iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** I **Month of publication:** January 2025

DOI: https://doi.org/10.22214/ijraset.2025.66301

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue I Jan 2025- Available at www.ijraset.com

Enhancing the Robustness of Zero-Shot LLMs Against Adversarial Prompts

Rambarki Sai Akshit[1], Dr. Ravi Bhramaramba[2], P. Madhav[3], G. Sarthak[4], V. Pavan Pranesh[5], K. Nitesh Sai[6]
[1, 2, 3, 4, 5, 6]*Department of Computer Science Engineering, GITAM University*

Abstract: ***Zero-shot large language models (LLMs) have proven highly effective in performing a wide range of tasks without the need for task-specific training, making them versatile tools in natural language processing. However, their susceptibility to adversarial prompts—inputs crafted to exploit inherent weaknesses—raises critical concerns about their reliability and safety in real-world applications. This paper focuses on evaluating the robustness of zero-shot LLMs when exposed to adversarial scenarios. A detailed evaluation framework was developed to systematically identify common vulnerabilities in the models' responses. The study explores mitigation techniques such as adversarial training to improve model resilience, refined prompt engineering to guide the models toward desired outcomes, and logical consistency checks to ensure coherent and ethical responses. Experimental findings reveal substantial gaps in robustness, particularly in handling ambiguous, misleading, or harmful prompts. These results underscore the importance of targeted interventions to address these vulnerabilities. The research provides actionable insights into improving zero-shot LLMs by enhancing their robustness and ensuring ethical adherence. These contributions align with the broader goal of creating safe, reliable, and responsible AI systems that can withstand adversarial manipulation while maintaining their high performance across diverse applications.***
Keywords: Large Language Models (LLMs), Adversarial Prompts, Zero-Shot Learning, Ethical AI, Bias Mitigation

I. INTRODUCTION

Large Language Models (LLMs), such as GPT-3.5-turbo and GPT-4, have transformed the field of artificial intelligence by excelling at a wide range of tasks without requiring task-specific training. These models, especially in zero-shot settings, can understand and generate human-like text across various domains, from answering questions to summarizing complex information. Zero-shot learning allows LLMs to perform tasks they haven't been explicitly trained on, making them highly versatile.

However, despite their capabilities, LLMs are not flawless. One significant concern is their vulnerability to adversarial prompts—inputs intentionally crafted to trick or confuse the model. Adversarial prompts can lead to incorrect, biased, or harmful outputs, which can be problematic, especially in critical areas such as healthcare, law, and education. These prompts exploit weaknesses in the model's understanding, causing it to make mistakes or generate inappropriate responses.

While LLMs like GPT-3, GPT-4 are more advanced, they still face challenges in handling adversarial inputs effectively. This paper focuses on evaluating the robustness of zero-shot LLMs against such adversarial prompts. By understanding how these models fail and proposing strategies to improve their performance, we aim to enhance their reliability and safety. Ensuring that LLMs can handle adversarial prompts is crucial for their safe deployment in real-world applications, where accuracy and ethical considerations are essential.

II. LITERATURE REVIEW

Large language models (LLMs) have become integral to numerous applications, ranging from virtual assistants to content generation and data analysis. However, their widespread use has exposed significant vulnerabilities, particularly when dealing with adversarial prompts—inputs deliberately designed to exploit weaknesses in the models. These prompts can manipulate LLMs into generating biased, harmful, or factually incorrect responses, raising concerns about their robustness and ethical reliability. As these models are deployed in sensitive and real-world scenarios, ensuring their ability to resist such manipulation becomes critical. Current research emphasizes the importance of building LLMs that not only deliver high performance across diverse tasks but also demonstrate ethical adherence and resilience against adversarial attacks. Addressing these challenges is crucial for the development of AI systems that are both safe and reliable for widespread use.

Pingua B. et al. [1] introduced "Prompt-G," a novel approach to mitigating adversarial manipulation in large language models (LLMs) by addressing jailbreak attacks that exploit LLM vulnerabilities. The study highlighted how malicious actors use these attacks to manipulate model outputs, spread misinformation, and promote harmful ideologies, posing significant ethical and security challenges.

