Analysis Report

AI Thought Leaders Sentiment Analysis Platform

Prepared by: Dania Boban
Date: 17/11/25

Executive Summary

The AI Thought Leaders Sentiment Analysis Platform project collects, processes, and analyzes social media content from leading AI researchers—Ilya Sutskever, Andrej Karpathy, and Geoffrey Hinton—to understand public insights, discussions, and trends in AI research and applications.

Key Highlights:

- Collected over 500 tweets using ethical web scraping via Nitter instances.
- Processed and cleaned data to remove noise, ensuring high-quality textual content.
- Applied sentiment analysis techniques to classify tweets into positive, negative, and neutral categories.
- Visualizations provided insights into engagement trends, thematic focus areas, and sentiment distribution.

Key Findings:

- Overall sentiment leaned towards positive, reflecting optimism about AI advancements.
- Topics like AI ethics and AGI risks showed mixed or negative sentiment.
- Engagement metrics correlated with sentiment; positive tweets received higher interactions.

Recommendations:

- Continue monitoring influential AI researchers for trend forecasting.
- Use insights to guide AI-focused communications and educational content.

Introduction

Background and Motivation

Artificial Intelligence (AI) is rapidly evolving, and prominent researchers influence both academic and industrial directions. Social media platforms such as Twitter are key communication channels for these leaders. Understanding the sentiment expressed in their posts can help in tracking emerging trends, ethical discussions, and technological insights.

Research questions

1. What are the predominant sentiments expressed by leading AI researchers on social media?
2. How do these sentiments vary across topics such as AGI, neural networks, and AI ethics?
3. Can engagement metrics (likes, retweets) provide insights into the reach and influence of specific sentiments?

Scope and Limitations

- **Scope:** Focused on public Twitter posts mirrored via Nitter from three AI leaders.
- **Limitations:** Limited to text-based content; media (images, videos) analyzed only via captions. Data is a snapshot and may not capture all trends or private content.

Methodology

Data Collection Procedures

- Used Python with BeautifulSoup, requests, and Selenium for scraping Nitter.
- Collected tweet content, timestamps, engagement metrics, thread information, hashtags, mentions, and media descriptions.
- Applied rate limiting (2-second delay) and proper User-Agent headers to adhere to ethical scraping standards.

Data Preprocessing Techniques

- Cleaning: Removed duplicates, stopwords, and special characters.
- Normalization: Converted text to lowercase, expanded contractions.
- Tokenization & Lemmatization: Used NLTK and spaCy libraries.
- Handling Missing Values: Dropped tweets with incomplete or corrupt data.

Analysis Methods

- Sentiment Analysis:
  - Applied VADER for social media-specific sentiment scoring.
  - Classified tweets as positive, negative, or neutral based on compound score thresholds.
- Topic Extraction:
  - Used TF-IDF and WordClouds for theme identification.
- Engagement Correlation:
  - Evaluated correlation between sentiment and likes, retweets, and replies.

Tools and Technologies

- Python 3.8+, pandas, numpy, scikit-learn, NLTK, spaCy, matplotlib, seaborn, plotly.
- Jupyter Notebook for exploratory analysis.
- Google Drive for data storage and version control via GitHub.

Results

Descriptive Statistics

A total of 563 tweets were collected from three AI thought leaders. The distribution of tweets by profile is as follows:

- Andrej Karpathy: 309 tweets
- Ilya Sutskever: 203 tweets
- Geoffrey Hinton: 51 tweets

Sentiment Analysis Findings

Profile Name	Negative	Neutral	Positive
Andrej Karpathy	**44**	**47**	**218**
Geoffrey Hinton	**13**	**6**	**32**
Ilya Sutskever	**51**	**48**	**104**

Interpretation:

