



THAKUR SHYAMNARAYAN
DEGREE COLLEGE



PROBLEM STATEMENT: 5. Deepfake Detection Using

TEAM NAME: Zephyrs

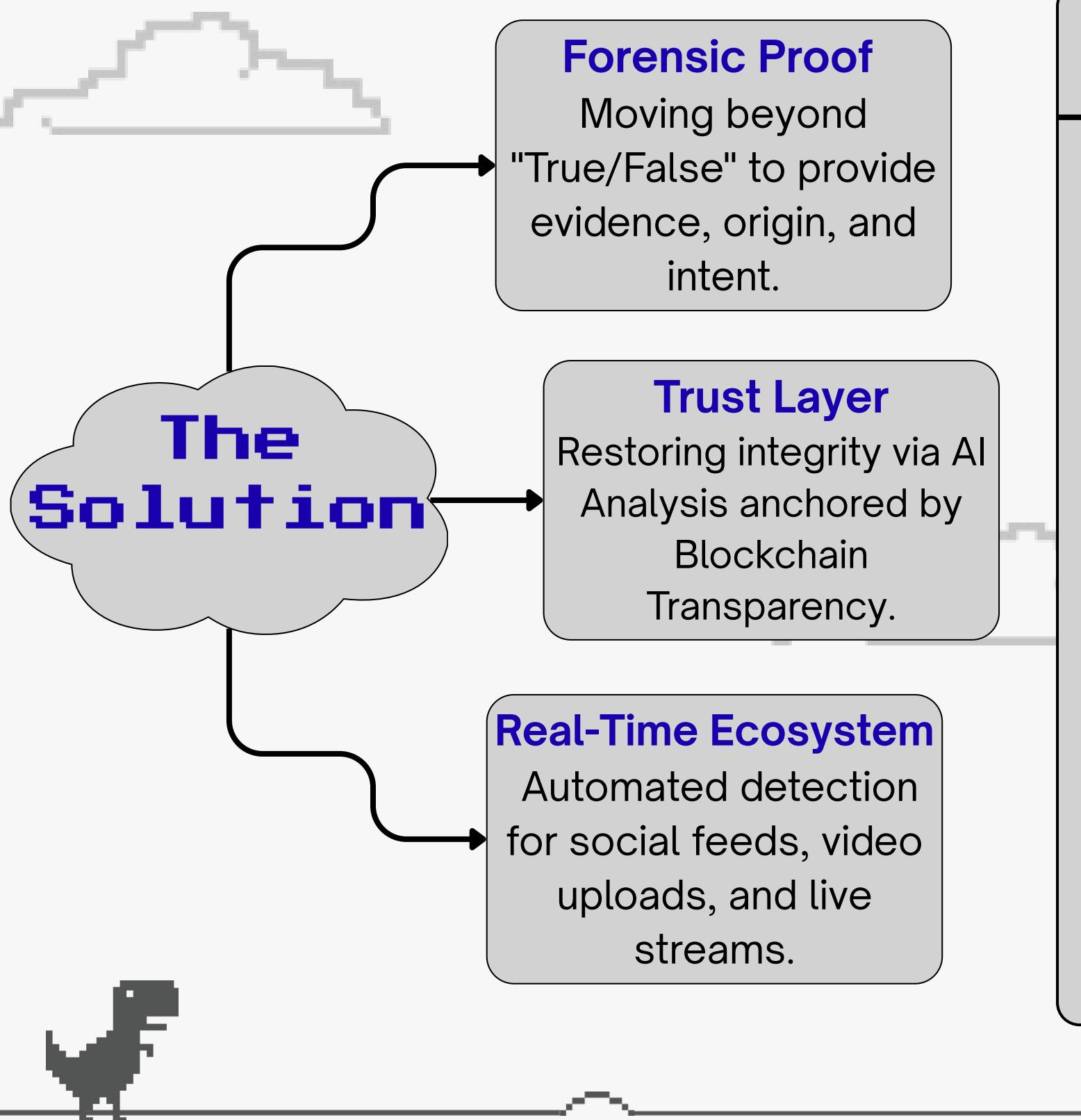
Generative AI

COLLEGE NAME: DJSCE





IDEA OVERVIEW



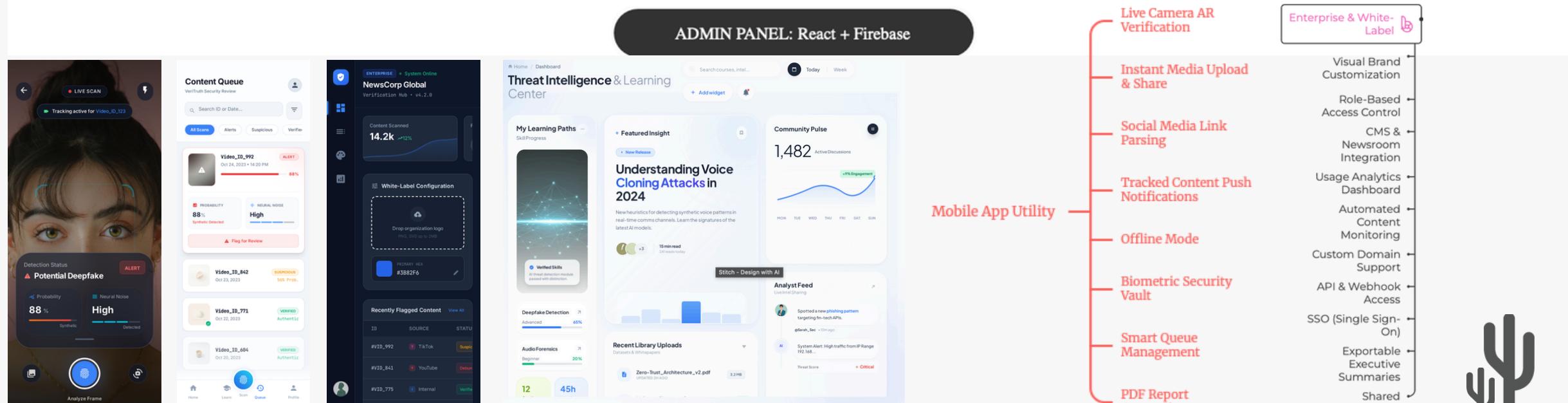
