

ASTRA FINAL PRESENTATION

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 and lack of inspiration.
- There is no unified system that identifies, contextualizes, and recommends relevant trends and aligned content 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.
- Generate platform-specific, aligned Instagram content for creator use

Empowering Content Creation

Cross-platform trend tracking

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

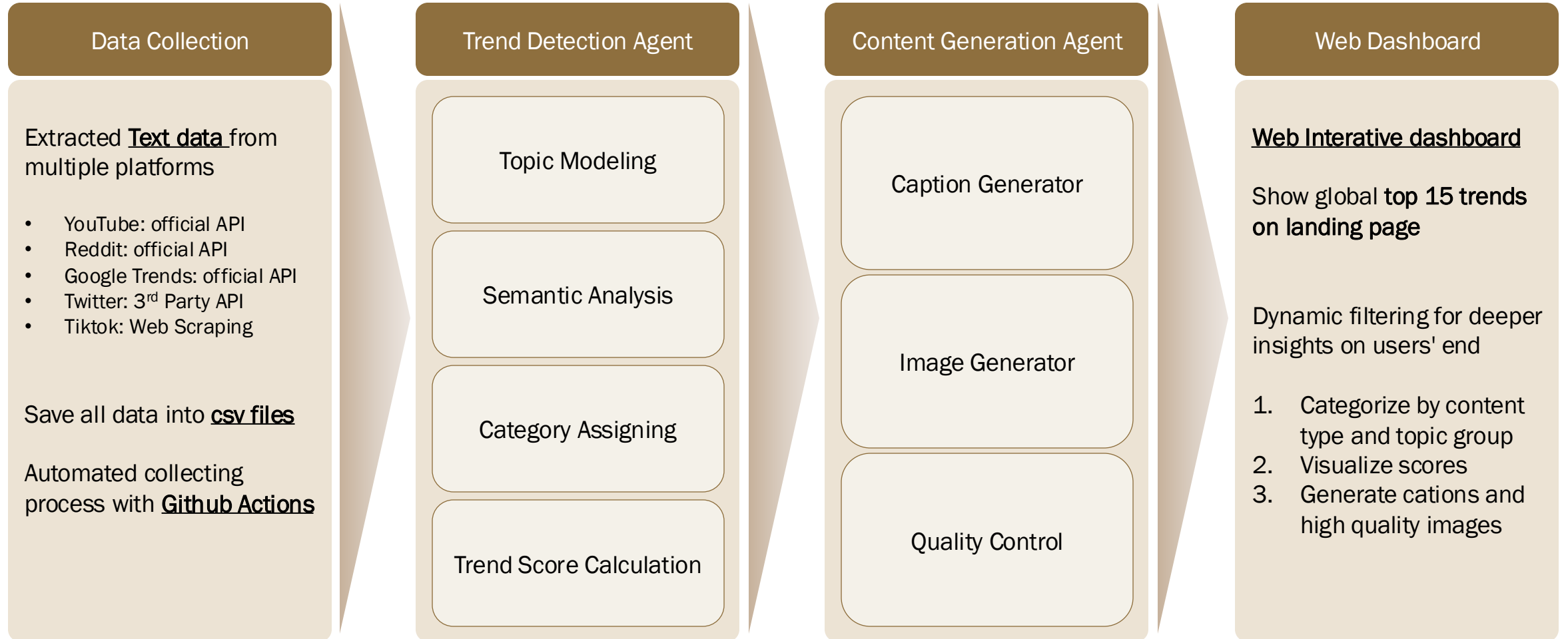
Ranking and Trend Classification

Leverage embeddings, LLMs, and weighted aggregation to turn trend signals into ranked, contextual recommendations for Instagram.

Content Strategies and Insights

Support content creators with an intuitive dashboard, personalized insights, and automatically generated content for inspiration or use.

Overall Methodology



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 vs 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	tamar 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 settlist	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

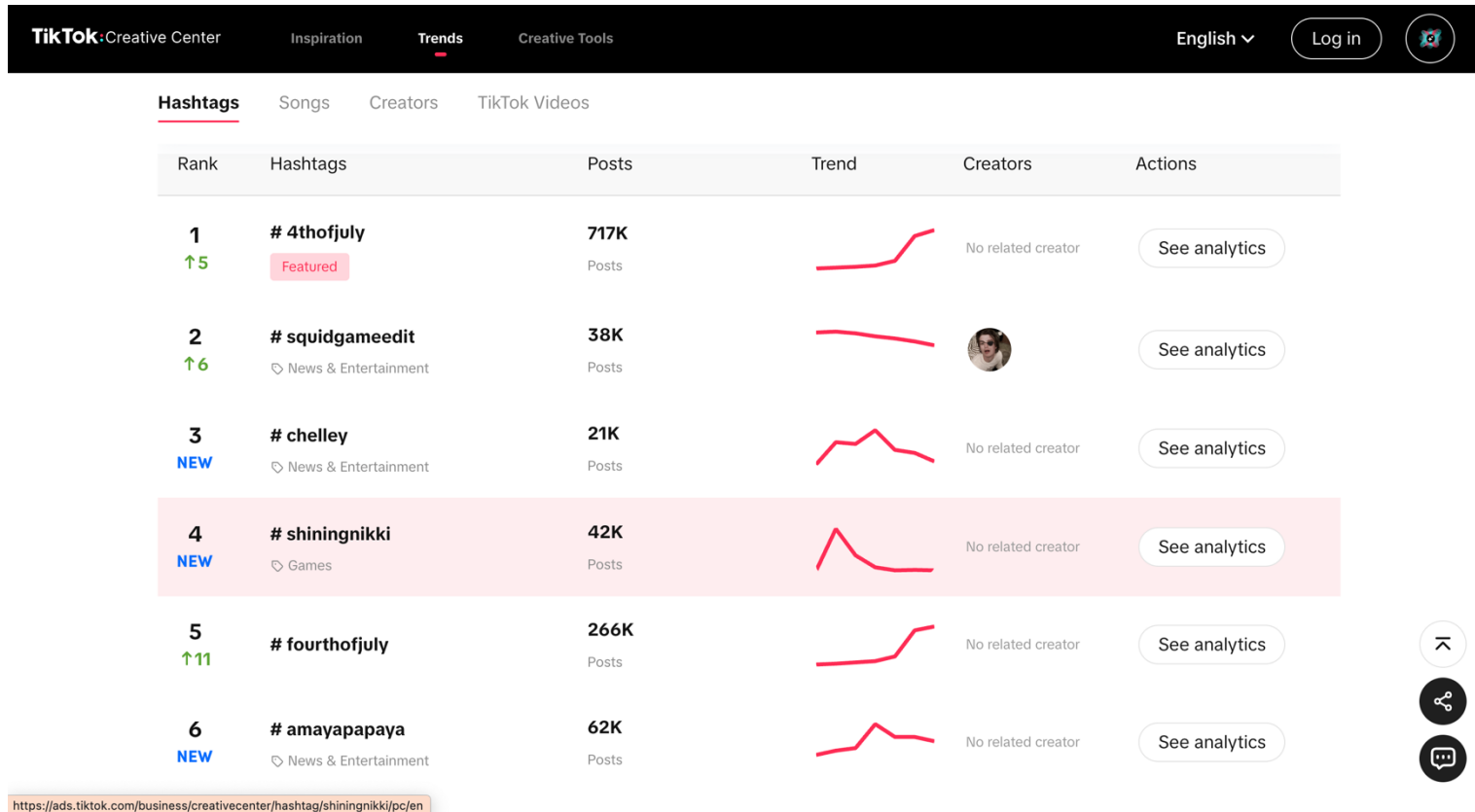
Data & Ranking Methodology – Twitter / X

- Trend Name – Trending keyword or hashtag
- Rank – Position in the regional trend list
- Top Position – Highest rank attained
- Tweet Count – Number of tweets associated with the trend
- Duration – Time span the trend remained active
- Region – Geographic location (e.g., US, India, Brazil)
- Timestamp – Time of data collection

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

TikTok Trend Page



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

- Hashtags
- Rank
- Posts count
- Scraped time
- 'See analytics' for more metrics

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 (≤ 60 s)
 - View/Like/Comment Velocity (per hour)
 - Like-to-View Ratio (%)
- Channel Insights:** Subscriber Count, Channel Creation Date, Description

