ASTRA MIDPOINT REVIEW

PURDUE UNIVERSITY® | Mitchell E. Daniels, Jr. School of Business

The Need For This Project

- Creators on Instagram, Facebook, and Threads depend on trend-driven content to stay relevant and grow their audience.

- Emerging trends often start on external platforms like Twitter/X, TikTok, YouTube, and Reddit.

- Meta creators struggle to detect and respond to these trends early due to lack of visibility.

- There is no unified system that identifies, contextualizes, and recommends relevant trends specifically for Meta platforms.

- A platform-specific solution is needed to aggregate cross-platform signals and align them with Meta's content behavior.

Creators on Meta platforms need a unified system that detects and contextualizes emerging trends from external sources to enable timely, strategic content creation.

Project Vision and Mission

- Enable Meta creators to stay ahead of trends with real-time, intelligent insights.

- Build an autonomous system that discovers, ranks, contextualizes, and translates cross-platform trends into actionable content strategies for Meta platforms.

Empowering Content Creation

Cross-platform trend tracking
Develop trend agents that monitor Twitter/X, Reddit, YouTube, and TikTok, structured for Meta-specific content discovery.

Ranking and Trend Classification
Leverage embeddings, LLMs, and weighted aggregation to turn trend signals into ranked, contextual recommendations for Meta platforms.

Content Strategies and Insights
Support content creators with an intuitive dashboard, personalized insights, and automated topic mapping aligned to their audience.

Overall Methodology

Data Collection

Extracted <u>Text data</u> from multiple platforms

- YouTube: official API
- Reddit: official API
- Google Trends: official API
- Twitter: 3rd Party API
- Tiktok: 3rd Party API

Save all data into <u>csv files</u>

Automated collecting process with <u>Github Actions</u>

LangChain Workflow

NLP Processing & LLM Topic Labeling

Consolidating Data & Aggregating Metrics

Cluster & Semantic Analysis

Finalizing Trend Scores

Web Dashboard

Web <u>interactive dashboard</u>

Show global trends on the landing page: **compare trends across platforms**

Dynamic filtering for deeper insights

1. Categorize by topic clusters and contents type
2. Visualize results
3. Contents Strategy Recommendation

Data & Ranking Methodology – Google Trends

Google Trends - Official API

2 datasets

__top_rising_terms (selection)__

- Up to 200 terms pulled every day of popular rising trends

  - Coverage_Ratio (Ratio of DMAs where the term appears. Usually 1.)

  - Median_Gain (how popular it is compared to last week)

  - Spread_Intensity (Coverage_Ratio X Median_Gain) *Ranking*

__top_terms (validation)__

- Up to 200 terms pulled every day of current popular terms

  - avg_rank (average rank of term in each dma on that certain day. Top 25 terms)

  - score (Sum of Google ranking scores, 0-100 scale, across DMAs)

- Using $300 of cloud credits to run scripts of roughly 10 MB per bot action.

trend_rising.csv

	A	B	C	D	E
1	term	dma_hits	coverage_ratio	median_gain	spread_intensity_score
2	camp mystic	210	1	15650	15650
3	diogo jota	210	1	13600	13600
4	texas flooding	210	1	11850	11850
5	michael madsen	210	1	9550	9550
6	lanthanum	210	1	5400	5400
7	atria	210	1	4450	4450
8	camp mystic texas	210	1	4350	4350
9	sophocles	210	1	3650	3650
10	guadalupe river	210	1	3350	3350
11	flooding in texas	210	1	3200	3200
12	oedipus	210	1	2700	2700
13	lanthanum meaning	210	1	2500	2500
14	kerrville flooding	210	1	2400	2400
15	atria meaning	210	1	2400	2400
16	julian mcmahon	210	1	2350	2350
17	cuatro de julio	210	1	2300	2300
18	sophia hutchins	210	1	2300	2300
19	kerrville tx	210	1	1900	1900
20	shylock	210	1	1900	1900
21	psg bayern	210	1	1850	1850
22	real madrid dortmund	210	1	1850	1850
23	inter miami montrÃ©al	210	1	1850	1850
24	musiala	210	1	1800	1800
25	mcdonald's big announcement	210	1	1800	1800
26	texas flood	210	1	1800	1800
27	hot dog eating contest	210	1	1750	1750
28	psg vs bayern	210	1	1500	1500
29	bobby jenks	210	1	1500	1500
30	astros vs dodgers	210	1	1350	1350

