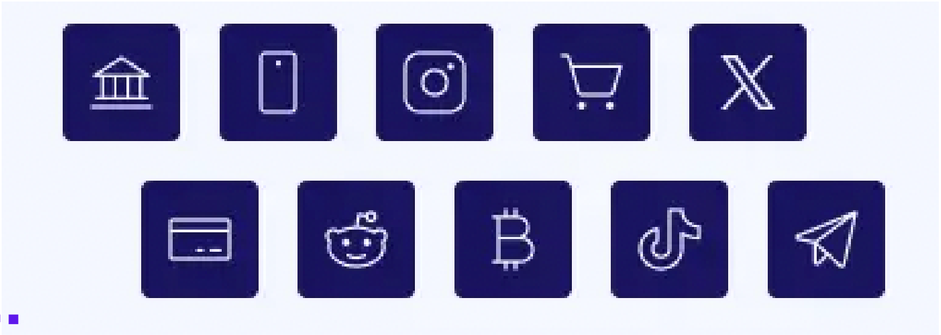
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

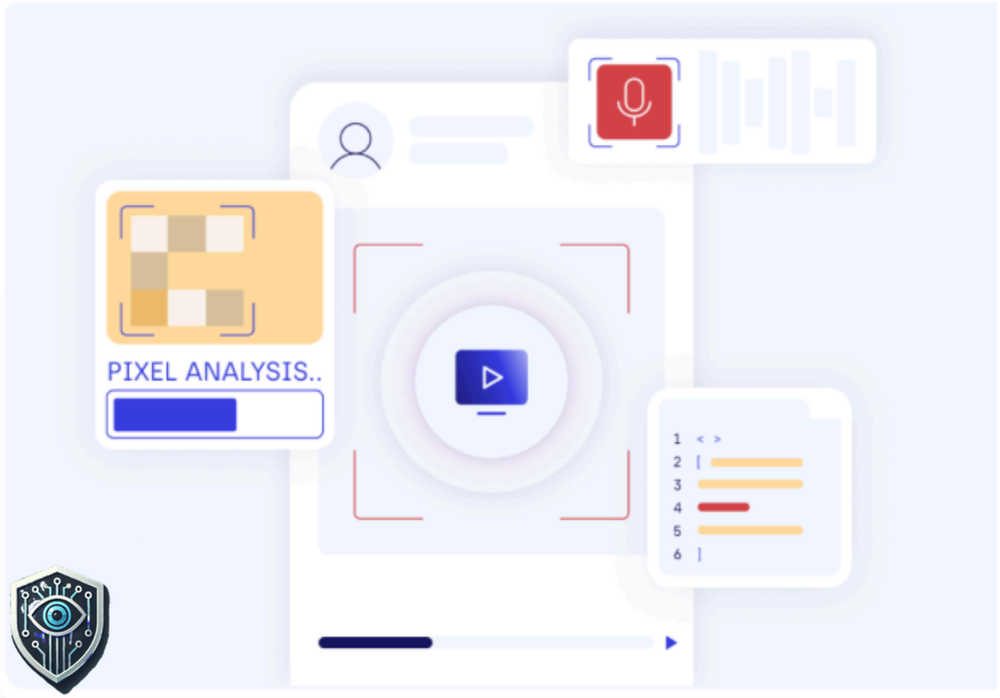
DeepShield

Detecting AI-generated deepfake videos and cloned voices in real-time!

Team: DeepEdu Explorers

Project Summary





Key Content:

- Face Manipulation: Identifies AI-altered faces, including deepfake swaps, lip-syncing, and face morphing, to combat deception in digital content.
- Generative AI: Detects synthetic images and videos generated by advanced AI models, ensuring authenticity in visual media.
- Voice Cloning: Recognizes AI-generated speech and cloned voices using deep learning, safeguarding audio integrity in calls and recordings.

DeepShield offers real-time detection, instant alerts, and an API for seamless integration. Our solution is available on web platforms and mobile applications, making deepfake detection accessible anytime, anywhere.

Project Background

Reansons

1. Democratization of Deepfake Technology

open-source tools (e.g., DeepFaceLab, Wav2Lip) can generate fake videos from requiring PhD-level skills to being achievable by ordinary netizens within 2 hours.

