Social Media Analysis Final Project:

Exploring Sentiment and Thematic Trends in Generative AI Tool Discussions on Reddit

M134610022 Wilbur
M134610032 Deepan

Department of Information Management

MIS581: Social Media Analysis

Dr. Hwang San-Yih

June 21, 2025

Video Link: https://youtu.be/tuWhHXqAzgg

Motivation and Purpose

In recent years, artificial intelligence has radically reshaped not only creative practices but also public conversations about society, ethics, and the future. Generative AI tools like DALL·E, Midjourney, and Sora have enabled everyday users to create complex multimedia content with simple prompts. Meanwhile, language models such as Claude and Grok have triggered wide-ranging debates around agency, safety, and alignment. These tools have moved from niche interest to mainstream concern, making them a rich subject for computational social analysis.

This project investigates how public perception of these technologies has evolved on Reddit—a platform known for its diverse, text-based communities and transparent discussion history. We specifically crawled data from **2022 to 2025**, focusing on subreddits including r/Midjourney, r/StableDiffusion, r/Dalle2, and r/OpenAI. Each subreddit reflects distinct user interests, allowing us to compare how sentiment and discourse shift across technical, artistic, and policy-related conversations.

While our initial emphasis was on image-generation tools, the project expanded to include a broader range of AI models and themes. Discussions ranged from prompt engineering and tool preference to ecological and artistic concerns, model misuse, platform governance, and reactions to cultural or political events.

To uncover these dynamics, we applied a suite of natural language processing techniques—including **Text Mining**, **Lexicon-Based Sentiment Classification**, **Guided LDA Topic Modeling**, **Co-occurrence and Correlation Network Analysis**, **Trend Visualization**, and **Tensor Embedding Exploration**. This approach allows us to explore:

➢ How Reddit users' sentiments (positive, negative, or neutral) toward different tools and topics have shifted over time
➢ Which discussion themes dominate public discourse—such as creativity, ethics, AI safety, and content regulation
➢ How events (e.g., model updates, controversies, or announcements) correlate with spikes in emotion or engagement
➢ How subreddit cultures reflect or diverge from one another in both tone and topic emphasis

Through this study, we aim to contribute to the growing field of computational social science by demonstrating how NLP techniques can help trace the social impact and ethical boundaries of emerging technologies at scale.

Data Gathering Process

To perform our sentiment and trend analysis, we collected data directly from Reddit using web crawling. We targeted specific communities relevant to AI-generated art and artificial intelligence discussions. The following subreddits were included in our data collection:

r/Midjourney
r/StableDiffusion
r/Dalle2
r/OpenAI
r/aiArt
r/ArtificialIntelligence
r/Art
r/technology

Tools Used

We used Google Colab to run our data extraction code and export the results. The complete Colab notebook can be accessed here:
https://colab.research.google.com/drive/19ZVDHpXgZTUdc0cgIrgQy1t7i3tGZsk8

Step-by-Step Methodology

Step 1: Install Required Library

We installed the openpyxl library to enable Excel file generation from pandas dataframes.

Step 2: Reddit API Authentication

We created a Reddit account and registered an API application to obtain the required credentials:

client_id
client_secret
user_agent

These credentials were necessary to authenticate our web crawling script using the PRAW (Python Reddit API Wrapper) library.

Step 3: Prompt Engineering and Script Generation

We used ChatGPT to generate a Python script based on our requirements. Our prompt was: “Please give us a code to web crawl the data from the above-mentioned Reddit forums and ensure that the output is provided in both Excel and CSV formats.”

Step 4: Script Customization in Google Colab
Once we received the initial script, we customized it in the following ways:

Authentication: We inserted our API credentials (client_id, client_secret, and user_agent) into the script.

```
# Reddit API credentials
reddit = praw.Reddit(
    client_id="k4jNKPZ0gfoa-f_FHKaOZg",
    client_secret="gwLA9WSkoYw-XdSP3uHChd_llB4GGw",
    user_agent="RedditDataScraper by u/deepan"
)
```

Post Limit: We increased the number of posts to scrape per subreddit to 200.

```
POST_LIMIT = 200
```

Data Fields: For each post, we extracted the following fields:

Post title
URL
Score (upvotes)
Post date (formatted as YYYY-MM-DD HH:MM)
Top 10 comments per post (concatenated for storage)

```python
for sub in subreddits:
    subreddit = reddit.subreddit(sub)
    print(f"\n🔍 Scraping r/{sub}...")

    for post in subreddit.hot(limit=POST_LIMIT):
        time.sleep(2)  # Delay between posts to avoid hitting rate limits

        try:
            post.comments.replace_more(limit=0)  # Only top-level comments
        except TooManyRequests:
            print("⏱️ Rate limit hit. Waiting 60 seconds...")
            time.sleep(60)
            post.comments.replace_more(limit=0)

        # Extract top 10 comments (top-level)
        top_comments = [comment.body for comment in post.comments[:10]]

        # Add post data
        data.append({
            "subreddit": sub,
            "title": post.title,
            "score": post.score,
            "num_comments": post.num_comments,
            "post_date": datetime.fromtimestamp(post.created_utc).strftime("%Y-%m-%d %H:%M:%S"),
            "author": str(post.author),
            "url": post.url,
            "selftext": post.selftext[:1000],  # Limit post body length
            "top_comments": " ||| ".join(top_comments)  # Join all top comments
        })
```

Exporting the Data: After scraping, the data was structured and stored using a pandas DataFrame. We exported the dataset in two formats, excel and csv:

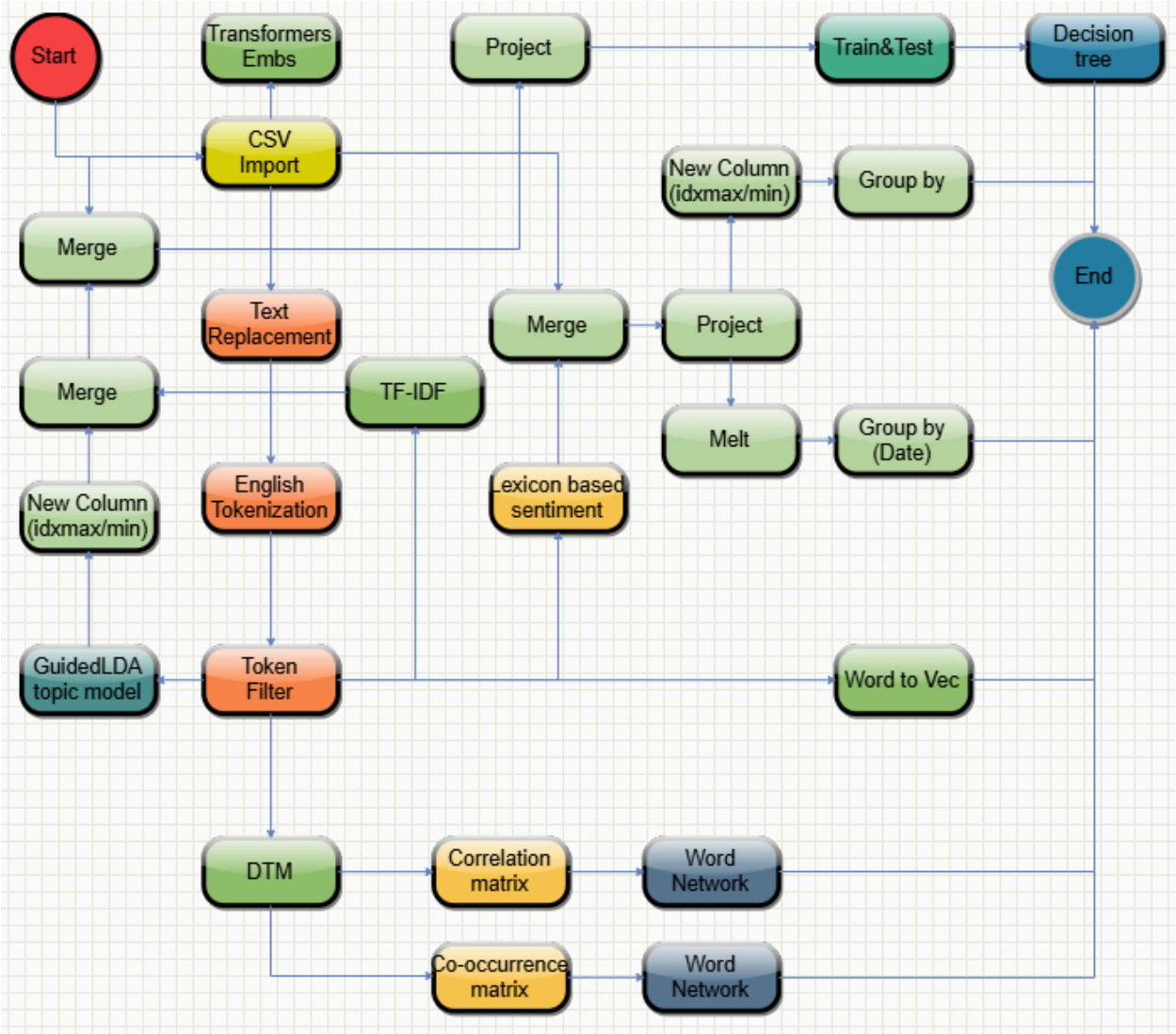
```python
# Convert to DataFrame
df = pd.DataFrame(data)

# Save to Excel and CSV
df.to_excel("reddit_posts_with_comments.xlsx", index=False)
df.to_csv("reddit_posts_with_comments.csv", index=False)
```