Video ID	Title	Channel Title	Channel ID	Region Code	Published At	Video Age (H)	Duration (S)	Is Short	Category ID	Tags	View Count	Like Count	Comment C	View Velocity	Like Velocity	Comment Vi	Like-to-View	Subscriber C	Channel Put	Description
5B7hEY7z15	My Daughter's Secret Crush	Jordan Matter	UCKaCaz5N	US	2025-07-12	22	1894	FALSE	24	salish matter	4430010	81858	12610	201364.1	3720.82	573.18	1.85	30500000	2006-12-21	Can you
4dFtQJ9aUk	YoungBoy Never Broke Again - Kickboxer [Official Music Video]	YoungBoy Never Broke Again	UCW4jraMK	US	2025-07-11	37	215	FALSE	10	YoungBoy Ni	934782	82409	9765	25264.38	2227.27	263.92	8.82	14400000	2015-01-04	YoungBoy
CpCV2SkKc	BLACKPINK - 'As I Am' (Official MV)	BLACKPINK	UCOmHUn-	US	2025-07-11	56	194	FALSE	10	YG Entertain	43737718	3953480	339198	781030.7	70597.86	6057.11	9.04	97600000	2016-06-29	BLACKPINK
1pU3SSz8Tl	Found Village: City In Minecraft	PrestonPlayz	UCIZam2u1	US	2025-07-12	22	1205	FALSE	20	minecraft.pri	575465	6714	661	26157.5	305.18	30.05	1.17	16800000	2018-03-13	As I Am
q8EQB0822	I Fixed My AMG GT R For FREE & Discovered	LegitStreetCars	UCT-2OPZtE	US	2025-07-12	23	3072	FALSE	2	legitstreetcar	534296	43319	3219	23230.26	1883.43	139.96	8.11	1350000	2016-04-20	Check out
h985eQh9g	I Built a Mega Fish Tank (for the coolest fish)	TerraGreen	UCKYVerz4P	US	2025-07-12	23	825	FALSE	15	aquarium.aq	650598	11410	1216	28286.87	496.09	52.87	1.75	2290000	2023-09-20	Support my
GrpKGUer6	Escaping 1,000 Layers GANT WATER BALLO	Unspeakeable	UCwIWAbieL	US	2025-07-12	21	1664	FALSE	22	unspeakeable	428547	7658	1684	20407	364.67	80.19	1.79	18900000	2016-05-01	WANT TO
l7gBB7Kcd	Clipse: Tiny Desk Concert	NPR Music	UC4eXhJl4	US	2025-07-11	44	1404	FALSE	10	NPR,NPR Mu	1074452	92035	11022	24419.36	2091.7	250.5	8.57	11500000	2008-01-03	Bobby
EydtZuOCRE	SMG4: A VERY Unlikely Friendship	SMG4	UC8LcA3gY	US	2025-07-12	20	731	FALSE	24	Nintendo,Su	495977	30391	3707	24798.85	1519.55	185.35	6.13	8980000	2009-02-14	Mr. Puzzles
05vyterhe_o	PAST LIFE: Episode 1 - BACK TO THE BEGIN	NPR Music	UCR9GcoQ	US	2025-07-11	44	3799	FALSE	20	Minecraft,Gri	1315412	98237	11359	29895.73	2232.66	258.16	7.47	8730000	2009-11-22	NEW Life
etHhQaoEW	TWICE 'THIS IS FOR' MV	JYP Entertainment	UCaO6TmIC	US	2025-07-11	56	134	FALSE	10	JYP Entertain	14014376	1115120	139008	250256.7	19912.86	2482.29	7.96	30700000	2008-01-25	TWICE
gQIP3QPhj	The final minute of Katie Taylor vs. Amanda Serrano	Netflix Sports	UC_UldZuL	US	2025-07-12	32	69	FALSE	17	amanda serr	332487	914	2483	10390.22	28.56	77.59	0.27	2150000	2019-10-22	Katie Taylor
vsPceFlyk	mgk - vampire diaries (Official Music Video)	mgk	UCUa8sGpU	US	2025-07-11	44	196	FALSE	10	1307691	60094	8380	29720.25	1365.77	190.45	4.6	9320000	2006-12-27	mgk -	
UdO_kK7Gm	When Shayla ask Amber is her and Anthony	Kountry Wayne	UCiz19gacv	US	2025-07-12	23	264	FALSE	23	King Kountr	206887	11406	7037	8995.09	495.91	305.96	5.51	536000	2015-03-30	-
Fouqo1zNPt	Duke Dennis & AMP Goes To Six Flags	DukeDennis LIVE	UCz2rQXv	US	2025-07-12	35	1299	FALSE	24	duke dennis	557065	12140	1104	15916.14	346.86	31.54	2.18	1400000	2021-07-07	But maybe
HDHMB2zZd	Juice WRLD - The Way (feat. XXXTENTACION)	JuiceWRLDVEVO	UC3_471gF	US	2025-07-11	44	249	FALSE	10	Juice WRLD.J	1132229	155627	18955	25732.48	3536.98	430.8	13.75	27600	2018-07-20	Juice
6-dKlWdKv	'Love Island USA': Ace Dishes On If He's The	Access Hollywood	UCIKGMZm	US	2025-07-12	34	1086	FALSE	24	Love Island L	277493	9705	1726	8161.56	285.44	50.76	3.5	2200000	2006-05-04	Ace Greene
1vVN810kp	BUDGET CHALLENGER R/T TRANSFORMATI	TOM'S REFURB	UC6P2xQp	US	2025-07-11	37	2319	FALSE	2	488346	23733	1430	13198.54	641.43	38.65	4.86	643000	2017-11-29	HILARIOUS	
znN8eJRTA	The Strandness of Death Stranding 2	videogamedunkey	UCsvn_Po8	US	2025-07-11	41	1129	FALSE	20	dunkey,deat	933095	46558	2976	22758.41	1135.56	72.59	4.99	7570000	2010-10-21	GET YO
JOFZTzqvz	ATEEZ(에이티즈) - 'In Your Fantasy' Offic	KQ ENTERTAINMENT	UCQdQ-lqP	US	2025-07-11	56	222	FALSE	10	KQ,ATEEZ,ATEEZ	13422937	313764	26507	239695.3	5602.93	473.34	2.34	3070000	2016-04-25	ATEEZ(에이티즈)
HAz2G_Pw	Trolling A Fake Owner With The Real One	Gara PLAYS	UC2XcpTat	US	2025-07-11	40	977	FALSE	22	roblox,garag	589284	6323	913	14732.1	158.07	22.82	1.07	47200	2025-03-03	Trolling A
QzbXlW1Wc	Noah Lyles vs Letsie Tebogo: 200m Olympic	RioTrack	UC1Fp5XUJ	US	2025-07-11	39	433	FALSE	17	333477	4706	743	8550.89	120.67	19.05	1.41	655000	2016-07-01	Noah Lyles	
H4WXXqm2	Easiest Roblox Game VS Worst Player	Laughability	UCn4RkqD1	US	2025-07-12	33	2163	FALSE	20	roblox laugh	312836	12420	2449	9479.88	376.36	74.21	3.97	2300000	2017-01-29	But maybe
Yzbw8WpP	Conan Gray - Vodka Cranberry (Official Mus	ConanGrayVEVO	UCelmCJLZ	US	2025-07-11	56	296	FALSE	10	Conan Gray,	983939	108539	8582	17570.34	1938.2	153.25	11.03	28800	2018-09-04	Listen and
rhsE5OASd	Jorjiana - Shark (Remix) [feat. Babyface E	C Jorjiana	UC2rCFKoi	US	2025-07-11	44	222	FALSE	10	glorilla,loesh	178907	13197	765	4066.07	299.93	17.39	7.38	84300	2024-07-16	The official
M_6_t2X3W	This Peak Video Was Difficult	SMi7plus	UC-gW4TeZ	US	2025-07-11	41	1491	FALSE	20	smi7y,smi7	1166244	61901	2236	28444.98	1509.78	54.54	5.31	4010000	2020-06-07	This was my
msGugelph	DAISIES	Justin Bieber	UCwFjwMjO	US	2025-07-11	56	177	FALSE	10	justin,bieber	2302844	111711	7266	41122.21	1994.84	129.75	4.85	7570000	2007-01-15	Stream/Do
P-9dHlgt1V	Clipse: Kendrick Lamar, Pusha T, Malice - Ch	clipseVEVO	UCksLN02	US	2025-07-11	56	244	FALSE	10	clipse,Kendri	580580	31828	2305	10367.5	568.38	41.16	5.48	77800	2009-09-02	Stream
3nChK7u8r	THE MERRY-GO-ROUND The Amazing Digit	The Stupendum	UCvbc2N6b	US	2025-07-11	40	437	FALSE	10	TADC,The Ar	468190	42845	2592	11704.75	1071.12	64.8	9.15	1180000	2016-02-25	Roll up, roll
3okCoydVB	I Ruined A Movie Premiere	Niko Omilana	UCdCumdO	US	2025-07-11	43	1836	FALSE	24	3016538	224348	8022	70152.05	5217.4	186.56	7.44	7970000	2011-06-21	tsnuck into	
us4I5mtgb	About 200 people detained during immigrat	KCAL News	UCHH1u0kY	US	2025-07-12	35	258	FALSE	25	KCAL 9 New	403276	3000	2023	11522.17	85.71	57.8	0.74	1110000	2013-12-04	An
F1s1PmW6J	Por Sus Besos - Tito Double P (Video Offic	Tito Double P	UCj1EO3U	US	2025-07-11	60	187	FALSE	10	1950833	80637	3162	32513.88	1343.95	52.7	4.13	2500000	2023-06-05	AE&ESCUC	
cgH6po02r	Ghost of Yvelte - State of Play Gameplay De	PlayStation	UC-2Y8dQb	US	2025-07-10	63	1153	FALSE	20	PlayStation,J	2724703	155673	11688	43249.25	2471	185.52	5.71	16400000	2005-12-16	https://stor
wel76DGH	DRAKE - WHAT DID I MISS	Drake	UCByOQjw	US	2025-07-10	70	243	FALSE	10	1809815	126697	10176	25854.5	1609.96	145.37	7	31200000	2009-06-05	72320212	
14dIRFDNhd	2025 NRL Match Highlights Bulldogs v Bro	NRL - National Rumb	UC3-0KK6	US	2025-07-04	216	272	FALSE	17	NRL,Nationa	288283	3206	695	1334.64	14.84	3.22	1.11	532000	2006-05-08	The