(MIT Technology Review's 2024 report)

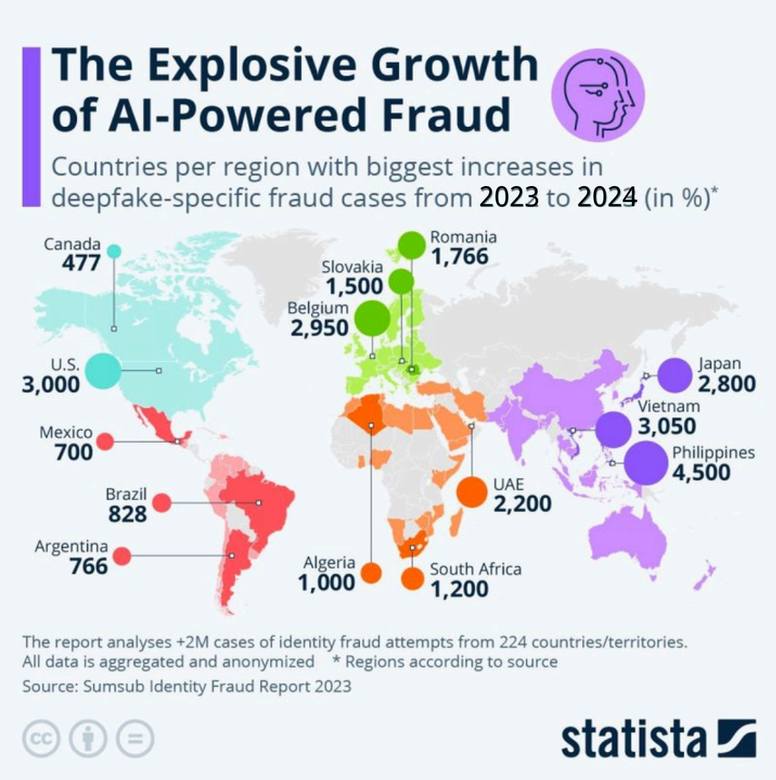
2. Surge in AI Voice Cloning

a 470% year-on-year increase in TTS-based voice fraud cases

(Tencent Security's "2023 Digital Financial Anti-Fraud White Paper")

Project Background

Alarming Surge in Global AI Fraud Metrics



Global AI Fraud Statistics (2023-2024)

Metric	Data	Source
AI Fraud Case Growth (2023)	• Asia-Pacific: +415% • Europe: +278% • North America: +193%	INTERPOL Global AI Crime Report 2024
Deepfake False Positive Rate	Industry Average: 6.8%	NIST FRVT 2023 Benchmark
Corporate Annual Cost	• SMEs: $280k < br >$ $\bullet Multinationals$:2.4M	PwC Cybercrime Cost Report 2024

Project Background

The Staggering Economic Toll of AI Scams

Global AI Scam Losses (2023-2024)

Loss Amount	Region	Source
$12.3B	Global Business Email Fraud	FBI IC3 2023 Annual Report
€1.8B ($1.94B)	EU Banking Sector	ENISA 2023 Threat Report
¥34.2B ($4.8B)	Asia-Pacific Social Media Scams	INTERPOL ASIA 2024 Cyber Report
£2.4B ($3.1B)	UK Pension Fraud	UK Finance 2023 Analysis
$760M	African Crypto Scams	Chainalysis 2024 Crypto Crime Report

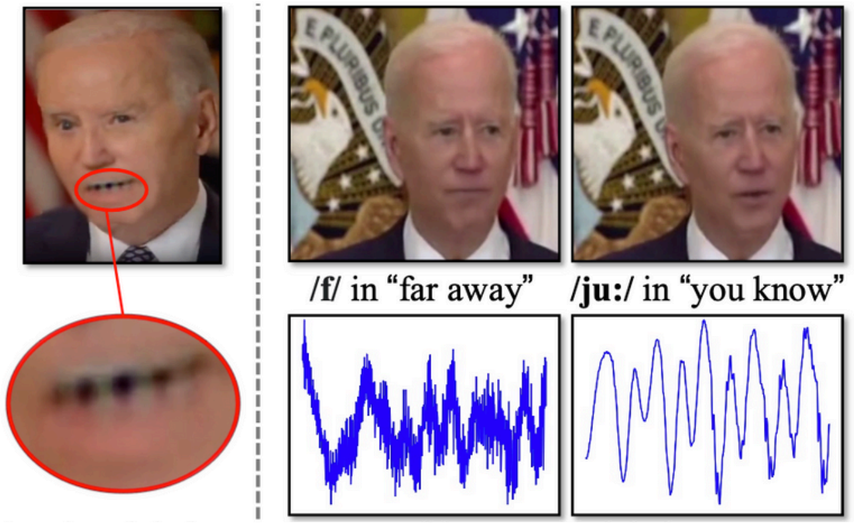
Pain Point Analysis

Pain Point	Critical Data	Source
1. Detection Lag		
Current systems fail to detect 72% of advanced deepfakes	NIST FRVT 2023 Benchmark	Real-time adversarial training
2. Attacker Evolution		
New AI fraud tools deploy 4.2x faster than defense updates	MIT Media Lab 2024 Report	Self-evolving detection engine
3. Cross-platform Spread		
83% of scams initiate on social media (vs 34% in 2020)	INTERPOL Global Crime Trend 2024	API integration with major platforms

Existing Problems

01. Fraud & Scams

- Criminals use deepfake videos and voice cloning to impersonate people in video calls and phone calls, tricking victims.