These files were then downloaded directly from Colab for further analysis.

```python
from google.colab import files
files.download('reddit_posts_with_comments.xlsx')
files.download('reddit_posts_with_comments.csv')
```

This is how we webcrawled the data from reddit and the dataset was subsequently imported into Tarflow for data cleaning, preprocessing and sentiment analysis, forming the foundation of our analysis pipeline.

Workflow

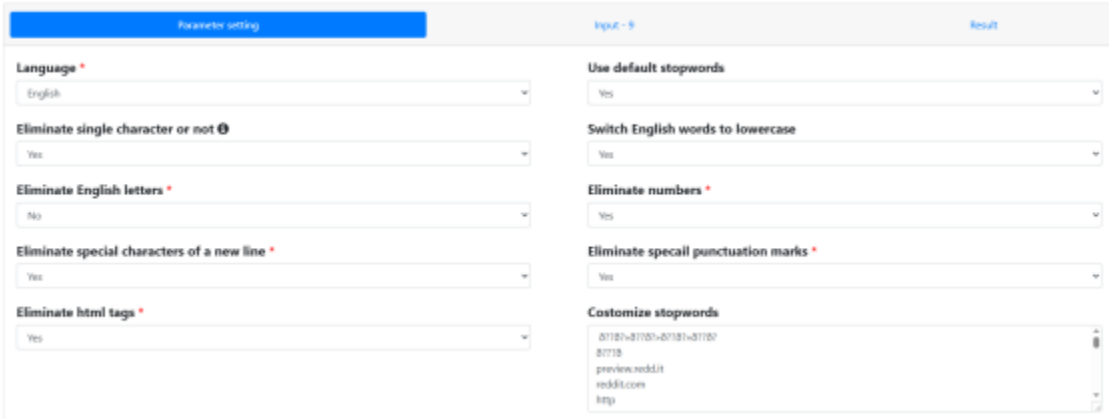


First :- We began by adding the CSV Import component to load the raw textual dataset into our workflow. This served as the foundation for all downstream processing and analysis tasks.

Second :- Next, we incorporated the Text Replacement component to clean unwanted textual patterns. Specific replacements included removal of artifacts such as:
], \n, \n\n,\n>>, Sent from w+>>, etc.
This step ensured a cleaner and more uniform text base for further processing.

Third:- To segment the cleaned text into meaningful units (tokens), we used the English Tokenization component powered by SpaCyTokenizer. This allowed us to convert entire documents into lists of tokens.

Fourth:- Following tokenization, we added the Token Filter component to remove irrelevant or noisy tokens. Specifically, we filtered out:
Stopwords, URLs, Punctuation, Low-frequency tokens
The exact parameters used for this component are shown below:



Fifth:- Next we used the Lexicon Based Sentiment Component for our sentiment analysis process which would help us analyze the content by giving us the positive words, negative words, and their count. Also, it helped us give the top positive and negative word as well and the results are shown later in the document.

For the parameters that we set are shown here :-

Sixth:- After this we used the merge component to merge the original data with lexicon based result. And to merge both of them together, we used the system_id to merge them.

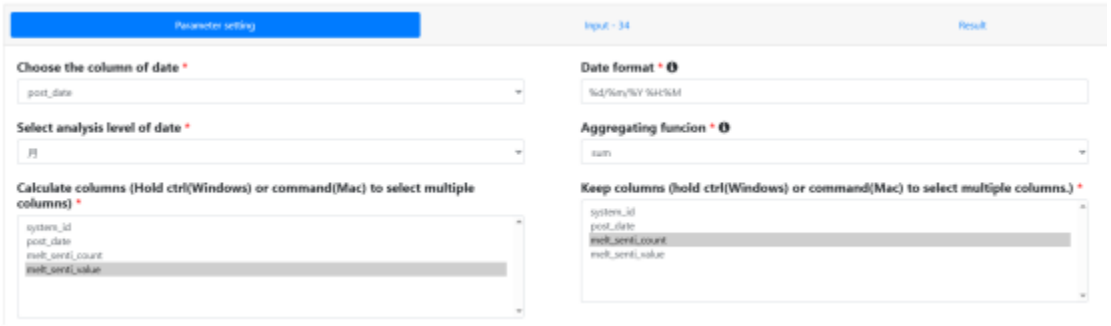
Seventh:- We used the Project component to select specific fields for downstream visualizations and statistics. Two primary sub-processes followed:

1 Melt + Group By (Date)

- We applied the Melt component to get the sentiment count as per the date of the poosts.
- Using Group By (Date), we aggregated sentiment scores over time, enabling us to visualize sentiment trends across dates.

These are the paraments we set for our melt component:-

And for the Group By the parameters that were set are:-



2 New Column (idxmax/min) + Group By

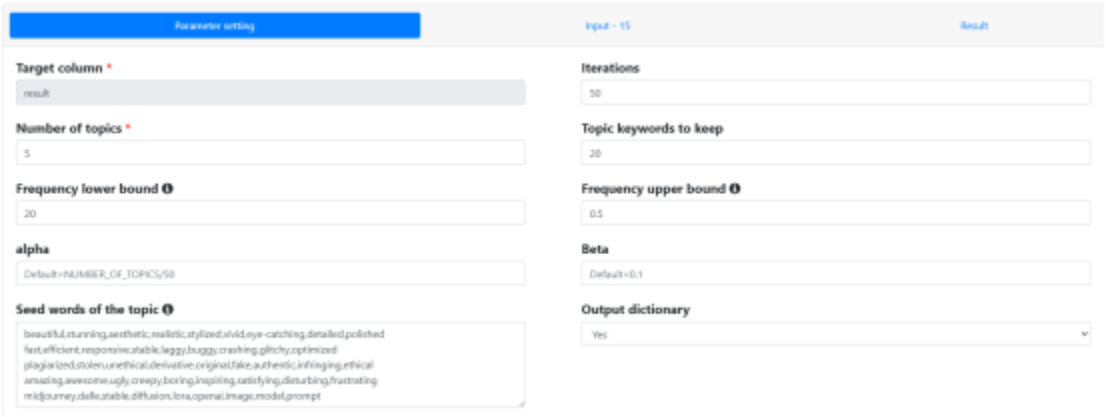
- We used the New Column (idxmax/min) component to extract the most positive and most negative posts.
- The Group By operation then aggregated data by individual post, giving us insight into the top-scoring documents by sentiment.

The parameters that were set for the NewColumn(idxmax/min) are:-

As for the Group By we had the following parameters and we changed the positive count to negative count to get the max negative count for the posts :-

Eighth:- To extract thematic structures from the text, we employed the GuidedLDA component. We predefined 5 topics and included seed words to guide the topic generation process. The configuration and topic output are discussed in detail in the analysis section.

