algorithms have been developed to solve the MAB problem, each with its own strengths and weaknesses. Below are some of the most prominent:

3.2.1 Upper Confidence Bound (UCB)

The UCB algorithm is one of the most well-known approaches for solving the MAB problem [2]. It selects arms based on the principle of optimism in the face of uncertainty, assigning each arm an upper confidence bound that balances exploration and exploitation. At each time step t, the algorithm selects the arm i that maximizes:

$$\text{UCB}_i(t) = \hat{\mu}_i + \sqrt{\frac{2\log t}{n_i}},$$

where $\hat{\mu}_i$ is the empirical mean reward of arm i, and n_i is the number of times arm i has been selected. UCB achieves logarithmic regret and is computationally efficient.

3.2.2 Thompson Sampling

Thompson Sampling is a Bayesian algorithm that maintains a posterior distribution over the rewards of each arm and selects arms based on sampling from this posterior [25, 9]. At each step, the algorithm: 1. Samples a reward θ_i for each arm i from its posterior distribution. 2. Selects the arm with the highest sampled θ_i.

Thompson Sampling is highly practical and performs well in empirical evaluations, often outperforming UCB in practice.

3.2.3 Exp3 for Adversarial Bandits

The Exp3 algorithm is designed for adversarial settings where reward distributions may change arbitrarily over time [3]. It assigns weights to arms and updates them using exponential weighting:

$$p_i(t) = \frac{e^{\eta S_i}}{\sum_{j=1}^{K} e^{\eta S_j}},$$

where S_i is the cumulative reward of arm i, and η is a learning rate parameter. Exp3 ensures robustness in non-stochastic environments.

3.3 Applications of MAB in Machine Learning and Beyond

The versatility of the MAB framework has led to its adoption across diverse fields, including:

- **Recommendation Systems:** Algorithms like UCB are used to recommend items to users based on past interactions [15].
- **Clinical Trials:** MAB methods are applied to allocate patients to treatments dynamically while balancing exploration of new treatments and exploitation of effective ones [26].
- **Adversarial Robustness:** MAB algorithms are increasingly being explored for adaptive adversarial defenses, allowing systems to allocate resources dynamically to counteract evolving attack strategies [5].

3.4 Challenges in MAB

Despite its success, the MAB framework faces several challenges:

1. **Scalability:** As the number of arms grows, traditional MAB algorithms may become computationally infeasible.
2. **Non-Stationary Environments:** In real-world applications, reward distributions may change over time, requiring algorithms to adapt dynamically [11].
3. **Exploration Costs:** In many applications, exploration incurs a cost, making it important to balance exploration and exploitation effectively.

These challenges motivate ongoing research into more efficient and adaptive bandit algorithms, particularly in the context of dynamic adversarial robustness.

4 State-of-the-Art Research Integrating MAB and Adversarial Robustness

The integration of multi-armed bandits (MAB) with adversarial robustness has gained significant traction in recent years, providing dynamic and adaptive solutions for defense strategies. The foundational work on contextual bandits by Slivkins [22] extended traditional MAB algorithms by incorporating contextual information to guide decision-making. This approach, particularly relevant in adversarial settings, enables the allocation of resources and defenses based on observed attack patterns and input features, highlighting its potential for adaptive adversarial robustness.

Building on this foundation, Zuo et al. [29] explored the application of cooperative multi-agent bandits in adversarial environments. Their study demonstrated how multi-agent setups could effectively address adversarial dynamics through collaborative decision-making. These findings underscore the versatility of MAB frameworks in adapting to complex and evolving threats.

4.0.1 Reinforcement Learning Hybrid Models

Bandit algorithms have been integrated with reinforcement learning (RL) frameworks to develop adaptive policies for adversarial defenses. While the foundational concepts of RL provide the basis for hybrid models, such as combining MAB for dynamic resource allocation and RL for long-term strategy optimization [23], recent work has proposed specialized frameworks to enhance robustness. For instance, the ERNIE framework [6] leverages Lipschitz continuity and adversarial regularization to mitigate sensitivity to environmental changes and adversarial actions. By controlling the policy's Lipschitz constant and reformulating adversarial regularization as a Stackelberg game, this approach ensures stability and robustness in multi-agent reinforcement learning (MARL) systems.