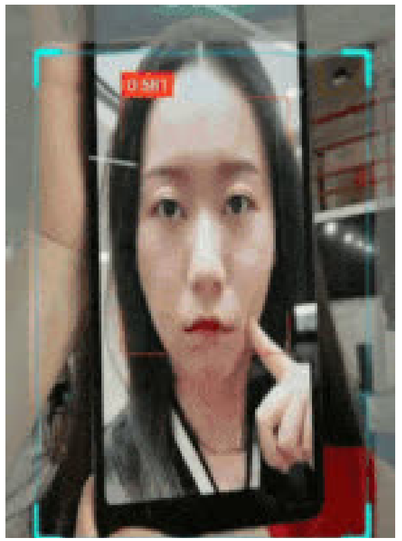


02. Misinformation & Disinformation

- AI-generated videos spread fake news, political propaganda, and misleading content.

03. Security Risks

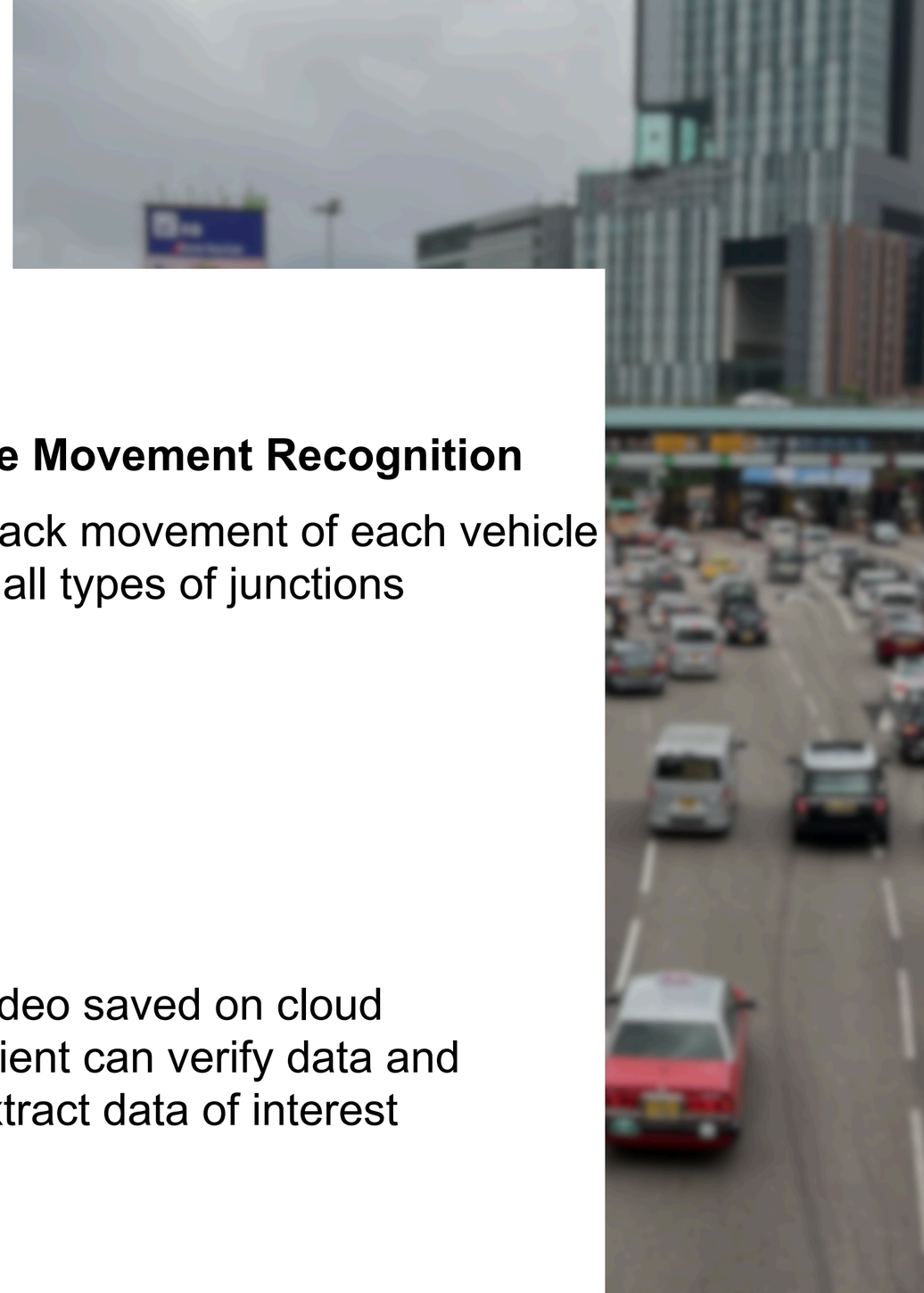
- Traditional authentication methods (e.g., facial recognition, voice authentication) are vulnerable to AI-generated forgeries.



04. Trust Issues in Digital Communication

- Users cannot easily differentiate between real and AI-generated content.