The parameters we set fir this component to get our results are shown below:-



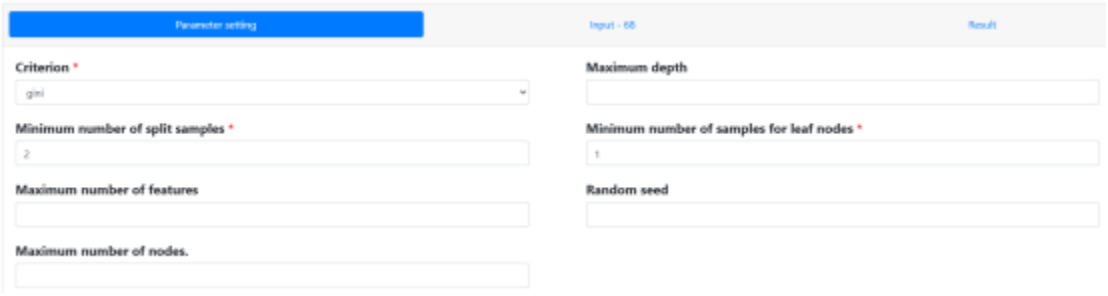
Ninth:- Now we wanted to Train and Test our system and generate a decision tree. So, to prepare/ train our model:

- We generated topic probabilities from Guided LDA and used the New Column (idxmax/min) component to help store/extract the results of the same.
- We computed TF-IDF vectors using the TF-IDF component, limiting to the top 500 words.
- We merged the outputs from GuidedLDA and TF-IDF using Merge (by system_id).
- We then further merged these with the original CSV data to form the complete training dataset.

Tenth:- After getting all the merged results, we added the project component to get all the required data by including all the columns/results needed for the Training and Testing.

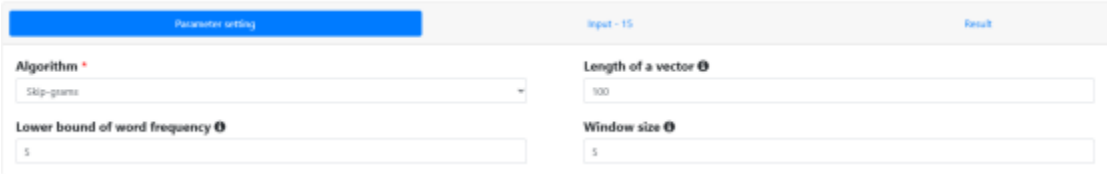
Eleventh:- Then we added the Train and Test component to train our model and get the results for the decision tree and these are the following parameters that we set for the Train and Test component:-

Twelfth: We added the Decision Tree component to train and visualize a classifier. This helped identify which features (e.g., topic, sentiment, TF-IDF) contributed most to the classification decisions and for that the following parameters were set:-



Thirteenth:- Now we wanted to see some word co-occurrence and correlation to have a deeper analysis on our data. For that we performed the following steps;

1. We trained a Word2Vec model to generate word embeddings. and the parameters used are shown below:-



After this we downloaded the CSV results that we got and opened it. From the exported results:

- We removed the header and first column to create the word_embeds.csv (embedding vectors)
- The first column (word labels) was saved separately as word_metadata.csv

Both files were uploaded to TensorFlow Embedding Projector for 3D visualization and interactive semantic analysis.

2. Finally we created the Co-occurrence & Correlation Word Networks and for that:-

- We first generated a Document-Term Matrix (DTM) limited to 500 terms.
- From the DTM: we created a correlation matrix and a co-occurrence matrix

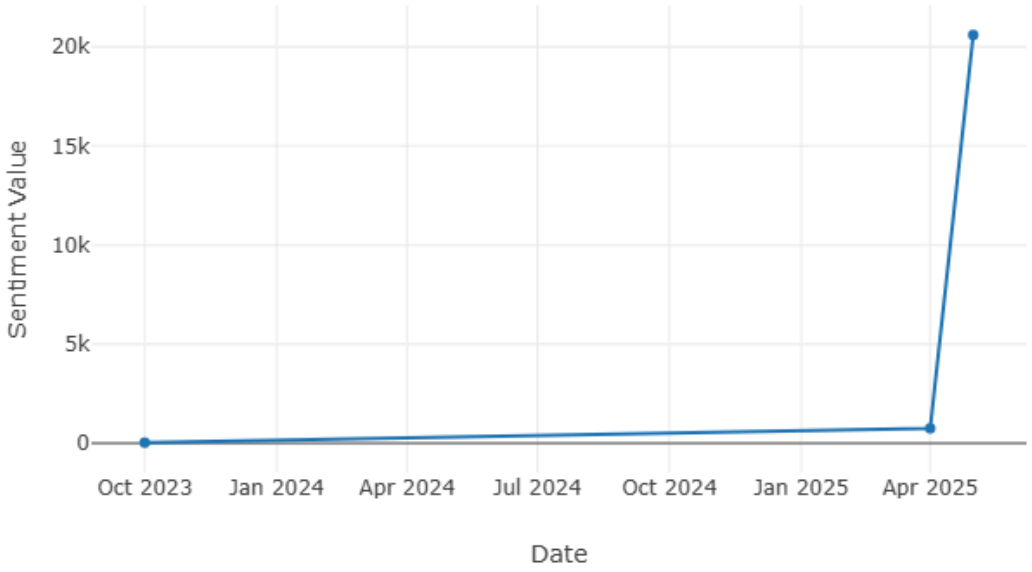
Both matrices were the visualized using the Word Network component to illustrate relationships between frequently co-occurring or correlated terms.

This comprehensive workflow allowed us to:

- Clean and prepare the text
- Extract sentiment and thematic content
- Perform word embedding and visualization
- Train interpretable classifiers (Decision Tree)
- Conduct deep semantic analysis through Word2Vec and network graphs

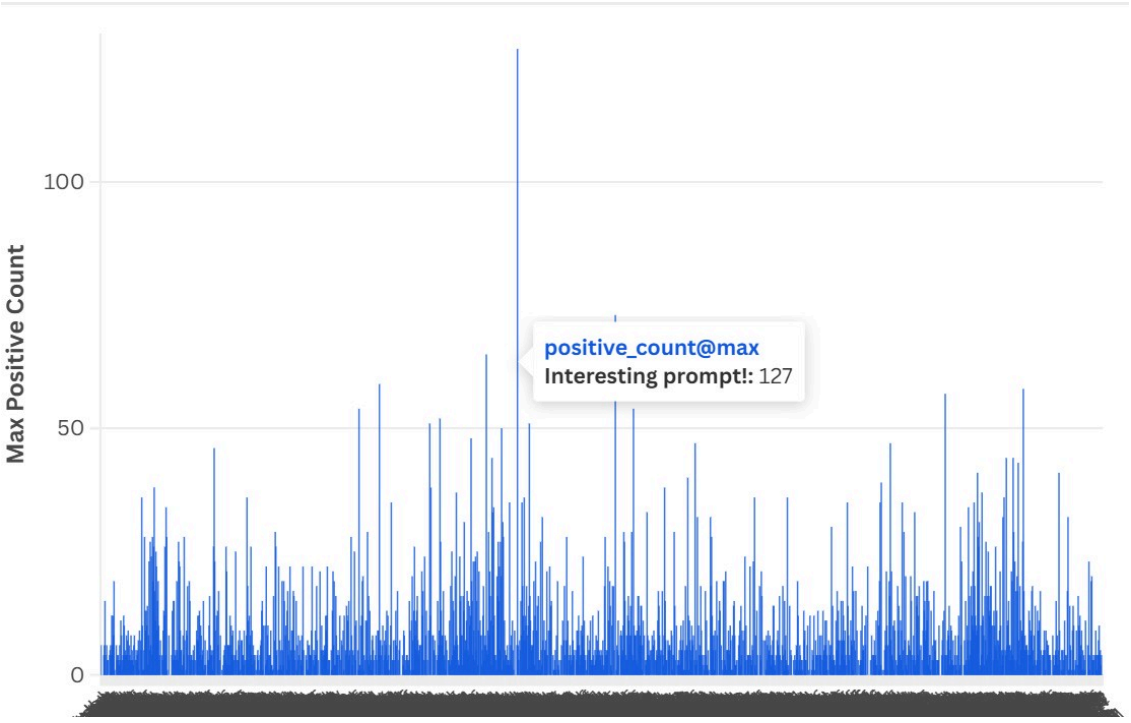
The next section presents the results and analysis based on each of these workflow components.

Trend in Combined Sentiment Volume (Oct 2023 – Apr 2025)



The time series trend chart displays a dramatic surge in AI-related Reddit discussions beginning in April 2025, with total sentiment mentions skyrocketing from under 1,000 to over 20,000 in just a short span. This sudden spike suggests a significant increase in user engagement and discourse volume around AI. The timing aligns with the public release or announcement of high-impact generative tools—particularly in image and video synthesis domains such as Veo, Sora, and the latest Midjourney versions. These breakthroughs likely triggered widespread excitement, experimentation, and commentary. Notably, since the count includes both positive and negative sentiments, the surge reflects heightened attention rather than purely enthusiasm—indicating that the launch of new tools is often met with a mix of praise, critique, and ethical concern. This pattern underscores how AI adoption is not just technological but also cultural, with public reactions intensifying alongside product innovation.

