**Previous Experience**

* Hack-A-Nite Hackathon at MAHE in March 2025. Finished in top 15.
* Developed an Enterprise-grade Generative AI platform for managing and simulating **self-sustaining space habitats**, featuring real-time telemetry, adaptive design, and resource optimization. It integrates a (GNNs, GANs, RL, PINNs), **quantum simulations** for threat prediction (solar storms, radiation, asteroids), and **3D digital twin visualization**.
* The system includes an AI-powered astronaut companion, gesture controls, voice interface, and a blockchain-backed resource ledger.

**Track**

AI vs AI – Detecting Deepfake & Misinformation

## Introduction

**The Problem**

**3.2 billion** fake-news stories shared every month (First Draft News, 2024).

**900%** annual growth in deepfake content.

**$78 billion** lost globally to misinformation campaigns.

Traditional detection systems are **reactive**, **single‑modal**, and often **outpaced** by new forgeries.

**The Gap**

Existing tools detect only one content type (text **or** audio **or** video).

Centralized verification is vulnerable to **tampering**.

**Latency** in model updates leaves openings for adversaries.

**Our Vision**  
A **real‑time**, **quantum‑enhanced**, **blockchain‑verified** truth‑detection network that stays **one step ahead** of misinformation.

**Visual Aid Suggestion**: A high‑impact infographic combining the four problem metrics in bold icons (e.g., globe for stories, waveform for deepfakes, dollar sign for losses, clock for latency).

## Solution Overview

**TruthShield AI = Detection + Verification + Protection**

| **Module** | **Description** |
| --- | --- |
| **Multimodal AI** | Detects fake **text** (RoBERTa), **audio** (Whisper), **images** (OpenCV + Dlib), **video** tampering. |
| **Quantum Graphs** | Uses **quantum walks** on information‑spread networks to spot coordinated misinformation clusters. |
| **ProofChain** | Stores cryptographic content fingerprints in a **blockchain ledger** (Polygon sidechain + IPFS batch). |
| **War Room SDK** | Plugin SDK for **Chrome**, **WhatsApp**, **Telegram**—enables real‑time, in‑browser content analysis. |
| **Dashboard** | React + Tailwind UI showing live alerts, risk scores, and forensic audit trails. |

“Quantum Graphs” leverages quantum‑speed pattern recognition to reveal spread dynamics invisible to classical analytics.

**Visual Aid Suggestion**: System architecture diagram with colored lanes showing data flow through Multimodal AI, Quantum Graphs, ProofChain, and Dashboard modules.

## Implementation Plan & Challenges

**48‑Hour MVP Sprints**

**Hours 0–12**:

MVP of **AI detectors**: CLI PoC for text, audio, image using RoBERTa, Whisper, OpenCV.

**Hours 12–24**:

Quantum‑graph prototype in PennyLane simulator; batch fingerprinting + IPFS storage.

**Hours 24–36**:

Real‑time backend (FastAPI + Redis caching + async workers) + React dashboard skeleton.

**Hours 36–48**:

Browser plugin demo; end‑to‑end test; deploy PoC on GCP Kubernetes.

**Key Risks & Mitigations**

**Integration Complexity**  
Mitigation: Standardize services in Docker containers; orchestrate via Kubernetes on GCP.

**Quantum Hardware Limits**  
Mitigation: Develop on **IBMQ simulators**; reserve real‑hardware runs for benchmarks only.

**Blockchain Gas Costs**  
Mitigation: Bundle 1,000 hashes off‑chain (IPFS), commit single Merkle‑root transaction on Polygon.

**Adversarial Drift**  
Mitigation: Continuous adversarial training pipeline; scheduled model retraining with new attack data.

**Data Privacy & Compliance**  
Mitigation: Anonymize user inputs; opt‑in agreements; adhere to GDPR, CCPA by design.

**Deliverables by 48 h**:  
Live web‑app PoC + browser extension + blockchain‑anchored verification demo.

**Visual Aid Suggestion**: A 2×2 timeline chart for the four sprint phases with icons, alongside a risk matrix plotting likelihood vs. impact for key challenges.

## Technical Approach & Tech Stack

**Layered Architecture**

**Input Layer**: Social‑media APIs (Twitter, Facebook), RSS, file uploads.

**AI Layer**:

Text: RoBERTa fine‑tuned on fake‑news corpora.

Audio: Whisper for transcription + custom tampering detector.

Vision: OpenCV preprocessing + Dlib face‑landmark analysis + custom forgery nets.

**Quantum Layer**: Qiskit + PennyLane for quantum‑walk analytics (hybrid simulator/hardware).

**Verification Layer**: IPFS for off‑chain content, Merkle proofs on Polygon sidechain.

**Application Layer**:

Backend: FastAPI, PostgreSQL, Redis

Frontend: React, Tailwind CSS

Plugins: WebExtension APIs, Web3.js

**Deployment**: Docker, Kubernetes, GitHub Actions CI/CD, GCP (Anthos).

**Visual Aid Suggestion**: Simplified layered diagram showing each stack layer with representative icons/logos (e.g. FastAPI, Qiskit, Polygon).

## Innovative Features

**DNA of Truth**: Content hashed at ingestion; Merkle‑tree root published on‑chain.

**Reality Simulator**: Reconstructs a “most likely original” by comparing multiple altered copies via weighted consensus.

**AI War Room**: Sandbox for adversarial testing—generate new fake content, measure detector robustness.

**Quantum Graph Analyzer**: Identifies misinformation super‑spreaders via quantum spectral clustering.

**Plugin SDK**: Easy integrators for Chrome, Telegram, WhatsApp—embed detection into any browser or chat platform.

**Visual Aid Suggestion**: Four icon‑based callouts (one for each feature), each with a 2‑sentence caption and minimalist graphic.

## Future Scope & Scalability

**Short‑Term (0–12 months)**

Launch **mobile app** for consumer alerts (iOS/Android).

Partnerships with top 5 social platforms; enterprise pilot with news agencies.

Government & election‑monitoring integrations.

**Long‑Term (12–36 months)**

**Global Truth Network**: Decentralized mesh of verification nodes worldwide.

**Quantum‑Native Processing**: Port critical pipelines fully onto quantum hardware as capacity grows.

**Predictive Misinformation**: AI models forecasting next‑week misinformation topics.

**Educational Suite**: Automated truth‑literacy tools for schools and universities.

**Market & KPIs**

**Market Size**: Estimated $24 billion by 2028.

**Year 1 Goals**: 50 enterprise clients; process 10 million items/day; achieve <5% false negatives.

**Year 2 Goals**: 100 million daily items; onboard 5 government bodies; reduce average detection latency <200 ms.

**Visual Aid Suggestion**: A combined bar chart of Year 1 vs Year 2 KPI goals, and a world map highlighting target regions for rollout phases.

With a scalable, modular architecture and clear KPI milestones, TruthShield AI is poised to become the global standard for trusted information verification.

## **Additional Visual Suggestions:**

****For each slide, consider adding:****

* **Architecture diagrams on technical slides**
* **Infographics showing the problem scale**
* **Screenshots of dashboard mockups**
* **Quantum circuit diagrams**
* **Market size charts**
* **Timeline visualization**
* **Before/after comparison graphics**