Review Form 3

Journal Name:	**Journal of Engineering Research and Reports**
Manuscript Number:	**Ms_JERR_131309**
Title of the Manuscript:	**Adversarial Threats to AI-Driven Systems: Exploring the Attack Surface of Machine Learning Models and Countermeasures**
Type of the Article	

General guidelines for the Peer Review process:

Artificial Intelligence (AI) generated or assisted review comments are strictly prohibited during peer review.

This journal's peer review policy states that **NO** manuscript should be rejected only on the basis of '**lack of Novelty'**, provided the manuscript is scientifically robust and technically sound.
To know the complete guidelines for the Peer Review process, reviewers are requested to visit this link:

https://r1.reviewerhub.org/general-editorial-policy/

Important Policies Regarding Peer Review

Peer review Comments Approval Policy: https://r1.reviewerhub.org/peer-review-comments-approval-policy/
Benefits for Reviewers: https://r1.reviewerhub.org/benefits-for-reviewers

Review Form 3

PART 1: Comments

	Reviewer's comment **Artificial Intelligence (AI) generated or assisted review comments are strictly prohibited during peer review.**	**Author's Feedback** *(Please correct the manuscript and highlight that part in the manuscript. It is mandatory that authors should write his/her feedback here)*
Please write a few sentences regarding the importance of this manuscript for the scientific community. A minimum of 3-4 sentences may be required for this part.	This manuscript makes a meaningful contribution to the field by addressing the growing security risks in AI systems and adversarial attacks. Reliable performance of machine learning models in important areas such as cybersecurity, healthcare and autonomous vehicles requires understanding their vulnerabilities. This study thoroughly analyzes adversarial attack methods, their effect on model performance and the effectiveness of defenses against them, considerably advancing AI security research. This research constructs a large foundation for considerably more strong AI models by revealing important flaws in existing safeguards as well as proposing revolutionary combined methods, thereby substantially increasing trust and safety in practical real-world AI applications.	
Is the title of the article suitable? (If not please suggest an alternative title)	The title accurately represents the manuscript's content. Wide-ranging research into adversarial attacks against AI systems is a meaningful focus and this area of study deserves important attention. Several minor modifications are necessary and replacing "Exploring the Attack Surface" with "Analyzing the Attack Surface" improves the text's technical accuracy.	
Is the abstract of the article comprehensive? Do you suggest the addition (or deletion) of some points in this section? Please write your suggestions here.	The abstract describes the problem and methodology used in the research. It also presents several key findings and offers specific recommendations. A thorough quantitative investigation, including the CIFAR-10 Adversarial Examples and MITRE ATLAS datasets, shows meaningful success rates of adversarial attacks and this investigation also shows the large effectiveness of several mitigation strategies. To improve the paper, the abstract must clearly present all statistically meaningful findings, such as p-values and confidence intervals. In addition, explaining how some findings relate to practical AI security would improve the overall results. Large proposed improvements might involve specifying many ways the results could considerably affect important AI security applications, including important cybersecurity systems, a collection of autonomous vehicle systems and a large number of financial applications.	
Is the manuscript scientifically, correct? Please write here.	The manuscript provides a strong scientific foundation, employing strict empirical analysis and statistically important evaluations and capitalizing on custom-built AI security datasets such as CIFAR-10 and MITRE ATLAS. This methodology adheres to best practices. These best practices are in adversarial machine learning. Improvements are required in two areas: First, the Chi-Square Goodness-of-Fit Test results need to specify the degrees of freedom, in addition to the exact p-value rather than "$p < 0.001$". Secondly, several limitations of adversarial training, including large overfitting risks and large computational costs, necessitate further discussion.	
Are the references sufficient and recent? If you have suggestions of additional references, please mention them in the review form.	The manuscript cites several recent references (2023–2025), covering AI security studies, adversarial attacks and deep learning robustness analyses, which is also grounded in highly reputable, peer-reviewed sources such as IEEE, Springer, Elsevier and ACM. A few recommendations for area of improvements is: Key foundational papers on FGSM, PGD and C&W attacks from 2014-2019, including Goodfellow et al. (2014), Madry et al. (2017) and Carlini & Wagner (2017), are important to consider. To provide large context for the recent improvements, cite several highly influential classic adversarial attack papers.	

Review Form 3

Is the language/English quality of the article suitable for scholarly communications?	The writing is precise and technically formatted. The only area of improvement is to remove any redundacy phrases like “Given AI’s expanding role in critical sectors, addressing these risks is imperative.” Suggestion is to remove “Given AI’s expanding role…” as it is implied.	
Optional/General comments		

PART 2:

	Reviewer’s comment	**Author’s comment** *(if agreed with reviewer, correct the manuscript and highlight that part in the manuscript. It is mandatory that authors should write his/her feedback here)*
Are there ethical issues in this manuscript?	*(If yes, Kindly please write down the ethical issues here in details)*	

Reviewer Details:

Name:	**Rianat Abbas**
Department, University & Country	**Baylor University, United States**