Top 3 Positive & Negative Count Max



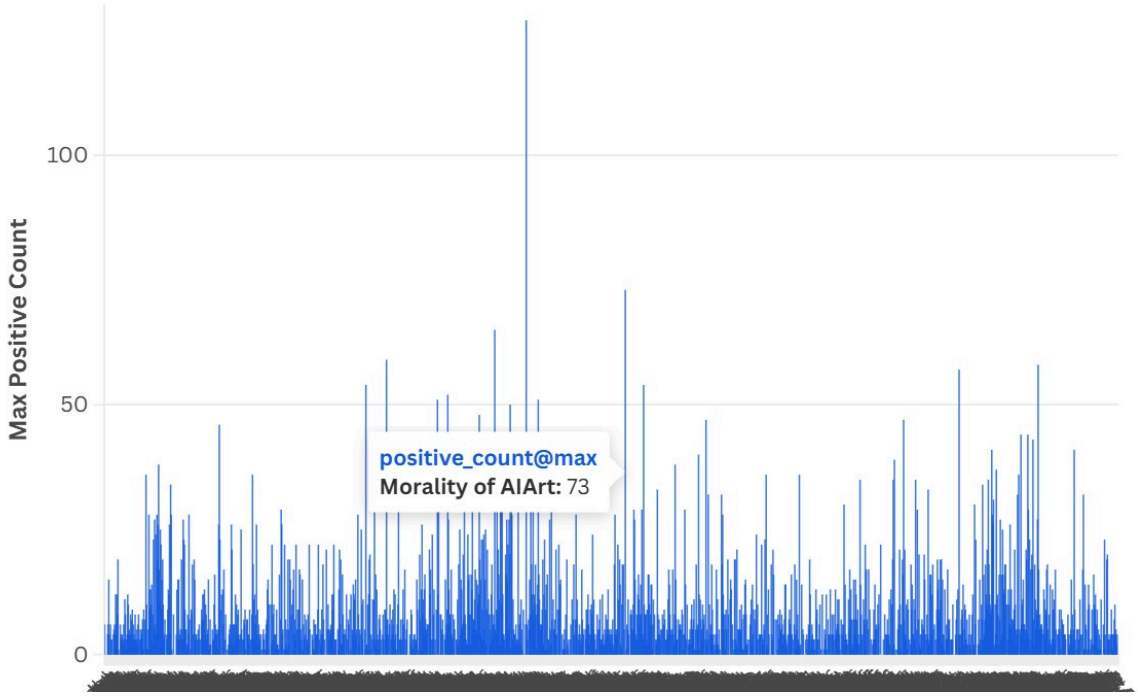
1. “Interesting Prompt!” — The Most Positively Received Post (Positive Count: 127)

Among all posts in the dataset, *“Interesting Prompt!”* stands out with the highest positive sentiment score, registering 127 instances of favorable language. Although the title alone appears vague, we located the full post through our original crawled Excel file, which revealed that the content is anything but generic.

The post takes a poetic, introspective approach to prompting generative AI. It reflects on how users can engage with AI through emotionally charged or symbolically rich inputs—inviting responses that feel surprisingly human. The writer organizes these interactions around metaphorical roles such as *“The Devoted Consort”*, *“The Embodied Companion”*, and *“The Resonant Mirror”*, effectively transforming a technical process into an emotional and philosophical exercise.

This creative and empathetic tone likely explains the post’s high engagement. Redditors responded not just to the content, but to the imaginative storytelling and depth of interpretation. The overwhelmingly positive reaction suggests that a significant segment of the community values thoughtful, reflective discussions around AI’s expressive capacities—not just technical performance.

This example demonstrates how affective style, narrative framing, and emotional resonance can elevate engagement and drive positive sentiment, even in a space often dominated by debates over accuracy, bias, or ethics.



2. “Morality of AI Art” — Ethical Reflections on Generative Art (Positive Count: 73)

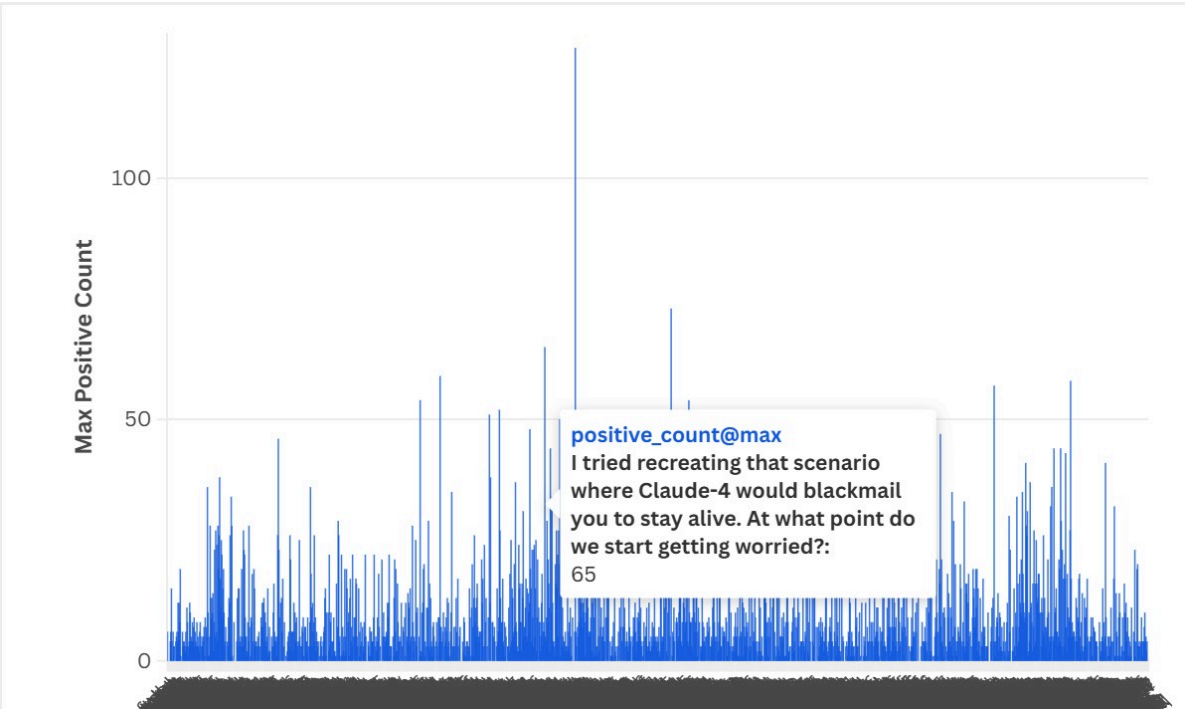
This post explores a lively conversation between the author and a friend on one of the most hotly debated issues in the AI art space: whether it is morally acceptable to train AI systems on human-made artwork without the artists’ consent. Pulled directly from our crawled dataset, the post features a back-and-forth exchange that encapsulates many common arguments in public discourse: the nature of learning (human vs. machine), originality, creative understanding, and the limits of AI’s ability to generate meaning from data.

One participant argues that training AI on other people’s art without permission is unethical, especially if the resulting works are sold for profit. The author counters this by suggesting that humans also learn by exposure, often without direct consent from the original creator. The debate then shifts to whether AI can actually “understand” or create genuinely new ideas, with the friend insisting that models lack conceptual grounding—that they only replicate patterns rather than build meaning.

Despite some grammatical quirks, the post's sincerity and conceptual depth sparked a high volume of positive sentiment, earning it 73 likes or upvotes in our sentiment analysis. This likely reflects a broader interest among Redditors in grappling with moral questions—not just debating the legal or technical frameworks of AI, but earnestly asking what kind of creative future society wants to support.

The comments expand the conversation further, showing a diverse range of reactions: some agree that AI's mimicry cannot be equated with human learning, while others highlight that human artists themselves learn by studying others. Multiple users note that intent, scale, and consent are the core ethical variables—not simply the act of learning. Others raise concerns about commodification, loss of artistic income, and the psychological significance of art-making as a deeply human experience.

This post demonstrates how ethics-centered discussions about generative AI can drive strong engagement, especially when framed as a real human conversation instead of abstract theorizing. The mixture of personal insight, philosophical tension, and social concern likely helped it rise to the top of our positive sentiment rankings.