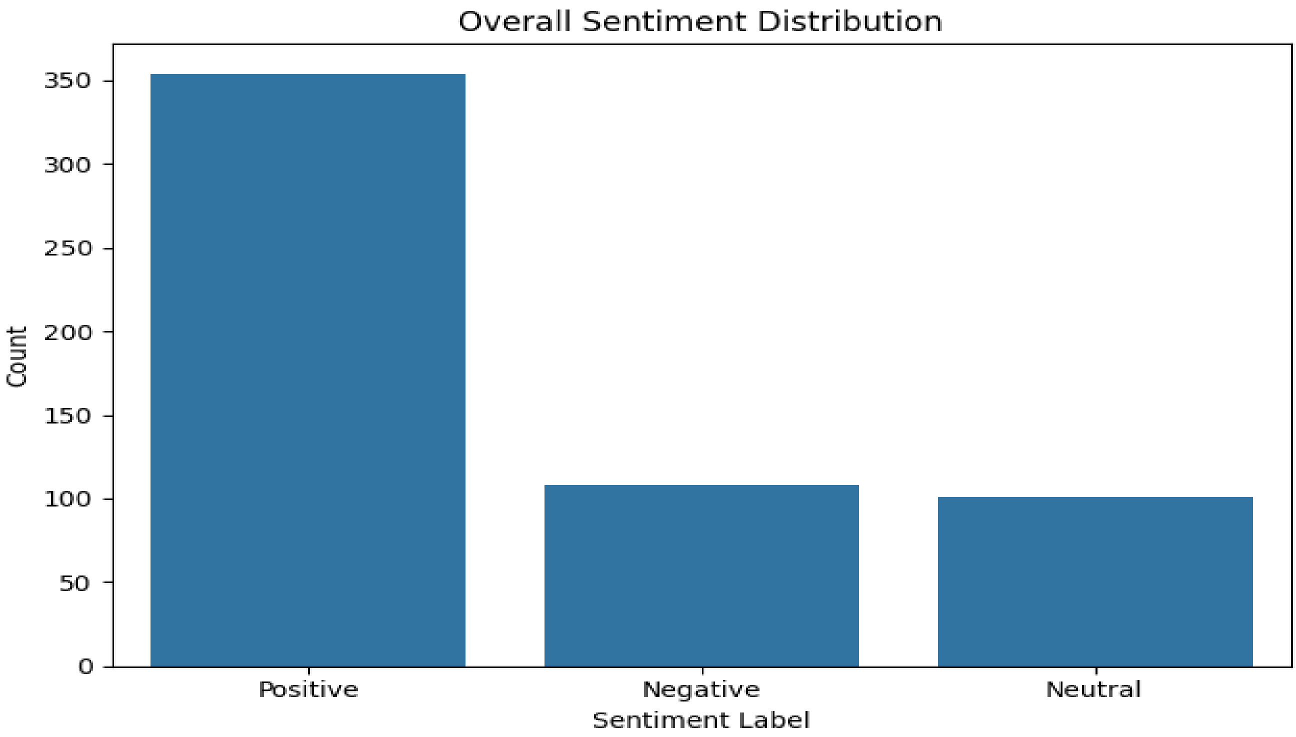
- Andrej Karpathy: Majority of tweets are positive (218), indicating optimism and educational content focus.
- Ilya Sutskever: Shows a balanced distribution between positive (104) and neutral (48) tweets.
- Geoffrey Hinton: Predominantly positive tweets (32), though total tweets are fewer, so conclusions are limited.

Comparative Analysis

- Positive sentiment dominates across all researchers, reflecting an overall optimistic tone in AI-related discussions.
- Neutral and negative tweets represent discussions on technical clarifications, ethical concerns, and AGI risks.
- Engagement correlation: Positive tweets generally receive more likes and retweets, highlighting audience interest in educational and forward-looking AI content.