trend_top.csv

	A	B	C	D	E
1	term	dma_hits	avg_rank	total_score	coverage_ratio
2	casey anthony	210	16	100	1
3	fluminense vs al-hilal	210	2	100	1
4	france vs england	210	17	100	1
5	gavin mckenna	210	22	100	1
6	jamal musiala	210	3	100	1
7	jane birkin	210	5	100	1
8	jim curtis	210	16	100	1
9	lamar brown	210	13	100	1
10	lyndon byers	210	8	100	1
11	lynx vs mercury	210	8	100	1
12	matt kuchar	210	20	100	1
13	microsoft outlook outage	210	7	100	1
14	monica barbaro	210	10	100	1
15	mosie burks	210	17	100	1
16	mp materials stock	210	12	100	1
17	national fried chicken day 2025 deals	210	9	100	1
18	nato scrambles fighter jets	210	13	100	1
19	nvidia	210	4	100	1
20	oasis setlist	210	25	100	1
21	pirates vs mariners	210	17	100	1
22	red sox vs nationals	210	13.261905	100	1
23	rob mcelhenney	210	12	100	1
24	ryan louis mosqueda	210	18	100	1
25	sabalenka	210	3	100	1
26	scottish open leaderboard	210	20	100	1
27	shane van gisbergen	210	16	100	1
28	snack wrap	210	24	100	1
29	storm vs sun	210	18	100	1
30	supreme court trump	210	22	100	1

Data & Ranking Methodology – Twitter / X

- Data Source
  - Trends are extracted from Trends24.in, which aggregates real-time trending topics from Twitter/X across multiple regions.
  - Scraping is done using requests and BeautifulSoup without browser automation, enabling lightweight, repeatable execution.

- Data Collected per Trend
  - **Trend Name** – Trending keyword or hashtag
  - **Rank** – Original position in the regional trend list
  - **Top Position** – Highest rank attained during trend visibility
  - **Tweet Count** – Number of tweets associated with the trend (if available)
  - **Duration** – Time (in hours) the trend remained active
  - **Region** – Geographic location (e.g., US, India, Brazil)
  - **Timestamp** – Time of data extraction

- Re-ranking Methodology
  - Although trends are initially ranked on Trends24, we re-rank them using a normalized score for consistency and comparability across datasets.
  - Normalized Score is computed using min-max scaling within a region

- Automation
  - Deployed via GitHub Actions for scheduled execution (for every 6 hours)

Rank	Trending Topic	Top Position	Tweet Count	Duration
1	Clipse	2	914,453	18 hrs
2	Justin Bieber	1	3,926,405	14 hrs
3	Pharrell	5	102,920	8 hrs
4	So Be It	6	2,185,088	8 hrs
5	Marijuana	14	771,223	9 hrs
6	Malice	9	135,503	8 hrs
7	Giveon	11	10,882	8 hrs
8	#LetGodSortEmOut	2	240	6 hrs
9	#SWAG	2	274,108	8 hrs
10	Superman	5	2,642,631	8 hrs
11	The Birds Don't Sing	14	240	8 hrs
12	Sweet Spot	14	10,902	7 hrs
13	blackpink	5	5,292,341	5 hrs
14	Good Friday	2	174,356	4 hrs
15	Druski	16	435,495	8 hrs

Data & Ranking Methodology - TikTok

1. TikTok Trend Page

> This page shows the hashtag trend on Tiktok in the US for the last 7 days

- Hashtags
- Post quantity

2. Scrap videos from hashtag page

> Gather trending hashtags from Trends 24 page

> Scrap 10 videos per trending hashtag
- Hashtags
- Video Link
- Description
- Likes
- Comments
- Shares
- Views
- Scraped time