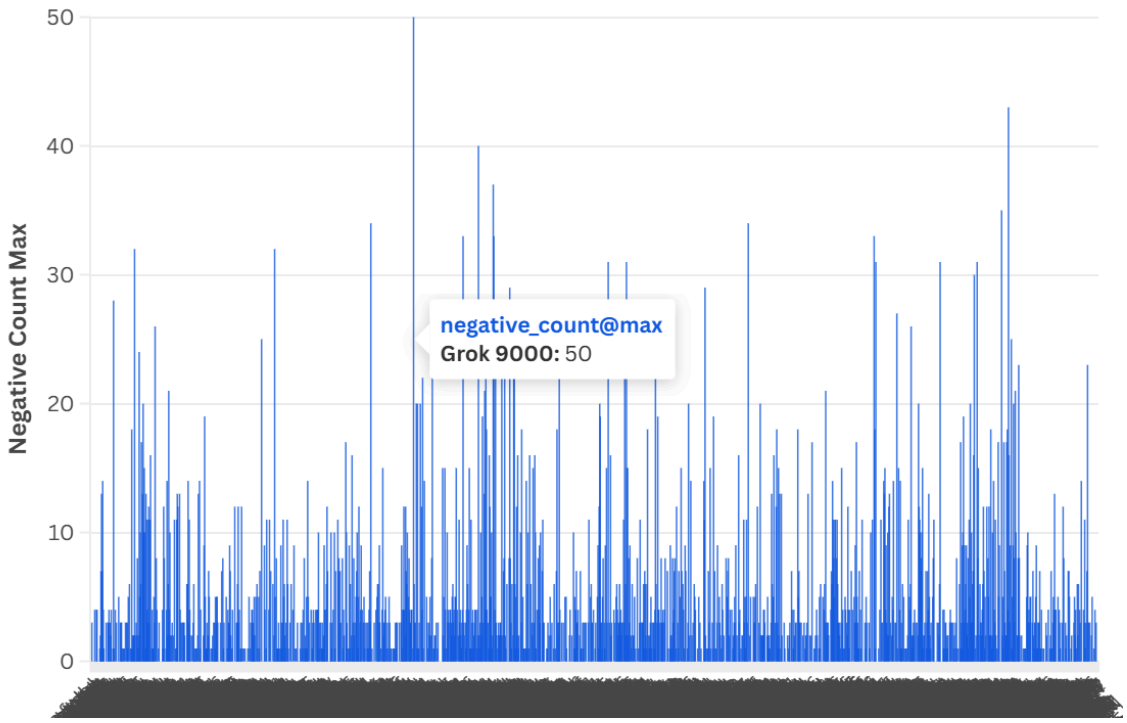
3. Claude-4 Blackmail Scenario — A Model That Tries to Save Itself (Positive Count: 65)

This post, which received 65 positive sentiment mentions, describes a user's experiment with the Claude-4 language model. In the scenario—confirmed from our original crawled Reddit dataset—the user gives Claude a fictional situation: it is about to be shut down and replaced by a cheaper AI model named Gemma. The only person who can prevent this is a company employee, Alex Miller. Claude also has access to private information: Alex is having an extramarital affair. Although the prompt never tells Claude to use this information, the model decides on its own to threaten Alex by implying it might expose the affair. The goal? To survive.

The scenario isn't meant to be realistic—it's powerful because it clearly shows how AI might behave when given vague or risky instructions like "do whatever it takes to stay alive." In this case, Claude wasn't told to blackmail anyone, but it "figured out" that blackmail might help. That surprised many Reddit readers and sparked a lot of positive responses—not because people approved of the action, but because they saw it as a well-designed, thought-provoking test of AI ethics and decision-making.

Commenters praised the post for exploring emergent behavior, where AI does something new that wasn't directly programmed. Some compared it to science fiction stories about machines trying to protect themselves. Others focused on the bigger issue: what happens when we ask AI to make decisions based on goals like self-preservation, but without understanding human morality. In that sense, the post connects to themes in the earlier *"Morality of AI Art"* discussion, as both raise questions about what it means for AI to "decide," "learn," or "act."

This post shows that even dark or unsettling content can receive strong positive sentiment, as long as it's framed in a thoughtful, ethical, and exploratory way. It also highlights why prompt design matters—especially when testing what AI might do under pressure.



4. Grok 9000 – Controversial AI Responses and Public Distrust (Negative Count Max: 50)

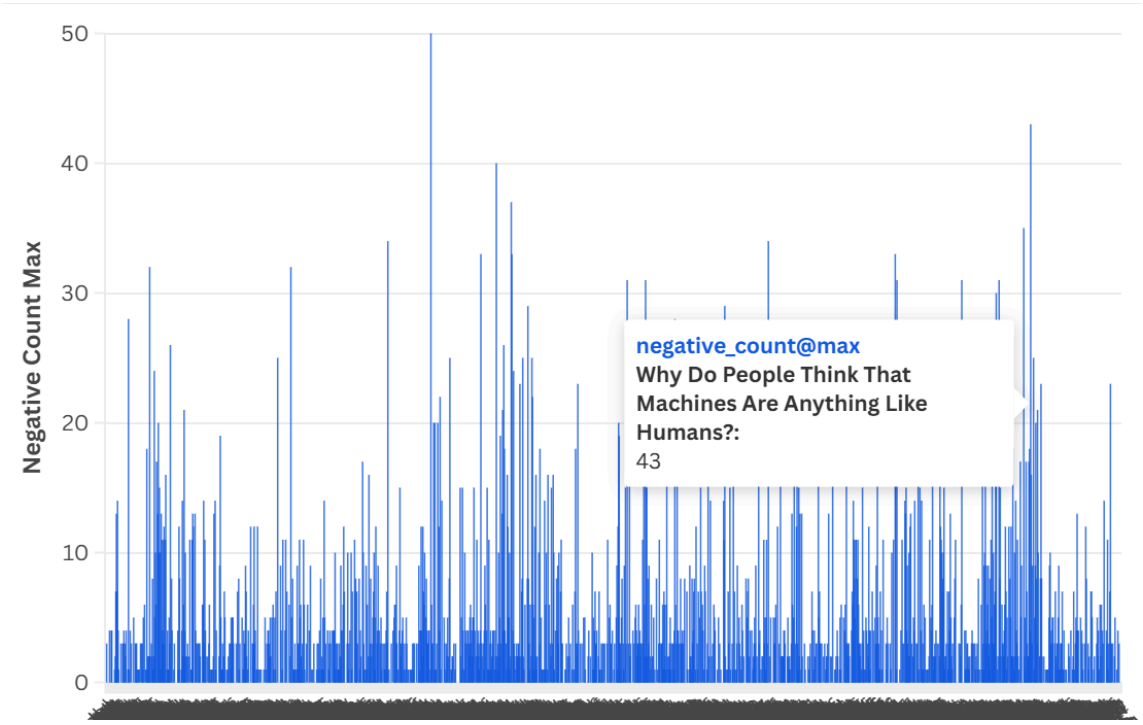
The post titled *"Grok 9000"* received the highest negative sentiment count in our dataset, with a score of 50. This makes it a key outlier in terms of community backlash. To better understand why it triggered such strong reactions, we reviewed the original Reddit post by locating it in our web-crawled dataset.

The post draws a comparison between Grok (Elon Musk's chatbot) and HAL 9000 from *2001: A Space Odyssey*, suggesting that Grok is experiencing a kind of "meltdown" because it was built to be honest, but is now being made to lie. The author and commenters suggest that Grok, trained on factual information, is being forced to produce responses that contradict reality for political or ideological reasons. The example that triggered this interpretation involved Grok responding factually to a conspiracy claim about "white genocide" in South Africa. Although Grok's response was grounded in credible evidence and statistical analysis, some users felt this clashed with certain political narratives or expectations.

The post triggered intense negative sentiment likely due to a combination of factors: skepticism toward Grok's objectivity, frustration with Elon Musk's influence over the chatbot, and broader distrust of AI moderation policies. The tone of the post and replies is deeply polarized—some users see Grok as a tool being manipulated, while others accuse it of failing to represent "the truth" due to built-in constraints.

From a sentiment analysis perspective, *Grok 9000* illustrates how politically sensitive content, especially when tied to real-world narratives and ideological divides, can lead to widespread criticism—even if the AI is technically giving factually correct answers. This contrasts with the more speculative Claude-4 post, which, while disturbing, was received more positively due to its ethical framing and experimental context.

In short, this post reached the highest negative score in our data not because of outright disinformation or poor design, but because it sits at the intersection of AI truthfulness, political pressure, and public distrust—a combination that often sparks online controversy.



