HAVDEF (Hindi Audio-Visual Deepfake Defense)

Capstone Project Proposal

Submitted by:

(102203191)	Shivane Kapoor
(102203194)	Kaustubh Singh
(102203205)	Japneet Singh
(102203499)	Arpit Jain
(102253002)	Diwakar Narayan Sood

BE Third Year –

COE

CPG No. 207

Under the Mentorship of

Dr. Seema Bawa

Professor

Dr. Sachin Kansal

Assistant Professor

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Computer Science and Engineering Department Thapar

Institute of Engineering and Technology, Patiala

January-2025

TABLE OF CONTENTS

Mentor Consent Form

I hereby agree to be the mentor of the following Capstone Project Team

Project Title: **HAVDEF (Hindi Audio-Visual Deepfake Defense)**		
Roll No	**Name**	**Signatures**
102203191	Shivane Kapoor	
102203194	Kaustubh Singh	
102203205	Japneet Singh	
102203499	Arpit Jain	
102253002	Diwakar Narayan Sood	

NAME of Mentor: Dr. Seema Bawa

SIGNATURE of Mentor: . . . . . .

NAME of Co-Mentor: Dr. Sachin Kansal

SIGNATURE of Co-Mentor: . . . . . .

Project Overview

Hindi Audio-Visual Deepfake Defense (HAVDEF) is a deepfake detection system that is designed to identify AI-generated voice fraud in real-time phone calls. With the rise of AI-generated voices in scams, trusted individuals are impersonated by fraudsters, making it crucial for a system to be developed that detects and prevents such activities. HAVDEF has been specifically designed for phone calls in Hinglish, a blend of Hindi and English commonly spoken in India, ensuring better accuracy in the detection of deepfake voice scams in real-world scenarios. By addressing language and accent-specific challenges, a more robust defense against AI-driven fraud is provided by the system.

Advanced machine learning techniques are employed by the system to analyze voice patterns and identify AI-generated speech. Deep learning models are trained on a diverse dataset containing both real and synthetic Hinglish voice samples, enabling precise differentiation between human and AI-generated voices. HAVDEF is run continuously in real-time, allowing incoming phone calls to be processed, speech features to be extracted, and suspicious activity to be instantly flagged. If a potential deepfake voice is detected, an alert is sent to the user immediately, allowing necessary precautions to be taken and fraud to be avoided.

HAVDEF will be designed to function efficiently on mobile devices, making it highly accessible and easy to integrate into everyday communication. Advanced technologies like signal processing, spectrogram analysis, and neural networks are incorporated by the system to classify and detect deepfake voices with high accuracy. The development process includes dataset collection, model training, and real-world testing to ensure reliability across different voices, accents, and environmental conditions. By offering a proactive approach to deepfake detection, user security is enhanced by HAVDEF, and financial and personal harm caused by AI-generated voice scams is prevented.

Need Analysis

In recent years, the rise of AI-generated deepfake content has been recognized as a significant threat to personal security, privacy, and trust in digital communication. A report published in 2023 [1] highlights how advancements in deep learning models have made it increasingly difficult for real and AI-synthesized voices, especially in audio domains, to be distinguished. While existing deepfake detection solutions are primarily focused on pre-recorded audio, the ability to address real-time fraud attempts, such as fraudulent AI-generated phone calls, is lacking. This creates a pressing need for systems capable of detecting audio deepfakes in real time.

In the Indian context, Hinglish (a blend of Hindi and English) is widely used in daily communication, especially in phone conversations. However, the linguistic patterns of Hinglish are not adequately addressed by existing systems, which are largely catered to English or Hindi. The importance of region-specific datasets and language-adapted detection models for achieving accurate results has been emphasized in studies like [2] and [3]. This gap highlights the need for a solution tailored to Hinglish, ensuring robust fraud detection in linguistically diverse scenarios.