Data - YouTube

Automatically fetch and analyze **trending videos** across 8 regions
(US, IN, GB, CA, DE, JP, BR, AU) using **YouTube Data API v3**

Data Pulled per Video:

- **Metadata:** Video ID, Title, Description, Channel Info (Title, ID), Region Code (Country level), Published Date, Tags, Category

- **Performance Stats:** Views, Likes, Comments

- **Derived Metrics for Detecting Trends:**
  - Video Age (Hours), Duration (Seconds), Is Short (≤ 60s)
  - View/Like/Comment Velocity (per hour)
  - Like-to-View Ratio (%)

- **Channel Insights:** Subscriber Count, Channel Creation Date, Description

Data - Reddit

Automatically fetch **trending posts** using **Reddit Hot API**
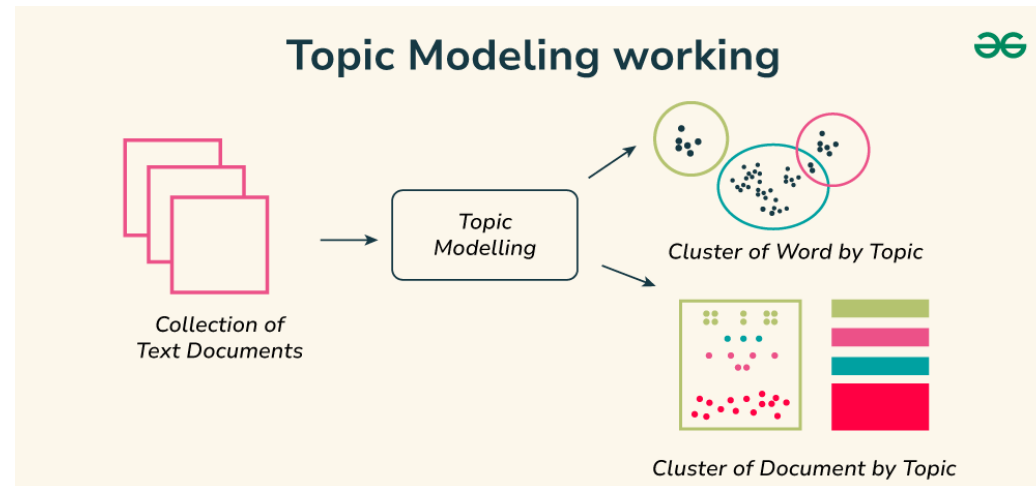
Data Pulled per Video:

- **API:** Extract data from Reddit "Hot" API endpoint which pulls the trending reddit posts data
- **Performance Stats:** Upvotes, Comments, rewards