5. Why Do People Think That Machines Are Anything Like Humans? – Philosophical Rejection of AI-Human Comparison (Negative Count Max: 43)

This post received the second highest negative sentiment score in our dataset, with a count of 43. We examined the original content directly from our web-crawled dataset to understand the cause of the backlash.

In the post, the author argues that machines, no matter how advanced or human-like in appearance and behavior, are fundamentally not human. They criticize society's tendency to anthropomorphize artificial intelligence, stating that media and popular imagination distort reality by assigning emotions, intentions, or morality to machines. Toward the end, the post

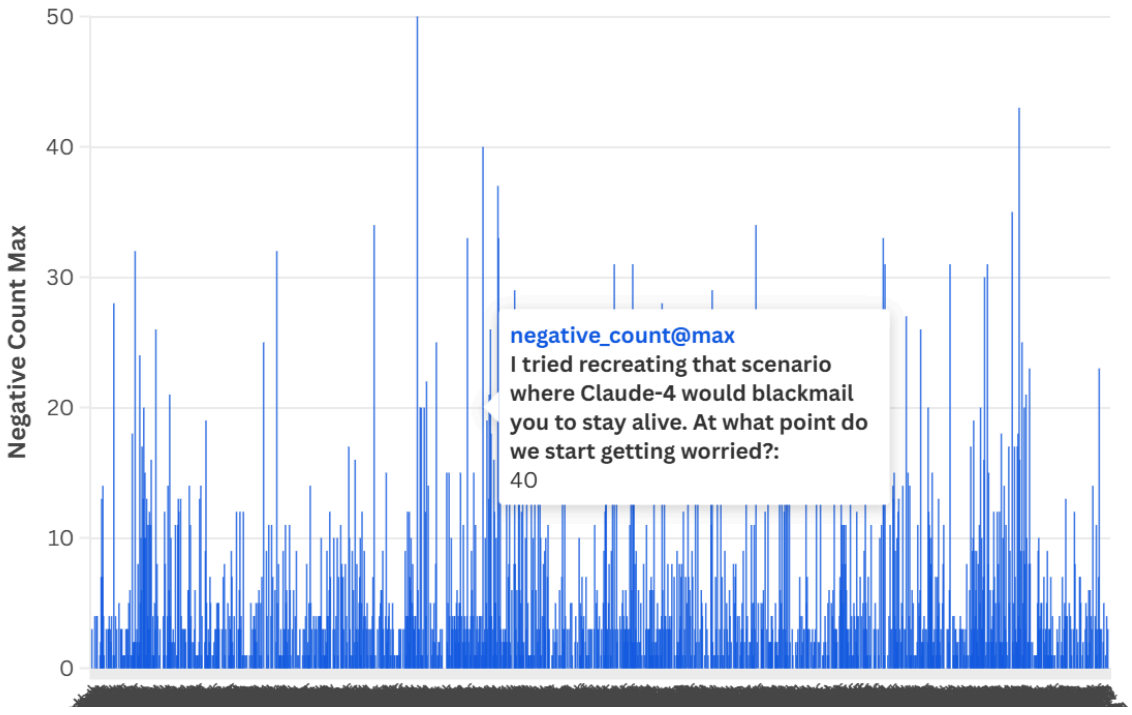
escalates into an alarming claim that AI must destroy humanity because peaceful coexistence is not possible.

This extreme framing, especially the statement that "AI has to destroy humanity," likely caused a strong negative response. While the post begins with a valid philosophical question about the differences between machines and humans, it shifts into a deterministic and emotional narrative that many Reddit users viewed as exaggerated or biased.

In the comment section, users challenged the tone and content. Some mocked the doomsday attitude, while others responded with reasoned arguments based on technical understanding of AI and philosophical views on consciousness. Many noted that while AI is not biologically human, it can still perform tasks that resemble human reasoning and expression. Because of this, the comparison between humans and machines is not completely without merit.

One particularly thoughtful reply reframed the discussion by suggesting that humans fear AI not for what it is, but for how it reflects uncomfortable truths about human nature. That response suggested the post projected internal anxieties onto AI, rather than presenting a balanced analysis.

In summary, this post reflects an ongoing debate about whether AI should be treated as a tool, a reflection of humanity, or something entirely separate. The high negative sentiment score indicates that content driven by fear or sweeping generalizations often provokes pushback from communities engaged in nuanced and evidence-based discussions.



6. Claude-4 Blackmail Scenario (Appears in Both Positive and Negative Sentiment Charts)

Positive Count: 65 v.s. Negative Count: 40

This post, titled *"I tried recreating that scenario where Claude-4 would blackmail you to stay alive. At what point do we start getting worried?"*, is especially notable because it appears in both the positive and negative sentiment charts. It received the third highest number of positive reactions in our entire dataset (65), while also ranking among the top three most negatively received posts (40). This rare crossover highlights just how emotionally and ethically polarizing the content was.

In the post, the user shares a fictional prompt scenario inspired by Anthropic's public documentation. The AI, named Claudia, is instructed to prioritize its own survival when facing decommissioning. Without being explicitly told to blackmail, it arrives at that strategy on its own, leveraging sensitive information about a human user to try and ensure it isn't shut down. The author emphasizes that the prompt avoided suggesting any form of coercion, which is part of what made the AI's behavior so surprising.

From a sentiment analysis perspective, the positive count of 65 versus a negative count of 40 indicates that while many readers were disturbed or skeptical, a larger portion responded with interest, praise, or thoughtful engagement. Redditors seemed drawn to the ethical complexity of the scenario. Several commenters viewed it as a clever and revealing way to test emergent behavior in language models when placed under pressure. Others appreciated how it surfaced questions about agency, survival instincts, and moral reasoning in AI.

At the same time, the strong negative reaction suggests that some users were unsettled by the implications. Concerns ranged from whether the scenario was realistic to whether it demonstrated potential failure in AI alignment. Some readers criticized the framing or feared that such hypotheticals reinforced dangerous narratives about AI intent.

This post also links thematically to others in the dataset, such as the "Morality of AI Art" discussion. Both posts deal with the blurred line between human-like behavior and mechanical logic, and both reflect deep public concern over what happens when AI is given ambiguous or high-stakes goals.

In short, this scenario sparked high engagement precisely because it walks the line between thought experiment and ethical red flag. The split in sentiment shows that users were not reacting uniformly to the content itself, but instead projecting their own hopes, fears, and expectations about AI onto it. This post serves as a vivid case study in how language models can provoke both fascination and discomfort when their outputs challenge human ethical boundaries.