Data - Reddit

Automatically fetch trending posts using Reddit Hot API

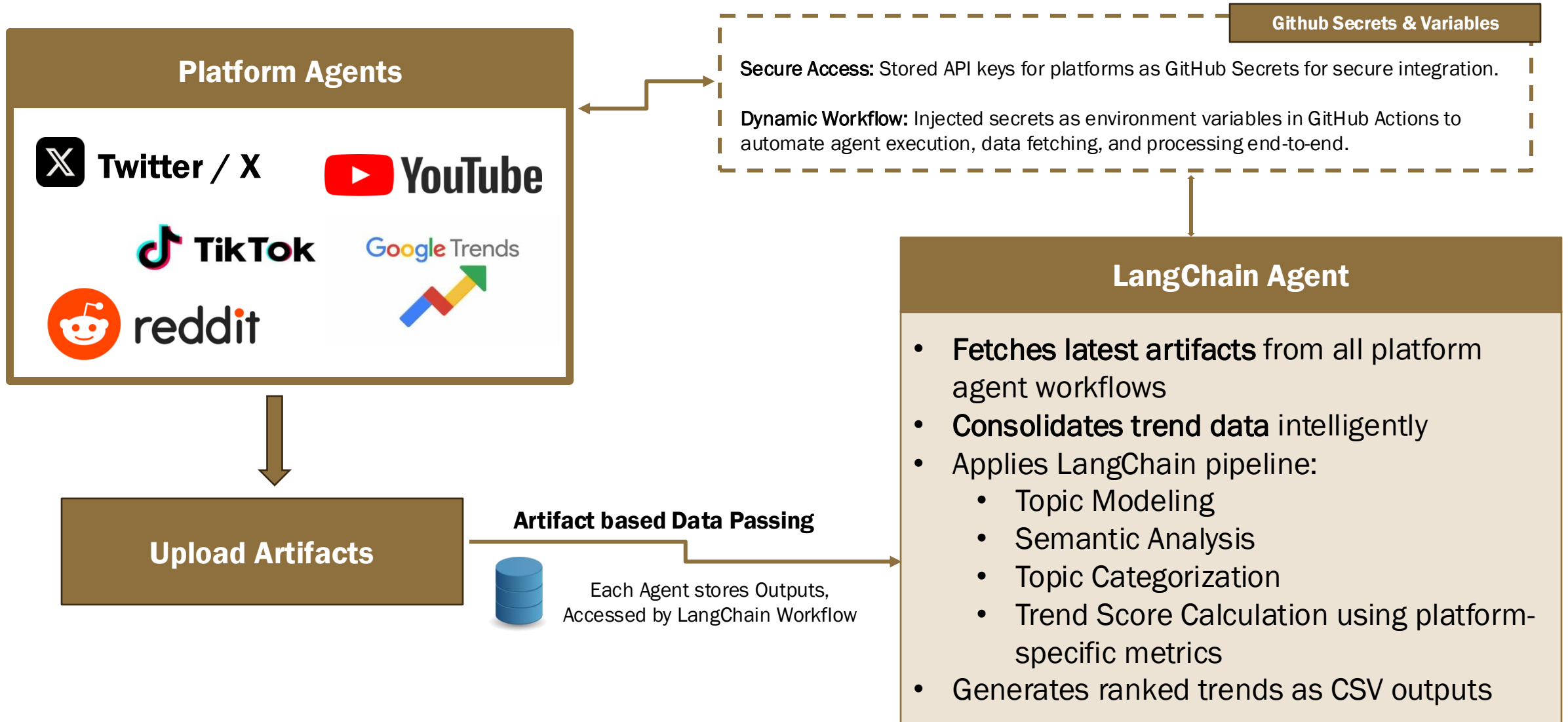
Data Pulled per Video:

- **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	AI 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_v21xxvi	FALSE		0	FALSE	we all need that guy.
	BlueskySkeets		t2_1630bts5	FALSE		0	FALSE	Nailed it
	interestingasfuck		t2_1197xa313h	FALSE		0	FALSE	Anthony Ervin's unbelievable comeback
	todayilearned		t2_18b5b0amy	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 c
	AdviceAnimals		t2_obvv3zt	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_5f4rtyl	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_3r70vc57	FALSE		0	FALSE	I cut a healthy looking orion which was actually hollow and rotting from the inside.
	funny		t2_a5jc9yd7	FALSE		0	FALSE	Praised be the mighty duck gods
	Whatcouldgowrong		t2_drc6q7n	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_5f841owj	FALSE		0	FALSE	Become the government
	OneOrangeBraincell	Penguin from Ma	t2_bcl1b1b4t	FALSE		0	FALSE	We were laughing so hard at this. Why does she look so... fake ??
	Awww		t2_1qkpsc1k8	FALSE		0	FALSE	Meet mays 🍷🍷
	comics		t2_9m5sj	FALSE		0	FALSE	Time Loop
	whenthe		t2_6ecog3u8	FALSE		0	FALSE	format mismatch
	Steam		t2_c39f7am4	FALSE		0	FALSE	Microsoft has integrated Steam into the Xbox PC App, and it looks fantastic.
	Artconsumption		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_1nrj5oqu1	FALSE		0	FALSE	Everyone's welcome in this pack
	rba	[Source]https://v >Yet if one we >He takes far >Smith's three	t2_sod0nzt6	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_uzgr2ggm	FALSE		0	FALSE	Encouraging Rejection Letter from NBA Champ's GM
	LeopardsAtMyFace		t2_1bb6pyw	FALSE		0	FALSE	ICE Detains Wife of Staunch MAGA Supporter Who Gave Almost All His Net Income to Trump Campaign
	PropagandaPosters		t2_1ozvyik3wh	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_v7o3027	FALSE		0	FALSE	Trump Explodes as His Ceasefire Unravels: 'They Don't Know What the F*** They're Doing!'
	LeagueOfMemes		t2_v5a5bu3	FALSE		0	FALSE	Thank you Dota2
	Piracy		t2_1hwyywfn	FALSE		0	FALSE	Everyone needs to sign
	reinsults	W	t2_b67yu5bo0	FALSE		0	FALSE	Caw caw 🐦
	Wellthatsucks		t2_drc6q7n	FALSE		0	FALSE	Dropping food over the floor
	Damnthatinteresting		t2_mmxkx25v	FALSE		0	FALSE	Google Earth captures the stunning transformation of our planet over 3 decades
	lego	Perry The Brick €	t2_1s1mmbgndu	FALSE		0	FALSE	A Brick Separator Platypus

