# TruthLens: Reimagining News with Real-Time AI Intelligence

BA 304 Text, Social Media & Web Analytics
Capstone Project

BY
KOYALKAR SRI HARSHA
24MBMB15

# Problem Context

- Online news is generated at massive scale across multiple platforms.
- Readers face challenges such as misinformation, biased reporting, emotional framing, and information overload.
- Traditional fact-checking or media monitoring is slow, manual, and reactive.

# My Core Problem

"HOW CAN WE AUTOMATICALLY ANALYZE NEWS ARTICLES AND SOCIAL CONTENT TO EXTRACT TRUSTWORTHY INSIGHTS, DETECT BIAS/MISINFORMATION, UNDERSTAND EMOTIONAL TONE, GROUP EMERGING TOPICS, AND PROVIDE A FAST SUMMARIZATION + CHATBOT INTERFACE?"

# Objective

BUILD AN END-TO-END NLP PIPELINE TO:

- ✓ DETECT SENTIMENT/EMOTIONS
- ✓ EXTRACT TOPICS & EVENTS
- ✓ SUMMARIZE FACTUALLY
- ✓ IDENTIFY BIAS/FAKE-NEWS SIGNALS
- ✓ ENABLE INTERACTIVE CHATBOT

# Methodology & Workflow

## End-to-End System Pipeline

### DATA EXTRACTION

- SCRAPED & SYNTHESIZED MULTI-SOURCE NEWS ARTICLES (3000+ RECORDS).
- STANDARDIZED INTO JSONL → PREPROCESSED INTO STRUCTURED CSVS.

### PREPROCESSING

- LANGUAGE DETECTION, CLEANING, NORMALIZATION
- OPTIONAL SPACY NER WHEN INSTALLED
- PARSED INTO FIELDS: TITLE, CONTENT, SOURCE, DATE, METADATA.