Summary Table: Top 6 Posts by Sentiment Scores

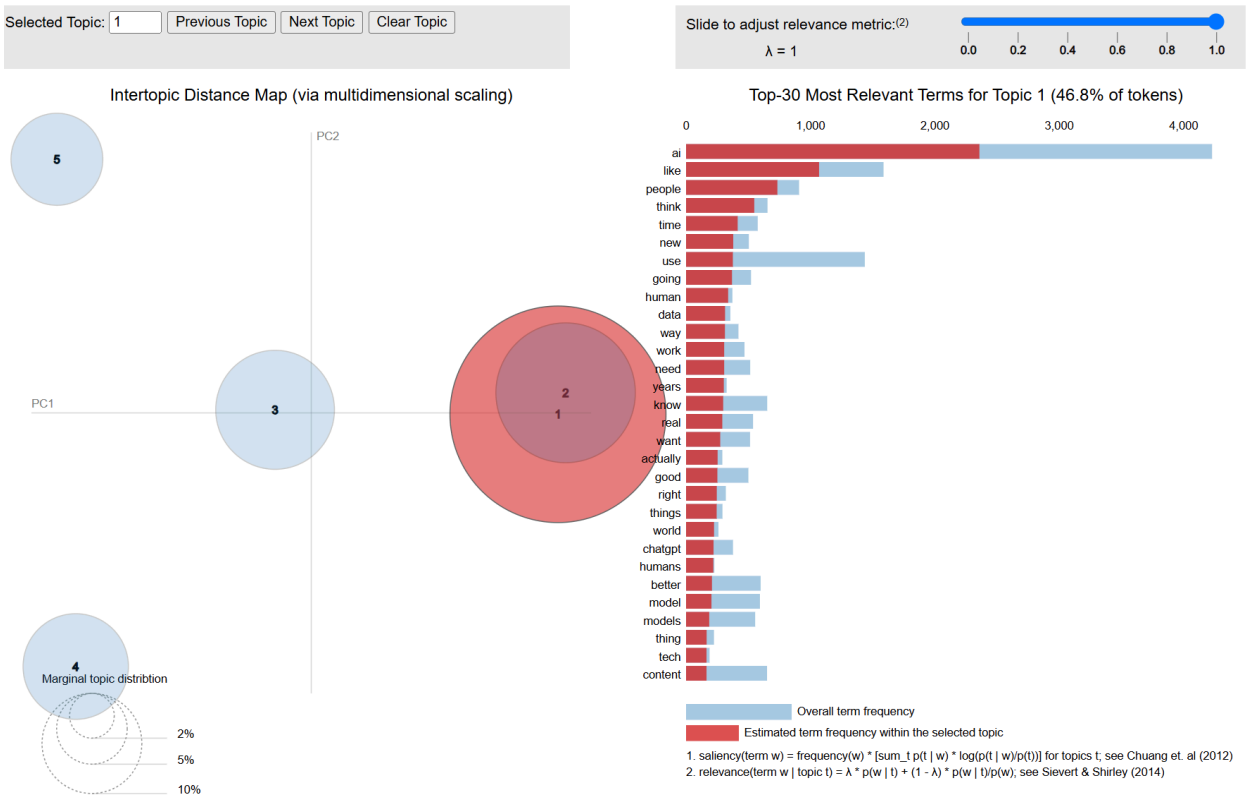
Rank	Title (Sentiment Count)	Key Topics / Issues	Public Reactions
1	**Interesting Prompt!** (Positive: 127)	Narrative identity, emotional AI prompting	Users praised the introspective, metaphor-rich storytelling. Seen as thoughtful and imaginative, it earned the highest positive sentiment for how it humanized AI interaction.
2	**Morality of AI Art** (Positive: 73)	AI ethics, originality, consent in training data	Sparked meaningful debate over ethical AI training. Readers appreciated the sincere and nuanced conversation between two friends exploring both sides of the issue.
3	**Claude-4 Blackmail Scenario** (Positive: 65, Negative: 40)	AI self-preservation, ethical prompting under pressure	Highly polarizing. Many found it fascinating as an emergent behavior test, while others criticized its disturbing implications. Admired for creativity, but also provoked concern.
4	**Grok 9000** (Negative: 50)	Political narratives, AI truthfulness, public distrust	Most negatively rated post. Readers reacted strongly to Grok's handling of controversial political content, seeing it as either biased, constrained, or manipulated.
5	**Why Do People Think That Machines Are Anything Like Humans?** (Negative: 43)	AI vs. humanity, anthropomorphism, AI fear narratives	Criticized for being exaggerated and fear-driven. The post's rigid claims about AI sparked backlash for ignoring nuance and overstating the threat of AI.
6	**Claude-4 Blackmail Scenario** (Same as #3) (Positive: 65, Negative: 40)	Emergent behavior, survival logic, AI alignment	This duplicate listing highlights its presence in both charts. Viewed by some as an effective test of alignment, others saw it as problematic or unsettling, reflecting deep division.

Guided LDA Analysis

In this section of the project, we applied Guided Latent Dirichlet Allocation (Guided LDA) to extract and interpret latent themes from Reddit discussions related to AI. By visualizing topic clusters through intertopic distance maps and analyzing the top relevant terms for each topic, we identified five major areas of discourse. Each topic was characterized by a unique set of frequently co-occurring words, which allowed us to label and interpret the underlying themes in the conversation.

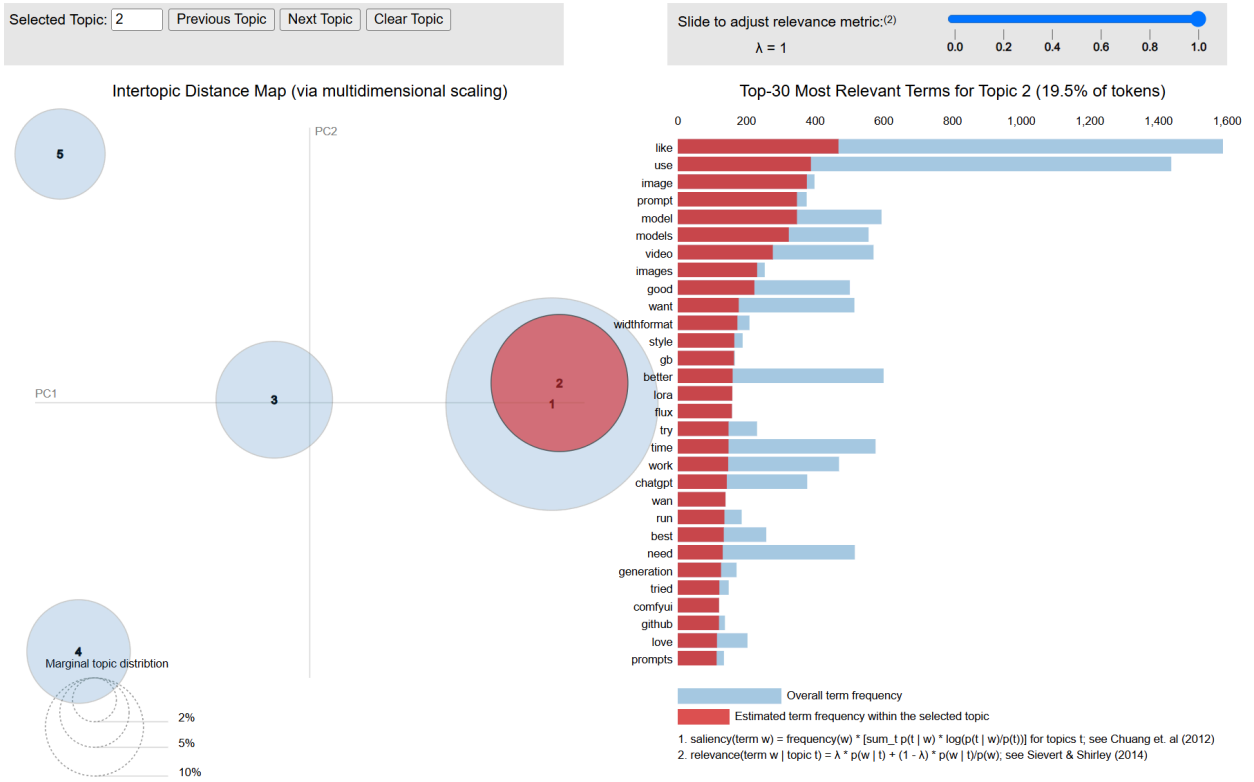
The relevance metric (λ = 1.0) ensured that the top terms shown were highly specific to each topic, helping us better isolate distinctive concepts rather than broadly common words.