In the context of adversarial multi-armed bandits, existing research has explored algorithms designed to mitigate adversarial effects while maintaining competitive performance. For example, Putta et al. [18] proposed a scale-free approach for adversarial multi-armed bandits that adapts to the scale and magnitude of adversarial losses without requiring prior knowledge of these parameters. Their algorithm employs adaptive learning rates and establishes regret bounds based on loss vector norms, providing significant insights into robust learning strategies under adversarial conditions. These advancements underscore the importance of adaptive mechanisms in adversarial settings, which is a core motivation behind the ARBO framework proposed in this paper.

4.1 Applications in Real-World Scenarios

The application of MAB-based adversarial robustness strategies spans several critical domains:

- **Cybersecurity:** Dynamic allocation frameworks have been explored to optimize defensive strategies in distributed networks under attack scenarios. While specific applications of UCB in this domain remain underexplored, MAB frameworks offer promising avenues for adapting defenses to evolving threats in real-time.
- **Adversarial Training:** Although bandit algorithms are extensively studied in adversarial contexts, their specific use in prioritizing adversarial examples for enhancing model robustness remains largely theoretical and requires further exploration in practical settings.
- **Recommender Systems:** MAB frameworks are widely used to adapt to user preferences and detect anomalous behaviors in recommendation systems. However, applying these methods to mitigate adversarial manipulations of rankings and profiles remains an open research challenge.

5 Identified Gaps and the ARBO Framework

While multi-armed bandit (MAB) frameworks have demonstrated potential in adversarial robustness, significant gaps remain in current research. This section identifies these gaps and introduces **Adaptive Robust Bandit Optimization (ARBO)**, a theoretical framework designed to address challenges in dynamic adversarial defense systems.

1. **Static vs. Dynamic Defenses:** Existing approaches often focus on static frameworks that fail to adapt to evolving adversarial strategies. For instance, Wang et al. [27] explored

stochastic bandits robust to adversarial attacks by introducing regret bounds dependent on attack budgets. However, these methods do not account for rapidly changing adversarial patterns in real-world scenarios.

2. **Reward Modeling Limitations:** Current algorithms often prioritize regret minimization over multi-objective rewards. For example, Sinha et al. [21] proposed Wasserstein-based metrics to enhance robustness, but integration of such metrics with multi-objective reward functions—balancing robustness, accuracy, and computational efficiency—remains underexplored.
3. **Scalability and Collaboration:** Multi-agent systems hold promise for distributed adversarial defenses but present challenges in resource allocation and scalability. While Zuo et al. [29] demonstrated collaborative multi-agent setups in adversarial environments, their methods lack generalizable frameworks for large-scale distributed systems.

5.1 Adaptive Robust Bandit Optimization (ARBO)

This work introduces **ARBO**, a novel framework leveraging advanced MAB methodologies to enhance adversarial robustness in machine learning systems. ARBO aims to dynamically adapt defenses, addressing gaps in static models, reward modeling, and scalability.

5.2 Framework Overview

ARBO proposes a dynamic system that balances exploration of new defense strategies and exploitation of proven methods. Its key components include:

1. **Hybrid Exploration Strategies:** Combines stochastic exploration (e.g., Exp3 [5]) with adversarial-specific methods, such as Wasserstein-based distance metrics, to target high-uncertainty regions.
2. **Multi-Objective Reward Modeling:** Integrates reward functions that account for clean-data accuracy, adversarial robustness, and computational efficiency [21].
3. **Dynamic Contextual Adaptation:** Employs contextual bandits to adapt defense mechanisms based on real-time observations [22].
4. **Multi-Agent Collaboration:** Extends the framework to multi-agent systems, enabling distributed and collaborative defenses, as explored by Zuo et al. [29].

