

Product Requirement Document (PRD)

Project Title: Reddit-Trend-Based News Ideation Platform

Prepared By: Rohan Dabas **Date:** June 18, 2025

1. Overview

The Reddit-Trend-Based News Ideation Platform is a web-based tool that helps content creators, journalists, and newsrooms identify high-potential news story ideas by analyzing trending Reddit discussions. It leverages Reddit data, performs sentiment and engagement analysis, categorizes topics, and uses the Gemini API to generate editorial-quality story angles, coverage reasons, and supporting references.

2. Goals & Objectives

- Automatically discover trending Reddit topics with high engagement.
 - Categorize and cluster similar posts into meaningful news topics.
 - Generate newsroom-ready story ideas with headlines, angles, urgency, and estimated reach.
 - Provide Gemini-generated reasons to cover a story based on editorial values.
 - Show supporting Reddit post references for transparency.
-

3. Core Features

3.1 Topic Discovery

- Use PRAW to fetch hot/trending posts from selected subreddits.
- Filter posts by engagement score (upvotes + comments).
- Cluster posts into topics using title similarity or subreddit overlap.

3.2 Topic Categorization & Sentiment

- Assign a category (e.g., Technology, Politics, Health) using keyword mapping or ML.
- Run sentiment analysis using TextBlob or another library.

3.3 Story Ideation Engine

- Select most engaging post per topic as the main post.
- Calculate detailed news values: timeliness, impact, conflict, etc.
- Estimate reach based on total topic engagement.
- Determine urgency (breaking, high, medium, low).
- Generate headline & angle using Gemini API (with fallback logic).
- Request Gemini API for 2-3 reasons a newsroom should cover this story.

3.4 Reference Generation

- Use top 3 Reddit posts in a topic as story references.
- Assign reference type: article, social_media, etc.
- Compute relevance score for each.

3.5 Frontend UI (React + Tailwind)

- **Dashboard View:** Summary cards for engagement, categories, post volume.
 - **Subreddit View:** Table showing post-level metrics.
 - **Topic Cards:** For each trending topic (title, category, sentiment).
 - **Story Idea View:** Shows headline, angle, reach, news value, urgency.
 - **Reference Cards:** Each with date, link, source, relevance bar.
-

4. Technical Stack

- **Frontend:** React + TypeScript + Tailwind CSS
 - **Backend:** Python (Flask)
 - **APIs:**
 - Reddit API (via PRAW)
 - Gemini API for LLM content generation
 - **NLP Tools:** TextBlob (sentiment), Regex (cleaning), Custom logic
 - **Deployment:** Local dev (with Vite & concurrently), future Netlify/Render option
-

5. User Stories

1. As a newsroom editor, I want to see trending stories with reach and urgency so I can pick what to cover quickly.
 2. As a journalist, I want ready-made headlines and angles to speed up my ideation process.
 3. As a reader, I want to understand where the story came from and verify sources.
 4. As a developer, I want clean APIs and components to extend the system easily.
-

6. Future Improvements

- Add clustering via embeddings (e.g., sentence-transformers).
 - Allow users to customize subreddits and timeframes.
 - Add tagging/labeling system for topics.
 - Track engagement changes over time.
 - Introduce story publishing workflow with CMS integration.
-

7. Known Issues & Assumptions

- Gemini API has token and format limits; error handling is in place.
- Estimated reach is heuristic-based, not precise.
- Topic clustering is currently simple (not AI-based).

- PRAW usage is rate-limited.
-

8. Appendix

Category Examples:

- Politics: Ukraine war, elections, legislation
- Technology: AI, Elon Musk, new gadgets
- Health: WHO announcements, vaccine updates

Engagement Score Formula:

- $(\text{upvotes} + \text{comments}) * \text{weight}$

Reach Estimation Formula:

- $\text{sum of engagement across topic posts} * 100$
-

End of PRD

Submitted by Rohan