**The College of Graduate Studies and the College of Information Technology Cordially Invite You to a**

**Master Thesis Defense**

Entitled

*A SECURE AND EFFECTIVE FRAMEWORK FOR KEY CONCEPT MINING FROM EDUCATIONAL CONTENT USING LARGE LANGUAGE MODELS*

By

Ashika Sameem Abdul Rasheed

ID: 700036309

Faculty Advisor

Prof. Mohammad Mehedy Masud

College of Information Technology

Date & Venue

12th November 2024, Tuesday

1:30 PM - 3:00 PM

Room 1012 , E1 Building

Online Link: https://uaeu-ac-ae.zoom.us/j/89321148959

Abstract

This thesis examines the use of Large Language Models (LLMs) in education, with a focus on improving performance and implementing strong security measures. The research has two main goals, namely, the development of an effective lecture summarization technique using LLMs and identifying and addressing security vulnerabilities in LLM applications according to OWASP (Open Web Application Security Project) guidelines. For the former goal, we have proposed an effective framework for fine-tuning LLMs using real lecture datasets and compared the performance of different LLMs. For the latter goal, we conducted a thorough review of the application dataflow of the proposed framework and revealed several vulnerabilities, categorized as high risk, medium risk, and low risk. We also propose countermeasures to these vulnerabilities and demonstrate their efficacy. Thus, this study suggests a framework for securely integrating LLMs into educational purposes, tackling critical security concerns while harnessing the models' efficiency.

**Keywords**: Large Language Models, NLP, Security, Risk Rating, OWASP, fine-tuning, defense mechanisms

**تتشرف كلية الدراسات العليا وكلية تقنية المعلومات بدعوتكم لحضور**

**مناقشة رسالة الماجستير**

العنوان

إطار عمل آمن وفعال لاستخراج المفاهيم الأساسية من المحتوى التعليمي باستخدام نماذج لغوية كبيرة

للطالبة

عاشقة ساميم عبد الرشيد

الرقم الجامعي: 700036309

المشرف

د. محمد مهدي مسعود

كلية تقنية المعلومات

المكان والزمان

E1 مبنى ، غرفة 1012

(3:00 – 1:30)

الثلاثاء، 12 نوفمبر 2024

الرابط

https://uaeu-ac-ae.zoom.us/j/89321148959

الملخص

تدرس هذه الأطروحة استخدام نماذج اللغة الكبيرة (LLMs) في التعليم، مع التركيز على تحسين الأداء وتنفيذ تدابير أمنية قوية. للبحث هدفان رئيسيان، وهما تطوير تقنية فعالة لتلخيص المحاضرات باستخدام نماذج اللغة الكبيرة وتحديد نقاط الضعف الأمنية في تطبيقات نماذج اللغة الكبيرة ومعالجتها وفقًا لإرشادات OWASP (مشروع أمان تطبيقات الويب المفتوحة). بالنسبة للهدف الأول، اقترحنا إطارًا فعالًا لضبط نماذج اللغة الكبيرة باستخدام مجموعات بيانات المحاضرات الحقيقية وقارنا أداء نماذج اللغة الكبيرة المختلفة. بالنسبة للهدف الأخير، أجرينا مراجعة شاملة لتدفق بيانات التطبيق للإطار المقترح وكشفنا عن العديد من نقاط الضعف، المصنفة على أنها عالية الخطورة ومتوسطة الخطورة ومنخفضة الخطورة. نقترح أيضًا تدابير مضادة لهذه نقاط الضعف ونوضح فعاليتها. وبالتالي، تقترح هذه الدراسة إطارًا لدمج نماذج اللغة الكبيرة بشكل آمن في الأغراض التعليمية، ومعالجة المخاوف الأمنية الحرجة مع الاستفادة من كفاءة النماذج.

**كلمات البحث الرئيسية** نماذج اللغة الكبيرة، معالجة اللغة الطبيعية، الأمان، تصنيف المخاطر، OWASP، الضبط الدقيق، آليات الدفاع .