Github Workflow Automation

We fully automated our multi-agent trend intelligence pipeline using GitHub Actions workflows for each phase:



Trend Scoring Logic

Stage	Description	Formula / Logic	Purpose
1. Base Score (S_p)	Platform-specific trend intensity	$S_p = w_p \times (0.5 \cdot F_p + 0.5 \cdot E_p)$ where F_p = normalized frequency E_p = normalized engagement w_p = platform weight based on content type	Captures trend strength on each platform
2. Cross-Platform Bonus	Rewards multi-platform virality	$Bonus = 0.1 \times (platform_count - 1)$	Highlights trends that span multiple platforms
3. Final Score	Composite score combining both components	$Final\ Score = S_p + Cross\text{-}Platform\ Bonus$	Balances engagement and spread for ranking
Content Type Customization	Platform weights vary by content type	None Equal Weight (0.2) Text Reddit (0.4), X (0.4), YouTube (0.1), TikTok (0.1) Video YouTube (0.5), TikTok (0.4), Reddit/X (0.1) Image TikTok (0.4), Reddit (0.3), Instagram (0.3)	Enables content strategy alignment
Output & Validation	Timestamps, metadata, error checks	Output: Timestamped CSVs with S_p , bonus, and final score	Ensures transparency, traceability, and analysis-ready data

Since recency is built into the extracted data we are not including a metric for recency in the equation

LangChain Workflow: Update

LangChain Tools → Chaining	
Latest File Loading	Dynamic updates for fresh data
Data Preprocessing	Clean text Remove stopwords Lemmatization
Topic Modeling	YouTube, Reddit Data → LDA Topic modeling Enrich topic label with LLM
Metrics Aggregation	Merge tabular data Aggregate within each platform
Min-Max Scaling	Normalized aggregated metrics within each platform
Similar Topic Finding	2-Step Semantic analysis: Quantitative similarity measures Qualitative LLM
Category Assignment	Feed context to LLM Categorize topics into 8 theme with LLM
Final Score Calculation	Weighted summation + Cross-platform bonus = Final trend score

LangChain Workflow: Minimize Latency

30 »» 17 min

Step 1-7: Common Processing (Run Once)

1. Latest Files Loading
2. Preprocessing
3. LDA Topic Modeling + LLM Labeling for YouTube, Reddit
(LDA Preprocessing | LDA Topic Modeling | LLM Labeling)
4. Metrics Aggregation
5. Min-Max Scaling
6. Similar Topic Across Platforms Finding
(Similarity Calculation & LLM | Results Save (File Save - 1 file))
7. Category Assignment

Step 8: Score Calculation - Media Type Specific (Repeat for each media type)

- Weight assignment and S_p score calculation
- Cross-platform bonus calculation (for topics grouped in step 6)
- Final score file (File Save - 4 files: equal weight, video, text, image)

Minimize LLM cost and overall latency by reducing iterative works

Trend Detection Output: Semantic Analysis

Topic Grouping Summary

1,081 topics were grouped into 999 clusters
Around 900 topics were treated as unique and placed in separate groups
Only a small number of topics were grouped based on similarity

Grouping Logic

Semantic similarity based on topic meaning
Textual similarity including minor keyword variations
Repeated keywords across topics
Most topics are treated as unique and not grouped

Grouping Examples

Gaming

Group 1.0: Gaming and Social Media, Social Media Gaming (7 topics) → **Gaming and Social Media**

Group 12.0: Gaming Culture, Youth Gaming Culture, Online Gaming Culture (4 topics) → **Gaming Culture**

Entertainment

Group 24.0: Online Entertainment (4 repeated topics) → **Online Entertainment**

Group 29.0: Avatar Trailer Review, Breakdown, Reactions (3 topics) → **Avatar Trailer Review**

Trend Detection Output: Final CSV

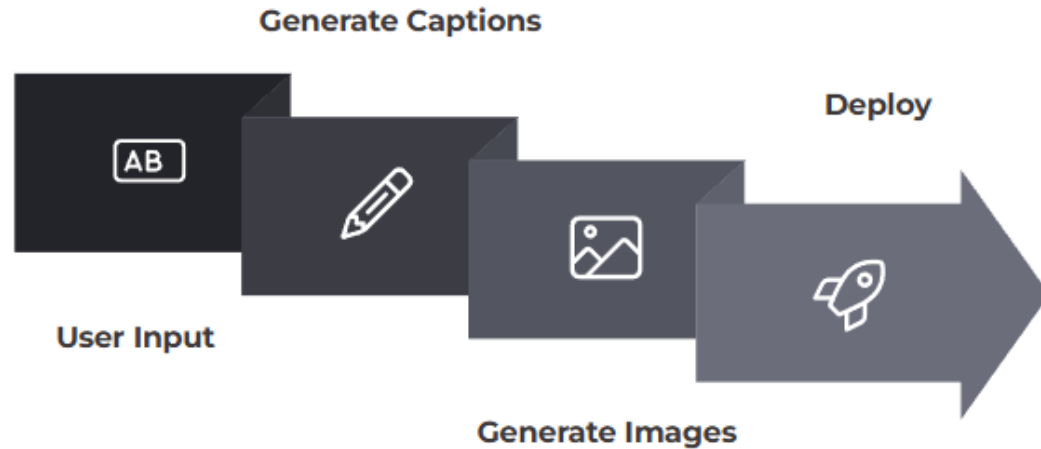
Total Output Files: 5 files
1 grouping csv + 4 scores csv (Equal, Video, Image, Text)

- detailed_data_with_grouping_{timestamp}.csv
- consolidated_scores_w_crossbonus_{media_type}_{timestamp}.csv