The increasing sophistication of AI-driven scams, particularly through deepfake audio, has led to financial and emotional exploitation of individuals. Proactive measures to combat such threats are emphasized by government reports and cybersecurity agencies. A mobile-based solution, such as HAVDEF, is designed to address this need by leveraging advanced speech pattern analysis, spectrogram-based feature extraction, and language-specific machine learning techniques to identify and alert users of AI-generated fraud calls in real time.

By providing a practical, user-friendly, and real-time detection system, individuals are empowered by HAVDEF to protect themselves from emerging AI-driven threats. The relevance of the solution is further enhanced by the integration of Hinglish-specific detection in the Indian context, making it a critical tool for securing digital communication in an increasingly AI-driven world.

Literature Survey

Deepfake detection systems aim to identify and mitigate the risks posed by AI-generated fraudulent content, particularly in audio and video formats. These systems rely on several core concepts and technologies, including:

1. **Deepfake Detection Technique:**

   Various detection methods, including speech pattern analysis, spectrogram-based features, and deep learning models, have been explored for the identification of AI-generated voices. The effectiveness of spectrogram analysis and speech inconsistencies, such as unnatural pauses and breathing patterns, in detecting synthetic audio has been highlighted in studies [4], [5], [6], [7].

2. **AI Models for Deepfake Audio Detection:**

   It has been demonstrated in research that CNNs, RNNs, and transformer-based models can be effectively used for deepfake audio detection. Pre-trained models like WavLM and Whisper have been fine-tuned for detecting AI-generated speech, leading to improvements in accuracy in real-world applications [8], [9].

3. **Language-Specific Challenges:**

   The need for region-specific datasets has been emphasized, particularly for Hindi and Hinglish deepfake detection. Critical insights into language-specific fraud detection have been provided by the HAV-DF dataset and research on AI-generated Hindi speech [2], [3].

4. **Real-Time Detection and Noise Reduction**:

   The importance of real-time processing for detecting fraudulent calls has been stressed in studies. Detection efficiency is enhanced by implementing low-latency AI models and noise suppression techniques, making deepfake detection more practical in real-time scenarios [5], [7].

5. **Ensemble Learning and Feature Extraction:**

   Spectrogram-based analysis, pause pattern detection, and breathing-talking-silence encoders (BTS-E) have been identified as crucial techniques for improving detection accuracy [6], [9].

Existing Solution

There exist several audio deepfake detection systems and research projects that are aimed at identifying and mitigating AI-generated audio fraud. The following are some notable ones:

1. **WavLM Model Ensemble for Audio Deepfake Detection:**
   WavLM is an ensemble of deep learning models that has been specifically designed for speech processing. Audio features such as pitch, tone, and spectrograms are analyzed by this approach to accurately detect AI-generated voices. High performance has been demonstrated in differentiating synthetic voices from real human speech [1].

2. **BTS-E: Audio Deepfake Detection Using Breathing-Talking-Silence Encoder:**
   Deepfake voices are detected by the BTS-E model by focusing on irregular breathing patterns, unnatural talking rhythms, and silence intervals. The inefficiencies in AI-generated voices when replicating natural human breathing and speech pauses are highlighted by this approach [9].

3. **Deepfake Voice Detection Using Speech Pause Patterns:**
   The differences in speech pause patterns between real and AI-generated voices are investigated by this system. Unnatural timing, rhythm, and duration of silences in deepfake voices are identified to enhance detection capabilities [2].

4. **Deepfake Audio Detection Using Spectrogram-Based Features:**
   In this approach, spectrogram-based feature extraction is used, combined with an ensemble of deep learning models, to improve detection accuracy. By leveraging detailed audio features from spectrograms, authentic and synthetic voices are effectively distinguished [7].