5.3 Proposed Workflow

The ARBO framework operates through the following stages:

1. **Initialization:** Establish a pool of defense strategies and a baseline reward model.
2. **Threat Detection:** Use contextual bandits to analyze input features and detect potential attack patterns.
3. **Defense Allocation:** Dynamically allocate resources to different defense mechanisms based on contextual observations.
4. **Feedback and Learning:** Continuously update the bandit algorithm and reward model based on observed outcomes.

5.4 Future Directions

While ARBO serves as a theoretical contribution, practical validation is necessary to realize its potential. Future work should focus on:

- **Empirical Validation:** Test ARBO on benchmarks such as MNIST, CIFAR-10, and ImageNet.
- **Optimization for Scalability:** Enhance computational efficiency in multi-agent systems.
- **Interdisciplinary Applications:** Integrate methods from optimization, reinforcement learning, and game theory to refine ARBO's adaptability.

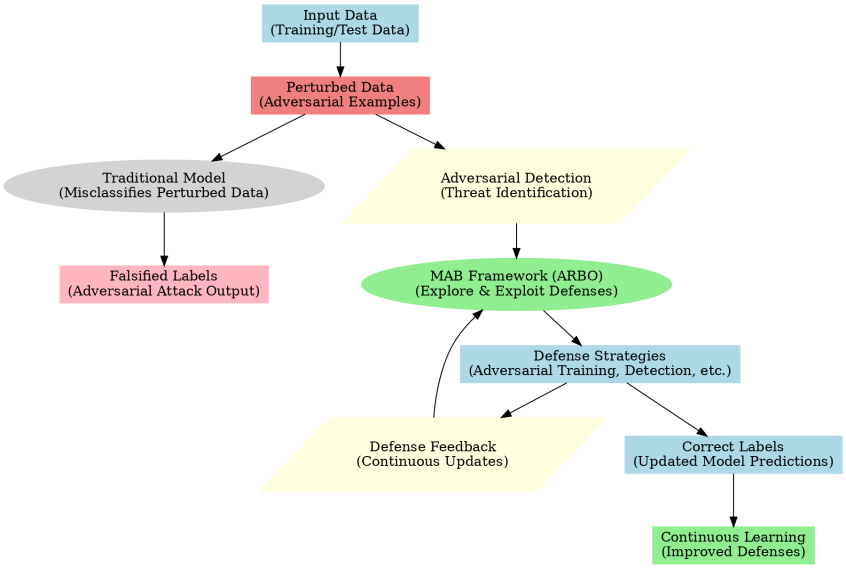

Figure 2: Adaptive Robust Bandit Optimization (ARBO) Pipeline. The diagram illustrates the flow of input data through the ARBO framework, integrating adversarial attack detection, the multi-armed bandit (MAB) decision-making system, and dynamically selected defense strategies.

The flow of the ARBO framework is illustrated in Figure 2. Input data, potentially containing adversarial perturbations, enters the pipeline where adversarial detection mechanisms identify potential threats. The ARBO framework, powered by a multi-armed bandit (MAB) system, dynamically explores and exploits various defense strategies to mitigate these threats. The selected defenses are applied to the data, enabling the model to produce accurate predictions while providing feedback to refine future defenses. This iterative feedback loop ensures continuous learning and adaptation, making ARBO an effective solution for evolving adversarial attacks.

6 Discussion and Future Work

The integration of multi-armed bandit (MAB) frameworks with adversarial robustness represents a frontier in machine learning security. As adversarial threats grow increasingly sophisticated, adaptive, scalable, and theoretically grounded approaches like ARBO (Adaptive Robust Bandit Optimization) are essential for tackling emerging challenges. This section outlines key advancements, ARBO's contributions, and directions for future research.

6.1 Anticipated Technological Advancements

Several advancements in the coming years are expected to influence MAB-based adversarial robustness:

6.1.1 Dynamic Defense Architectures

Future defense systems will prioritize real-time adaptability to adversarial environments. The work of Besbes et al. [4] on non-stationary MAB frameworks demonstrates methods for adapting to shifting reward distributions, providing a theoretical foundation for ARBO's contextual bandit-based dynamic defenses.

