Gen AI Exchange Hackathon

Google Cloud H2S

Team Name : AegisX

Team Leader Name : AMOGH SUNIL

Problem Statement : Society faces a critical information warfare asymmetry. The World Economic Forum has identified misinformation and disinformation as one of the top global risks. Malicious actors launch campaigns with speed and scale, while traditional fact-checking is reactive and slow, leaving public discourse and democratic stability vulnerable to manipulation.

Brief about the prototype:

AegisX is an AI-powered system that shifts the paradigm from reactive debunking to proactive, predictive inoculation against misinformation.

It operates in two stages:

1. **The Watchtower (Predictive Forecasting):** Uses real-time data analysis and anomaly detection to identify the statistical fingerprints of emerging misinformation campaigns before they go viral.
2. **The Shield (Generative Inoculation):** Leverages the Google Gemini API to automatically generate and disseminate "pre-bunking" content. This content inoculates the public against the underlying manipulation techniques, not just a single false claim.

How is it different from existing solutions?

- Existing solutions (e.g., traditional fact-checking) are **reactive**. They address a lie after it has spread. Aegis-X is **proactive**. It identifies and counters narratives at their inception.
- It moves beyond debunking a single fact to **inoculating against the entire manipulation tactic** (e.g., scapegoating, false dichotomy). This builds long-term cognitive resilience.

How will it solve the problem?

- It closes the **scalability and speed gap**. By using AI for detection and generative AI for counter-messaging, it can match the velocity of attackers, shifting the advantage to the defenders.

USP (Unique Selling Proposition):

- **Predictive Inoculation.** We don't just debunk lies; we vaccinate minds against the techniques used to spread them.

List of Features:

- **Real-time Narrative Monitoring:** Ingests and analyzes data from high-velocity sources like social media APIs, GDELT, and news feeds.
- **AI-driven Threat Scoring:** Anomaly detection scores emerging narratives on velocity, emotional intensity, and signs of inorganic amplification.
- **Automated "Pre-bunking" Content Generation:** Uses the Gemini API to create targeted, easy-to-understand content (short videos, infographics, social media threads) explaining the manipulation tactic in use.
- **Technique-Focused Library:** Identifies and catalogs campaigns based on tactics like scapegoating, hyper-emotive language, and decontextualization.
- **Analyst Threat Dashboard:** An interactive interface for journalists and researchers to explore emerging threats and inoculation campaigns.

Data Ingestion

Social Media APIs

GDELT

News Feeds

BERTopic

The Wathtower (Analysis Engine)

BERTopic identifies emerging conversational clusters and narratives

IsolationForest detects anomalies, flagging narratives with unnatural velocity, sentiment polarization, or network behavior

Threat Alert Generated

The Shield (Generative Engine)

Threat context (narrative + tactic) is sent as a prompt to Google Gemini

Gemini generates "inoculation content"

For instance, a 90-second videe script explaining the tactic of scapegoating. using the flagged narrative as a timely example