Solution



Computational Traffic Survey

- Counting number of vehicles
- Pedestrian counting
- Sorting vehicles into 10 classes with 90% confidence level

Vehicle Movement Recognition

- Track movement of each vehicle in all types of junctions

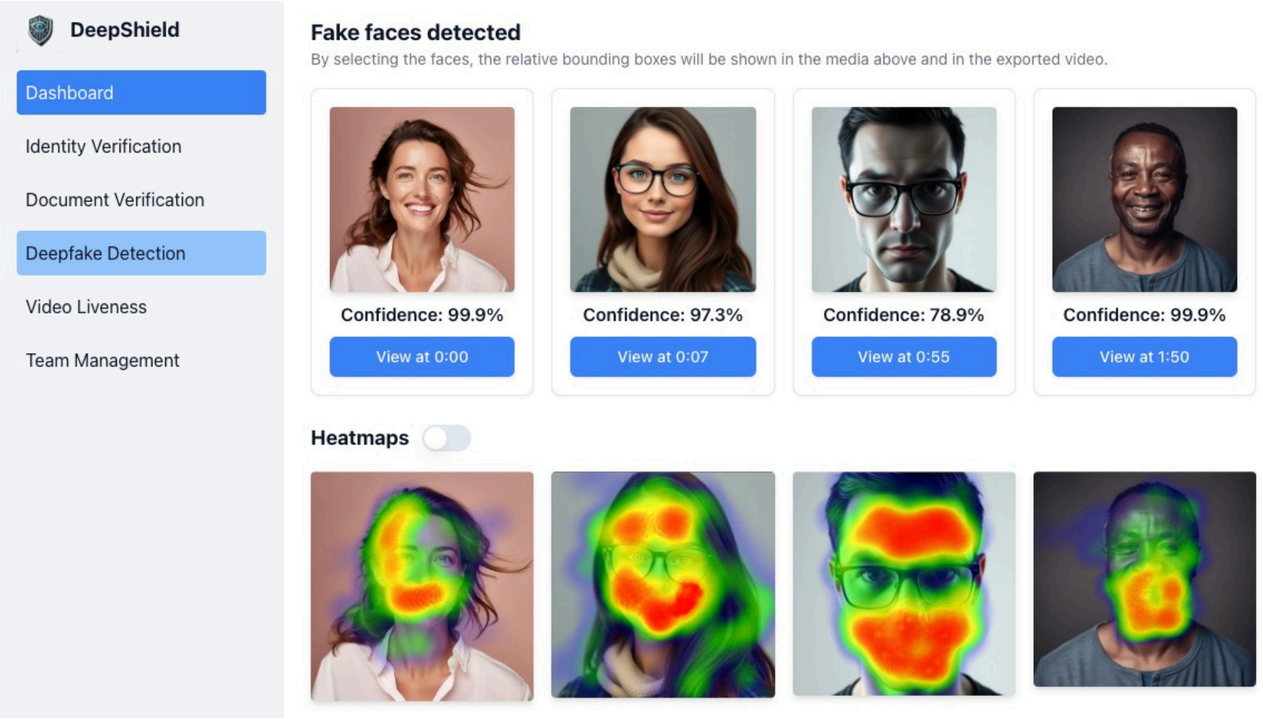
Automated Data Processing

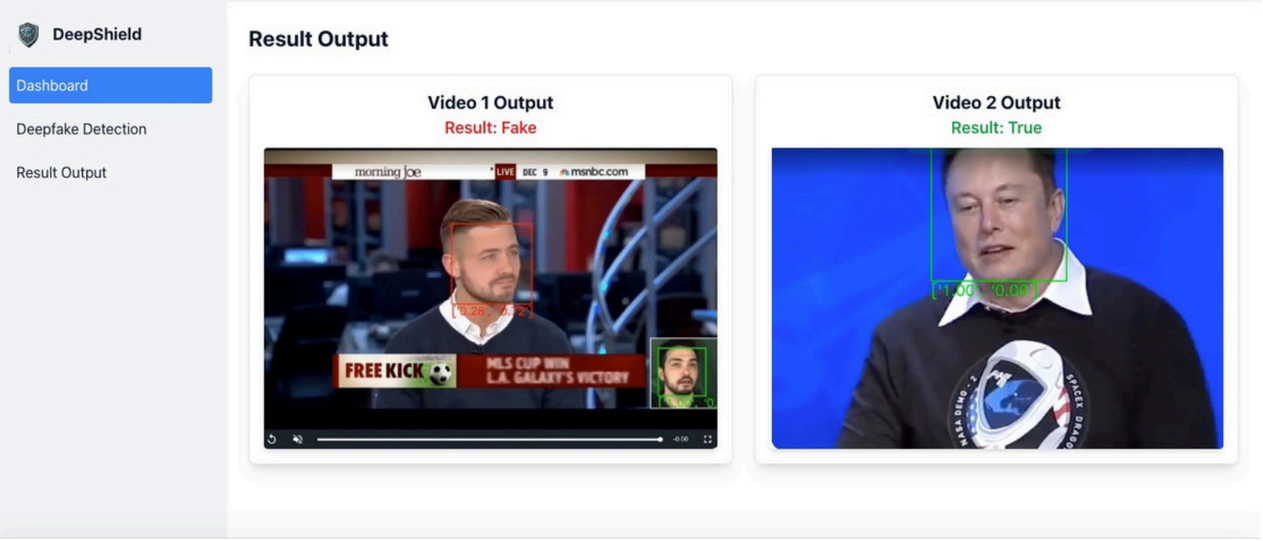
- PCU unit counting
- Peak hour identification
- Intersection flow balance