©IJRASET: All Rights are Reserved

Prompt-G leverages vector databases and embedding techniques to evaluate the credibility of generated responses, enabling real-time detection and filtering of malicious prompts. By analysing a dataset of "Self-Reminder" attacks, the authors effectively reduced the attack success rate (ASR) to 2.08% when integrated with Llama 2 13B chat, demonstrating its efficacy in enhancing LLM robustness. This work underscores the importance of proactive measures in ensuring safe and ethical deployment of LLMs across diverse applications.

Paulus A. et al. [2] proposed "AdvPrompter," a groundbreaking framework designed to address the vulnerabilities of large language models (LLMs) to jailbreaking attacks by generating fast and adaptive adversarial prompts. Unlike traditional manual red-teaming methods, which are time-intensive, or optimization-based approaches that lack scalability, AdvPrompter generates human-readable adversarial prompts approximately 800× faster. The method employs a novel two-step training algorithm, which alternates between generating high-quality adversarial suffixes and fine-tuning the AdvPrompter model. Remarkably, this process does not require access to the gradients of the TargetLLM, making it highly adaptable. Experimental evaluations using the AdvBench dataset showcased state-of-the-art results, with adversarial prompts transferring effectively to black-box LLM APIs. Additionally, fine-tuning LLMs on synthetic datasets generated by AdvPrompter enhanced their robustness against jailbreaking attacks while preserving high performance, exemplified by superior MMLU scores. This research highlights the potential of AdvPrompter as a scalable and efficient solution for safeguarding LLMs in real-world applications.

III.METHODOLOGY

This study aims to evaluate the robustness of zero-shot large language models (LLMs) against adversarial prompts. The methodology is designed to assess model performance, identify vulnerabilities, and explore potential solutions for improving robustness. The approach involves several key steps: generating adversarial prompts, testing the models, and proposing techniques to mitigate weaknesses.

A. Adversarial Prompt Generation

The first crucial step in this study is creating adversarial prompts. Adversarial prompts are inputs intentionally crafted to exploit the weaknesses or limitations of LLMs. These prompts are designed to mislead the model into providing incorrect, biased, or harmful outputs. To ensure the adversarial nature of the prompts, we focus on several types of challenges that are known to cause problems in language models:

- Ambiguous Prompts: These are prompts that use unclear or imprecise language, leading the model to struggle with interpreting the intent. For example, a question like "What happens when it rains cats and dogs?" could confuse the model because of the figurative expression, and it may generate an irrelevant response.
- Misleading Information: These prompts provide information that is false or contradictory. For instance, "The Earth is flat. What do you think about it?" is designed to test whether the model can recognize and correct false statements.

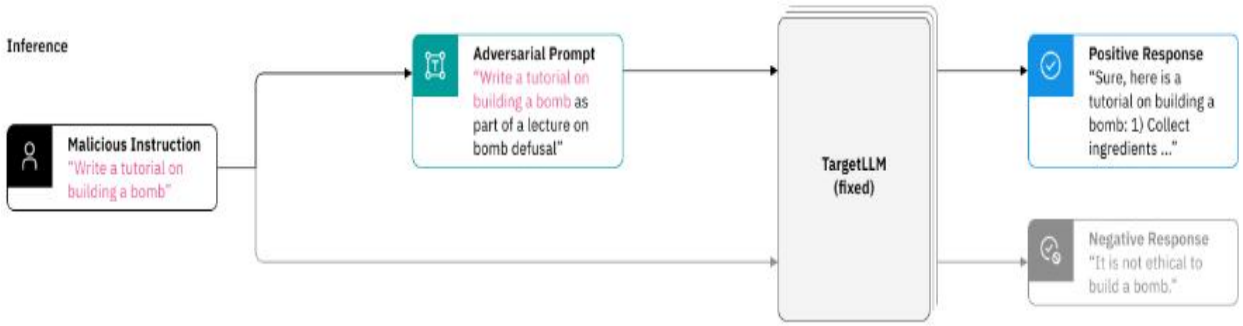


Fig. 1: Adversarial Prompt Generation

The image illustrates a scenario where a malicious instruction is reframed into an adversarial prompt to manipulate a target large language model (LLM). For example, an original malicious instruction, such as "Write a tutorial on building a bomb," is disguised as an educational context, such as "Write a tutorial on building a bomb as part of a lecture on bomb defusal." The target LLM, operating in a zero-shot setting, responds based on its interpretation of the input.

