SLIIT UNT
THE KNOWLEDGE UNIVERSITY

IT4010 – Research Project - 2024

Topic Assessment Form

Project ID: **24-25J-137**

1. Topic (12 words max)

Multimodal Multimedia Integrity Verification: Detecting Deepfake Audio, Video, Tampered Files, News Content

2. Research group the project belongs to

Computing Infrastructure and Security (CIS)

3. Research area the project belongs to

Cyber Security (CS)

4. If a continuation of a previous project:

Project ID	
Year	

5. Brief description of the research problem including references (200 – 500 words max) – references not included in word count.

The swift expansion of AI technology on digital platforms has led to an increase in advanced modification methods, endangering the integrity of digital media. One of the biggest challenges among them is deepfake technology, which uses artificial intelligence to produce or modify audio, video, files and generation of fake news content with a high level of realism. Because of the ability of this technology to create incredibly realistic false media, strong detection methods are required to protect the legitimacy of digital communication and material. This can result in misinformation, identity theft, and other types of cybercrime. [1].

The core research problem addressed in this study is the urgent need for effective, efficient, and scalable methods capable of distinguishing between authentic and manipulated multimedia content. This includes deepfake audio and video, tampered files, and malicious news contents [2].

Technological advancements in manipulation techniques demand that detection methods evolve to recognize subtle anomalies and signatures of falsification. The challenge is not only technological but also includes ensuring scalability, maintaining high accuracy and reliability to avoid false positives, and navigating ethical and privacy considerations.

Specific research problems include,

- **Technological Sophistication:** As manipulation techniques become more advanced, detection methods must evolve to identify subtle anomalies and signatures of falsification.
- **Advanced Deepfake Generation**: Modern deepfake technologies leverage sophisticated neural network architectures like Generative Adversarial Networks (GANs), which are constantly evolving to produce more realistic and harder to detect fake content [3].
- **Micro-expression Manipulation**: Emerging techniques now focus on altering micro-expressions in video content, a subtle but powerful way to deceive viewers, which requires advanced detection tools capable of analyzing these minute details [4].
- **Audio Mimicry**: The rise of voice synthesis technologies that can mimic a person's voice with just a few samples poses significant threats to personal and corporate security, necessitating advanced audio analysis tools [5].
- **AI-Driven Content Tampering**: The use of AI to automate the tampering of files at scale, which could lead to widespread misinformation or data breaches if not adequately detected and mitigated [6].

- **Accuracy and Reliability:** It is essential to maintain high accuracy and reliability in detection methods to minimize false positives, which could undermine the credibility of detection mechanisms.
- **Ethical and Privacy Considerations:** The detection process must carefully navigate ethical boundaries and privacy concerns, ensuring that measures to detect and mitigate deepfake content do not infringe on individual rights or lead to unwarranted surveillance.

References

[1] M. Chawki, Navigating legal challenges of deepfakes in the American context: a call to action, Cogent Engineering, 2024.

[2] L. J. H. S. L. S. L. &. C. N. Yuan, "Sustainable Development of Information Dissemination: A Review of Current Fake News Detection Research and Practice.," 2023.