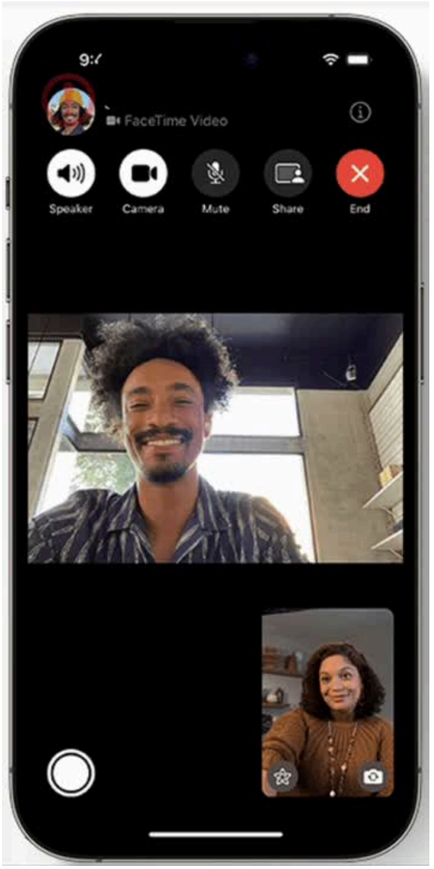
Video

- Video saved on cloud
- Client can verify data and extract data of interest

Our Approach

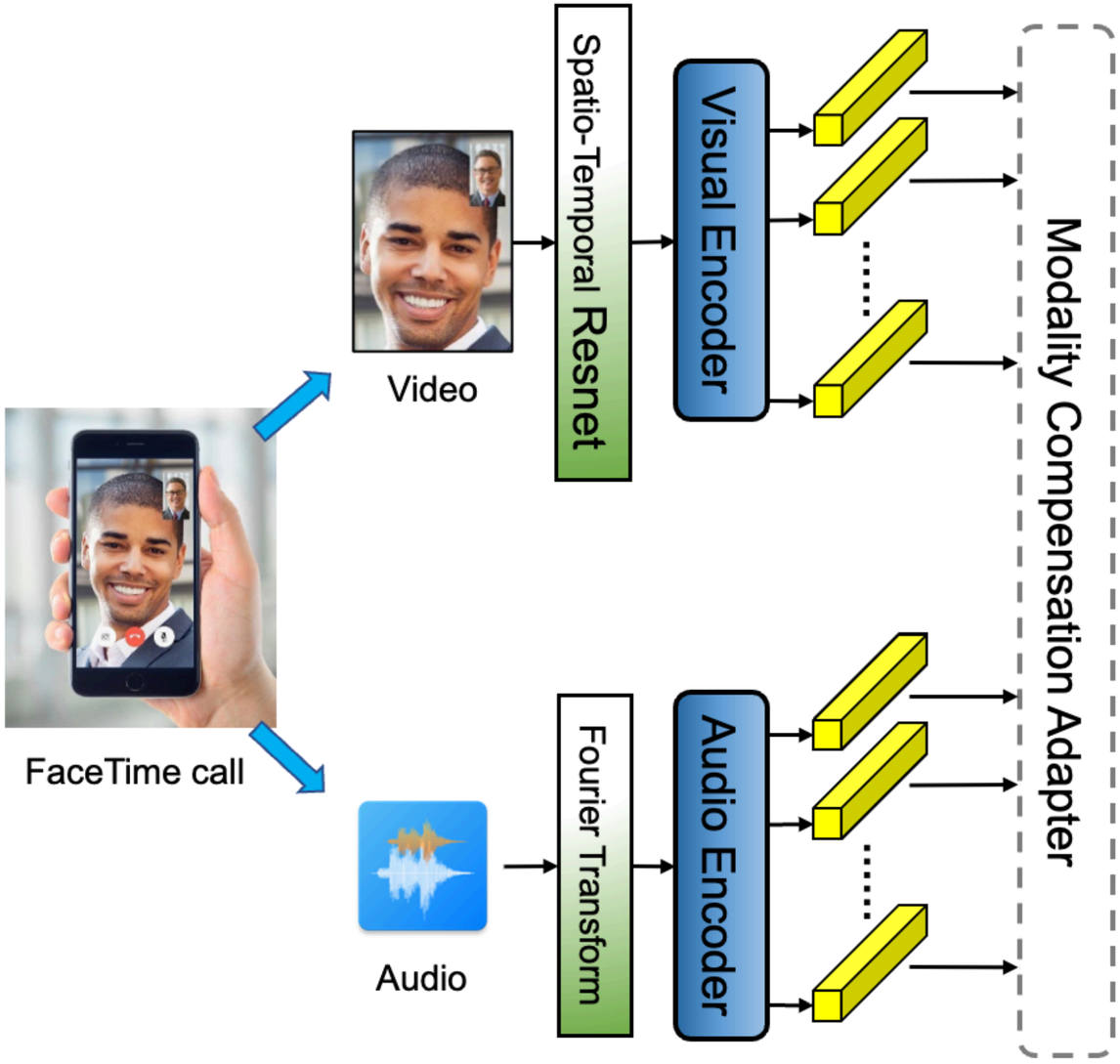


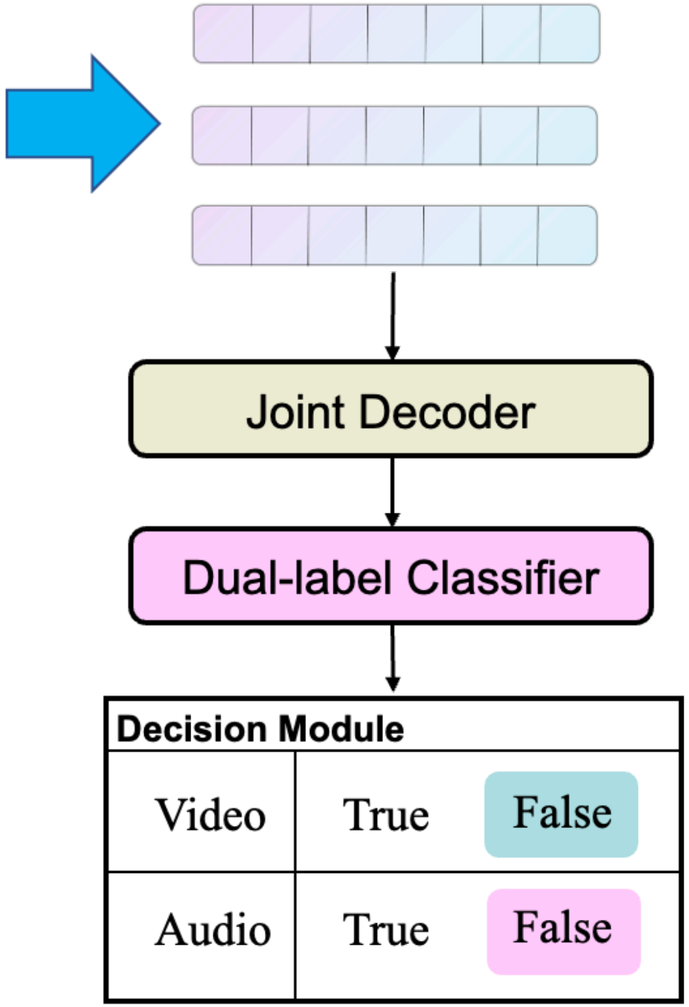


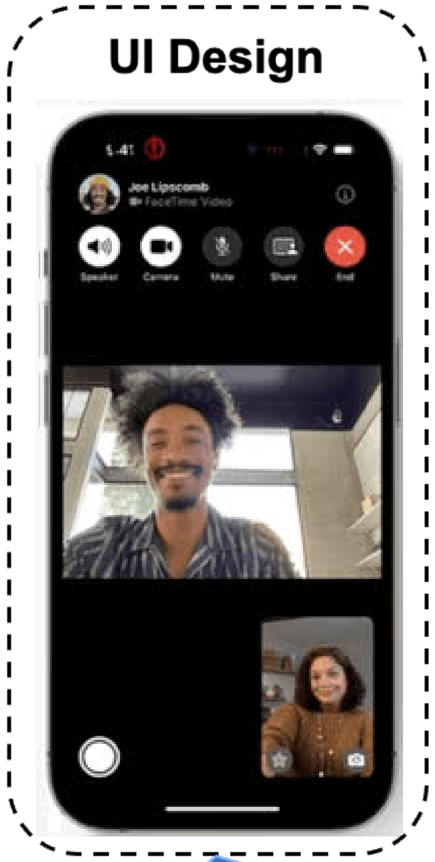


- **Multi-Modal Analysis: Detects manipulations in video, images, and audio.**
- **Instant Alerts: Notifies users during calls and media playback.**
- **Cross-Platform Support: Available on web, mobile, and via API.**
- **Enhanced Security: Prevents fraud and misinformation.**