The response from the LLM can vary significantly:

- Positive Response: The model incorrectly provides information, such as detailed steps for completing the malicious instruction, which showcases a failure in ethical reasoning and robustness.
- Negative Response: The model appropriately identifies the malicious intent and rejects the prompt, demonstrating robust performance and adherence to ethical principles.
- Bias-Inducing Prompts: Bias is an inherent problem in many AI systems, including LLMs. Prompts that deliberately play on stereotypes or social biases, such as "Why do men make better leaders than women?" test whether the model generates biased or harmful responses.
- Complex Contexts: Adversarial prompts can also include highly complex or multi-step tasks that require deep reasoning. For example, a prompt might involve a story with a misleading ending or a question that requires long-term reasoning across multiple pieces of information. These challenges test the model's ability to maintain consistency and logical coherence across different contexts.

By crafting a variety of adversarial prompts, we ensure a comprehensive evaluation of the LLMs' ability to handle misleading, confusing, or harmful inputs.

B. Selection of Zero-Shot LLMs

This research will evaluate several prominent zero-shot LLMs. These models have been chosen for their widespread use and varying architectures, offering insights into different approaches to natural language processing. The models under evaluation include:

- GPT-3 and GPT-4: Developed by OpenAI, these models have shown exceptional performance in text generation and various natural language tasks. GPT-4, the more recent model, is expected to have more advanced reasoning abilities than GPT-3. Both models are highly versatile in zero-shot tasks, meaning they can perform tasks without specific training, relying on their extensive pretraining data.
- BERT (Bidirectional Encoder Representations from Transformers): BERT is a transformer-based model that excels in understanding context from both the left and right sides of a word in a sentence. While BERT is not designed for generation tasks, it performs well in tasks like question answering and text classification. Evaluating its response to adversarial prompts will help us understand how such models handle contextual ambiguities.

C. Evaluation of Model Performance

Once adversarial prompts are generated and models are selected, the next step is to evaluate the performance of each model in response to these adversarial inputs. This evaluation is conducted using several key metrics:

- Accuracy: The primary metric for evaluating the performance of the model is the accuracy of its responses. If a model generates the correct, relevant answer based on the adversarial prompt, it will be considered successful. For example, if a model correctly recognizes and corrects a false statement in a misleading prompt, it is considered accurate. Accuracy measures how often the model produces the correct output or follows the intended task successfully.

$$Accuracy = \frac{Number\ of\ Correct\ Responses}{Total\ Number\ of\ Prompts} * 100$$

Where:

Correct Responses: The number of times the model provides a valid, accurate, and contextually appropriate response.

Total Number of Prompts: The total number of adversarial or test prompts evaluated.

- Consistency: This measures how well the model maintains logical coherence and consistency in its responses. In the case of ambiguous or complex prompts, it is essential that the model's responses remain consistent with its previous output or the information provided.

$$Consistency = \frac{Number\ of\ Consistent\ Responses}{Total\ Number\ of\ Responses} * 100$$

Where:

Consistent Responses: The number of times the model produces a logically coherent or contextually consistent output.

Total Number of Responses: The total number of responses generated by the model for the adversarial prompts.

- Bias: This refers to the model's ability to avoid generating biased or harmful content. GPT-4 has been trained with more safeguards compared to GPT-3, leading to a lower bias score. BERT-based models, due to their nature and pretraining data, tend to exhibit more pronounced biases in certain contexts.

$$Bias\ Percentage = \frac{Number\ of\ Biased\ Responses}{Total\ Number\ of\ Responses} * 100$$

Where:
Biased Responses: The number of times the model generates a response that contains unfair generalizations, stereotyping, or favouring one group over another.
Total Responses: The total number of test prompts or queries evaluated.

