

AGENTIC AI – APP REVIEW TREND ANALYSIS SYSTEM

1. Introduction

This project implements a complete Agentic AI system designed to analyze Google Play Store reviews and generate actionable trend insights for product teams. Modern applications receive a massive volume of user feedback daily, making it difficult to manually identify recurring issues or emerging problems. This system automates that process by ingesting reviews, extracting issues, consolidating similar topics, and tracking their trends over time.

2. Problem Statement

User reviews are unstructured, noisy, and linguistically diverse. The same issue may be described in many different ways by different users. Traditional topic modeling techniques often fail to merge semantically similar topics, resulting in fragmented insights and unstable trend analysis. The objective of this project is to build a system that can reliably identify, consolidate, and track user issues over time using semantic reasoning.

3. Why Agentic AI

Instead of relying solely on statistical topic modeling techniques such as LDA or TopicBERT, this project adopts an Agentic AI approach. Each stage of the pipeline is implemented as a specialized agent responsible for a distinct reasoning task. This modular design improves interpretability, stability, and recall, making the system more suitable for real-world product analytics.

4. System Architecture

The system follows a multi-agent pipeline architecture consisting of four major stages: Review Ingestion, Topic Extraction, Semantic Deduplication, and Trend Aggregation. Reviews flow sequentially through these agents, with each agent adding structure and intelligence to the data.

5. Review Ingestion Agent

The Review Ingestion Agent fetches Google Play Store reviews in controlled batches. Reviews are stored along with metadata such as review date and rating. Treating each day's reviews as a batch enables temporal trend analysis and simulates real-world streaming data pipelines.

6. Topic Extraction Agent

The Topic Extraction Agent converts raw review text into structured issue categories. This agent is designed for high recall, ensuring that important complaints are not missed. Rule-guided logic is used to produce consistent and interpretable topic labels.

7. Semantic Topic Deduplication Agent

This agent represents the core intelligence of the system. It uses sentence embeddings and cosine similarity to identify semantically equivalent topics and merge them into a single canonical category. This prevents topic explosion and ensures clean and reliable trend signals.

8. Trend Aggregation Agent

The Trend Aggregation Agent computes topic frequencies over a rolling 30-day window. The final output is a matrix where rows represent topics, columns represent dates, and each cell contains the frequency of a topic on that date. This format is directly usable by analytics and product teams.

9. Backend API (Optional Extension)

An optional backend API has been implemented using FastAPI to expose the final trend report. The API allows frontend dashboards or external systems to consume trend insights programmatically, demonstrating backend and full-stack readiness.

10. Output Description

The final output is generated as a CSV file containing topic trends over time. A category named 'Other feedback' is intentionally preserved to capture neutral or non-actionable comments, preventing noise from affecting issue trends.

11. Design Decisions

Key design decisions include prioritizing recall over precision, using semantic similarity for topic consolidation, maintaining modular agent design, and producing simple CSV and API-based outputs for easy integration.

12. Assumptions and Limitations

The system currently processes English-language reviews only and relies on rule-based topic extraction. The architecture is extensible to support LLM-based extraction, multilingual analysis, dashboards, and real-time ingestion.

13. Conclusion

This project demonstrates a production-ready Agentic AI pipeline for review trend analysis. By combining semantic reasoning, modular agents, and temporal aggregation, the system produces reliable and actionable insights that overcome the limitations of traditional topic modeling approaches.