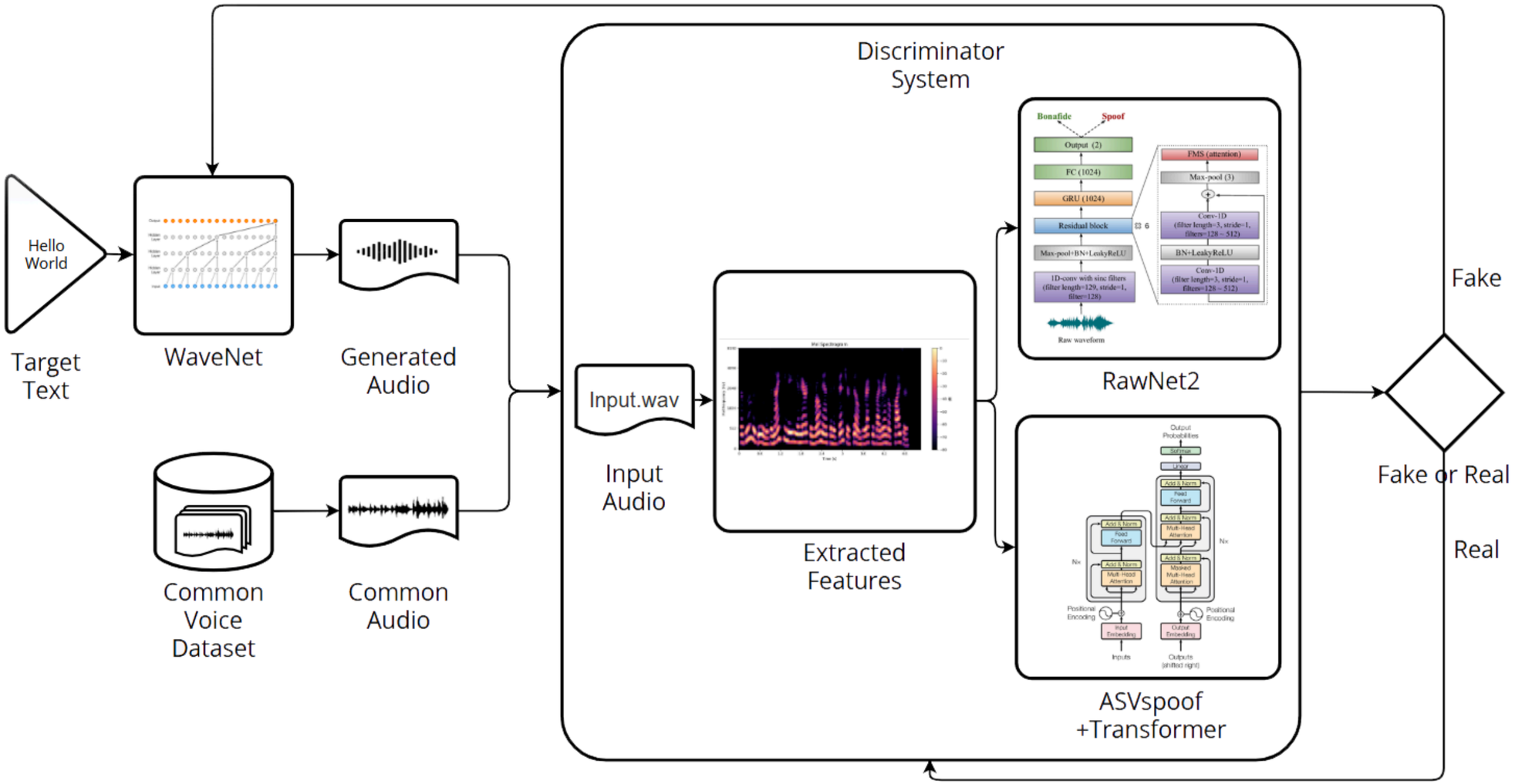
Technology Video







Technology Audio



Experiment

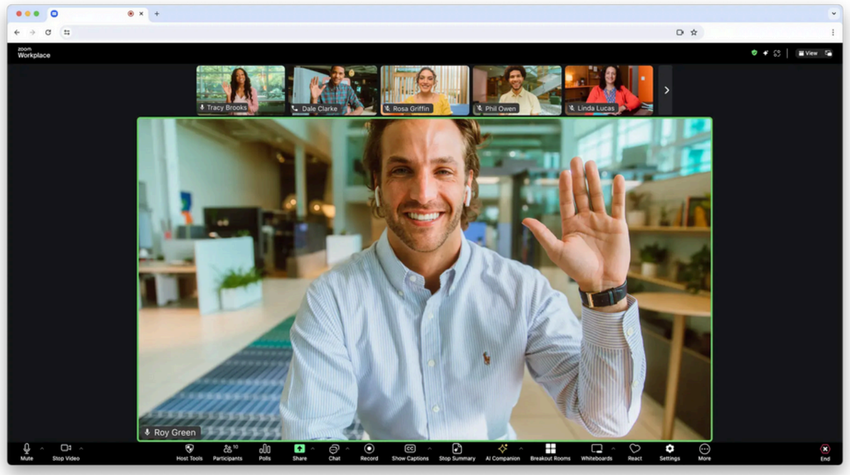
AIGC Video Detect Result

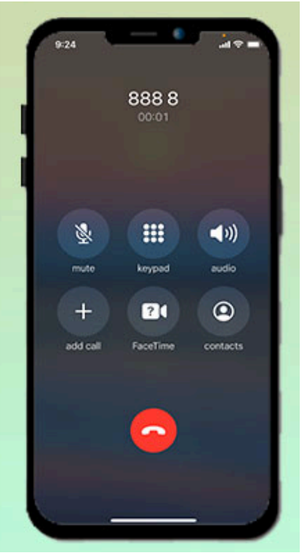
Methods	CrossFakeAVCeleb					CrossDFDC				
Variants	OF1	CF1	WF1	VACC	AACC	OF1	CF1	WF1	VACC	AACC
w/o TAM&FCD	65.77	54.07	57.12	74.73	49.34	55.89	29.23	53.00	58.46	**94.65**
w/o TAM	67.50	55.53	**59.68**	76.41	**49.55**	58.28	32.67	55.77	59.19	94.61
Ours	**68.57**	**55.83**	58.77	**80.00**	48.89	**58.57**	**36.97**	**56.75**	**61.42**	94.07

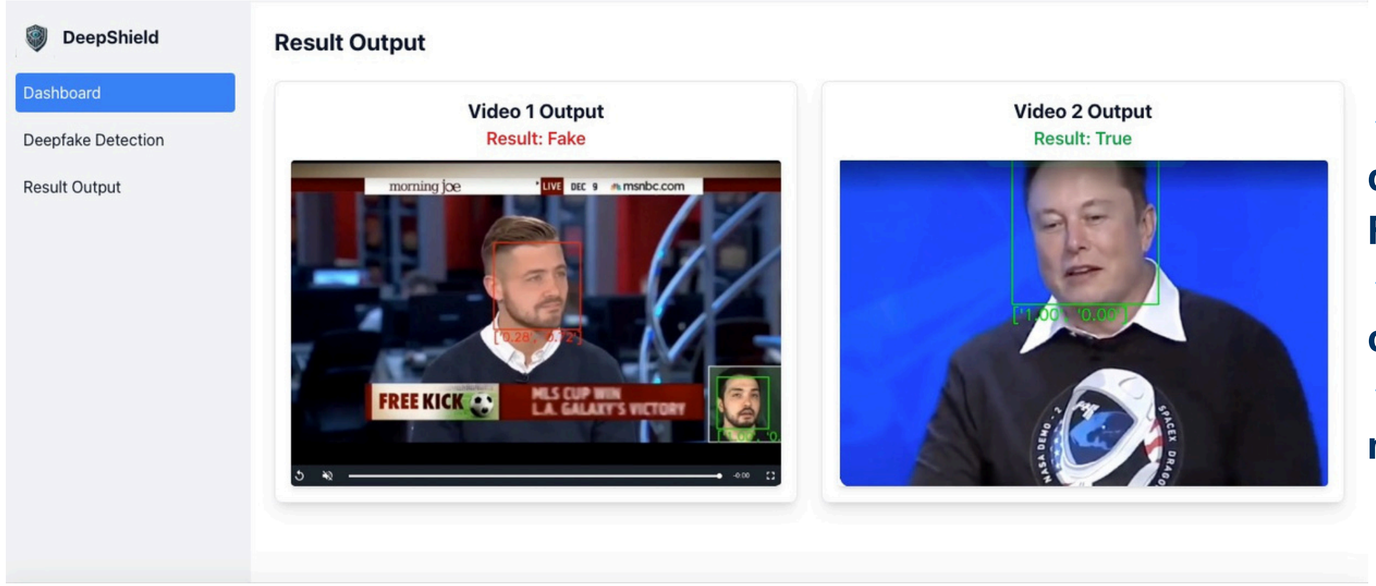
AIGC Audio Detect Result

Compensation	OF1	CF1	VF1	AF1	WF1
None	88.96	86.02	89.62	82.40	88.57
Video	87.79	84.41	88.50	80.17	87.57
Audio	**90.24**	**87.80**	**90.84**	**84.61**	**90.24**

Application







- Video Calls & Chats – Detects deepfakes in Zoom, WhatsApp, and FaceTime.
- Social Media – Flags AI-generated content to prevent misinformation.
- Security & Compliance – Ensures media authenticity for investigations.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

HACKTHON CONFERENCE 2025

DeepShield

Exposing Deepfakes, Protecting Reality.

💡 What's Next?

🚀 Deploying DeepShield API for third-party integrations.

📱 Developing an Android/iOS app prototype for real-time deepfake detection.

🤝 Seeking partners in cybersecurity, social media, and telecom industries.