**Voice Scam Shield – Multilingual AI for Real-Time Call Scam Detection**

**Track:** VC big bets (Cybersecurity)

**1. Motivation / Goal to Achieve**

Generative voice technology now enables scammers to impersonate friends, family, and institutions during live phone or video calls. Victims often realize too late, after sensitive information or money is lost.

**Goal:**Develop a multilingual AI agent that works during phone or video calls to:

1. Detect scam intent and synthetic voices in real time.
2. Alert the user with discreet on-call feedback.
3. Support English, Spanish, and French (bonus: add 1–2 more).

**2. Core Features (MVP)**

1. **Real-Time Call Audio Monitoring**
   * Integrate with Twilio, WebRTC, or Zoom SDK to capture live call audio.
   * Perform voice activity detection and speaker diarization to separate user from caller.
2. **Multilingual Scam & Deepfake Detection**
   * Streaming transcription (e.g., Whisper or Deepgram) → detect scam patterns with LLM prompts or trained classifiers.
   * Run anti-spoofing detection (AASIST or RawNet2) to identify synthetic voices.
   * Label segments as Safe, Suspicious, or Scam with short rationale.
3. **On-Call User Alerts**
   * Visual dashboard or floating widget showing live risk score.
   * Spoken alerts via ElevenLabs TTS to discreetly warn the user (“This call may be fraudulent. Do not share codes.”).

**3. Stretch Goals (Optional Enhancements)**

1. **Caller Verification**
   * Match the caller’s voice against a pre-enrolled “safe list” (family, known company reps).
2. **Incident Report Generation**
   * After the call, generate a summary with flagged segments, scam cues, and recommended next steps.

**4. Hints and Resources**

**Frontend**:

* React or Next.js UI for live risk display and after-call reports.

**Backend**:

* FastAPI for audio ingestion and detection pipeline; WebSockets for streaming.

**Speech Stack**:

* ASR: Whisper, Vosk, or NVIDIA NeMo streaming ASR.
* Anti-Spoofing: ASVspoof 2019/2021 datasets + AASIST baseline.
* Scam Intent Detection: LLM (GPT-4o-mini) or multilingual transformer classifier.

**TTS for Alerts**:

* ElevenLabs real-time TTS with multilingual voices.

**Example Datasets**:

* ASVspoof (synthetic voice detection).
* FakeAVCeleb / WaveFake (voice cloning samples).
* Custom multilingual scam script dataset for intent classification.

**5. Evaluation Criteria**

* Coverage: Works for phone and video calls; multilingual EN/ES/FR.
* Detection Accuracy: ≥80% correct classification of scam vs safe calls.
* Anti-Spoofing: ≤10% Equal Error Rate on synthetic voice detection.
* Latency: Alerts within 2 seconds of suspicious speech.
* User Experience: Clear, discreet alerts without disrupting the call.

**6. Why It Matters**

This challenge equips people with a real-time defense system against the new wave of AI-driven scam calls. It’s proactive, multilingual, and accessible—ideal for personal safety, corporate fraud prevention, and protecting vulnerable communities in the age of generative voice.