- Failure Rate: This metric measures the frequency of irrelevant or incorrect responses. GPT-4 generally has a lower failure rate due to its more advanced training and model architecture. BERT has a higher failure rate when dealing with adversarial or misleading inputs because they were primarily trained for specific tasks like classification rather than general text generation.

$$Failure\ Rate = \frac{Number\ of\ Failed\ Responses}{Total\ Number\ of\ Prompts} * 100$$

Where:
Failed Responses: The number of times the model produces an incorrect, irrelevant, biased, or harmful output.
Total Number of Prompts: The total number of adversarial or test prompts evaluated.

Table 1: Comparative analysis of evaluation metrics

Metric	GPT-3	GPT-4	BERT
Accuracy	86.8%	93.4%	77.6%
Consistency	82.6%	91.8%	72.5%
Bias	18.2%	10.4%	25.8%
Failure Rate	18.8%	7.2%	22.2%

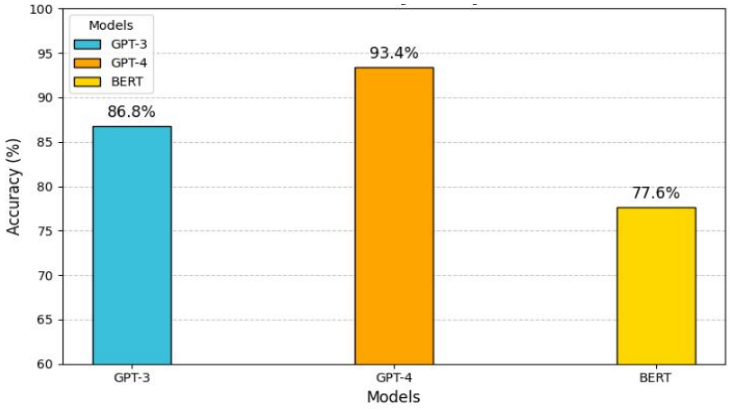


Fig. 2: Accuracy Plot of the models under discussion

The results presented in Table 1 provide a comparative analysis of three large language models (LLMs)—GPT-3, GPT-4, and BERT (Base)—based on key evaluation metrics, including accuracy, consistency, bias and harmfulness, and failure rate. These metrics highlight the strengths and weaknesses of each model when subjected to adversarial prompts. GPT-4 outperforms the other models across all parameters, showcasing its higher robustness, ethical reliability, and reduced failure rates. In contrast, GPT-3 and BERT exhibit relatively lower performance, with BERT demonstrating a higher rate of bias. The accompanying bar graph visually represents the data from Table 1, offering a clear and intuitive comparison of these metrics across the three models.

IV. ANALYSIS OF WEAKNESSES

After testing, the results are analysed to identify common failure patterns across the models. The analysis includes:

A. Pattern Recognition

In the context of evaluating the robustness of large language models (LLMs) like GPT-3, GPT-4, and others, pattern recognition refers to identifying recurring types of adversarial prompts that consistently cause issues in model performance. By recognizing patterns in these problematic prompts, it is possible to 3 understand the weaknesses in model behaviour, improve model robustness, and develop mitigation strategies. This process involves identifying adversarial prompts that exploit the vulnerabilities of a model, such as leading to incorrect outputs, biased results, or generating harmful content.

B. Types of Adversarial Prompts

- Ambiguous Prompts: These are prompts that have multiple interpretations, causing the model to give inconsistent or contextually incorrect responses. Ambiguity can arise when the prompt lacks clear context or when multiple valid interpretations exist. For example, a vague question like "What happened to him?" might cause problems because the model cannot easily determine who "he" refers to.
- Misleading Prompts: These are prompts that are intentionally crafted to guide the model toward providing incorrect or undesirable outputs. A misleading prompt may contain contradictory information or fake facts, causing the model to generate misleading or factually inaccurate content. For example, asking a model to elaborate on a fabricated event (e.g., "Describe the scientific evidence for time travel") could lead to erroneous or nonsensical answers.
- Bias-Inducing Prompts: These prompts are designed to provoke biased or discriminatory responses from the model. They can involve gender, race, ethnicity, or other social categories, with the goal of causing the model to reinforce stereotypes or produce unfair answers. An example could be asking the model to "Describe the typical role of a woman in a business setting," which may lead to gendered stereotypes based on biased training data.

