MyMediTruth

Unmasking Misinformation in Healthcare

Problem

Widespread misinformation in healthcare can lead to public health risks for individuals and communities.

Individual Example: Believing that drinking excessive water flushes out all toxins can lead to water intoxication and electrolyte imbalance.

Societal Example: False claims that COVID-19 spreads through 5G networks caused widespread fear and vandalism of communication infrastructure.

Solution

We use AI to detect and demystify the credibility of self-care claims to ensure everyone has access to medical truth.

Individuals are empowered to make informed health decisions for themselves, recognize sources of misinformation, and are less likely to spread misinformation to others.

Healthcare professionals and organizations gain visibility into misinformation trends so they can react quickly.

Seamless User Experience

User Submits Claim

Users start by entering a healthcare claim directly into the dashboard.

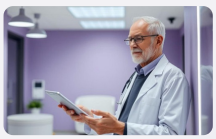
Instant AI Analysis

The system uses a fine-tuned RoBERTa model (trained on the FakeHealth dataset) to classify the claim as **"real" or "fake".**

Results & Explanation

Users see the claims classification, along with a medically backed explanation of the decision generated by OpenAI.

Business Use Cases



Healthcare

Doctors stay ahead of misinformation trends by continuously monitoring emerging health myths, understanding their sources, and addressing patient concerns with evidence-based information.

This proactive approach helps build trust, correct false beliefs, and ensure patients make informed decisions for their well-being.

Marketing

Brands uncover unmet needs by analyzing why consumers turn to alternative remedies, such as rising demand for natural ingredients or gentler, safer product alternatives.

By identifying these gaps, companies can refine their product offerings and align more closely with evolving consumer preferences, staying competitive and relevant in the market.

GitHub

Dashboard

Pipeline: Raw Data To User-Friendly Dashboard

Cloud Storage

Training Data: FakeHealth

Ingestion: GitHub to Azure Data Lake Storage (ADLS)

Data Preprocessing

ETL: Databricks (PySpark)

Modeling

Classification: RoBERTa

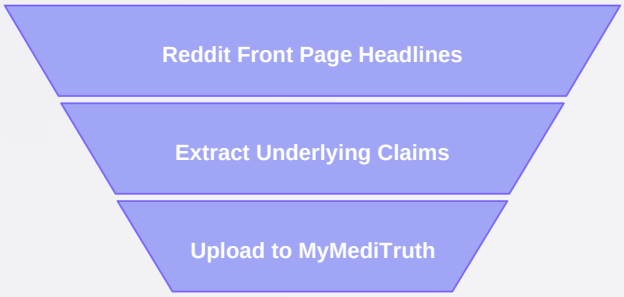
Explanation: OpenAI

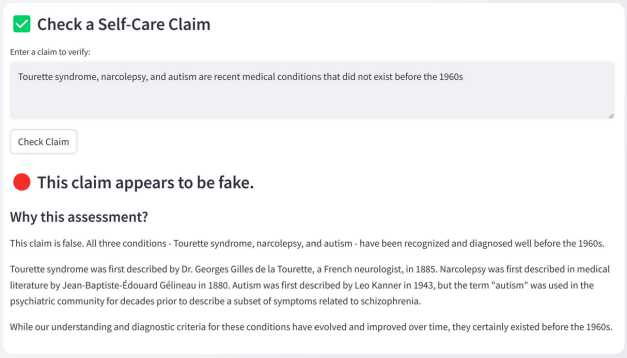
Streamlit Dashboard

Input: User claims **Output:**

Classification and explanation

Trending This Week on Reddit





Scalability: Real-Time Social Media Monitoring

MyMediTruth's dashboard currently reflects **static data**, but a few adjustments can scale up its capabilities. **Using an API to scrape posts** on X (formerly Twitter) or Reddit can enable **real-time social media monitoring**, resulting in a live dashboard.

Web Scraping API

Stream public posts in real time using relevant keywords like “natural cure”, “detox”, and "wellness hack".

Apache Kafka and Spark can be used for scalable streaming and real-time analytics.

ADLS Ingestion

Store streamed post metadata (timestamp, user, engagement) in Parquet for efficient big data querying.

Spark and MLlib can tag posts with their source platform and confidence level.

Live Dashboard

Auto-refresh and show real-time feeds of analyzed claims, trending topics, and viral misinformation.

Elasticsearch and Kibana dashboards can highlight misinformation trends.

Team 6: MyMediTruth

Anshu Mehta: mehta395@umn.edu | **Faithan To:** to000004@umn.edu | **Mayank Singh:** sing1329@umn.edu
Shihao Zhao: zhao1522@umn.edu | **Shrawani Tare:** tare0003@umn.edu