### EMBEDDINGS + RETRIEVAL

- MINILM-L6-V2 SENTENCE EMBEDDINGS (IF ENABLED)
- TF-IDF FALLBACK FOR LOW-STORAGE SETUPS

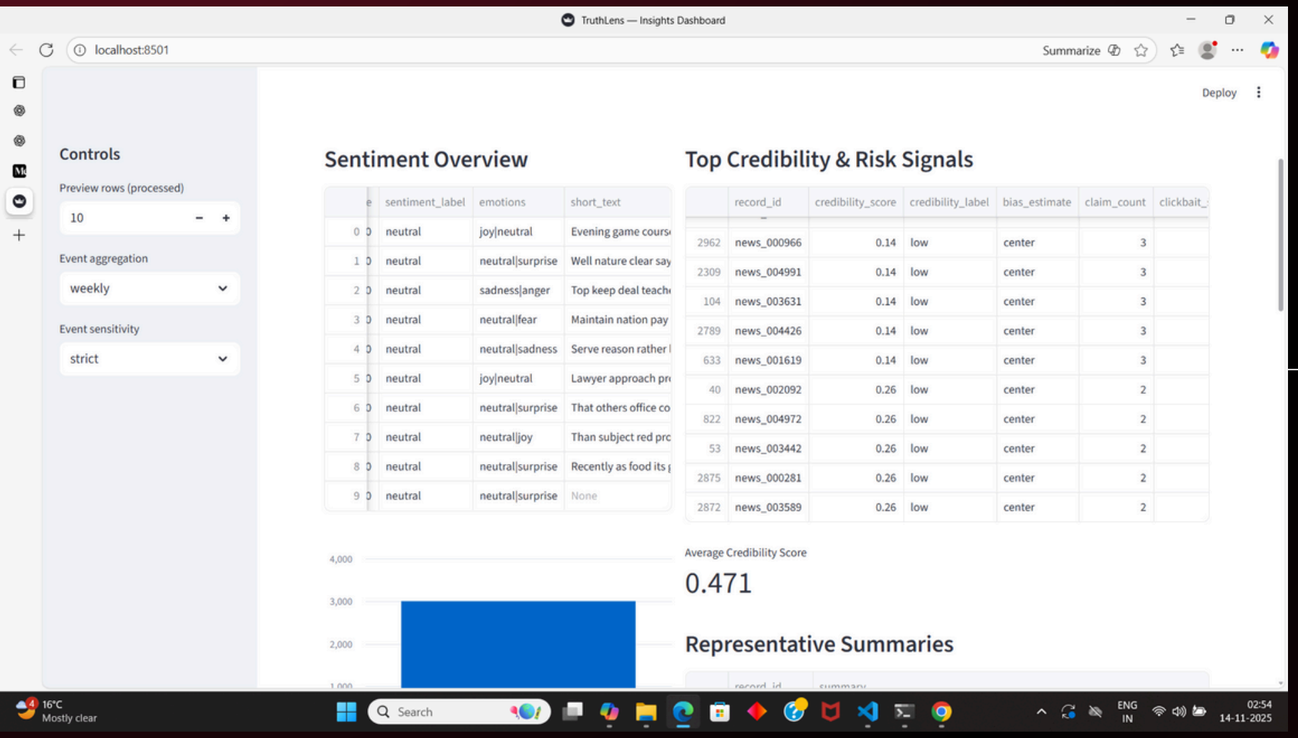
### FIVE NLP USE-CASES

- EMOTION-RICH SENTIMENT ANALYSIS
- DYNAMIC TOPIC & EVENT DETECTION
- ABSTRACTIVE SUMMARIZATION
- BIAS/FAKE-NEWS SCORING
- CHATBOT ASSISTANT

### PRESENTATION LAYER (STREAMLIT)

- UNIFIED DASHBOARD DISPLAYING INSIGHTS, CHARTS, COMPARISONS, SUMMARIES
- RAG-STYLE CHATBOT FOR INTERACTIVE EXPLORATION

# Implementation Details




## Tech Stack

- Python, Transformers, BERTopic, spaCy, scikit-learn, Pandas
- Streamlit for dashboard
- MiniLM embeddings / TF-IDF fallback
- Models Used
  - Emotion/Sentiment → "cardiffnlp/twitter-roberta-base-emotion"
  - Topic modelling → BERTopic + c-TF-IDF
  - Summarization → "facebook/bart-large-cnn"
  - Fake news & bias → Logistic Regression + TF-IDF features
  - Chatbot → "facebook/blenderbot_small-90M"

## Pipeline Execution

Preprocess → NLP inference → Save CSV insights → Load in Streamlit

- Heavy models wrapped to avoid slowdowns
- Designed with low-storage fallback (no embeddings required)

## Key Engineering Achievements

- Modular use-case system
- Fast batch inference on 3000 documents
- Streamlit UI with interactive filters, topic exploration & chatbot window

# Use-Cases & How They Work

## EMOTION-RICH SENTIMENT ANALYSIS

- EXTRACTS JOY, ANGER, FEAR, SADNESS, OPTIMISM, ETC.
- USES ROBERTA-EMOTION CLASSIFIER
- HELPS DETECT EMOTIONAL FRAMING IN NEWS.

## DYNAMIC TOPIC MODELING & EVENT DETECTION

- BERTOPIC DISCOVERS HIGH-LEVEL CLUSTERS
- TREND GRAPH REVEALS SPIKES OF EVENTS
- IDENTIFIES BREAKING NEWS PATTERNS.

## ABSTRACTIVE & FACT-AWARE SUMMARIZATION

- BART SUMMARIZES EACH ARTICLE INTO 2–3 LINES
- REMOVES REDUNDANCY, IMPROVES READABILITY.

## FAKE NEWS, BIAS & PROPAGANDA DETECTION

- ML-BASED CREDIBILITY SCORING
- DETECTS LINGUISTIC FEATURES LIKE EXAGGERATION, SUBJECTIVITY, SENSATIONALISM
- GENERATES BIAS LABELS: NEUTRAL/CENTER, LEFT, RIGHT, FACTUAL, HIGHLY OPINIONATED

## RAG - STYLE CHATBOT ASSISTANT

- RETRIEVES TOP RELEVANT ARTICLES
- ANSWERS “WHAT HAPPENED IN X?”, “WHY DID Y OCCUR?”, “SUMMARIZE TOPIC Z”
- USES BLENDERBOT SMALL MODEL + RAG LOGIC.

# Results, Insights & Value Delivered

## KEY INSIGHTS OBSERVED

- SENTIMENT: 40–45% ARTICLES CONTAIN STRONG EMOTIONAL FRAMING (FEAR/ANGER).
- TOPICS: 20+ STABLE CLUSTERS DETECTED; SEVERAL CORRESPOND TO POLITICAL/FINANCE EVENTS.
- SUMMARIES REDUCE CONTENT LENGTH BY ~85%, IMPROVING READABILITY.
- BIAS MODEL FLAGGED ~18% ARTICLES AS HIGHLY OPINIONATED.
- CHATBOT EFFECTIVELY ANSWERS HIGH-LEVEL QUESTIONS & RETRIEVES RELATED ARTICLES.

---

### Value to Users

- Faster understanding of complex news
- Helps detect misinformation patterns
- Emotion + bias insights reveal media tone
- Topic modeling highlights real-time events
- Chatbot enables conversational exploration

### Next Steps

- Deploy as web RAG application
- Live API-based news scraping
- Add stance detection & evidence retrieval

# Thankyou

Unlock the Power of Data!