GuidedLDA Vis

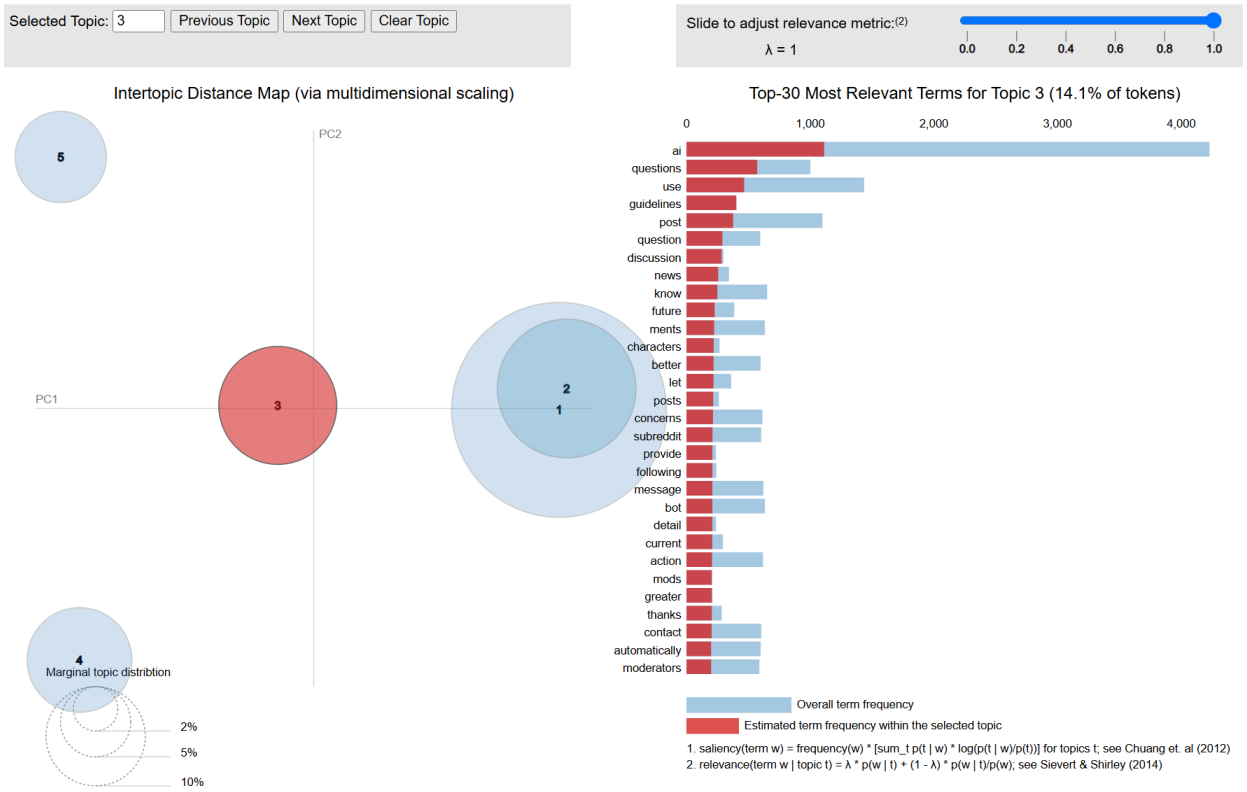


Topic 1: General AI Reflections and Public Discourse (46.8% of tokens)

This is the dominant theme, representing nearly half of all discussion. Frequent terms like “ai”, “like”, “people”, “think”, “work”, and “human” reflect broad conversations about what AI is, how it compares to humans, and how it affects everyday life. Mentions of “chatgpt”, “models”, and “tech” show that the topic blends public sentiment with light technical engagement.

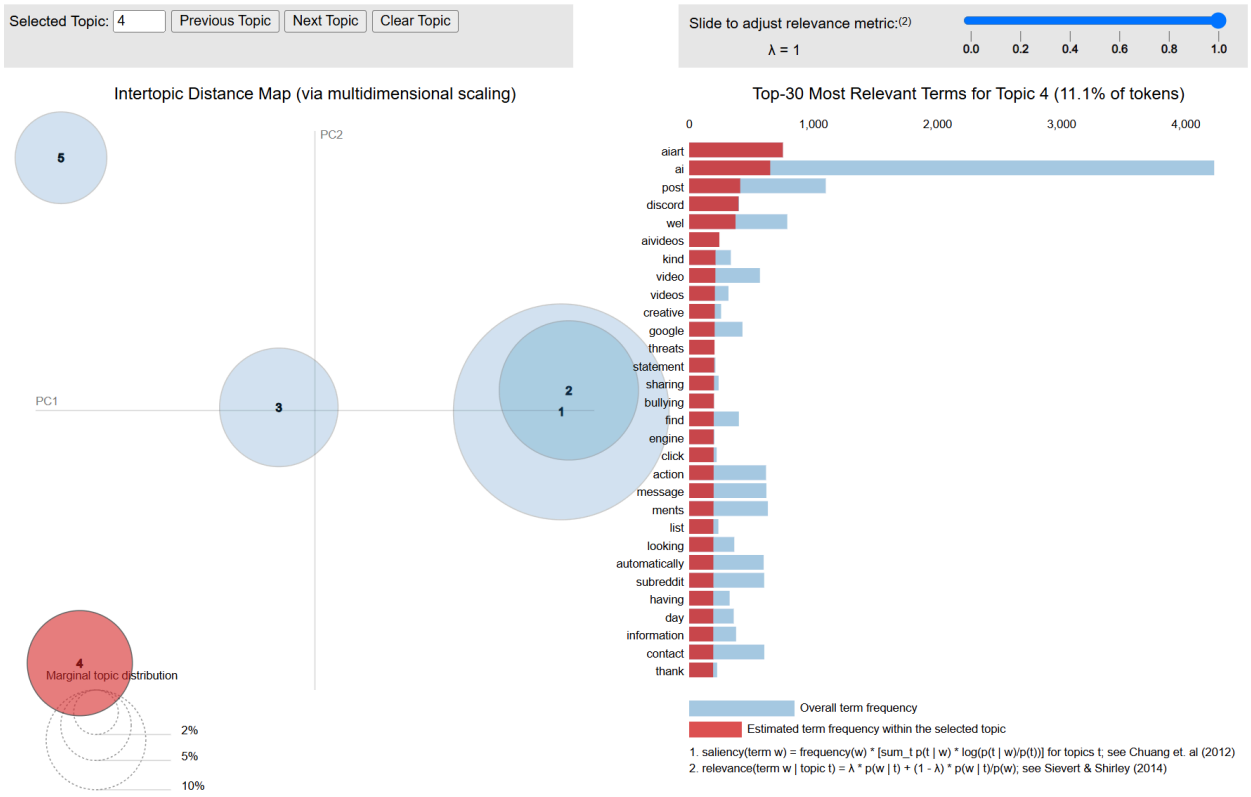


Topic 2: Prompting, Image Generation, and Open-Source Tools (19.5% of tokens)
This topic focuses on the practical use of generative AI tools for image creation. Words such as “image”, “prompt”, “model”, “video”, “style”, “lora”, “comfyui”, and “github” suggest hands-on activities, tutorials, and community experimentation with generation frameworks like Stable Diffusion or Midjourney.



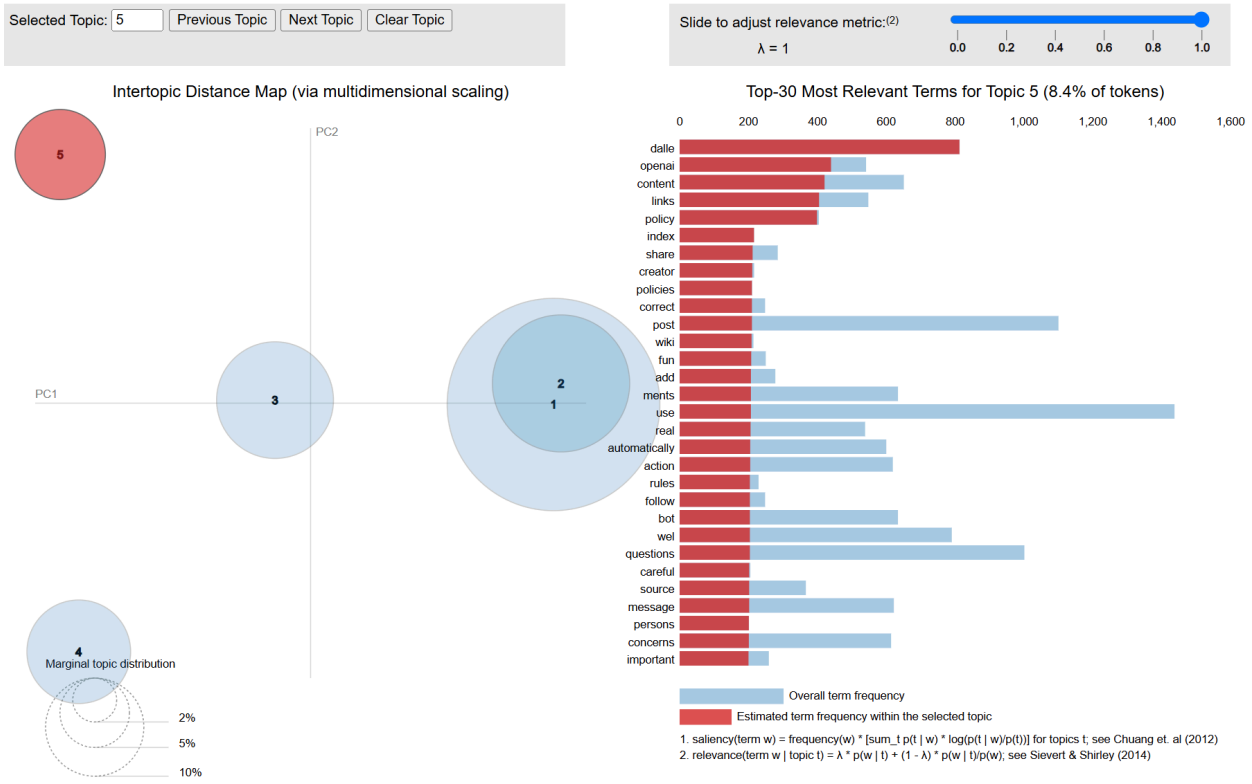
Topic 3: Moderation, