TrendPilot – Week 4 Status Report

Project Overview & Current Progress

TrendPilot is a modular, agent-based system designed to automate end-to-end professional content creation for platforms such as LinkedIn. The overall architecture consists of **four core modules**:

1. **Trending Topic Identification**
2. **Post Generation**
3. **Engagement Prediction**
4. **Visual Generation & Publishing Support**

At this stage of the project, **primary development focus is on the Trending Topic Identification module**, which serves as the foundation for all downstream components.

Current Project Status

- **30,000+ records collected** from public news sources to support trend modeling
- **Exploratory Data Analysis (EDA) completed** to understand topic distribution, domain coverage, and noise patterns
- **Feature extraction and model development currently in progress**, focusing on trend relevance, frequency, and readiness signals
- Modular architecture validated for seamless downstream integration

Trending Topic Identification Module

Module Overview

The Trending Topic Identification module is the **first and most critical component** of the TrendPilot architecture. Its goal is to autonomously identify high-momentum, domain-relevant topics suitable for professional social media platforms such as LinkedIn.

This module establishes the **signal quality** for all downstream agents, including post generation, engagement prediction, and visual synthesis.

Problem Addressed

Professionals often struggle to determine:

- Which topics are currently trending
- Which trends are relevant to their professional domain
- Which topics have sufficient momentum to justify content creation

Manual trend discovery is:

- Time-consuming
- Subjective
- Often outdated

This module automates the process using **real-time public data sources** and **objective ranking signals**.

Data Sources

To ensure freshness, reliability, and reproducibility, only **free and publicly accessible APIs** are used:

- **NewsAPI.org** – Provides top headlines across domains such as technology, business, and science
- **GNews API** – Used as a secondary source to reduce single-source bias and improve coverage

Together, these sources provide:

- High-volume content
- Near real-time updates
- Domain-level categorization

The aggregated dataset currently exceeds **30,000 articles**, forming a strong empirical base for trend modeling.

Approach & Methodology

1. Domain Selection

Users begin by selecting a broad domain (e.g., Technology, Business, Sports).
This ensures contextual alignment before deeper personalization is applied in later stages.

2. Headline Aggregation

For the selected domain:

- Top headlines are fetched from multiple APIs
- Article fields are normalized (title, description, content)
- Duplicate, empty, or low-signal records are removed

This produces a clean, unified corpus of current events.

3. Topic Abstraction Using LLM

Raw headlines are often too granular or noisy for direct use. To address this:

- Each article is processed using an LLM-based topic extraction step
- The model summarizes content into a concise topic phrase (3–8 words)
- Each phrase captures the core event, technology, organization, or issue

This step converts unstructured news data into **post-ready topic candidates**.

4. Trend Readiness Output

At this stage, the system produces:

- Multiple topic phrases per domain
- Each topic supported by multiple independent news sources

These outputs are passed downstream for:

- Trend scoring
- Ranking
- Automated or manual selection

Alignment with Overall Architecture

This module corresponds to the **first stage of the TrendPilot pipeline**:

Identify Trending Topics
→ Generate Multiple Posts
→ Predict Engagement
→ Generate Visuals
→ Publish or Review Final Content

By isolating trend discovery as an independent agent, the system remains:

- Modular
- Scalable
- Easy to enhance without impacting downstream modules

Current Outcomes

- Fully automated, domain-based trend discovery
- Multi-source validation of trending topics
- Clean, concise topic phrases suitable for LinkedIn content
- Strong foundational input for engagement prediction models

Limitations & Ongoing Work

- Current trend scoring emphasizes **headline frequency**
- **Velocity, recency, and momentum-based features** are under active development
- Social engagement signals (likes, shares) will be incorporated when APIs permit

- Database persistence and scheduled execution planned for the next phase

Conclusion

The Trending Topic Identification module has successfully established a **robust, data-driven foundation** for TrendPilot. With large-scale data ingestion, completed EDA, and ongoing feature engineering and model development, the system is well-positioned to support intelligent content generation and engagement prediction in subsequent phases.