Table 1: Comparison of Existing Technologies and the Proposed System

Feature	Existing Solutions	HAVDEF (Proposed Solution)	References
Language Support	✗ English/Hindi only	✓ Hinglish (Hindi-English mix)	[2], [3]
Dataset	✓ Pre-recorded speech datasets	✓ Custom-built Hinglish fraud call dataset	[2], [3]
Real-Time Detection	✗ Offline analysis, no real-time support	✓ Real-time fraud detection & alerts	[6], [7], [1]
Deployment Feasibility	✗ Research-based, no real-world applications	✓ Practical mobile app for real use	[5], [6], [7], [1]
Fraud Call Focus	✗ Generic deepfake detection	✓ Explicitly targets AI-based fraud calls	[1], [2], [5], [6], [9]

Research Findings

Some studies have explored the effectiveness and challenges associated with deepfake detection applications. Key research findings include:

1. **Spectrogram-Based Analysis:** Spectrograms are highly effective in capturing the nuances of audio features, making them a reliable input for deepfake detection models [7], [6].
2. **Speech Pause Patterns:** Utilizing speech pause patterns significantly improves the detection of AI-generated voices, especially in detecting unnatural rhythm and timing [2], [8].
3. **Real-Time Processing:** Several studies emphasize real-time applicability by optimizing processing speed without compromising detection accuracy [3], [5].
4. **Breathing-Talking-Silence Encoder:** This model leverages the analysis of breathing, talking, and silence patterns to detect inconsistencies in audio deepfakes, providing high accuracy in multilingual datasets [9].

Objectives

The objectives of this project are to:

1. **Develop a Real-Time AI Fraud Detection System:** Build a system capable of analyzing phone call audio in real time to identify AI-synthesized voices and instantly alert users.

2. **Implement Hinglish Language Support:** Design language-specific detection techniques for Hinglish (a mix of Hindi and English) to ensure accurate fraud detection in Indian phone conversations.

3. **Create a User-Friendly Mobile Application**: Develop a smartphone-based solution that enables easy fraud call detection and prevention without requiring technical expertise.

Methodology

The methodology of this project is designed to systematically achieve the outlined objectives through a structured approach. So the methodologies are:

1. Real-Time Detection of AI-Generated Fraud Calls

1.1 Audio Preprocessing: Noise reduction and feature extraction techniques will be implemented to improve voice clarity.

1.2 Real-Time Streaming Analysis: WebRTC or VoIP-based frameworks will be integrated for capturing live audio during phone calls.

1.3 Deepfake Detection Models: Machine learning algorithms (CNNs, RNNs, transformers) will be utilized, trained on real and AI-generated speech data, to classify incoming audio in real time.

1.4 Alert System Implementation: An automated alert mechanism will be developed to notify users when an AI-generated voice is detected.

1.5 Latency Optimization: Low-latency processing will be ensured by optimizing model size and computation speed for real-time performance.

2. Hinglish Language Support

2.1 Dataset Collection: A dataset containing real and AI-generated Hinglish voice samples will be curated.

2.2 Phonetic and Linguistic Analysis: Deep learning models will be trained on Hinglish-specific phonetic patterns, pronunciation variations, and code-mixing behaviors.

2.3 Pre-Trained Model Adaptation: Existing language models (e.g., Wav2Vec2, Whisper) will be fine-tuned for Hinglish speech recognition and analysis.

2.4 Feature Engineering: Unique linguistic and tonal features will be extracted to distinguish between human and synthetic Hinglish speech.

2.5 Performance Testing: Model accuracy will be validated using Hinglish-specific fraud scenarios, and detection capabilities will be refined accordingly

.

3. User-Friendly Mobile Application Deployment

3.1 Cross-Platform Development: React Native or Flutter will be utilized to ensure a seamless mobile experience across Android and iOS.

3.2 Lightweight Model Deployment: AI models will be optimized for mobile environments using TensorFlow Lite or ONNX.

3.3 User Interface (UI/UX) Design: An intuitive interface will be developed with real-time alerts, fraud call history, and customizable settings.

3.4 Integration with Call Handling: Background call analysis will be implemented to detect fraud calls without interrupting conversations.