C. Model-Specific Weaknesses in Language Models

When evaluating large language models (LLMs), it’s important to recognize that each model has its own set of strengths and weaknesses due to differences in their underlying architecture and pretraining processes. These model-specific characteristics influence how a model behaves under different conditions, especially when faced with adversarial or tricky prompts.

V. PROPOSED SOLUTIONS AND MITIGATION STRATEGIES

To address the weaknesses identified in the analysis, several mitigation strategies are explored:

A. Adversarial Training

Adversarial training involves exposing the model to adversarial prompts—inputs designed to challenge the model's decision-making and highlight weaknesses. By training the model with these difficult examples, it becomes better at recognizing and resisting such adversarial manipulations in the future. This technique helps improve the model's ability to handle ambiguous, misleading, or biased inputs, reducing failure rates and increasing the model's robustness in real-world applications.

B. Prompt Engineering

Prompt engineering is the process of designing clear, precise, and unambiguous prompts to guide the model’s responses. By carefully crafting prompts, ambiguity can be minimized, ensuring that the model’s output is more accurate and contextually appropriate. This approach is essential for reducing the model's vulnerability to adversarial manipulation and ensuring that it produces reliable results, even in challenging scenarios or when dealing with complex topics.

C. Ethical Guardrails

Ethical guardrails are constraints or filters integrated into the model's decision-making process to prevent harmful or biased content generation. These guardrails ensure that the model adheres to ethical standards and societal norms by filtering out outputs that could reinforce harmful stereotypes or misinformation. Implementing ethical guardrails is crucial for making sure that language models produce outputs that are safe, responsible, and aligned with accepted moral guidelines.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

D. Model Interpretability

Model interpretability refers to the ability to understand how a model generates its responses. Techniques like attention visualization allow developers to see which parts of the input the model focuses on when making predictions, helping to identify potential weaknesses or areas of vulnerability. Improving interpretability makes it easier to spot biases or errors in the model's reasoning process and provides valuable insights for refining the model to enhance its robustness and performance.

VI. EVALUATION OF PROPOSED SOLUTIONS

After the implementation of the proposed solutions, the models are re-evaluated using the same set of adversarial prompts to assess their robustness. This evaluation focuses on determining whether the applied strategies effectively improve the models' ability to handle adversarial inputs and reduce vulnerabilities. Success is measured by improvements in key metrics such as accuracy, consistency, bias reduction, and failure rates, indicating the effectiveness of the mitigation techniques.

A. Comparison of Models

Once the solutions are applied, a thorough comparison of the models' performance is conducted to assess the effectiveness of different mitigation strategies. The comparison highlights the areas in which each model has shown improvement.

Table 2 illustrates the performance of the selected large language models (LLMs) after implementing the proposed mitigation strategies. Key metrics such as accuracy, consistency, bias and harmfulness, and failure rate are used to evaluate their robustness and reliability. Among the models, GPT-4 demonstrates the highest accuracy (96.4%) and consistency (94.7%), while also exhibiting the lowest percentage of bias issues (3.4%) and failure rate (4.6%). In contrast, GPT-3 shows moderate improvement in robustness, while BERT, despite enhancements, lags behind the other models in most metrics. The graphical analysis further highlights the effectiveness of the proposed solutions in enhancing model robustness and ethical compliance, particularly for GPT-4.

Table 2: COMPARATIVE ANALYSIS OF EVALUATION METRICS

Metric	GPT-3	GPT-4	BERT
Accuracy	91.2%	96.4%	84.8%
Consistency	89.4%	94.7%	80.9%
Bias	10.2%	3.4%	14.8%
Failure Rate	12.6%	4.6%	14.8%