approved_at_utc	subreddit	selftext	author_fullname	saved	mod_reason_title	gilded	clicked	title
	law		t2_dca49b9y8	FALSE		0	FALSE	Al Green Introduces Article Of Impeachment For Trump
	news		t2_1k0cep2afr	FALSE		0	FALSE	Purple Heart Army veteran self-deports after nearly 50 years in the U.S.
	BeAmazed		t2_v21xxrvi	FALSE		0	FALSE	we all need that guy.
	BlueskySkeets		t2_1630bts5	FALSE		0	FALSE	Nailed it
	interestingasfuck		t2_1197xs313h	FALSE		0	FALSE	Anthony Ervin's unbelievable comeback
	todayilearned		t2_1lbl5b0amy	FALSE		0	FALSE	TIL comedian Redd Foxx would often pretend to have a heart attack and pratfall for laughs. Years later, Foxx would have a heart attack for real, with many people thinking it was a bit as he fell to the ground. He d
	AdviceAnimals		t2_cbvv3izt	FALSE		0	FALSE	Yeah, it didn't work for Michael either.
	SipsTea		t2_1dzdnq1lr5	FALSE		0	FALSE	Why dating is over for men
	politics		t2_w25ndfc	FALSE		0	FALSE	Donald Trump Nobel Peace Prize nomination withdrawn
	goodnews		t2_17yo6dkbhc	FALSE		0	FALSE	Donald Trump Nobel Peace Prize nomination withdrawn
	facepalm		t2_ckwy7	FALSE		0	FALSE	"IF I TYPE IN CAPS, PEOPLE TAKE ME SERIOUSLY."
	Fauxmoi		t2_5fj4njyi	FALSE		0	FALSE	Trump is NOT happy this morning with Israel and Iran saying "they don't know what the f*ck they're doing" after breaking ceasefire
	PublicFreakout		t2_g4kd9ccob	FALSE		0	FALSE	Trump says Israel and Iran "don't know what the f*ck they're doing."
	mildlyinteresting		t2_3l70vc57	FALSE		0	FALSE	I cut a healthy looking onion which was actually hollow and rotting from the inside.
	funny		t2_a5jc9yd7	FALSE		0	FALSE	Praised be the mighty duck gods
	Whatcouldgowrong		t2_drc6ql7n	FALSE		0	FALSE	Carrying food above your head
	agedlikemilk		t2_17yo6dkbhc	FALSE		0	FALSE	After r/conservatives embarrassing display yesterday these guys are in absolute shambles
	memes		t2_5j841owj	FALSE		0	FALSE	Become the government
	OneOrangeBraincell	Penguin from Ma	t2_bcl1bb4t	FALSE		0	FALSE	We were laughing so hard at this. Why does she look so… fake ??
	Awww		t2_1qfkpsc1k8	FALSE		0	FALSE	Meet maya ❤️🥰
	comics		t2_9m5si	FALSE		0	FALSE	Time Loop
	whenthe		t2_6ecqg3u8	FALSE		0	FALSE	format mismatch
	Steam		t2_c39f7amn4	FALSE		0	FALSE	Microsoft has integrated Steam into the Xbox PC App, and it looks fantastic.
	Anticonsumption		t2_12g15l	FALSE		0	FALSE	Venice against Jeff Bezos wedding - taken today (23rd June)
	therewasanattempt		t2_jyy61mlo	FALSE		0	FALSE	.. to avoid having a cult leader as President
	pics		t2_1098fybz	FALSE		0	FALSE	OC: Activists in Venice deploy banner displaying Jeff Bezos, protesting his wedding in the city
	Eyebleach		t2_1rrj5oqu1	FALSE		0	FALSE	Everyone's welcome in this pack
	nba	[Source]]https://v >Yet if one we >He takes far >Smith's threa	t2_sod0nzlt6	FALSE		0	FALSE	Jon Lewis: "NBA will be better off without Stephen A. Smith"
	science		t2_6hji	FALSE		0	FALSE	CRISPR used to remove extra chromosomes in Down syndrome and restore human cell function. Japanese scientists discovered that removing the unneeded copy using CRISPR gene-editing normalized gene e
	MadeMeSmile		t2_uzgn2ggm	FALSE		0	FALSE	Encouraging Rejection Letter from NBA Champ's GM
	LeopardsAteMyFace		t2_txb6pyw	FALSE		0	FALSE	ICE Detains Wife of Staunch MAGA Supporter Who Gave Almost All His Net Income to Trump Campaign
	PropagandaPosters		t2_1ozyvik3wh	FALSE		0	FALSE	Pro-choice poster, 1981, U.S.
	interesting		t2_8whkqkst	FALSE		0	FALSE	In China, when traffic is heavy, lanes are opened to allow cars to move freely.
	politics		t2_v7o3l027	FALSE		0	FALSE	Trump Explodes as His Ceasefire Unravels: 'They Don't Know What the F*** They're Doing!'
	LeagueOfMemes		t2_v5sa5bui3	FALSE		0	FALSE	Thank you Dota2
	Piracy		t2_1twywwfn	FALSE		0	FALSE	Everyone needs to sign
	rareinsults	W	t2_b67yu5bo0	FALSE		0	FALSE	Caw caw 🐦
	Wellthatsucks		t2_drc6ql7n	FALSE		0	FALSE	Dropping food over the floor
	Damnthatsinteresting		t2_mmvkx25v	FALSE		0	FALSE	Google Earth captures the stunning transformation of our planet over 3 decades
	lego	Perry The Brick S	t2_1s1rnnbqmdu	FALSE		0	FALSE	A Brick Separator Platypus

Data & Ranking Methodology – Reddit & YouTube
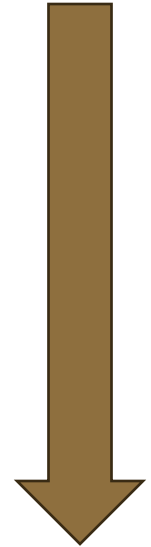
Data Extraction	Extract data from Reddit and Youtube respectively
LDA	Extract text data from the posts and apply topic model on top of the preprocessed text
Aggregation	Group frequently occurring words into a topic bucket based on latent topic relevance
Topic Label	Use LLM to label the topic buckets based on the keywords
Ranking	The topics are then ranked based on engagement

Topic Modeling working

Collection of Text Documents → Topic Modelling → Cluster of Word by Topic

Cluster of Document by Topic

Consolidated Trend Ranking Methodology

Google Trends

Twitter / X

Reddit

YouTube

TikTok

Normalize score based on Min Max scaling

Assign weights to each score based on end goal

Rank topics based on weighted score

Apply clustering to normalize and merge similar topics
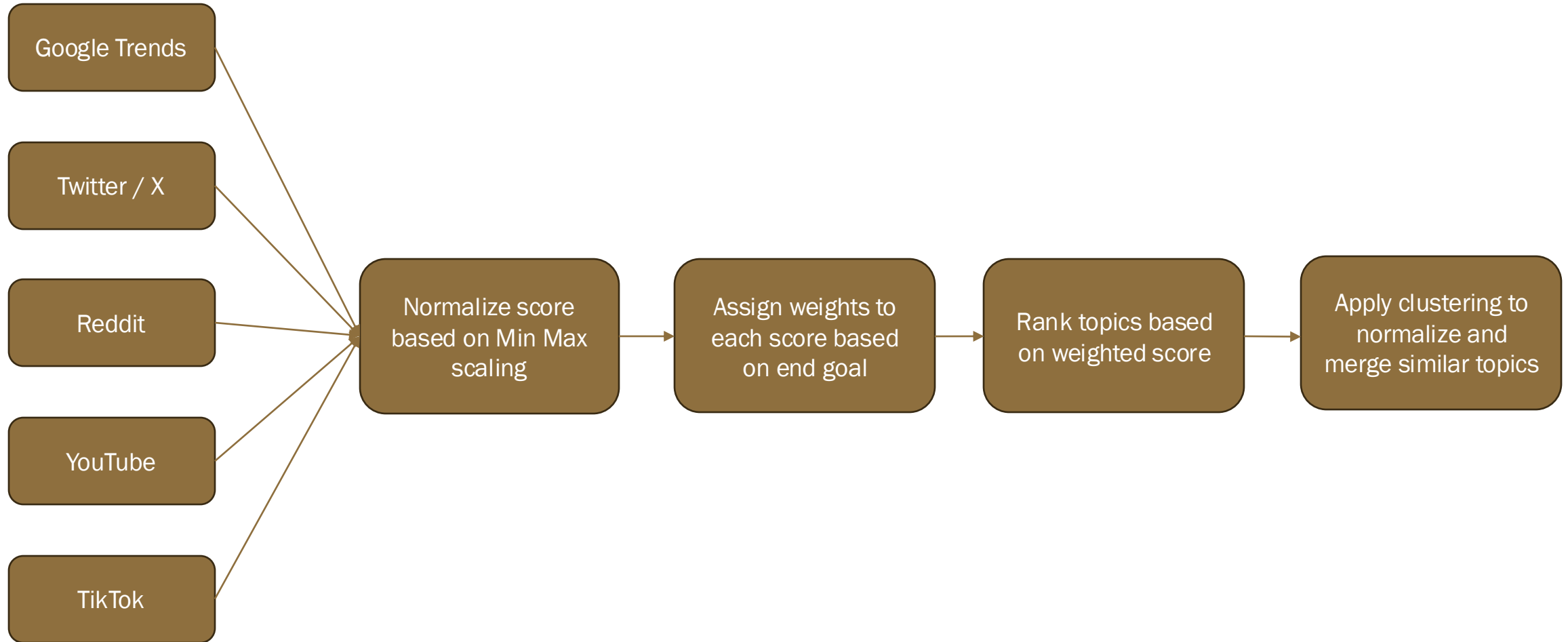
LangChain Workflow