Pre-bunking content is pushed to public-facing channels. siclad media accou-

Dissemination

Dedicated social media accounts, partner news outlets, an-

AEGIS·X

Dashboard

Live Threat Feed

Narrative: Hurricane Milton Conspiracy

Threat Score 9.1 / 10

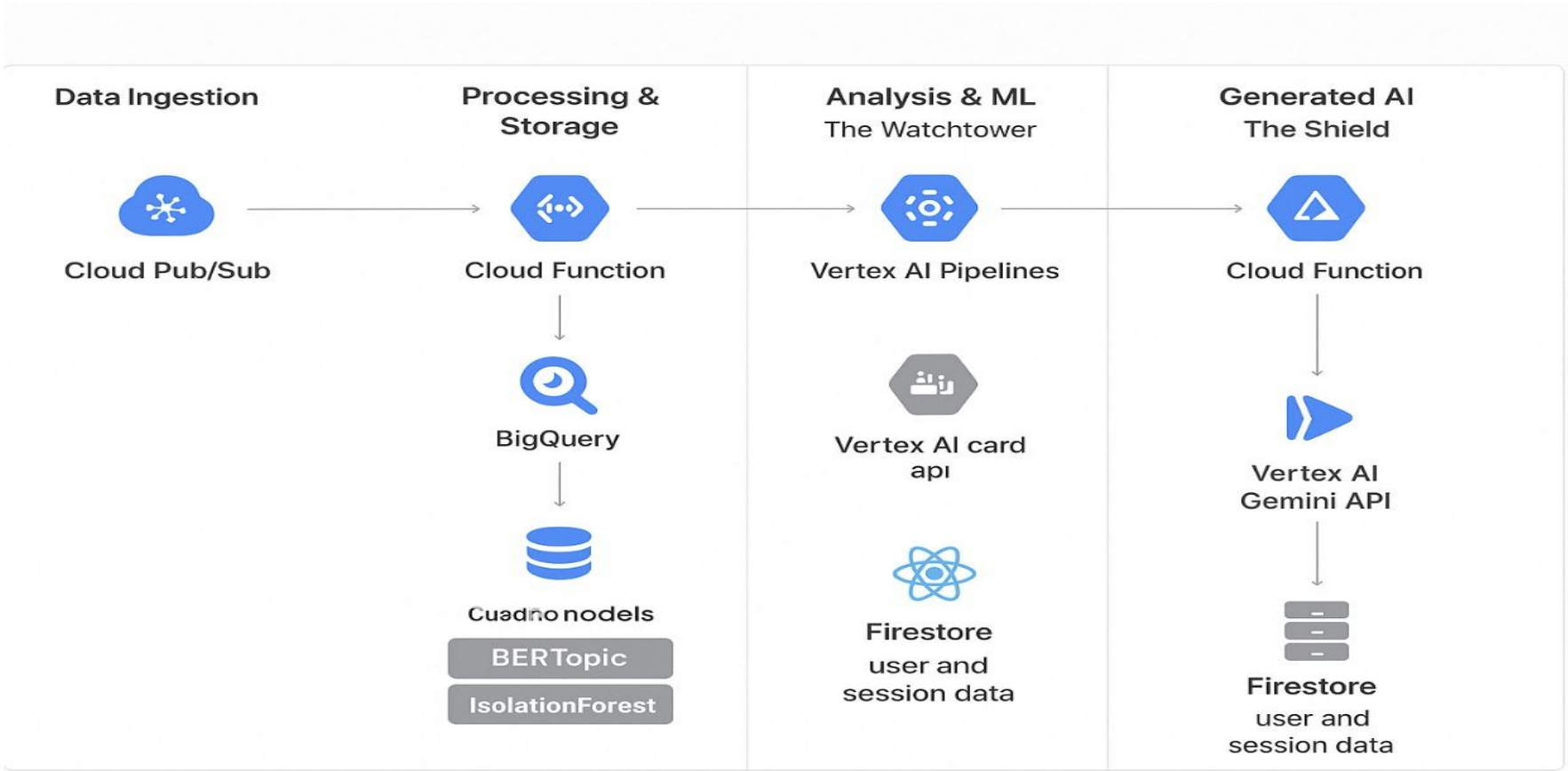
Detected Tactic
Scapegoating

Narrative Analysis

Generated Inoculation Content

How to Spot a Disaster Conspiracy Theory

You might see claims that recent hurricanes were man-made. This is a common tactic to sow distrust.


Data Ingestion
Cloud Pub/Sub
Processing & Storage
Cloud Function
BigQuery
BERTopic
IsolationForest
Analysis & ML
The Watchtower
Vertex AI Pipelines
Vertex AI card api
Firestore
user and session data
Generated AI
The Shield
Cloud Function
Vertex AI Gemini API
Firestore
user and session data

Technologies to be Used

- **Cloud Platform:** Google Cloud Platform (GCP)
- **AI / ML:** Vertex AI (Pipelines, Training, Prediction), TensorFlow, Scikit-learn
- **Generative AI:** Vertex AI Gemini API
- **Data Stack:** BigQuery, Cloud Pub/Sub, Cloud Dataflow
- **Compute:** Cloud Run, Cloud Functions
- **Core Libraries:** BERTopic, scikit-learn (for IsolationForest), pandas, NLTK
- **Backend:** Python (FastAPI)
- **Frontend:** React.js, D3.js (for visualizations)

Estimated Implementation Cost

This prototype is designed for cost-efficiency using serverless and managed GCP services.