6.1.2 Scalability to Large-Scale Systems

Scalability remains a critical challenge as adversarial robustness extends to distributed applications, such as federated learning. While specific federated bandit frameworks are underexplored, ARBO's architecture can lay the groundwork for scalable multi-agent systems by building on distributed bandit algorithms [4].

6.1.3 Reward Optimization and Trade-offs

Balancing adversarial robustness, accuracy, and computational efficiency is essential. Sinha et al. [21] illustrate the utility of Wasserstein-based metrics in optimizing for robustness while aligning with real-world data distributions, a key aspect of ARBO's reward modeling.

6.1.4 Interdisciplinary Integration

Advances in MAB research increasingly integrate methods from optimization, game theory, and reinforcement learning. This trajectory aligns with ARBO's hybrid exploration strategies, which bridge theoretical and applied research.

6.2 ARBO's Role in Shaping Future Research

The ARBO framework addresses critical gaps in MAB research and exemplifies the characteristics required for next-generation adversarial defenses:

- *Dynamic Adaptation:* ARBO integrates hybrid exploration strategies and contextual bandit mechanisms, enabling real-time defenses against adaptive attacks.
- *Scalable and Collaborative Defenses:* With decentralized multi-agent collaboration, ARBO offers a blueprint for robust defenses in distributed and networked systems.
- *Application Flexibility:* ARBO's multi-objective reward modeling ensures adaptability across domains with varying robustness and performance requirements.
- *Blueprint for Future Work:* ARBO provides a theoretical foundation for the development of advanced algorithms and experimental validation frameworks.

6.3 Future Research Directions

While ARBO offers a significant theoretical contribution, the following directions are critical for practical realization:

1. *Experimental Validation:* ARBO must be tested on standard adversarial benchmarks (e.g., MNIST, CIFAR-10, ImageNet) to evaluate its effectiveness in diverse scenarios.
2. *Optimization for Scalability:* Research is needed to enhance ARBO's computational efficiency in large-scale and multi-agent systems.
3. *Reinforcement Learning Integration:* Extending ARBO with reinforcement learning techniques can enable long-term strategic adaptation to adversarial threats.
4. *Ethical and Practical Deployment:* Deployment in sensitive applications, such as healthcare and finance, requires addressing ethical considerations and regulatory compliance.

7 Conclusion

Adversarial robustness, especially through the Adaptive Robust Bandit Optimization (ARBO) framework, represents a critical frontier in ensuring the reliability and security of machine learning systems deployed in sensitive and high-impact domains. These domains include automated healthcare diagnostics, autonomous navigation systems, and algorithmic decision-making in financial markets. This technology is of interest due to the pressing need to counteract evolving adversarial threats, which increasingly exploit the vulnerabilities of static defense mechanisms.

Several key technological aspects of ARBO highlight its value and potential impact:

- **Hybrid Exploration Techniques:** Combining stochastic exploration with adversarial heuristics represents an underexplored area with significant implications for improving defense robustness. Recent advancements in Wasserstein-based metrics have provided a foundation for aligning defenses with real-world distributions, as shown in the work by Sinha et al. [21].

- **Multi-Objective Reward Modeling:** By optimizing for robustness, accuracy, and computational efficiency simultaneously, ARBO addresses the trade-offs critical to large-scale implementations. The use of Wasserstein metrics for robustness ensures effective handling of distributional shifts [21].
- **Dynamic Contextual Adaptation:** Leveraging contextual bandits allows ARBO to adapt defense strategies based on real-time inputs, building on the theoretical framework established by Slivkins [22].
- **Collaborative Multi-Agent Systems:** Distributed defenses in federated environments exemplify ARBO's scalability. However, further investigation is required to identify concrete applications of multi-agent frameworks under adversarial conditions, as this area remains underexplored.

References

[1] Anish Athalye, Nicholas Carlini, and David Wagner. Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples. In *International Conference on Machine Learning*, pages 274–283. PMLR, 2018.

[2] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multi-armed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

[3] Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing*, 32(1):48–77, 2002.

