|  |  |
| --- | --- |
| **Project Code** |*Kod Projek* | **BITU 3973** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | UNIVERSITI TEKNIKAL MALAYSIA MELAKA  FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY | | | | | | | | | | | | | | | |
| PROJEK SARJANA MUDA 1: PROPOSAL FORM | | | | | | | | | | | | | | | | | | | |
| **A** | **TITLE OF PROPOSED PROJECT** |*TAJUK PROJEK YANG DICADANGKAN* | | | | | | | | | | | | | | | | | | |
| AI for Social Engineering Attack Detection | | | | | | | | | | | | | | | | | | | |
| **B** | **DETAILS OF STUDENT** |*BUTIRAN PELAJAR* | | | | | | | | | | | | | | | | | | |
| Name | | NURSYUHADAH BINTI AHMAD SUDERMAN | | | | | **Program:** | |  | |  | |  | |  | |  | |  |
| **BITC** |  | | **BITD** | |  | | **BITI** | |  |  |
|  | |  | |  |
| **Matric No.** | | B032310114 | | | | | **BITE** |  | | **BITM** | |  | |  | | |  |
|  | |  | |  | | |
| **Handphone No.** | | 0107914486 | | | | | **BITS** |  | | **BITZ** | | x | |  | | |  |
|  | |  | |
|  |  | |  | |  | |  | | |  |
| **Semester** | | 2 | **Session** | | 2024/2025 | **Email Address** | | nursyuhadah1609@gmail.com | | | | | | | | | | | |
| **C** | **PROJECT INFORMATION** | MAKLUMAT PROJEK | | | | | | | | | | | | | | | | | | |
| **(i)** | **Executive Summary of Project Proposal [Maximum 300 words]** | | | | | | | | | | | | | | | | | | |
| *(Please include the background of the project, problem statements, objectives, and expected outcomes/ proposed solution from the project)*  Social engineering attacks exploit human psychology to manipulate individuals into divulging sensitive information, making them one of the most significant cybersecurity threats today. With advancements in Artificial Intelligence (AI), machine learning models are increasingly being explored for detecting and mitigating such attacks. However, existing AI-based social engineering detection systems face challenges such as data scarcity, high false-positive rates, and adversarial attacks. This research aims to explore the effectiveness of AI techniques in detecting social engineering attacks, identify current limitations, and propose recommendations for improvement. It will investigate various AI-based detection methods, analyze key linguistic and behavioral features used in social engineering detection, and evaluate their effectiveness across different attack types such as phishing emails, fake messages, and impersonation attempts. Additionally, the study will assess case studies of AI implementations in cybersecurity solutions and explore emerging trends, including adversarial training and reinforcement learning. The expected outcome of this research is a comprehensive analysis of AI-driven social engineering attack detection, providing insights into its effectiveness and areas for enhancement. The findings will contribute to cybersecurity knowledge and offer recommendations to improve AI-powered detection strategies in the future. | | | | | | | | | | | | | | | | | | | |
| **(ii)** | **Detailed Proposal of the Project** | | | | | | | | | | | | | | | | | | |
| **(a) Introduction** *(Project Background and Problem Statements)* | | | | | | | | | | | | | | | | | | |
| Social engineering attacks have become increasingly sophisticated, exploiting human psychology to manipulate individuals into divulging sensitive information. These attacks, which include phishing, pretexting, and baiting, pose significant risks to individuals and organizations. Traditional cybersecurity solutions often focus on technical defenses, leaving human vulnerabilities as an exposed attack surface.  Artificial Intelligence (AI) has emerged as a potential solution for detecting social engineering attacks by analyzing linguistic, behavioral, and contextual patterns. However, existing AI-based detection methods face several challenges. These include the scarcity of high-quality datasets for training AI models, the high rate of false positives in detection, and the adaptability of attackers who continuously refine their tactics to bypass security measures. Additionally, adversarial manipulation of AI models poses a serious risk, as attackers can craft messages designed to evade detection.  This research explores the role of AI in social engineering attack detection, assesses its current effectiveness, and identifies key areas requiring improvement. The study will employ a combination of **literature review, comparative analysis, and case study research** to evaluate the effectiveness of AI in social engineering detection. A **literature review** will examine prior studies, government reports, and industry findings on AI-driven detection mechanisms. **Comparative analysis** will assess various AI models based on accuracy, robustness, and scalability. Finally, **case study research** will analyze real-world AI implementations in cybersecurity solutions to understand their strengths and limitations. By synthesizing these findings, this research aims to contribute valuable insights and recommendations for enhancing AI-driven security measures against social engineering threats. | | | | | | | | | | | | | | | | | | |
| **(b) Objectives of the Project** | | | | | | | | | | | | | | | | | | |
| 1. To investigate existing AI models used for detecting social engineering attacks. 2. To assess the effectiveness of linguistic, behavioral, and contextual features in AI-based detection. 3. To identify key challenges and limitations of AI in detecting social engineering attacks. | | | | | | | | | | | | | | | | | | |