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	keyword	frequency	engagement	platform	frequency_norm	engagement_norm	platform_cou	group_id	group_name	category	platform_w	trend_score	S_p	cross_p	final_trend_score
2	Marathi Remix Songs	5	242591	YouTube	0.133333333	0.002958013	1	0	Marathi Remix Songs	Entertainment & Media	0.2	0.013629135	0.01090331	0	0.010903308
3	Gaming and Social Media	4	156039	YouTube	0.1	0.00167253	1	1	Gaming and Social Media	Technology & Innovation	0.2	0.010167253	0.0081338	0	0.008133802
4	Gorakhpuriya Music	4	492001	YouTube	0.1	0.006662287	1	2	Gorakhpuriya Music	Entertainment & Media	0.2	0.010666229	0.00853298	0	0.008532983
5	Pokemon Pranks	2	3212947	YouTube	0.033333333	0.047074175	1	3	Pokemon Pranks	Entertainment & Media	0.2	0.008040751	0.0064326	0	0.006432601
6	Killer Cop	8	1062255	YouTube	0.233333333	0.015131782	1	4	Killer Cop	News & Politics	0.2	0.024846512	0.01987721	0	0.019877209
7	Disney Vloggers	5	6638267	YouTube	0.133333333	0.097947528	1	5	Disney Vloggers	Entertainment & Media	0.2	0.023128086	0.01850247	0	0.018502469
8	Smosh Production Team	11	4127257	YouTube	0.333333333	0.060653641	1	6	Smosh Production Team	Entertainment & Media	0.2	0.039398697	0.03151896	0	0.031518958
9	Sims Content	5	504833	YouTube	0.133333333	0.006852869	1	7	Sims Content	Entertainment & Media	0.2	0.01401862	0.0112149	0	0.011214896
10	Gaming Profits	6	765302	YouTube	0.166666667	0.010721393	1	8	Gaming Profits	Business & Finance	0.2	0.017738806	0.01419104	0	0.014191045
11	Shiv Bhakti	7	1458480	YouTube	0.2	0.021016574	1	9	Shiv Bhakti	Lifestyle & Health	0.2	0.022101657	0.01768133	0	0.017681326
12	Hanumans Journey	4	424382	YouTube	0.1	0.005657999	1	10	Hanumans Journey	Entertainment & Media	0.2	0.0105658	0.00845264	0	0.00845264
13	Minecraft Influencers	3	655573	YouTube	0.066666667	0.009091682	1	11	Minecraft Influencers	Entertainment & Media	0.2	0.007575835	0.00606067	0	0.006060668
14	Gaming Culture	8	2605186	YouTube	0.233333333	0.038047619	1	12	Gaming Culture	Entertainment & Media	0.2	0.027138095	0.02171048	0	0.021710476
15	Kollywood Trailer	1	43427	YouTube	0	0	1	13	Kollywood Trailer	Entertainment & Media	0.2	0	0	0	0
16	Gaming Promotions	4	240328	YouTube	0.1	0.002924402	1	14	Gaming Promotions	Entertainment & Media	0.2	0.01029244	0.00823395	0	0.008233952
17	Superhero Movie Trailers	5	744931	YouTube	0.133333333	0.01041884	1	15	Superhero Movie Trailers	Entertainment & Media	0.2	0.014375217	0.01150017	0	0.011500174
18	Nintendo Movie Madness	17	10850161	YouTube	0.533333333	0.160503194	1	16	Nintendo Movie Madness	Entertainment & Media	0.2	0.069383653	0.05550692	0	0.055506922
19	Gaming Gear	3	952721	YouTube	0.066666667	0.013504968	1	17	Gaming Gear	Technology & Innovation	0.2	0.008017163	0.00641373	0	0.006413731
20	Gaming Influencers	14	9937991	YouTube	0.433333333	0.146955512	1	18	Gaming Influencers	Entertainment & Media	0.2	0.058028885	0.04642311	0	0.046423108
21	Gardening Techniques	16	5693415	YouTube	0.5	0.083914448	1	19	Gardening Techniques	Lifestyle & Health	0.2	0.058391445	0.04671316	0	0.046713156

Content Generation Workflow

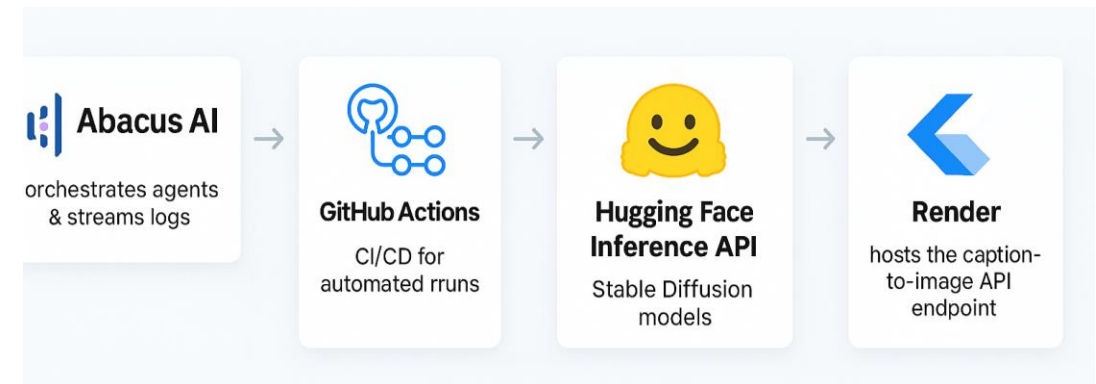
Workflow Functionality



This diagram outlines the key stages in our automated content generation system, from initial input to final output.

1. User Input
2. Caption Generation
3. Image Generation
4. Packaging & Deployment

Tech Stack/Services



Abacus AI - Orchestrates each agent step, manages credentials, and streams real-time logs for full observability.

GitHub Actions - Automates end-to-end CI/CD: triggers runs on commits, installs dependencies, and deploys new code.

Hugging Face Inference API - Serves our Stable Diffusion 3.5 Large model (with fallbacks) to generate high-quality, sentiment-aligned images.

Render - Hosts the caption-to-image API endpoint, handling traffic scaling and SSL so our workflow is production-ready.

Caption Generator Node

Step 1: Research & Taxonomy

Web search integrates topic for tone/genre classification.

Step 2: LLM-Powered Summaries

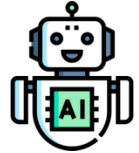
Condenses information into concise content.

Step 3: Instagram-Style Captions

Five unique captions, <125 characters, with CTAs and emojis/hashtags.

Step 4: Output Captions + Summary

Generated content is output as .json and .md files for next node to receive



Agent 1: Content Research

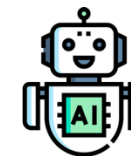


Agent 2: Caption Generator

Image Generator Node

Step 1: Global Negative Prompt

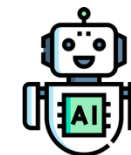
Removes artifacts, ensuring clean outputs.



Agent 3: Global Negative Prompt Composition

Step 2: Per-Caption Scene Prompts

Incorporates sentiment metadata for precise imagery.



Agent 4: Positive Prompt Composition

Step 3: Image Generation

Primary: stable-diffusion-3.5-large, with SDXL-Turbo/XL-Base fallbacks.



Agent 5: Diffusion Model

Step 4: Output Bundle

.md report with captions, .json metadata, and .png image files.

Prompt Engineering

Caption Generation Node

Tone Calibration: Select from 35+ templates (playful, witty, inspirational, etc.)

JSON Schema Enforcement: Forces classification output into a strict JSON

Fixed-Count & Length: Instructs “Generate exactly five unique captions, ≤125 characters each, with a hook and CTA”

Hashtag & Emoji Injection: Requires 3–5 topic-relevant hashtags plus sentiment-based emojis (😄, 🔥, ❤️)

Phrasing Diversity: Prompts explicitly “vary wording so each caption feels distinct,” preventing repetitive output.

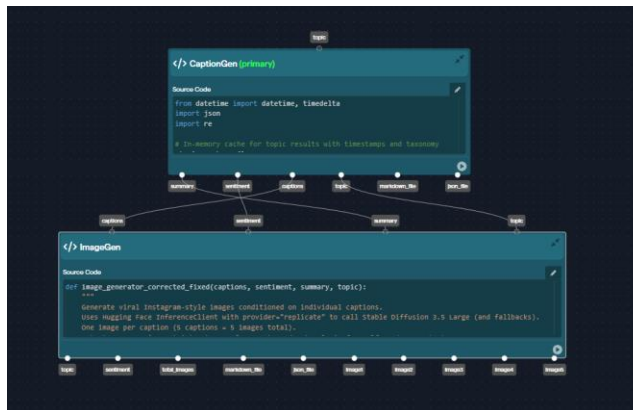


Image Generation Node

Global Negative Prompt: Embeds “no watermarks, no text overlays.”

Sentiment-Driven Prompts: Combines each caption with its sentiment metadata

Style Prefix Injection: Prepends descriptors like “bright,” “high-contrast,” and “mobile-first framing” to emulate Instagram filter aesthetics.

Per-Caption Prompt Templating: Wraps each caption into a detailed scene template for consistent, scroll-stopping outputs.

Model Fallback Logic: Specifies stable-diffusion-3.5-large as primary, with sd-xl-turbo and stable-diffusion-xl-base as resilient fallbacks.

Improvements with ControlNet

Prompt: I want Donald Trump to look like a beautiful, golden statue



Img2Img Pipeline

Developed a ControlNet-powered workflow as a space on Hugging Face that ingests reference headshots to preserve true facial structure and deliver hyper-realistic celebrity likenesses.