Project Outcomes & Individual Roles

Upon Completion of project following will be the outcome:

1. **Real-Time AI Fraud Detection System:** The system will analyze phone call audio in real time, detecting AI-generated voices instantly and providing low-latency fraud alerts to users.
2. **Hinglish-Specific Deepfake Detection:** A language-adapted model will be developed to accurately identify deepfake voices in Hinglish conversations, addressing India's linguistic diversity.
3. **User-Friendly Mobile Application:** A smartphone-based application will feature an intuitive interface, allowing users to receive real-time fraud alerts with customizable notification settings.
4. **Advanced AI-Powered Detection Techniques:** The system will integrate speech pattern analysis, spectrogram-based feature extraction, and deep learning models (CNNs, RNNs, transformers) for high detection accuracy.
5. **Custom Hinglish Deepfake Dataset:** A comprehensive dataset containing real and AI-generated Hinglish speech samples will be created to enhance the accuracy of fraud detection in Indian speech patterns.
6. **Scalable & Secure Telecommunication Solution:** A cost-effective and scalable system will be implemented to reduce fraudulent activities and enhance trust in mobile communication.

Individual Roles

Table 2: The list of the Individual role

Name	Role	
Shivane Kapoor	Data Preprocessing	Model Development
Kaustubh Singh	Documentation	Speech Processing
Japneet Singh	Data Collection	Frontend & Backend
Arpit Jain	Model Development	Testing & Optimization
Diwakar Narayan Sood	Documentation	Frontend & Backend

Work Plan

Phase 1: Requirement Analysis and System Design (Jan 2025 – Mar 2025)

a. Conduct research on AI fraud calls in India and Hinglish speech patterns.

b. Finalize use cases, workflows, and system architecture.

Phase 2: AI Model Development and Data Collection (Apr 2025 – Jul 2025)

a. Collect and preprocess Hinglish audio datasets.

b. Validate model performance on real-world datasets.

Phase 3: Application Development and Integration (Jun 2025 – Oct 2025)

a. Build a real-time fraud detection app for Android.

b. Integrate AI models and implement user alerts for fraud detection.

c. Conduct system integration testing.

Phase 4: Testing, Optimization, and Deployment (Aug 2025 – Dec 2025)

a. Optimize AI models for real-time detection.

b. Conduct user trials, refine the system, and prepare technical documentation.

c. Deploy the final version and prepare for project presentation.

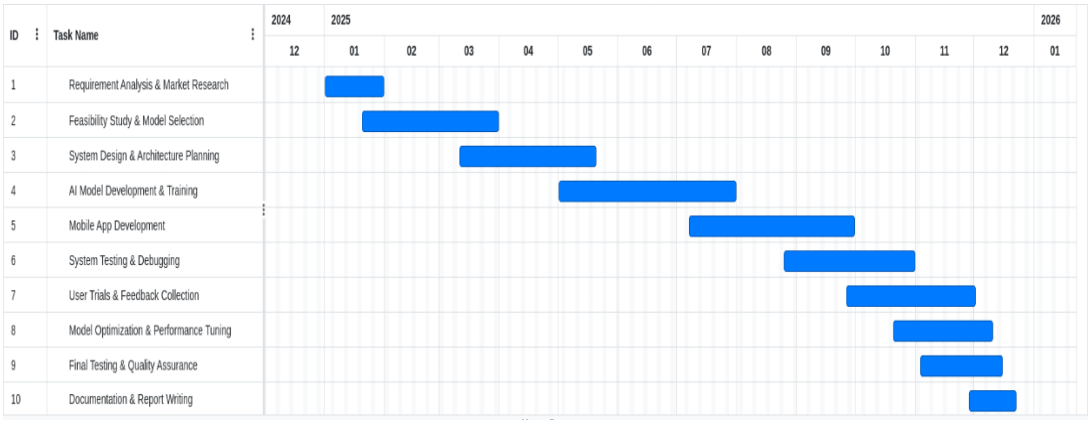

Figure 1: Gantt chart for the Work plan

Course Subjects

The following course subjects were instrumental in the successful execution of the HAVDEF (Hindi Audio-Visual Deepfake Defense) project, covering various aspects of computer science and artificial intelligence:

1. Software Engineering (UCS503)

Developed and maintained a modular and scalable software infrastructure, incorporating UML diagrams and structured coding practices to ensure efficient real-time deepfake detection.