| **(c) Scope of the Project** | | | | | | | | | | | | | | | | | | |
| * **AI Techniques:** Explore machine learning classifiers (e.g., SVM, Random Forest) and deep learning models (e.g., BERT, GPT) for detecting social engineering attacks. * **Feature Analysis:** Identify key linguistic, behavioral, and contextual indicators used in detection models. * **Challenges & Limitations:** Examine issues such as data scarcity, false positives, and adversarial attacks. * **Case Studies:** Analyze real-world AI-based social engineering detection tools and compare their effectiveness. * **Future Trends:** Investigate upcoming AI advancements, including adversarial training and reinforcement learning, for enhancing detection strategies. | | | | | | | | | | | | | | | | | | |
| **(d) Expected Outcome/ Proposed Solution** | | | | | | | | | | | | | | | | | | |
| * Provide a detailed analysis of AI-driven social engineering detection systems. * Identify strengths and weaknesses of current AI techniques. * Offer recommendations for improving AI-based detection models. * Contribute to the understanding of emerging AI trends in cybersecurity | | | | | | | | | | | | | | | | | | |
| **D** | **REFERENCES** | *RUJUKAN* | | | | | | | | | | | | | | | | | | |
| *State your references (Minimum 10 references)* | | | | | | | | | | | | | | | | | | | |
| **1** | Ai, L., Kumarage, T., Bhattacharjee, A., Liu, Z., Hui, Z., Davinroy, M., Cook, J., Cassani, L., Trapeznikov, K., Kirchner, M., Basharat, A., Hoogs, A., Garland, J., Liu, H., & Hirschberg, J. (n.d.). Defending Against Social Engineering Attacks in the Age of LLMs. Retrieved July 27, 2024, from <https://arxiv.org/pdf/2406.12263> | | | | | | | | | | | | | | | | | | |
| **2** | Adam, M., Yusra al-Amod, Yousif, O., & Ibrahim, J. (2011). Awareness of Social Engineering Among IIUM Students. 1(9), 409–413. <https://www.researchgate.net/publication/339943503_Awareness_of_Social_Engineering_Among_IIUM_Students> | | | | | | | | | | | | | | | | | | |
| **3** | Jritm Www, & Jritm. (2024). AI-Based Social Engineering Detection. Commerce and Management (JRITM), 1(2). <https://jritm.org/Upload/AI%20Based%20Social%20Engineering%20Detection.pdf> | | | | | | | | | | | | | | | | | | |
| **4** | Fakhouri, H. N., Basim Alhadidi, Omar, K., Sharif Naser Makhadmeh, Hamad, F., & Halalsheh, N. Z. (2024). *AI-Driven Solutions for Social Engineering Attacks: Detection, Prevention, and Response*. <https://doi.org/10.1109/iccr61006.2024.10533010> | | | | | | | | | | | | | | | | | | |
| **5** | Ozen, I., Subramani, K., Vadrevu, P., & Perdisci, R. (2024). *SENet: Visual Detection of Online Social Engineering Attack Campaigns*. ArXiv.org. <https://arxiv.org/abs/2401.05569> | | | | | | | | | | | | | | | | | | |
| **6** | Ali, S. (2024, December 1). *The Role of AI in Social Engineering Attack Prevention: NLP-Based Solutions for Phishing and Scams*. <https://doi.org/10.13140/RG.2.2.22981.97765> | | | | | | | | | | | | | | | | | | |
| **7** | Lenn, P., Supervisor, Dipti, K., Sarmah, M., Elhajj, L., & Mariot. (2025). *MSc Computer Science Final Project Leveraging Generative Pre-trained Transformers for the Detection and Generation of Social Engineering Attacks: A Case Study on YouTube Collusion Scams*. <https://essay.utwente.nl/104867/1/Perik_MA_EEMCS.pdf> | | | | | | | | | | | | | | | | | | |
| **8** | View of AI’s Revolutionary Role in Cyber Defense and Social Engineering. (2024). Itscience.org. <https://jurnal.itscience.org/index.php/ijmdsa/article/view/4752/3537> | | | | | | | | | | | | | | | | | | |
| **9** | Alsufyani, A. A., & Alzahrani, S. M. (2021). Social Engineering Attack Detection Using Machine Learning: Text Phishing Attack. Indian Journal of Computer Science and Engineering, 12(3), 743–751. <https://doi.org/10.21817/indjcse/2021/v12i3/211203298> | | | | | | | | | | | | | | | | | | |
| **10** | Camilo López, Pérez, J., & Rivera, E. (2022). Strengthening Digital Security Against Social Engineering Attacks Using AI-Powered Behavioral and Predictive Detection Systems. *International Journal of Intelligent Automation and Computing*, *5*(2), 92–105. <https://research.tensorgate.org/index.php/IJIAC/article/view/153> | | | | | | | | | | | | | | | | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **E** | **Declaration by student** |*Akuan pelajar* | | | | | | | | | | | | | | | |
| **(i)** | **Date:** | |  | | **Student’s Signature:** | | | | |  | | | | | | |
| 28-02-2025 | |  | | | | |  | |
|  | |  | | | | |  | | | | |  | |
| **E** | **RECOMMENDED BY SUPERVISOR**  *PERAKUAN OLEH PENYELIA* | | | | | | | | | | | | | | | |
| **(ii)** |  | | | | |  |  |  |  | |  |  |  | | | |
| **Recommended** | | | | |  | √ |  |  | |  |  |  | | | |
|  | | | | | | | | | | | | | | | |
| **Not Recommended** | | | | |  |  |  |  | |  |  |  | | | |
|  | | | | |  |  |  |  | |  |  |  | | | |
| **Comments:** | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | |
| **Supervisor’s Name:** | | | | | | | | | | | | | | | |
| Dr. Zaheera binti zainal abidin | | | | | | | | | | | | | | | |
| **Signature & Stamp:** | | | | | | | | | | | | | | | |
|  |  | | | | | | | | | | | |  | | |
|  | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | |
| **Date:** | | | 09-03-2025 | | | | | | | | | | | |  |
|  | | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | | |