**Group Project Proposal (Improvement for Ethics Risk)**

Course code: COMP 3511B

Team No.: 16

|  |  |  |  |
| --- | --- | --- | --- |
| Student ID #1: |  | Student ID #4: |  |
| Student Name #1: | Wenyu HAN | Student Name #4: | Feng YE |
| Student ID #2: |  | Student ID #5: |  |
| Student Name #2: | Yuhang HU | Student Name #5: | Siyu ZHOU |
| Student ID #3: |  |  |  |
| Student Name #3: | Lee Ling LEE |  |  |

### Guidelines

1. Proposal submit should be within xx pages with diagrams/tables for this assignment. Please also submit presentation slide for 15min (10min presentation & 5min Q&A) together with the proposal.

2. Make sure that you have carefully read and fully understood the questions before answering them. Answer the questions fully but concisely and as directly as possible.

3. Answer all questions in your own words. Do not copy any text from the casebook, readings or other sources. **The assignment must be your group work only.**

|  |
| --- |
| **Plagiarism declaration:** |
| **1. I know that plagiarism is wrong. Plagiarism is to use another’s work and pretend that it is one’s own.**  **2. This assignment is my own work.**  **3. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.**  **4. I acknowledge that copying someone else’s assignment (or part of it) is wrong, and declare that my assignments are my own work.** |

|  |
| --- |
| Executive Summary of Proposal (within this page)  1. **Current State**   AI technology has revolutionized industries, but it has also been used for commercial fraud, affecting economic activities and triggering negative public opinion. A multinational company's Hong Kong branch suffered a Deepfake scam, where an employee was tricked into a HKD$200 million transfer. The employee later became aware of the scam and reported it to the police. The organization was defrauded due to employees' weak exposure to AI fraud and an imperfect regulatory system. To prevent future tragedies, the company should implement advanced cybersecurity technologies for fraud detection and a reliable financial transaction program.   1. **Proposal Initiative**   - **Vision and Planning (Plan)**: To prevent AI fraud, a strong communication system, including cybersecurity teams, AI ethics officers, and biometrics, needs immediate action.  - **Do**: AI detection tools (comply with ISO/IEC 42001) will be used to analyze human behaviors in video conferences, and staff trained to detect AI-related fraud (Deepfake). Financial transactions should be processed after two verification processes (comply with ISO/IEC 27001).  - **Check**: Hiring third-party audit with relevant certificates to assess the effectiveness of the plan.  - **Act**: Refining AI detection algorithms, updating policies, and training programs.  - **Competitive Advantage**: By complying with ISO/IEC 27001 and ISO/IEC 42001, this plan offers competitive advantages: secure data management, gaining trust from society and users, and minimizing fraud losses.  - **Technical/managerial requirements**: The plan integrates some authentication technology and needs staff training and cybersecurity checks for system compatibility and potential threats.   1. **Plan of action and criteria for success**   The plan above is developed to combat Deepfake incidents by improving AI systems and algorithms. This includes training employees to use AI systems and software ethically, ensuring they respect privacy and data protection laws. The plan also includes fostering a supportive environment for employees to report anomalies without fear of repercussions. Human decision-making will be crucial, with a protocol allowing flagged transactions to be reconsidered if they don't align with human morality. The plan also includes establishing a protocol for flagged transactions that do not align with human morality. The success of this plan will be evaluated through increased employee proficiency and stakeholder feedback.   1. **Conclusion**   The proposed plan aims to combat AI fraud, particularly Deepfake scams, to also prevent future tragedies and regain public trust, by implementing advanced cybersecurity measures, AI detection tools, and a reliable financial transaction program. It emphasizes strong communication systems, employee training, and authentication technology integration. Success will be measured by reduced fraud attempts, improved employee proficiency, and positive stakeholder feedback, while monitoring ethical performance indicators to ensure AI alignment. By implementing this plan, the company can achieve secure data management and minimize AI fraud losses, gaining a competitive advantage and establishing trust in society and users. |

|  |
| --- |
| Current state In recent years, the innovative development of artificial intelligence technology has brought new business opportunities and revenues to various industries, but it has also been used by many people with ulterior motives as a tool to escalate commercial fraud. The Deepfake AI algorithm, which generates convincing audio and video through digital processing, has been utilized frequently in identity theft fraud. It has affected the economic activities of organizations and brought about negative public opinion.  The Hong Kong branch of a multinational company had been involved in an AI deepfake scam. An employee in the finance department received a message from the company’s UK CFO asking him to carry out a series of secret transactions. The employee had also suspected early on that this was a phishing message, but after the employee was invited to participate in the team video conference and realized that the company's CFO and other employees were present, he let his guard down. During the conference, the scammer asked the victim to introduce himself, but there was no actual interaction, and the fake image on the screen mainly issued commands before the conference abruptly ended. The fraud team used public video and audio of the financial officer and other employees to replicate the appearance and voice of the targeted individual, luring the employee step by step with a prepared script. Unfortunately, the employee did not realize the problem in time and followed the instructions to make fifteen transfers in five Hong Kong bank accounts totaling HK$200 million. A week later, the employee realized it was a scam after making inquiries with the company's headquarters and reporting cases to the police for investigation.  Reviewing the whole incident, the weak alertness of employees to AI fraud and the imperfect regulatory system of the company’s financial department are the main factors that caused the organization to be defrauded. As a multinational company, cross-border financial transactions are part of daily operations, and the company needs to have a strict auditing and monitoring system to ensure the use and flow of each fund. In this case, the action of a single employee to conduct a large amount of money transactions should not have happened. Meanwhile, the employees and leaders of the finance department need to be sensitive to such behavior, rather than discovering the situation only after a week. Online meetings are common for companies of this nature, but employees do not receive sufficient cybersecurity training to develop a sense of vigilance, and blind trust in the video content without further verification and confirmation of the so-called “secret transitions” in a timely manner.  The development of artificial intelligence has increased the credibility of fraudulent content, and employees need to develop a strong sense of security in a complex cybersecurity situation. Also, introduce a series of technologies that can efficiently detect AI fraud, and deploy a more sophisticated and reliable financial transaction program to prevent similar tragedies from happening in the future. |

|  |
| --- |
| Proposed initiative **Vision and Planning (Plan)**: Through the dual approach of ISO/IEC 27001 and ISO/IEC 42001, our goal extends not only to strengthen our communications systems against deepfake fraud, but also to ensure that our AI technologies are developed, deployed, and managed responsibly and ethically. This also includes setting up a specific group called cybersecurity to handle these incidents and having AI ethics officers oversee compliance with ethical standards. We are also going to involve advanced biometric technologies to send the warning or take immediate action if the Deepfake or other AI technologies are suspected.  **Do**: Implement speech recognition and other AI behavior analysis tools that comply with ISO/IEC 42001 standards to ensure ethical use. We will involve voice recognition software or other AI detection tools to analyze the behaviors during video conferences. Also, set up mandatory training for staff to detect the deepfake. In addition to the deepfake detection, if a financial transaction is initiated, the person conducting the transaction must complete two separate verification processes to confirm their identity. For example, after entering the password, they also need to enter a one-time PIN code or provide a fingerprint or facial recognition, which will be strictly implemented due to the ISO/IEC 27001 emphasis on information security.    **Check**: Monitor the efficiency and effectiveness of the above plans. Hire a third-party security audit to assess your company's defenses against deepfakes. The goal is to evaluate how well the organization protects against threats like deepfakes and whether it uses AI ethically and responsibly. The third-party auditors should have relevant certifications and qualifications about ISO 27001 and ISO 42001. Choose an audit firm that understands our industry and the specific challenges we are facing, especially when it comes to deepfakes. They should be able to adapt their audit approach to meet our unique needs.  **Act**: Update and refine algorithms that improve AI systems for better detection and responsiveness. Update policies and training programs to address any security vulnerabilities discovered.    **Competitive Advantage**: By complying with ISO/IEC 27001 and ISO/IEC 42001, we position ourselves as a leader in not only protecting data but also in the ethical use of AI. This dual compliance not only minimizes fraud losses but also enhances our social and user trust, presenting a unique competitive and ethical advantage.    **Technical or managerial requirements**: For the technical requirement, it includes the integration of multi-factor authentication technology, dedicated internal servers for storing encrypted biometric data, and compatibility of these systems with various communication platforms. For the managerial requirements, this plan requires mandatory and comprehensive staff training. The cybersecurity group needs to check the systems, reports, potential threats, etc. |

|  |
| --- |
| Plan of action and criteria for success In order to solve related ethics problems that may arise when using the above proposed solution in this company to avoid Deepfake incidents from happening again, we draw up a detailed plan of improvement action. The plan includes ensuring AI systems and software we use to protect individual privacy and obey data protection laws by redesigning or improving algorithms, especially training employees on how to use these AI systems and software, which aims to double guarantee that AI cannot make a violation of ethics.    In the first half of the plan, technologies will be pivotal in initial detection efforts. Related personnel need to be responsible for testing these new technologies. First, some employees should be trained to use these AI systems and softwares. After training, with the use of these new technologies, these employees need to participate in new verification processes and take more steps in their regular workflows. For example, they need to involve voice and facial recognition software to analyze the behaviors of the other side during the video conferences. Managers need to foster a supportive environment where employees feel confident to report any anomalies without fear of repercussions, ensuring swift and ethical incident response. Jobs will be more complex and technical because of the use of AI.    In the second half of the plan, people will be responsible for critical decision-making. This is because sometimes AI's mind doesn't match that of humans. Then some related ethical problems, including personal privacy leakage and data leakage, will be generated. For example, when we use voice and facial recognition software to analyze the behaviors of people involved during video conferences and ask participants to enter a related password with a one-time PIN code or a fingerprint to complete two separate verification processes, their personal data such as their facial features, fingerprints, and passwords may be stored by AI. According to data protection laws, this is not a secure method. Therefore, AI should keep the data secure, maintain the data only for as long as is necessary to accomplish the purpose, and use accurate, fair and representative data sets. Developers should ensure that data sets are broad and inclusive to reduce algorithmic bias when “teaching” AI and give users control of their data. After improving AI algorithms, employees who have been trained before can double check that AI follows ethics. For this, we can establish a protocol where any flagged transaction which is regarded as not matching human morality will be reconsidered.    After finishing the plan, employees can be upskilled. Their technical skills on designing and improving algorithms, using AI systems and softwares, and non technical skills on communicating with others, organizing training or activities can be improved.  In conclusion, the success of this plan will be evaluated through a decrease in successful fraud attempts, increased employee proficiency in managing the new system, and positive stakeholder feedback. Ethical performance indicators will also be monitored, ensuring that the deployment of AI aligns with our ethical commitments and the broader community's expectations. |