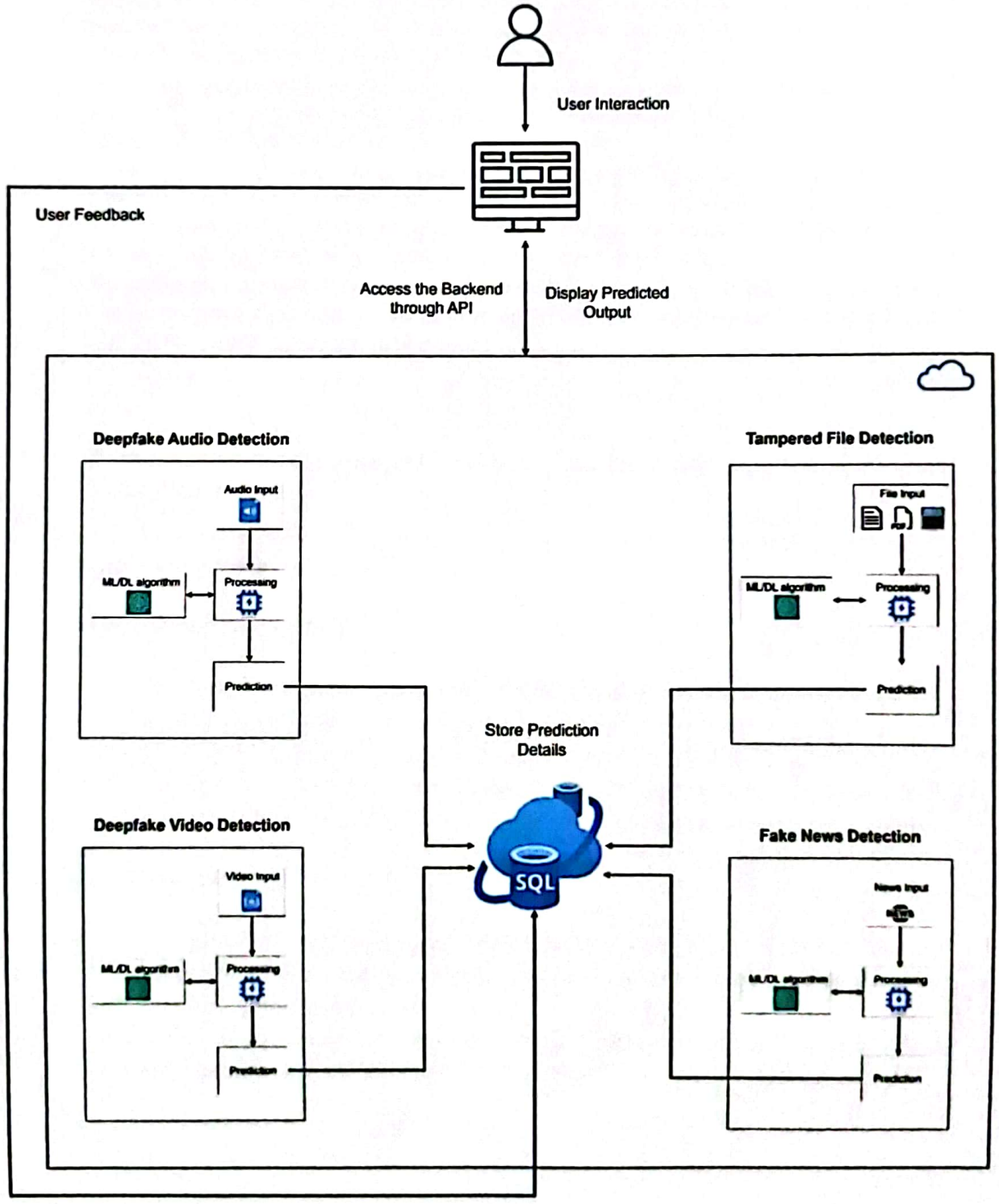
[3] R. V.-R. R. F. J. M. A. &. O.-G. J. Tolosana, "Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection.," in *Information Fusion*, 2022.

[4] T. T. N. C. M. L. S. &. Y. G. Nguyen, "Deep Learning for Micro-expression Recognition: A Survey.," 2021.

[5] Y. Z. J. S. S. &. C. X. Jia, "Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis. Neural Information Processing Systems," 2021.

[6] R. H. A. B. Y. F. A. &. C. Y. Zellers, "Defending Against Neural Fake News. Advances in Neural Information Processing Systems (NeurIPS)," 2021.

6. Brief description of the nature of the solution including a conceptual diagram (250 words max)



Our planned solution is an advanced, scalable, and reliable Multimodal Multimedia Integrity Verification system designed to detect deepfake audio, video, tampered files, and fabricated news content. This desktop application will be an invaluable tool for investigations by police and other legal authorities.

The system will leverage sophisticated machine learning techniques and neural network architectures to identify subtle anomalies and falsifications in multimedia content, ensuring high accuracy while minimizing false positives.

It will incorporate psychoacoustic analysis and bio-inspired algorithms for deepfake audio detection, hybrid neural network architectures for deepfake video detection, a combination of traditional forensic and deep learning methods for tampered file detection, and advanced Natural Language Processing techniques for fake news detection.

Comprehensive datasets of genuine and manipulated samples will train and refine these models, maintaining ethical standards to avoid privacy infringements and ensuring trust in digital communication and content.

7. Brief description of specialized domain expertise, knowledge, and data requirements (300 words max)

1. Deep Fake Audio Detection

Expertise and Knowledge:

- Psychoacoustic analysis and audio signal processing are essential, with a deep understanding of machine learning techniques specific to audio.
- Proficiency in using Convolutional Neural Networks (CNNs) for spectral feature extraction and Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, for capturing temporal dependencies in audio signals.