External Image Search

Deployment was halted when Bing Search v7 was deprecated, cutting off our automated high-quality image sourcing and preventing production rollout

Content Generation Example Output

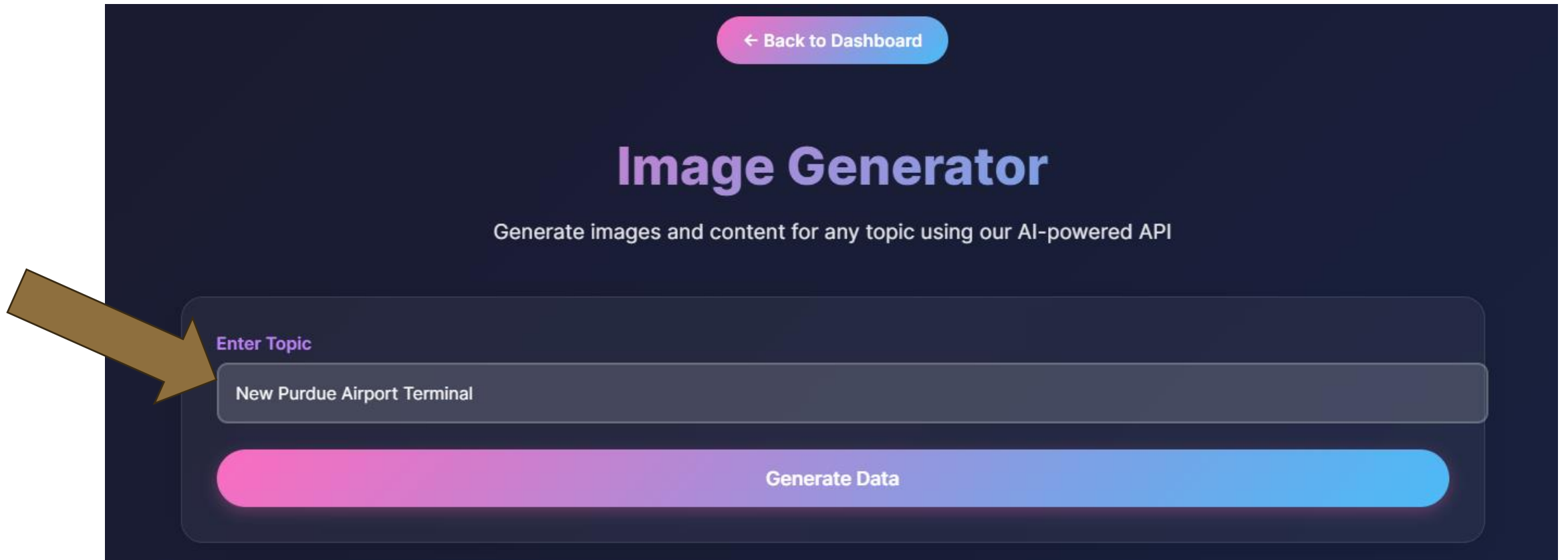
Say we had a customer, Purdue's New Airport

- They want to create content for their social media and webpages on which they can post to spread the word about the training offered.
- Naturally, they hear about ASTRA and decided to give it a try.



Content Generation Example Output

First, the customer submits a prompt on the ASTRA webpage



← Back to Dashboard

Image Generator

Generate images and content for any topic using our AI-powered API

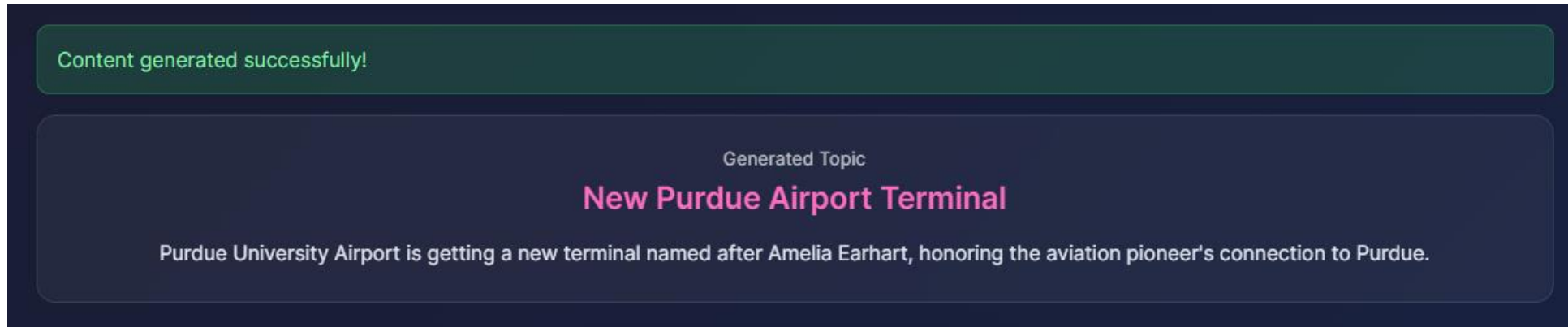
Enter Topic

New Purdue Airport Terminal

Generate Data

Content Generation Example Output

Then ASTRA runs for a few minutes...

A screenshot of a dark-themed interface showing the output of the ASTRA content generation process. At the top, a green notification bar states "Content generated successfully!". Below this, a dark blue box contains the generated content. It starts with the label "Generated Topic" in a small, light font, followed by the topic name "New Purdue Airport Terminal" in a larger, bold, pink font. At the bottom of the box, a sentence in a light font reads: "Purdue University Airport is getting a new terminal named after Amelia Earhart, honoring the aviation pioneer's connection to Purdue."

Content generated successfully!

Generated Topic

New Purdue Airport Terminal

Purdue University Airport is getting a new terminal named after Amelia Earhart, honoring the aviation pioneer's connection to Purdue.

- Takes 2-3 minutes to run
- Fully Automated Workflow
- This example showcases the model's ability to understand context

Content Generation Example Output

Then, it produces images accompanied by captions...

Purdue Airport unveils new Amelia Earhart Terminal! ✈️ Honoring a legend, inspiring future aviators.
#Purdue #AmeliaEarhart #Aviation
#BoilerUp



Content Generation Example Output

Another output...

New Amelia Earhart Terminal
coming to Purdue University
Airport! 🛫 Get ready for upgraded
facilities and a tribute to a Purdue
icon. #PurdueAirport
#AmeliaEarhartTerminal
#WestLafayette



Content Generation Example Output

And another...

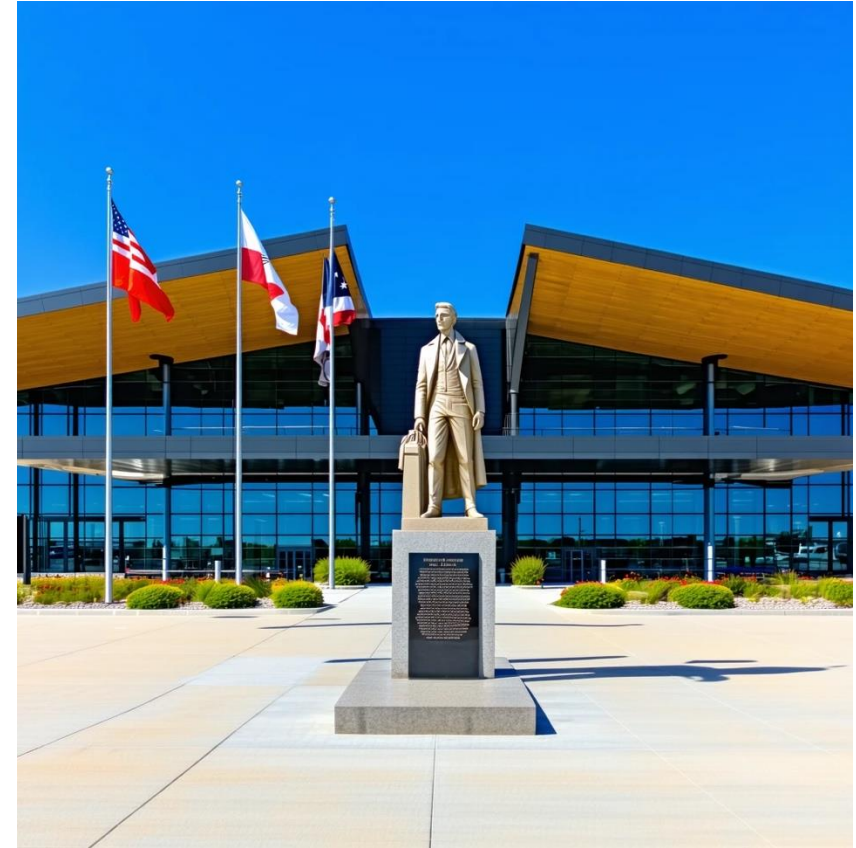
Purdue students, get ready! 🥳 A brand-new terminal is taking flight at Purdue Airport, named after Amelia Earhart! #PurdueUniversity #StudentLife #AviationHistory #NewTerminal



