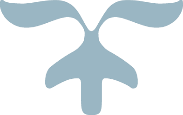


INDRISTRIAL ATTACHMENT

SEMINAR REPORT

CSE-420(Dual)



# Name: SADIA AKTER

# ID: 2125051059

Batch: 50(Autumn 2021)

Section: 7B1

Semester: Autumn 2024

Seminar Topic: Prompt Engineering and Jailbreaking LLM Models

Name of the keynote speaker: Raihan Alam

Principal Software Engineer, Microsoft

# Date of Seminar: 07, October 2024

### **Title: Exploring Prompt Engineering and Jailbreaking LLM Models**

#### **Introduction**

In the world of Artificial Intelligence (AI), many new technologies are shaping how we interact with machines. One of the most exciting areas is **Prompt Engineering**, which is a method used to guide AI models like ChatGPT, Gemini, and other Large Language Models (LLMs) to produce useful results.

On October 7, 2024, our university organized an online seminar to help students and professionals learn more about this growing field. The seminar was titled **"Prompt Engineering and Jailbreaking LLM Models"** and was conducted through the Zoom platform. The speaker, **Raihan Alam**, explained the basics of Prompt Engineering, how LLMs work, and the risks and benefits of jailbreaking these models.

This seminar was an excellent opportunity for attendees to gain knowledge about a field that is changing how businesses and individuals use AI in daily tasks.

### **About the Keynote Speaker**

The speaker for this seminar was **Raihan Alam**, who is a **Principal Software Engineer** at **Microsoft Australia**.

* **Professional Background:** Raihan Alam is an expert in Prompt Engineering and has extensive experience working with advanced AI systems.
* **Role in the Seminar:** As the only speaker, he shared his insights on LLM models and how they can be utilized in different industries.
* **Engaging Presentation:** Mr. Alam’s energetic and clear explanation of complex topics kept the audience engaged throughout the one-and-a-half-hour session.

His ability to simplify technical topics made the seminar enjoyable and easy to follow for all participants.

### **Purpose and Goals of the Seminar**

The main purpose of this seminar was to introduce attendees to **Prompt Engineering** and explain how it connects with the growing field of LLMs. The goals included:

1. **Educating Participants:** Explain what Prompt Engineering is and how it works.
2. **Highlighting the Importance of LLMs:** Discuss the role of LLMs in AI and their impact on industries like e-commerce, customer service, and more.
3. **Discussing Security Concerns:** Provide an understanding of how jailbreaking LLMs works and its ethical implications.
4. **Encouraging Innovation:** Inspire participants to explore the potential of Prompt Engineering as a career path.

The seminar aimed to make everyone more aware of the opportunities and challenges in this exciting area of AI.

### **Key Points Discussed**

During the seminar, Mr. Alam covered several important topics related to Prompt Engineering and LLMs. Some of the key points included:

#### **1. What is Prompt Engineering?**

Prompt Engineering is the process of creating specific inputs or prompts to guide AI models to produce the desired output. It is an essential skill for working with generative AI, as it helps users get accurate and useful responses from models like ChatGPT.

#### **2. Trends in Large Language Models (LLMs)**

Mr. Alam talked about the current trends in LLMs, highlighting their increasing size and capabilities. He explained how LLMs are being used in industries such as healthcare, education, and e-commerce to improve efficiency and automate tasks.

#### **3. How Prompt Engineering Works**

He provided a detailed explanation of how to design prompts effectively. Key points included:

* Understanding how LLMs interpret language.
* Using examples and clear instructions in prompts to improve results.

#### **4. Prompt Ejection**

Prompt ejection occurs when AI models are manipulated to reveal unintended responses. This technique can be used for creativity, testing, or even exploitation.

#### **5. Improving Prompt Security**

The speaker shared ways to secure prompts so they cannot be easily exploited. This is particularly important when using AI in sensitive areas like banking or healthcare.

#### **6. Impacts of AI Chatbots**

AI chatbots, powered by Prompt Engineering, are transforming industries by automating customer interactions and improving user experiences. For example, businesses can use chatbots to handle customer queries more efficiently.

#### **7. Jailbreaking LLMs**

Jailbreaking LLMs refers to bypassing restrictions on these models to explore their full capabilities. While this can help developers test and improve models, it also raises concerns about ethical usage and security risks.

### **New Information Gained**

The seminar was filled with new and exciting information, including:

1. **How Prompt Engineering Works:** I learned the steps involved in crafting prompts to guide AI systems.
2. **Current Trends in AI:** Understanding the advancements in LLMs gave me a broader view of how AI is evolving.
3. **Ethical Concerns of Jailbreaking:** The session highlighted the risks of manipulating AI systems, such as exposing sensitive data.
4. **Applications of Prompt Engineering:** I discovered how this skill can be used to create better AI applications for different industries.

This knowledge broadened my understanding of how AI works in practical scenarios.

### **Bridging the Gap Between Academics and Industry**

This seminar helped bridge the gap between what we learn in classrooms and how technology is used in the real world:

* **Classroom Learning:** Our academic curriculum focuses on programming and algorithms but often lacks coverage of emerging topics like Prompt Engineering.
* **Industry Relevance:** The seminar showed how Prompt Engineering is already being used in industries to solve real-world problems.

By participating in the seminar, I gained practical knowledge that complements the theoretical lessons from my coursework, making me feel more prepared for industry challenges.

### **Suggestions for Improving Academic Programs**

To prepare students for careers in AI and technology, academic programs should:

1. **Include Emerging Topics:** Add courses on Prompt Engineering, LLMs, and ethical AI practices.
2. **Encourage Real-World Projects:** Let students work on AI-based projects that mimic industry scenarios.
3. **Host More Seminars:** Bring in industry experts to share the latest advancements and insights.
4. **Focus on Ethics and Security:** Teach students about the ethical implications of working with AI systems.

Making these changes will ensure that students are ready to take on the challenges of the ever-evolving tech industry.

### **Seminar Summary**

The seminar took place online on October 7, 2024, and began at 9:00 AM. **Raihan Alam**, a Principal Software Engineer at Microsoft Australia, delivered a detailed presentation on Prompt Engineering and its role in LLMs.

The session covered topics such as how Prompt Engineering works, current trends in AI, and the ethical issues surrounding jailbreaking LLMs. Mr. Alam also provided practical examples and tips for using prompts effectively.

At the end of the seminar, a Q&A session allowed participants to ask questions and clarify their doubts.

### **Concluding Thoughts**

The seminar was highly informative and engaging. It introduced me to a new field in AI and helped me understand its practical applications and challenges. Mr. Alam’s presentation was clear and inspiring, making it easy to follow even complex topics.

I believe more seminars like this should be organized to help students stay updated with the latest advancements in technology. These sessions not only expand our knowledge but also inspire us to explore new and innovative career paths in the tech industry.

**The End**