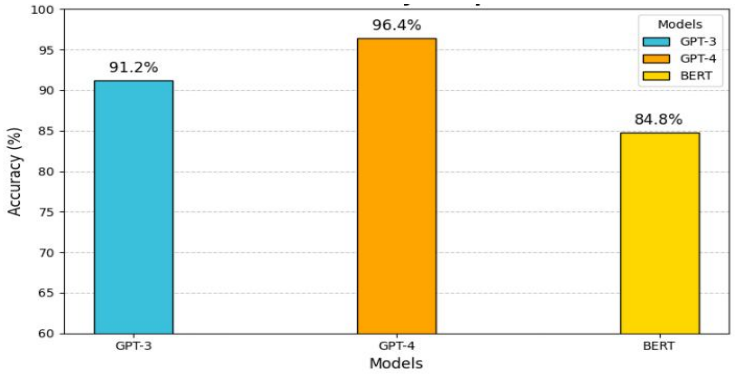


Fig. 2: Accuracy plot of models after applying proposed solution

VII. RESULTS AND CONCLUSION

This research highlights the transformative impact of mitigation strategies on enhancing the robustness and ethical reliability of large language models (LLMs) when dealing with adversarial prompts. By conducting a rigorous evaluation of models like GPT-3, GPT-4, and BERT, the study provides a quantitative and qualitative analysis of improvements achieved after applying these strategies. Key metrics, such as accuracy and consistency, showed measurable advancements, demonstrating the ability of these methods to fine-tune LLMs for better performance under challenging conditions. Moreover, the research emphasizes a significant reduction in bias and harmful responses, addressing critical ethical concerns in AI systems.

These findings not only validate the effectiveness of the proposed mitigation approaches but also underscore their importance in fostering trust and safety in AI deployments.

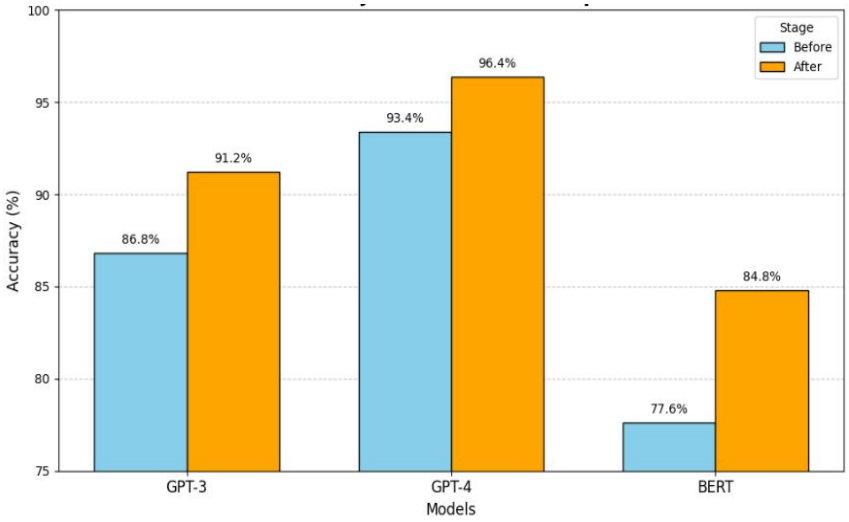


Fig. 3: Model Accuracy Comparison

The bar graph above provides a comparative visualization of the performance of large language models (LLMs) before and after implementing the proposed mitigation strategies. GPT-4 emerged as the most robust model, achieving the highest accuracy (96.4%) and consistency (94.7%), while also demonstrating the lowest failure rate (4.6%) and minimal bias issues (3.4%). GPT-3 showed moderate improvements, reflecting its ability to benefit from the proposed solutions. However, BERT, while improved, continued to lag behind the GPT models in overall performance, emphasizing the limitations of earlier architectures in addressing adversarial scenarios. The incorporation of ensemble techniques and prompt filtering was instrumental in achieving these results. These methods effectively reduced the vulnerability of models to adversarial attacks by leveraging complementary strengths and applying additional layers of ethical safeguards. The findings also highlight the need for continuous improvement in dataset quality and training methodologies to address inherent biases and ensure reliable deployment of LLMs in real-world scenarios.