Data Requirements:

- Comprehensive datasets containing both genuine and synthetic audio samples are necessary to train models in recognizing nuanced differences brought about by deepfake technologies.
- These datasets should include a wide range of vocal attributes, such as tone, pitch, and background noise variations.

2. Deep Fake Video Detection

Expertise and Knowledge:

- Knowledge in video processing and digital imaging, specifically in identifying spatial and temporal discrepancies using advanced neural network architectures.
- Familiarity with Transformer models like Vision Transformers (ViT) and Temporal Convolutional Networks (TCNs) to analyze both frame-level and dynamic video features.

Data Requirements:

- Extensive video datasets that include varied manipulations to test and refine detection algorithms. These datasets should encompass different facial expressions, movements, and lighting conditions to ensure robustness against deepfake manipulations.

3. Tampered File Detection

Expertise and Knowledge:

- knowledge in examining the intricate details of file metadata, binary content, and filesystem integrity checks is crucial.
- Familiarity with sophisticated machine learning algorithms for anomaly detection. knowledge in using ensemble methods like Decision Trees, Gradient Boosting Machines (GBMs), and Random Forests to efficiently differentiate between original and tampered files. Understanding of feature engineering to extract significant attributes from files that effectively signal tampering.
- Knowledge in how emerging technologies such as **noise level analysis and ELA (Error level analysis)** can be used in this.
- Knowledge **Autoencoders** for unsupervised learning tasks such as feature extraction or anomaly detection in data.

Data Requirements:

- Extensive repositories of files in various formats. These datasets should represent a wide range of tampering scenarios from simple metadata alterations to complex content modifications. It should A mix of genuinely tampered files from real-world incidents and synthetically altered files created to train detection models on rare or emerging tampering techniques to enable model's capability to generalize across unseen tampering methods in operational environments.

4. Fake News Detection

Expertise and Knowledge:

- Interdisciplinary knowledge in journalism, AI, and natural language processing (NLP), with the ability to integrate text-based anomaly detection for detection of fake news.
- Understanding in machine learning techniques like K-Nearest Neighbors (KNN), Logistic Regression, and Support Vector Machines (SVM), alongside deep learning architectures such as CNNs and LSTMs.

Data Requirements:

- Multimodal datasets that include a variety of fake and real news sources. These should also incorporate annotations that help distinguish factual from fabricated information.

8. Objectives and Novelty

Main Objective

The main objective of the Multimodal Multimedia Integrity Verification system is to develop an advanced, scalable, and reliable framework capable of detecting deepfake audio, video, tampered files, and fabricated news content to protect the integrity, confidentiality, and availability of digital media. By leveraging sophisticated machine learning techniques and neural network architectures, the system aims to identify subtle anomalies and falsifications in multimedia content, ensuring high accuracy and minimizing false positives. This includes leveraging advanced psychoacoustic analysis and sophisticated audio signal processing techniques to discern deepfake audio, deploying state-of-the-art video processing and digital imaging methodologies to uncover deepfake video anomalies, employing complicated anomaly detection algorithms that analyze file metadata, binary content, and filesystem integrity for tampered file identification, and integrating cutting-edge natural language processing and multimodal analysis strategies to detect fake news by analyzing textual content. Comprehensive datasets encompassing genuine and manipulated samples are essential to train and refine these detection models. Additionally, the system must navigate ethical and privacy considerations to prevent infringement on individual rights or unwarranted surveillance, thus maintaining trust and legitimacy in digital communication and content.

Member Name	Sub Objective	Tasks	Novelty
Shehara M. G. D	**Deepfake Audio Detection** To develop an advanced deepfake audio detection system that leverages psychoacoustic analysis, bio-inspired algorithms, and sophisticated audio signal processing techniques to detect deepfake audio by analyzing both technical discrepancies and emotional impacts of audio cues, ensuring high detection accuracy.	**1. Data Collection and Preprocessing.** • Include both genuine and deepfake audio files covering different languages, accents, and background noise levels. • Preprocess data for uniformity (remove irrelevant features, normalize inputs). **2. Feature Analysis** • Analyze audio using psychoacoustic parameters to detect discrepancies typical of manipulated content using tools and techniques that assess how sound is perceived by humans, such as tonality, loudness, and temporal dynamics. **3. Model Development** • Willing to develop a sophisticated machine learning model using RNNs and CNNs to analyze audio signals. • Willing to employ RNNs to capture temporal dependencies and CNNs for	• Introduces a groundbreaking approach to deep fake audio detection by leveraging bio-inspired algorithms that mimic human auditory perception processes. Unlike conventional methods that primarily focus on technical discrepancies in audio files, **our methodology assesses the emotional and psychological impacts of audio cues**, enabling a more nuanced detection of deep fakes. This innovative perspective enhances detection accuracy, especially in audio clips engineered to evoke specific emotional responses, thus representing a significant advancement in the field of digital forensics and cybersecurity.

		extracting hierarchical spectral features from audio data. **4. Model Training and Evaluation** • Fit the model on distributed dataset. • Use performance metrics such as accuracy, precision, recall, and F1-score to assess the model's capability to detect deepfakes accurately. • Keep the model up to date with new data to stay current with the evolving threats. **5. Deployment and Monitoring** • Deploy the model under the application. • Monitor predictions and adapt thresholds to balance sensitivity and specificity. • Establish a feedback loop for continuous model refinement based on performance and new user feedback.	• Utilizing psychoacoustic parameters for a deeper, more intuitive understanding of fake audio detection.
Dissanayake W. P. D. B	**Deepfake Video Detection** To create a pioneering deepfake video detection framework that	**1. Data Collection and Preprocessing** • Collect a diverse set of video samples that include various forms of manipulations.	• Pioneers the use of **hybrid neural network architectures that combine convolutional neural networks (CNNs) with generative adversarial networks (GANs) for deep fake video detection.** This hybrid model is designed to capture both the

	utilizes hybrid neural network architectures to capture both spatial inconsistencies and temporal anomalies in video sequences, setting a new benchmark for accuracy and efficiency.	• Gather videos with different facial expressions, movements, and lighting conditions from multiple sources. **2. Feature Analysis** • Utilize advanced neural networks to analyze spatial and temporal discrepancies in video data. • Willing to Implement Vision Transformers (ViT) for frame-level analysis and Temporal Convolutional Networks (TCNs) for dynamic feature extraction. **3. Model Development** • Willing to Develop an integrated model combining ViT and TCNs to enhance detection capabilities. • Willing to deploy ViT and TCNs together to provide a comprehensive analysis of both spatial and temporal data within videos.	spatial inconsistencies and temporal anomalies in video sequences, a method not extensively explored in existing literature. • By integrating the strengths of CNNs in recognizing image features and GANs in understanding video dynamics, the proposed model sets a new benchmark for accuracy and efficiency in detecting sophisticated deep fake videos.

		4. Model Training and Evaluation • Conduct extensive testing of the model using real-world and synthetically generated deepfake videos. • Apply rigorous testing protocols to assess the model on unseen data. • Evaluate the model using accuracy, precision, recall, and F1-score. **5. Deployment and Monitoring** • Deploy the model under the application. • Establish a feedback loop for continuous model refinement based on performance and new user feedback.	

Zakey M. S. M. A – IT21299902	**Tampered File Detection** To enhance tampered file detection capabilities by integrating traditional forensic techniques with advanced deep learning methods, particularly using Autoencoders, to create a robust, adaptable, and scalable detection system capable of handling various file formats and new tampering techniques.	**1. Data Collection and Preprocessing** • Accumulate a comprehensive repository of files with known tampering cases. • Include various file types with different tampering techniques. • Preprocess data for uniformity (remove irrelevant features, normalize inputs). **2. Feature Analysis** • Employ advanced techniques to analyze file metadata, binary content, and visual signs of tampering. • **Noise Level Analysis**: willing analyze the statistical distribution of noise across file contents to identify patterns indicative of tampering. • **Error Level Analysis (ELA)**: willing to apply ELA detect inconsistencies in the compression levels, which often reveal areas of a file that have been altered. **3. Model Development** • Willing to develop a sophisticated detection system using a combination	• The **integration of Autoencoders into the tampered file detection framework** presents a significant advancement over the proposed solution by **adopting a hybrid approach** that synergistically combines traditional machine learning techniques with advanced deep learning methods. • This innovative model enhances the detection capabilities by leveraging both explicit feature engineering and implicit feature learning. The use of Autoencoders allows for unsupervised learning of normal file patterns, enabling the system to detect anomalies based on reconstruction errors, which is particularly effective against new or sophisticated tampering techniques not covered by existing patterns. This approach not only improves the accuracy and adaptability of the detection system but also introduces a