1. **Project Charter**

|  |  |
| --- | --- |
| **Business Case** | **Milestone** |
| An employee from the finance department was exposed to an AI deepfake scam with no awareness of double-checking the command of making transactions and confirming the identity of people involved (members in the video conference).  The company did not detect the scam on time (a week later) and was not sensitive to telecommunication fraud.  The technology applied in the company was outdated (failed to detect fake images and voices).  **Problem Statement:**  - Weak alertness of employees to AI fraud  - Insufficient cybersecurity training of employees  - Imperfect regulatory system in the company’s financial department  - Lack of strict auditing and monitoring system for ensuring the use and flow of each fund  - Outdated technology (biometric & cybersecurity) | |  |  | | --- | --- | | **Date** | **Milestone & Release** | | Jul 19 | Project Start (Define project charter) | | Jul 30 | Current Status AI-IS Study | | Oct 13 | Document Issues in Current System | | Oct 25 | To-Be Map | | Nov 15 | Expected Benefits Agreed by Company | | Nov 25 | Implement short term To-Be Map | | Nov 30 | HR: Cybersecurity Training | | Dec 12 | AF: Develop Regulatory Auditing System | | Dec 18 | IT: Analysis of AI deepfake | | Dec 20 | Work with IT on To-Be Map | | Jan 20 | IT: Establish Cybersecurity Department | | Jan 30 | IT: Plan and Develop Optimized Algorithm and Technology | | Feb 10 | IT: Third-party Cybersecurity Audit | |
| **Goal** | **Scope** |
| - To provide employees with sufficient training on cybersecurity and knowledge of related fraud.  - To establish a strict and systematic process on making transactions in financial department (requesting, reporting, documenting, multiple confirmations)  - To equip the company with advanced technologies and refined algorithms and build a robust telecommunication system. | **In:**  - Staffs that lack of cybersecurity knowledge  - Poorly-supervised financial department  - Outdated technology to cyber scam  **Out:**  - Employees with sufficient training in required area  - A regulatory auditing system in financial dept  - Enhanced technology and algorithms for detecting cyber fraud (a robust telecommunication system) |
| **Team** | **Benefits (+) & Cost (-)** |
| Team Leader: Tracy  Sponsor: EU  Members:  BSC (engineering): Vivian  CSO (strategy): Ivy, Lena  IT: Mark, Tom, Jerry | **Tangible:**  **(-)** Development of Optimized Algorithm  **(-)** Third-party Cybersecurity Audit  **(-)** Recruitment of cybersecurity specialists on training  **(-)** Expenses onestablishment of new department  **Intangible:**  **(+)** Improving trust and reputation in industry & society  **(+)** Minimizing losses caused by fraud (future)  **(-)** Time needed for employees to get familiar with the new telecommunication system |

|  |
| --- |
| Reference (If any) Brad,M. (2021, October 28). *4 Ways to Preserve Privacy in Artificial Intelligence*.  <https://www.boozallen.com/s/solution/four-ways-to-preserve-privacy-in-ai.html>  InfoSec. (2024, April 2). *InfoSec: Deepfake*.  <https://www.infosec.gov.hk/en/knowledge-centre/deepfake>  *ISO/IEC 27001:2022*. (2022, October). ISO. <https://www.iso.org/standard/27001>  *ISO/IEC 42001:2023*. (2023, December). ISO. <https://www.iso.org/standard/81230.html>  Kong, H. (2024, February 4). *‘Everyone looked real’: multinational firm’s Hong Kong office loses*  *HK$200 million after scammers stage deepfake video meeting*. South China Morning Post.  <https://www.scmp.com/news/hong-kong/law-and-crime/article/3250851/everyone-looked-real-multinational-firms-hong-kong-office-loses-hk200-million-after-scammers-stage> |

**END**