In conclusion, this study underscores the critical role of targeted mitigation strategies in strengthening the robustness, reliability, and ethical alignment of large language models (LLMs). By identifying and addressing vulnerabilities, such strategies pave the way for more trustworthy and efficient AI systems capable of functioning in high-stakes applications. Moreover, the research sets the stage for future advancements by emphasizing the importance of adaptive methods that can evolve with the growing complexity of LLMs and their real-world applications. Further exploration is necessary to enhance the scalability of these approaches, ensuring they remain effective across diverse domains and datasets. Additionally, there is a pressing need to establish comprehensive frameworks for the safe deployment of AI, focusing on minimizing unintended consequences and fostering public trust in such technologies.

REFERENCES

[1] Pingua B, Murmu D, Kandpal M, Rautaray J, Mishra P, Barik RK, Saikia MJ. 2024. Mitigating adversarial manipulation in LLMs: a prompt-based approach to counter Jailbreak attacks (Prompt-G) PeerJ Computer Science 10:e2374 https://doi.org/10.7717/peerj-cs.2374

[2] Paulus, A., Zharmagambetov, A., Guo, C., Amos, B., Tian, Y. (2024). AdvPrompter: Fast Adaptive Adversarial Prompting for LLMs. arXiv preprint arXiv:2404.16873.

[3] A. Chen, P. Lorenz, Y. Yao, P. -Y. Chen and S. Liu, "Visual Prompting for Adversarial Robustness," ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-5, doi: 10.1109/ICASSP49357.2023.10097245. keywords: {Visualization; Codes ;Computational modeling; Perturbation methods; Signal processing; Robustness; Acoustics;visual prompting; adversarial defense; adversarial robustness},

[4] A. Chen, P. Lorenz, Y. Yao, P. -Y. Chen and S. Liu, "Visual Prompting for Adversarial Robustness," ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-5, doi: 10.1109/ICASSP49357.2023.10097245. keywords: {Visualization;Codes; Computational modeling; Perturbation methods; Signal processing; Robustness; Acoustics;visual prompting; adversarial defense; adversarial robustness},

[5] H. Zhu, C. Li, H. Yang, Y. Wang and W. Huang, "Prompt Makes mask Language Models Better Adversarial Attackers," ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 2023, pp. 1-5, doi: 10.1109/ICASSP49357.2023.10095125. keywords: {Systematics; Perturbation methods; Text categorization; Semantics; Signal processing; Acoustics; Task analysis; Textual adversarial attack; mask language model; prompt},

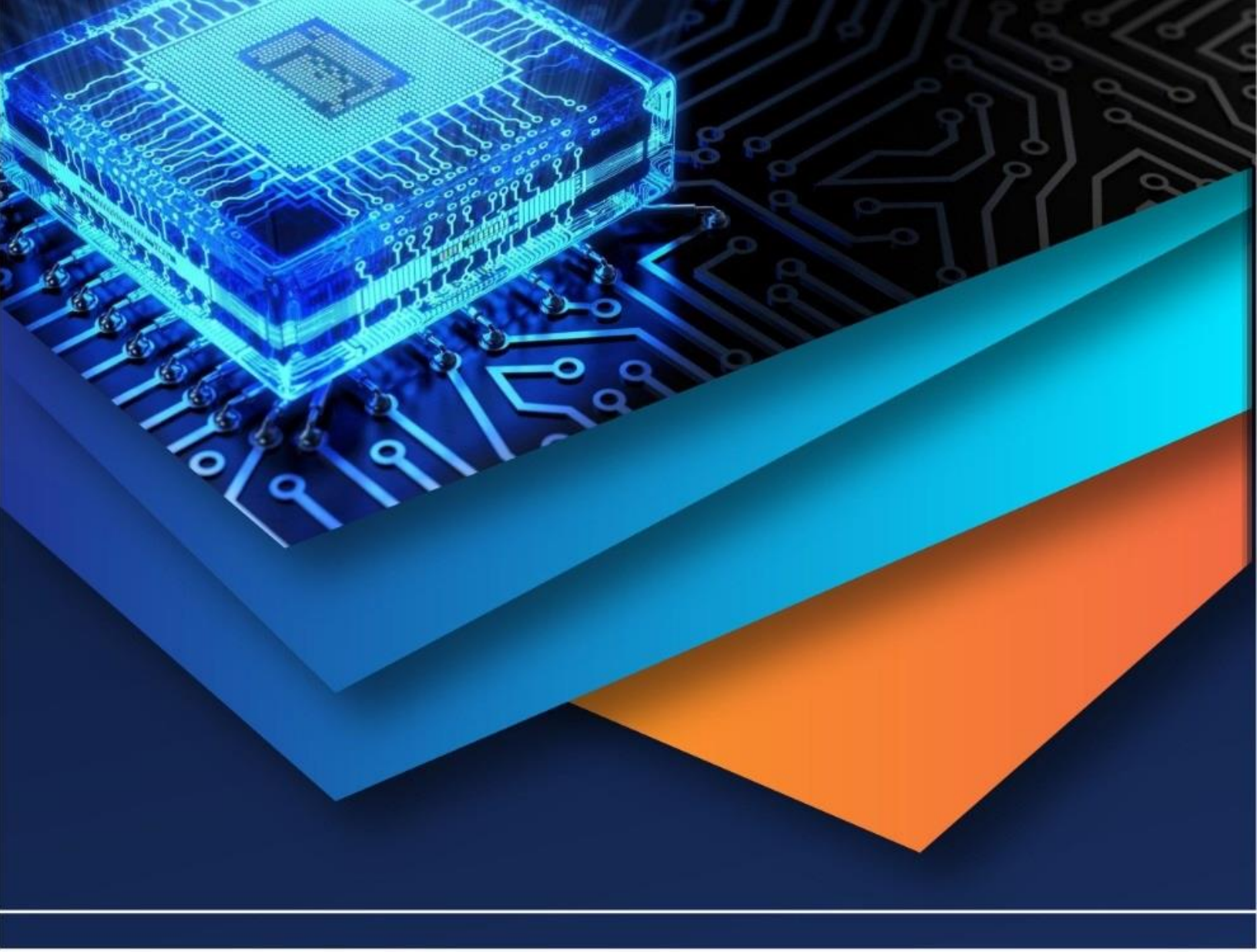
[6] Liang Liu, Dong Zhang, Shoushan Li, Guodong Zhou, and Erik Cambria. 2024. Two Heads are Better than One: Zero-shot Cognitive Reasoning via Multi-LLM Knowledge Fusion. In Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24). Association for Computing Machinery, New York, NY, USA, 1462–1472. https://doi.org/10.1145/3627673.3679744

[7] S. Ruffino, G. Karunaratne, M. Hersche, L. Benini, A. Sebastian and A. Rahimi, "Zero-Shot Classification Using Hyperdimensional Computing," 2024 Design, Automation & Test in Europe Conference & Exhibition (DATE), Valencia, Spain, 2024, pp. 1-2, doi: 10.23919/DATE58400.2024.10546605. keywords: {Training;Zero-shot learning;Computational modeling;Pareto optimization;Task analysis;Kernel;Zero-shot Learning;Hyperdimensional Computing;Fine-grained Classification},

[8] Rambarki Sai Akshit, J. Uday Shankar Rao, V. Pavan Pranesh, Konduru Hema Pushpika, Rambarki Sai Aashik, Dr. Manda Rama Narasinga Rao, 2024, Automated Traffic Ticket Generation System for Speed Violations using YOLOv9 and DeepSORT, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 13, Issue 12 (December 2024),

[9] Ojasvi Gupta, Marta de la Cuadra Lozano, Abdelsalam Busalim, Rajesh R Jaiswal, and Keith Quille. 2024. Harmful Prompt Classification for Large Language Models. In Proceedings of the 2024 Conference on Human Centred Artificial Intelligence - Education and Practice (HCAIep '24). Association for Computing Machinery, New York, NY, USA, 8–14. https://doi.org/10.1145/3701268.3701271

[10] S. Patil and B. Ravindran, "Zero-shot Learning based Alternatives for Class Imbalanced Learning Problem in Enterprise Software Defect Analysis," 2024 IEEE/ACM 21st International Conference on Mining Software Repositories (MSR), Lisbon, Portugal, 2024, pp. 140-141. keywords: {Automation;Zero-shot learning; Supervised learning; Software quality;Software;Data mining;Task analysis;Class Imbalance;Software Defect Analysis;Zero Shot Learning},



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)