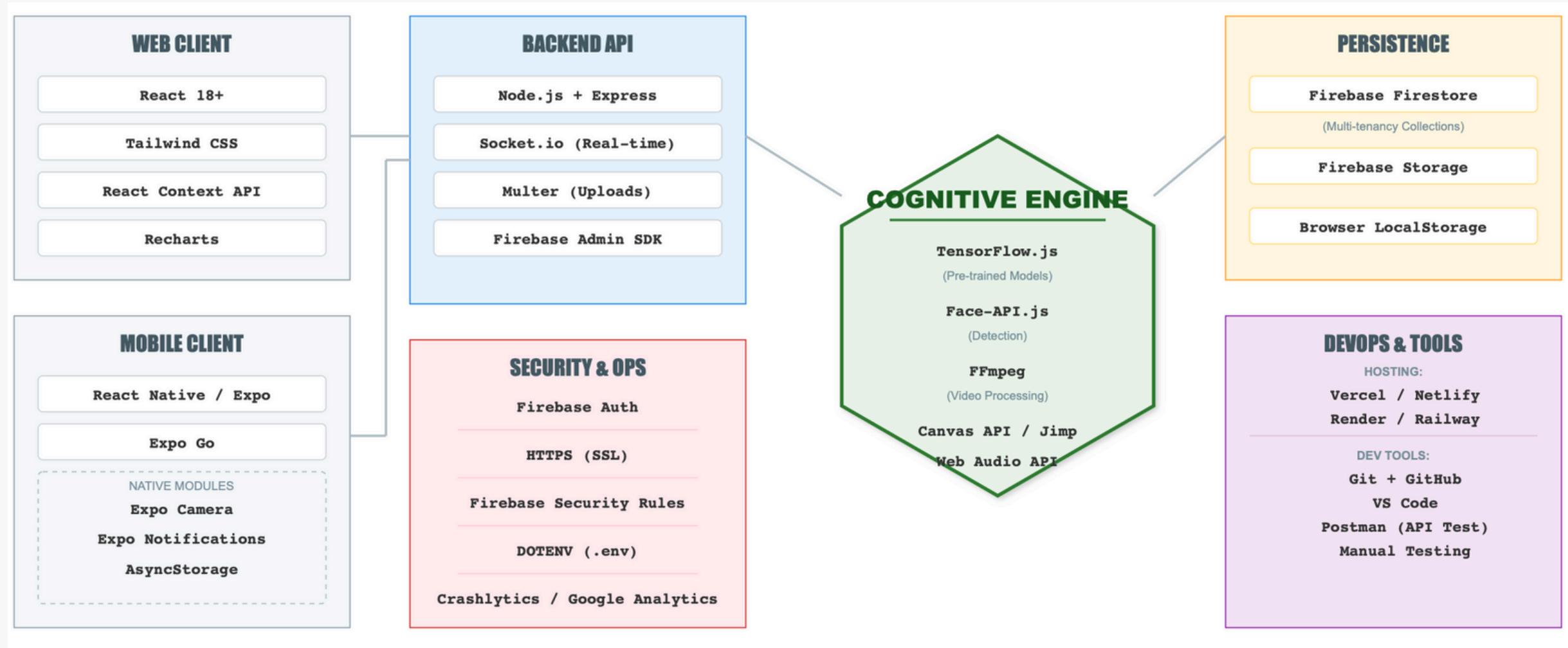
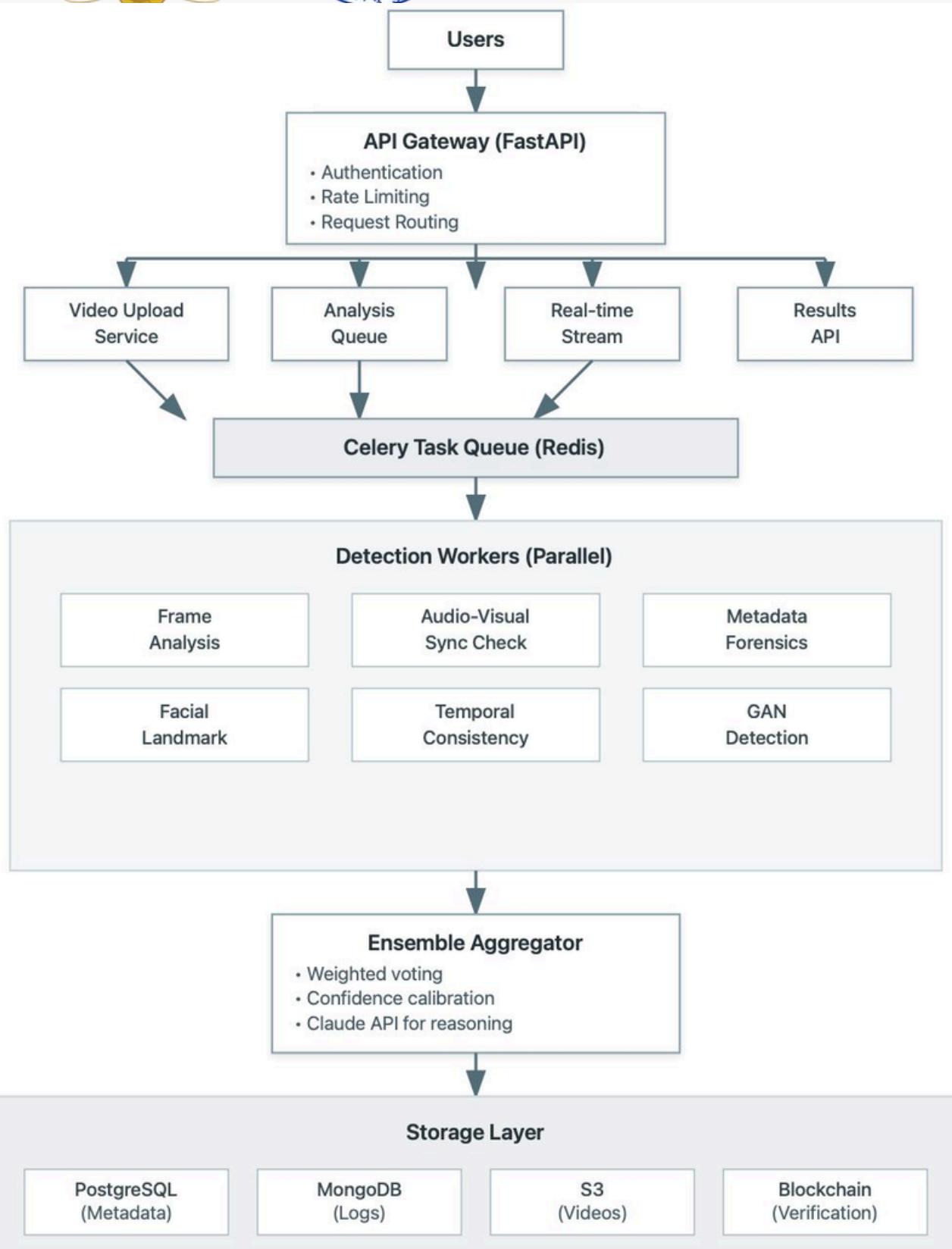
Key Innovations