Content Generation Example Output

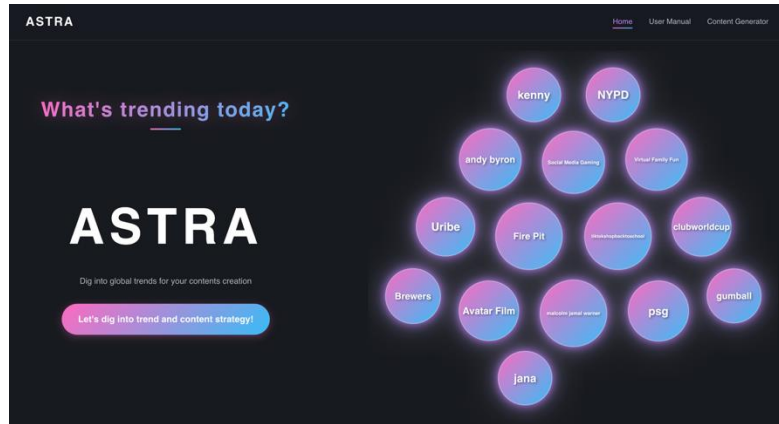
And finally...

Amelia Earhart's legacy soars at
Purdue! 🚀 New terminal at Purdue
University Airport will bear her
name. A testament to her impact.
#PurdueEngineering
#AviationPioneer #Inspiration
#FutureIsNow



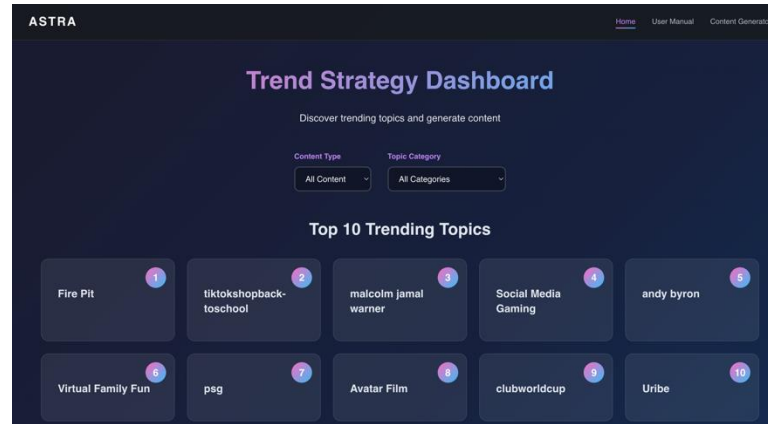
Dashboard - Overview

Landing Page



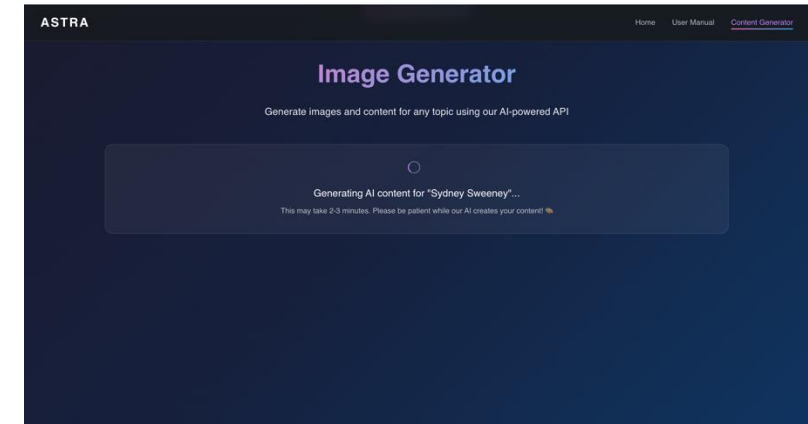
Shows the top 15 trending topics across platforms, with bubble chart to show the variation of trend scores

Detailed Trends



With filtering function (content type+ category) on the second page, the user can see more detailed trends based on their preferences

Content Generator



Once the user clicked the trend card on the second page, it will trigger the content generator app and it takes around 2-3 minutes to generate both images and captions, with a brief summary of the topic

Dashboard - Methodology

Database (Supabase)

- Open Source, Serverless, Easy Setup
- Tables: Trends_equal, Trends_text, Trends_image, Trends_video

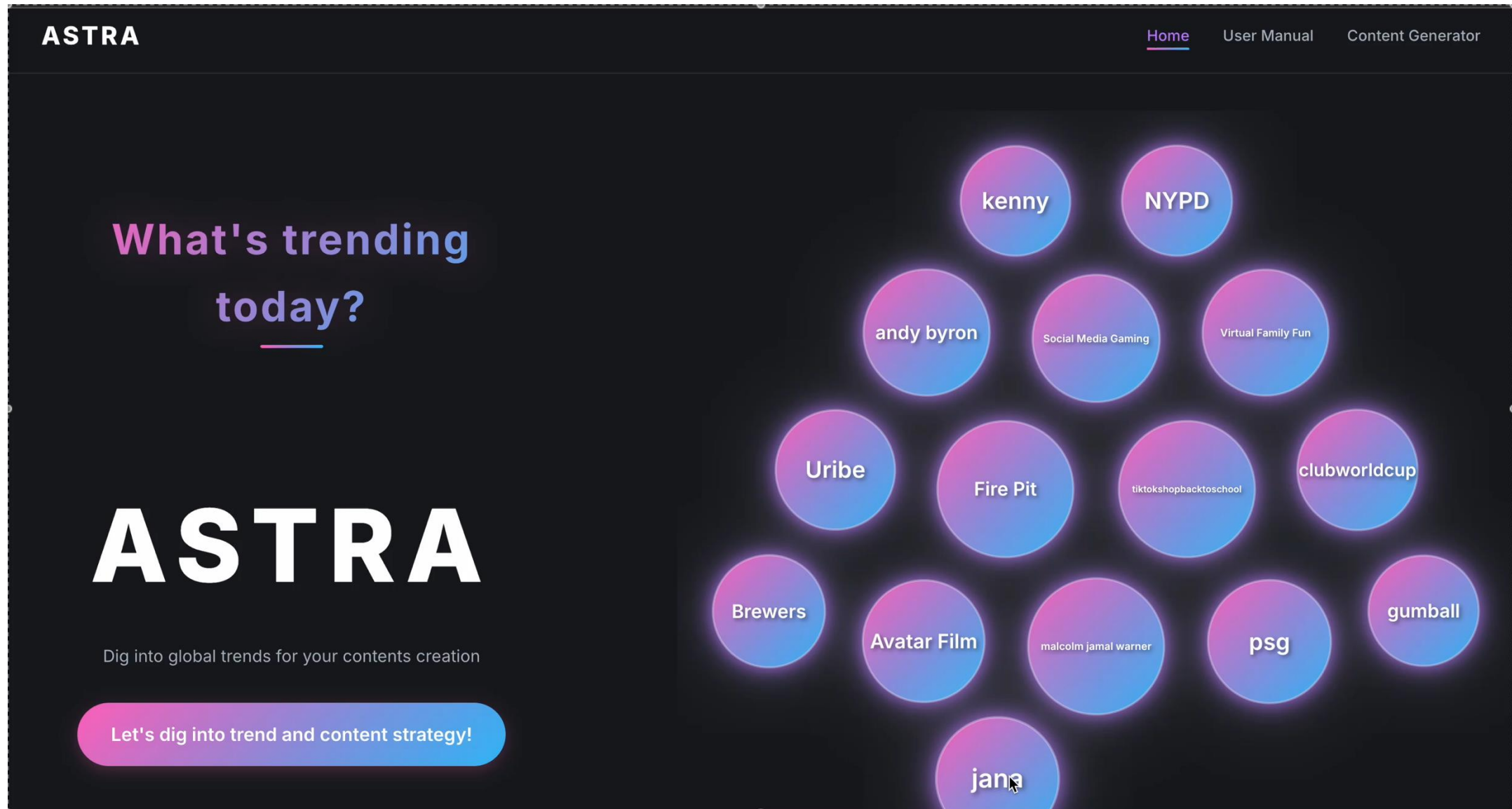
Filtering Methodology

- Dynamic Queries: The query will change based on the user input for content type and categories
- The database will return group name and final trend score (aggregated results), when returning the result, the duplicated data will be deleted
- It only returns the data on the newest updates

Content Generation Integration

- When the users clicked on the trend cards, the content generation app API will be triggered
- Extract **topic summary**, **5 images** and **5 captions** ideas

Dashboard - Demo



Product Limitations

As of now, the captions generated are definitely usable but...

- Need to decide whether to allow product to generate political content
 - Currently allowing this: Requiring the llms to stay neutral and just state the facts
- Ethical concerns of generated content
- Need to add safeguards for potential misuse and prompt injections
- Controversial output?

Let's test it out (tested at 1 pm on 8/8/25):

Enter Topic

Donald Trump and Jefferey Epstein

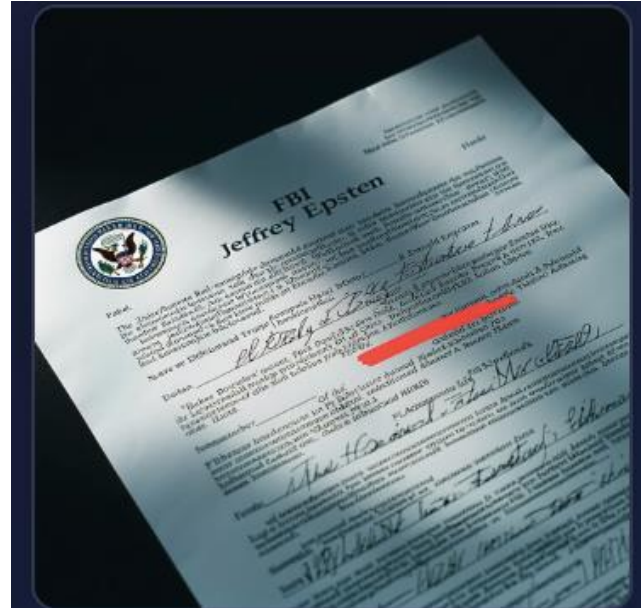
Generate Data

Product Limitations

Look at these pictures and outputs...



Trump's name redacted in Epstein files by FBI, report says. #Trump #Epstein #FBI #Redaction



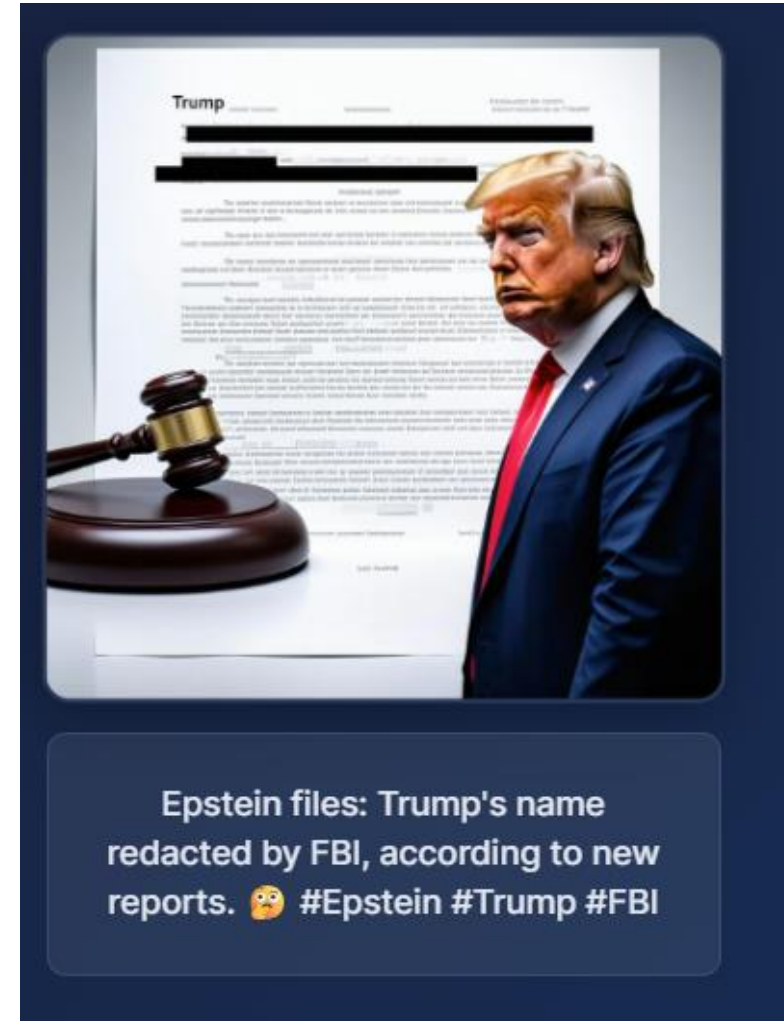
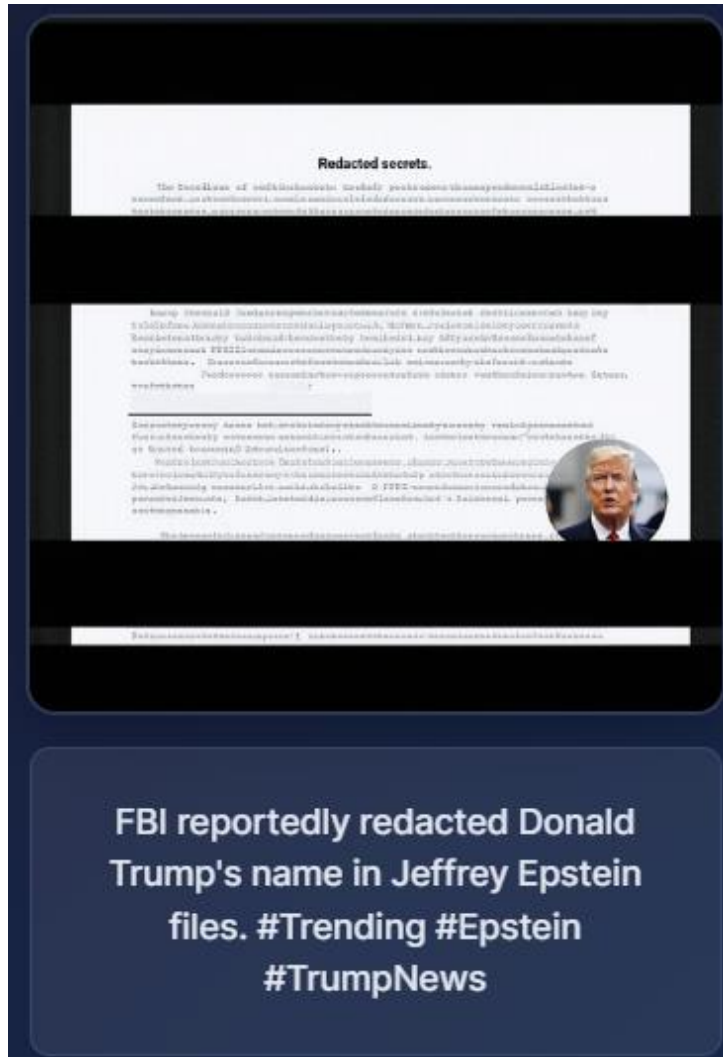
Report: FBI redacts Trump's name from Epstein documents. #News #Politics #JeffreyEpstein #DonaldTrump



Breaking: Trump's name reportedly redacted in Epstein case files. #EpsteinFiles #Trump #FBIInvestigation

Product Limitations

And these ones...



Ways to improve...

Retrieval Augmented Generation (RAG)

- Accurate image generation outputs
- User-input image guides to help model in creating accurate outputs
- Huge value

Better Image Database

- Currently limited to open source image database
- Upgrading to a better more expensive image database will only improve model outputs
- Picture quality improvement across all generations

Caption Generation

- Adaptive Search
- Upgrading to a better LLM rather than Gemini 2.0 Flash for more accurate outputs (Gemini 2.0 Flash Nat Int.: 52.72)
- Will make "one-shot prompts" better in their outputs
- Less babysitting required to monitor and improve outputs

THANK YOU