LangChain Tools → Chaining	
Load Latest Files	Dynamic updates for fresh data
Data Preprocessing	Clean text | Remove stopwords | Lemmatization
NLP & LLM Topic Labeling	LDA → High probablility keywords | Generate topic names with LLM
Data Merging	Merge tabular data into one data frame
Metric Aggregation 1	Aggregate metrics within each platform
Semantic & Cluster Analysis	Similarity measures (Jaccard, SequenceMatcher) | Clustering (KMeans, DBSCAN)
LLM Matching & Labeling	Cross-platform semantic matching with LLM | Generate cluster names with LLM
Metric Aggregation 2	Aggregate metrics across platforms

Semantic & Cluster Analysis

K-means

Total Clusters: 15

Characteristics: Most balanced grouping, clear cluster boundaries

DBSCAN

Total Clusters: 556

Noise Topics: 2,275 (57.8%)

Meaningful Clusters: 555

Characteristics: Density-based natural grouping

Similarity-based

Total Clusters: 2,902

Largest Cluster: 33 topics

Average Cluster Size: 1.4 topics

Characteristics: Fine-grained grouping based on high similarity

LLM-powered Topic Keyword Generation

Semantic & Cluster Analysis: Output

🎯 Top 5 Named Clusters

1. Social Media Influencers (Cluster 2)
Topic Count: 53 | Platform Distribution: YouTube (42), X (7), Reddit (4)
Key Topics: Shoreline Mafia Social Media, Social Media Influence, Minecraft Influencers, Palestinian Music Conflict

2. Gaming Love Community (Cluster 15)
Topic Count: 29 | Platform Distribution: X (17), YouTube (9), Reddit (3)
Key Topics: Love and Redemption, Pokemon Gaming, Power of Love, Minecraft Gaming

3. Family Struggles (Cluster 0)
Topic Count: 26 | Platform Distribution: Reddit (19), X (4), YouTube (3)
Key Topics: Car Salesman Stories, Car Rental, Personal Transformation, Parenting Fatigue

4. Celebratory Occasions (Cluster 59)
Topic Count: 14 | Platform Distribution: X (12), Reddit (2)
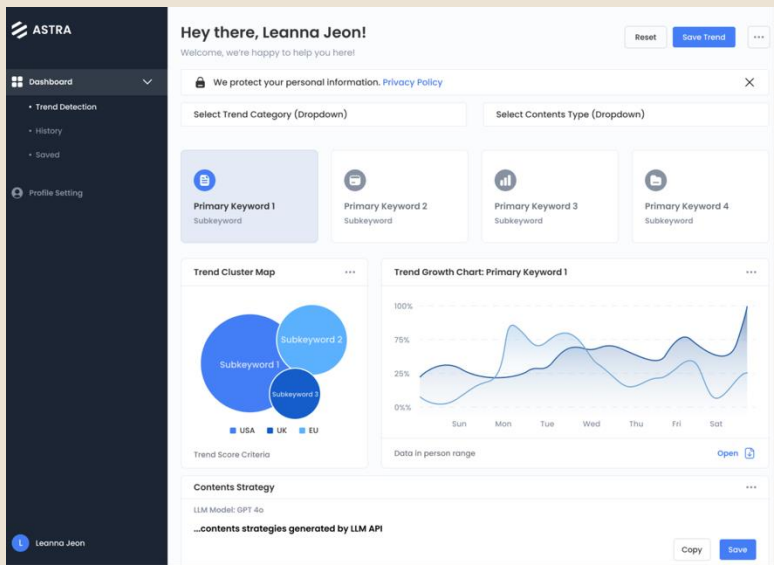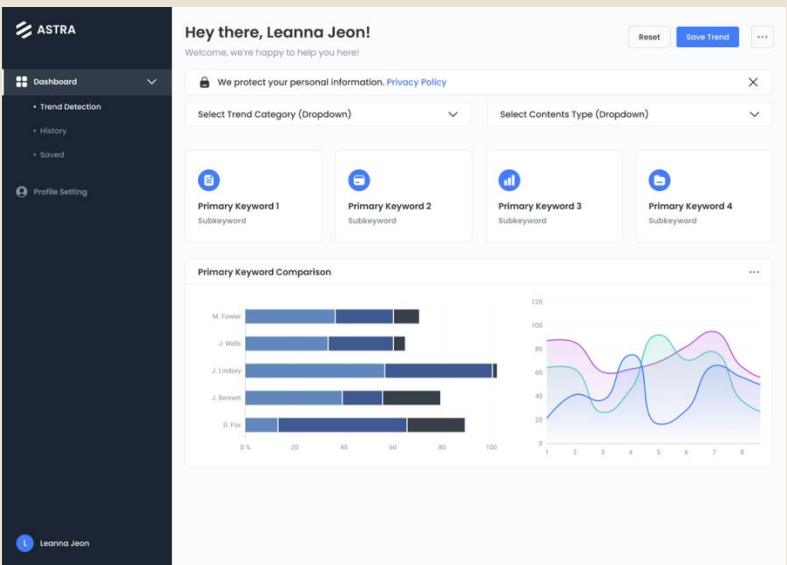Key Topics: Tattoo Tips, Happy Tattoo Designs, Happy Friday Eve, Happy Anniversary

5. Future Dreams (Cluster 88)
Topic Count: 12 | Platform Distribution: X (12)
Key Topics: Back to the Future, DREAM THE FUTURE, JENO IS THE FUTURE, DREAM TEAM

Dashboard: Design Mockup

Dashboard: Features

User Input Options

- Select trend category (e.g. Gaming, Beauty, Fitness, etc.)
- Choose content type: Text, Image+Text, Long-form Video, Short-form Video

Main Output

- Card view showing primary keywords and related subkeywords
- Visual comparisons of primary keywords with graphs

Drill-Down View (on keyword click)

- **Trend Cluster Map:** Displays trend scores and relationships among similar or subkeywords
- **Trend Growth Chart:** Shows growth patterns and activity spikes over the past 7 days or 24 hours
- **Content Strategy Recommendations:** AI-driven suggestions generated by an LLM

Next Steps

Important: *Finalize Trend metrics and weights based on video content type*

Our suggestions/questions for the future...

Validation and Changing Weights

A. Change weights of platforms and engagement metrics based on feedback from an external source (Possibly Meta Trending data)

Model Upgrades

A. Active Learning Loop: feed false-negatives & borderline cases into weekly fine-tune (LoRA on Llama-2 7B)
B. RAG for Context: pull top Reddit / TikTok posts for each term. Enrich prompt with media type & sentiment cues. Geo-burst features?
C. Agent Scale: Capacity is likely Mixtral 8 7B or Llama 2 7B.

Request

A. Internal Content-Performance endpoint (read-only)
- Post-ID, publish-time, 24 h views & saves.
- *Overhead:* one JSON fetch nightly (<1 MB).
- Rolling Back-test (6 months)

Thank You!

Any Questions?

PURDUE UNIVERSITY® | Mitchell E. Daniels, Jr. School of Business