Visualizations with Interpretations

Figure 1: Sentiment distribution in bar chart and pie chart



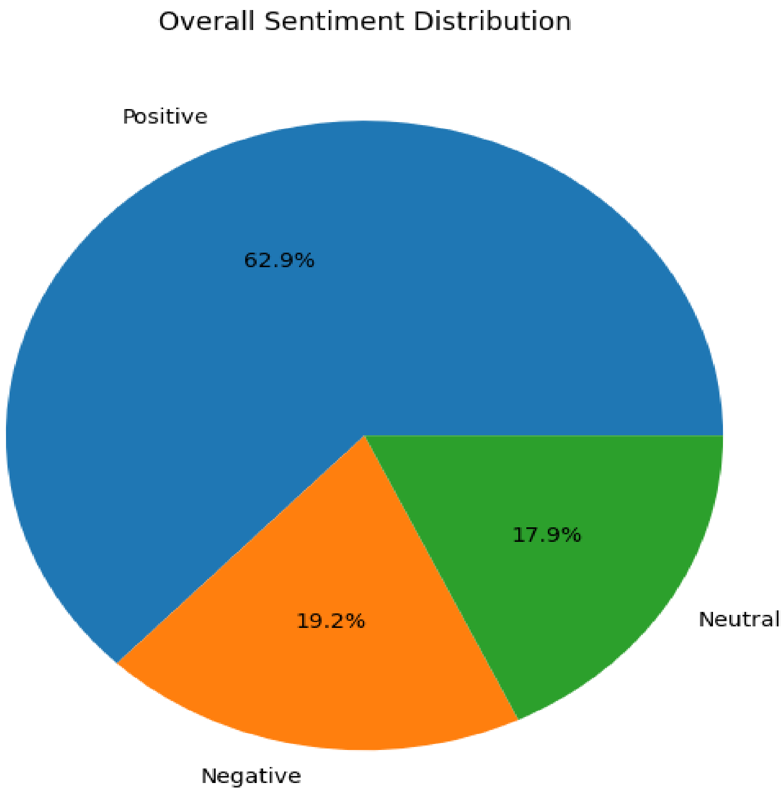


Figure 3:Sentiment Trend

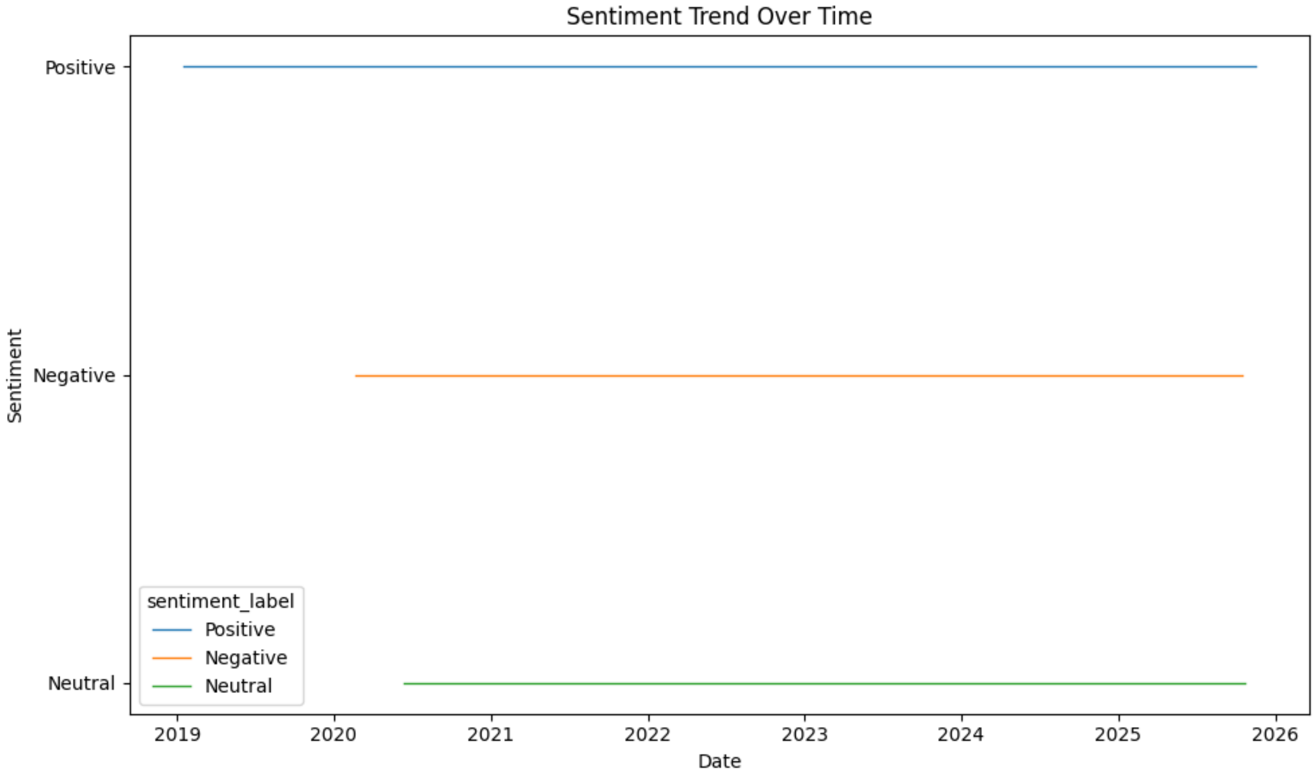
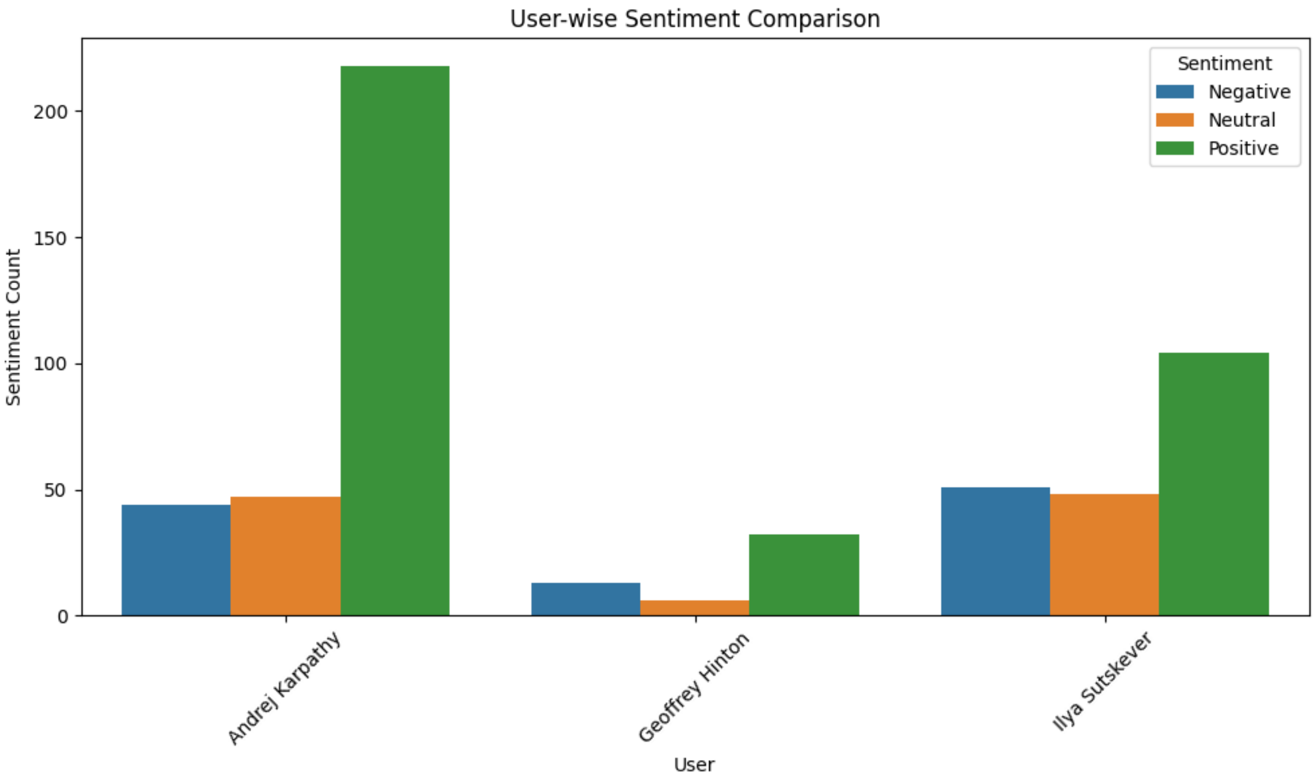