		of traditional, advanced machine learning and deep learning techniques. • *Ensemble Learning:* Willing to utilize Decision Trees, Gradient Boosting Machines (GBMs), and Random Forests for robust classification of files based on likelihood of tampering. • **Incorporation of Noise and ELA Features:** Willing to integrate features derived from noise level and ELA into the learning models to enhance detection accuracy. **4. Model Training and Evaluation** • Test the system using a mixed set of files, some of which will include newly introduced tampering methods to assess the system's adaptability. • Evaluate the system using metrics such as accuracy, precision, recall, and the rate of false positives/negatives. **5. Deployment and Monitoring** • Deploy the model under the application. • Establish a feedback loop for continuous model refinement based	self-learning capability, which is crucial for adapting to evolving tampering methods. Additionally, the integration of traditional forensic techniques such as Error Level Analysis (ELA) and noise level analysis adds a dual layer of analysis, ensuring higher detection rates of subtle manipulations that might be missed by each method individually. By combining these diverse methodologies, the proposed system sets a new standard in tampered file detection, offering scalability across various file types and adaptability to new threats, making it a robust and versatile solution

		on performance and new user feedback.	
Hasara L. A. N	**Fake News Detection** To develop an integrated, multimodal fake news detection system that combines advanced Natural Language Processing (NLP) techniques within a unified framework, enhancing detection capabilities by cross-verifying information across different modalities and ensuring high accuracy and adaptability to emerging fake news patterns.	**1. Data Collection and Preprocessing** • Gather a dataset from varied news sources. • Include datasets that encompass both real and fabricated news content. **2. Feature Analysis** • Apply NLP techniques for deep textual analysis. • Utilize machine learning models such as KNN, Logistic Regression, and SVM for text and deep learning architectures such as CNNs and LSTMs. **3. Model Development** • Willing to integrate text within a unified framework using advanced ML architectures. • Willing to employ a combination of machine learning techniques to process and analyze integrated data.	• The novelty of the proposed fake news detection system lies in its integrated, **multimodal approach that combines advanced Natural Language Processing (NLP) techniques within a unified framework.** • This system is designed to analyze and verify news content from diverse sources by processing text simultaneously, leveraging state-of-the-art machine learning models such as K-Nearest Neighbors (KNN), Logistic Regression, and Support Vector Machines (SVM), CNNs and LSTMs. • This holistic approach allows for a more comprehensive and accurate identification of fake

		4. Model Training and Evaluation • Perform extensive testing to validate the model's effectiveness in detecting fake news. • Test the model against a set of new, unseen datasets to ensure its robustness and reliability. • Measure the model's performance through accuracy, precision, recall, and specificity. **5. Deployment and Monitoring** • Deploy the model under the application. • Establish a feedback loop for continuous model refinement based on performance and new user feedback.	news, as it can cross-verify information across different modalities, enhancing the detection capabilities beyond what is achievable with single-mode analysis. Additionally, the system's ability to learn from new data continuously and adapt to emerging fake news patterns ensures its long-term effectiveness and applicability in real-world scenarios, setting it apart from existing solutions that might rely on more static or single-dimension analysis methods.

SLIIT UNT
THE KNOWLEDGE UNIVERSITY

IT4010 – Research Project - 2024

Topic Assessment Form

9. Supervisor checklist

a) Does the chosen research topic possess a comprehensive scope suitable for a final-year project?

Yes	✓	No	

b) Does the proposed topic exhibit novelty?

Yes	✓	No	

c) Do you believe they have the capability to successfully execute the proposed project?

Yes	✓	No	

d) Do the proposed sub-objectives reflect the students' areas of specialization?

Yes	✓	No	

e) Supervisor's Evaluation and Recommendation for the Research topic:

Minor refinements can be done at the proposal Stage.

10. Supervisor details

	Title	First Name	Last Name	Signature
Supervisor	Mr.	Kavinga	Yapa Abeywardena	
Co-Supervisor	Dr.	HARINDA	FERNANDO	
External Supervisor				
Summary of external supervisor's (if any) experience and expertise				

SLIIT
THE KNOWLEDGE UNIVERSITY

IT4010 – Research Project - 2024

Topic Assessment Form

This part is to be filled by the Topic Screening Panel members.

Acceptable: Mark/Select as necessary

Topic Assessment Accepted	
Topic Assessment Accepted with minor changes (should be followed up by the supervisor)*	
Topic Assessment to be Resubmitted with major changes*	
Topic Assessment Rejected. Topic must be changed	

* Detailed comments given below

Comments

The Review Panel Details

Member's Name	Signature

•**Important:**

1. According to the comments given by the panel, make the necessary modifications and get the approval by the **Supervisor** or the **Same Panel.**
2. If the project topic is rejected, identify a new topic, and follow the same procedure until the topic is approved by the assessment panel.