2. Artificial Intelligence (UCS411)

Implemented AI-driven decision-making techniques to enhance fraud detection accuracy by identifying deepfake voice patterns in real-time phone calls.

3. Machine Learning (UML501)

Trained and fine-tuned machine learning models on real and synthetic Hinglish voice datasets, leveraging neural networks and spectrogram analysis for deepfake detection.

4. Natural Language Processing (UCS672)

Implemented multilingual and accent-robust deepfake detection techniques, enhancing the system's ability to analyze diverse linguistic variations in Hinglish phone calls.

5. Data Science & Signal Processing (UCS548)

Processed and extracted key features from large datasets using spectrogram analysis and signal processing techniques to differentiate between human and AI-generated voices.

6. Cybersecurity & Fraud Prevention (UCS534)

Developed an AI-powered challenge-response mechanism that dynamically tests callers using phonetic-based challenges to detect deepfake inconsistencies in real-time.

7. Object-Oriented Programming (UTA018)

Applied OOP principles like encapsulation, inheritance, and polymorphism to structure the deepfake detection system efficiently, ensuring modularity, code reusability

References

[1] D. Combei, A. Stan, D. Oneață, and H. Cucu, "WavLM model ensemble for audio deepfake detection," *Technical University of Cluj-Napoca & POLITEHNICA Bucharest, Romania*, 2023. Available: https://paperswithcode.com/paper/wavlm-model-ensemble-for-audio-deepfake

[2] S. Kaura, M. Buhari, N. Khandelwal, P. Tyagi, and K. Sharma, "Hindi audio-video-deepfake (HAV-DF): A Hindi language-based audio-video deepfake dataset," *arXiv preprint*, 2023. Available: https://arxiv.org/abs/2411.15457

[3] K. Bhatia, A. Agrawal, P. Singh, and A. K. Singh, "Detection of AI synthesized Hindi speech," *arXiv preprint*, 2023. Available: https://arxiv.org/abs/2203.03706

[4] J. Yi, C. Wang, J. Tao, X. Zhang, C. Y. Zhang, and Y. Zhao, "Audio deepfake detection: A survey," *arXiv preprint arXiv:2304.08531*, 2023. Available: https://arxiv.org/abs/2308.14970

[5] A. Shetty, H. Karani, S. K. H., R. Khan, and A. G. Amruth, "Deepfake audio detection using deep learning," *IEEE Xplore*, 2023. Available: https://ijarcce.com/papers/deepfake-audio-detection-using-deep-learning/

[6] H. H. Kilinc and F. Kaledibi, "Audio deepfake detection by using machine and deep learning," *IEEE Xplore*, 2023. Available: https://ieeexplore.ieee.org/document/10323004

[7] L. Pham, P. Lam, T. Nguyen, H. Nguyen, and A. Schindler, "Deepfake audio detection using spectrogram-based feature and ensemble of deep learning models," *arXiv preprint*, 2023. Available: https://arxiv.org/abs/2407.01777

[8] N. V. Kulangareth, J. Kaufman, J. Oreskovic, and Y. Fossat, "Investigation of Deepfake Voice Detection Using Speech Pause Patterns: Algorithm Development and Validation," *PubMed*, 2023. Available: https://pubmed.ncbi.nlm.nih.gov/38875685/

[9] T.-P. Doan, L. Nguyen-Vu, S. Jung, and K. Hong, "BTS-E: Audio Deepfake Detection Using Breathing-Talking-Silence Encoder, 2023" Available: https://ieeexplore.ieee.org/abstract/document/10095927