- Hybrid Ensemble Engine:** Fuses Visual, Audio, & Temporal models.
- Explainable AI (XAI):** Generates Heatmaps proving exactly where manipulation occurred.
- Semantic Lip-Sync:** Detects Phoneme Mismatches in mouth movements.
- Source DNA Search:** Finds Original Authentic Footage via Vector Search.
- Immutable Ledger:** Blockchain-anchored verification for tamper-proof trust.
- Compression Resilience:** Robust on Low-Quality viral videos (WhatsApp/TikTok).
- Voice Biometrics:** Tracks Micro-Tremors to verify Human Audio.
- Live Sentinel:** Real-time protection for Zoom/Meet calls.



TECHNICAL APPROACH

HACKVISION



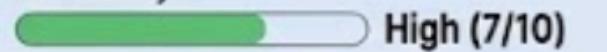


FEASIBILITY & PRACTICALITY



LIVE VERIFICATION SCANNER

Feasibility:



Key Technical Points: Uses React Native Vision Camera combined with Skia or TensorFlowJS.

Constraint: Full deepfake detection processes are too GPU-heavy to run natively on mobile devices.

Solution: Adopting a "Hybrid Approach" — using a local Face Mesh for immediate UI feedback, while offloading heavy analysis to a Cloud API.

REAL-TIME PUSH ALERTS

Feasibility:



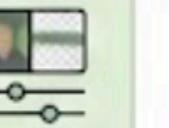
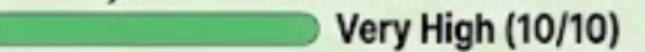
Key Technical Points: Firebase Cloud Messaging (FCM).

Utility: Allows users to subscribe to specific video 'Hash IDs'.

Workflow: Analysis Complete → Server Trigger → User Notification (e.g., "Video #992 is FAKE").

EDUCATIONAL DEEPFAKE SIMULATOR

Feasibility:



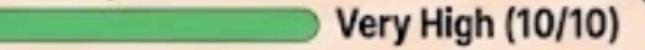
Key Technical Points: Simple Image Layering (Opacity Manipulation).

Advantage: Does NOT require local AI generation, making it ethical and highly performant.

Solution: Stacking static image layers (Original, Glitch, Blur) and controlling their alpha (transparency) values with sliders.

ENTERPRISE WHITE-LABELING

Feasibility:



Key Technical Points: React Context API + Dynamic Theming.

Mechanism: App fetches a JSON configuration file (containing Logo URL, Hex Codes) upon launch.

Benefit: Enables instant rebranding for enterprise clients like CNN, BBC, or Reuters.

CONTENT VERIFICATION QUEUE

Feasibility:



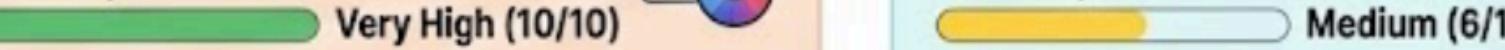
Key Technical Points: Standard REST or GraphQL List View.

Constraint: Major social platforms (TikTok/X) actively block automated scrapers.

Solution: Relying on 'User Reported' links and community flagging mechanisms rather than auto-crawling content.

FORENSICS ANALYSIS REPORTS

Feasibility:



Key Technical Points: Complex spectral charts are too cramped to read effectively on mobile screens.

Solution: Use simplified Progress Bars and Radar Charts for the mobile app, reserving full Spectral Graphs for the Web Dashboard.

Tech Stack: D3.js or Victory Charts.

- Brand Shielding:** Instantly detects and flags deepfake smear campaigns,
- Zero-Cost Rebranding:** The architecture allows major media houses to launch their own branded verification app in minutes

Operation Pract

- Hybrid Performance:** Splits processing between local "Face Meshes" and Cloud APIs
- Passive Protection:** "Real-Time Push Alerts" eliminate manual checking
- Privacy-First Architecture:** Local-first processing ensures that immediate visual feedback





IMPACT AND BENEFIT



The Urgent Crisis

- Explosive Growth:** Deepfakes rose **19%** in Q1 2025 (surpassing all of 2024).
- Human Failure:** Manual detection success rate is only 24.5%.
- Weaponized:** **98%** of deepfakes are non-consensual imagery (NCII).

Economic Shield

- Stop CEO Fraud:** Prevents incidents like the recent **\$25M** Arup scam.
- Market Defense:** Mitigates \$40B in projected fraud losses (2027).
- Save Costs:** Avoids the avg. \$500k loss per corporate attack.

Market Opportunity & Growth

- CAGR:** **47.6%** (2024-2034).
- Market Size:** **\$5.6B** projected by 2034.
- Preventable Fraud:** ~**60%** of \$40B with TruthGuard.

Operational Speed

- Real-Time:** Detection in **<60** Seconds (vs. hours manually).
- Election Integrity:** Instant monitoring for Political Misinformation.
- Trust Standard:** Restores credibility for News & Media.