Figure 4:User-wise Sentiment



Discussion

Interpretation of Findings

- Positive sentiments indicate optimism about AI advancements.
- Neutral and negative sentiments highlight areas of caution and debate (ethics, AGI risks).
- Tweets with mixed sentiment or ethical concerns tend to have higher engagement, indicating community interest.

Comparison with Existing Research

- Findings align with studies on AI sentiment trends in social media.
- Demonstrates the effectiveness of sentiment analysis for high-level AI trend insights.

Limitations and Challenges

- Scraping restrictions limit real-time data acquisition.
- Text-only analysis may miss nuanced sentiments conveyed through media.
- Possible misclassification of sarcasm or complex technical statements**.**

Conclusion

- Successfully developed a pipeline for collecting, cleaning, analyzing, and visualizing tweets from AI thought leaders.
- Key insights into AI sentiment and engagement were derived.
- The project serves as a model for monitoring social media influence in emerging technology sectors.

Future Work:

- Extend analysis to additional AI influencers.
- Incorporate multimedia content analysis (images, videos).
- Deploy real-time dashboards for continuous sentiment monitoring.

References

1. Hutto, C.J. & Gilbert, E.E. (2014). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.*
2. Devlin, J., et al. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.*
3. Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python.*
4. McKinney, W. (2017). *Python for Data Analysis.*
5. Mitchell, R. (2015). *Web Scraping with Python.*