Primary Cost Drivers:

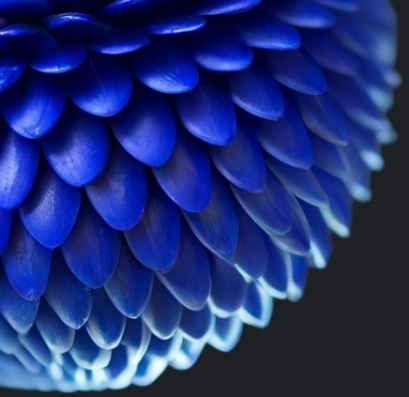
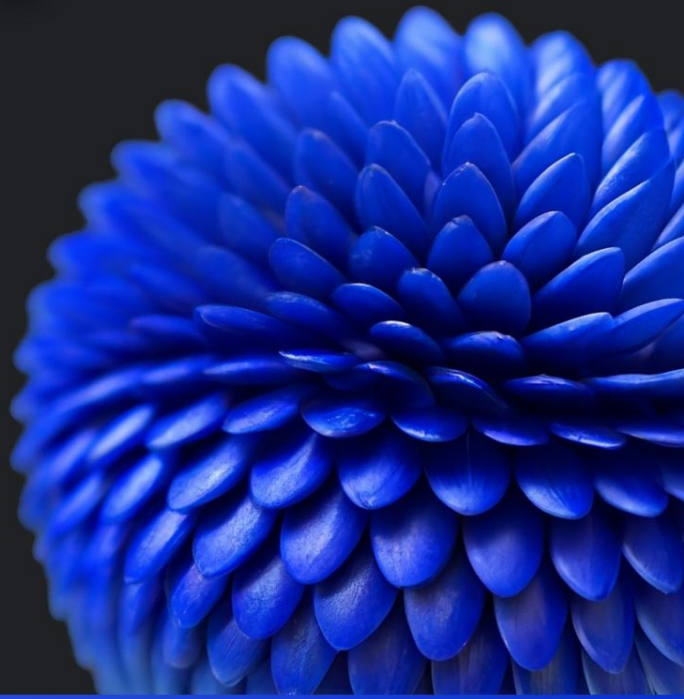
- **Vertex AI Prediction:** Costs scale with the volume of data analyzed per minute.
- **Gemini API Calls:** Billed per 1,000 characters of input (prompt) and output (generated content).
- **BigQuery:** Costs for data storage and analysis queries.

How Aegis-X Aligns with Judging Criteria

We designed Aegis-X specifically to meet the core goals of this hackathon. Here's how we align with the key evaluation criteria:

- **Use of Generative AI (40%):** The heart of our solution is the strategic and meaningful integration of the Gemini API. We aren't just using GenAI as an add-on; it is the engine of our "Shield" system, generating tailored, real-time inoculation content. This deep integration is where we believe our project truly excels.
- **Business Impact & Idea (30%):** We are tackling a critical, real-world problem that has been identified as a top global risk. Our solution is not just theoretical; the dashboard you saw provides a clear, intuitive interface for analysts, and our model has a direct and measurable impact on a multi-billion dollar problem.
- **Presentation & Vision (30%):** Throughout this presentation, we've aimed to clearly articulate our vision. We have a functional prototype, comprehensive documentation, and a clear roadmap for how Aegis-X can be scaled for real-world deployment to protect communities from harmful misinformation.
- **Innovation & Creativity:** Our core innovation is the paradigm shift from *reactive fact-checking* to *proactive inoculation*. This is a fundamentally new and more effective approach to building societal resilience against information warfare.

Google Cloud H2S

Gen AI Exchange Hackathon

Thank you