[4] Omar Besbes, Yonatan Gur, and Assaf Zeevi. Stochastic multi-armed bandit problem with non-stationary rewards. *Operations Research*, 62(4):954–970, 2014.

[5] Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. In *Foundations and Trends in Machine Learning*, volume 5, pages 1–122, 2012.

[6] Alexander W. Bukharin, Yan Li, Yue Yu, Qingru Zhang, Zhehui Chen, Simiao Zuo, Chao Zhang, Songan Zhang, and Tuo Zhao. Robust multi-agent reinforcement learning via adversarial regularization: Theoretical foundation and stable algorithms. *arXiv preprint arXiv:2310.10810*, 2023.

[7] Nicholas Carlini, Florian Tramèr, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Luca Melis, Andy Roberts, Dawn Song, Andreas Terzis, et al. On evaluating adversarial robustness. In *IEEE Symposium on Security and Privacy (SP)*, pages 317–332. IEEE, 2019.

[8] Nicholas Carlini and David Wagner. Adversarial examples are not easily detected: Bypassing ten detection methods. *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*, pages 3–14, 2017.

[9] Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. In *Advances in neural information processing systems*, pages 2249–2257, 2011.

[10] Jeremy M Cohen, Elan Rosenfeld, and J Zico Kolter. Certified robustness to adversarial examples via randomized smoothing. *arXiv preprint arXiv:1902.02918*, 2019.

[11] Aurélien Garivier and Olivier Cappé. The kl-ucb algorithm for bounded stochastic bandits and beyond. *arXiv preprint arXiv:1102.2490*, 2011.

[12] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

[13] Alexey Kurakin, Ian Goodfellow, and Samy Bengio. Adversarial examples in the physical world. In *arXiv preprint arXiv:1607.02533*, 2016.

[14] Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.

[15] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. pages 661–670, 2010.

[16] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. In *arXiv preprint arXiv:1706.06083*, 2017.

[17] Jan Hendrik Metzen, Tim Genewein, Volker Fischer, and Bastian Bischoff. On detecting adversarial perturbations. In *International Conference on Learning Representations*, 2017.

[18] Sudeep Raja Putta and Shipra Agrawal. Scale-free adversarial multi armed bandits. In *Proceedings of The 33rd International Conference on Algorithmic Learning Theory*, pages 910–930. PMLR, 2022.

[19] Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952.

[20] Ali Shafahi, Mahyar Najibi, Amir R Ghiasi, Zheng Xu, John P Dickerson, Christoph Studer, Larry S Davis, and Tom Goldstein. Adversarial training for free! *Advances in Neural Information Processing Systems*, 32, 2019.

[21] Aman Sinha, Hongseok Namkoong, and John C Duchi. Certifying some distributional robustness with principled adversarial training. In *International Conference on Learning Representations*, 2018.

[22] Aleksandrs Slivkins. *Introduction to Multi-Armed Bandits and their Applications*. Foundations and Trends in Machine Learning, 2019.

[23] Richard S Sutton and Andrew G Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2018.

[24] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv preprint arXiv:1312.6199*, 2013.

[25] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

[26] Sofia S Villar, Jack Bowden, and James MS Wason. Multi-armed bandit models for the optimal design of clinical trials: Benefits and challenges. *Statistical Science*, 30(2):199–215, 2015.

[27] Xuchuang Wang, Jinhang Zuo, Xutong Liu, John CS Lui, and Mohammad Hajiesmaili. Stochastic bandits robust to adversarial attacks. *arXiv preprint arXiv:2310.05308*, 2024.

[28] Eric Wong and J Zico Kolter. Provable defenses against adversarial examples via the convex outer adversarial polytope. In *International Conference on Machine Learning*, pages 5283–5292. PMLR, 2018.

[29] Jinhang Zuo, Zhiyao Zhang, Xuchuang Wang, Cheng Chen, Shuai Li, John C. S. Lui, M. Hajiesmaili, and Adam Wierman. Adversarial attacks on cooperative multi-agent bandits. *ArXiv*, abs/2311.01698, 2023.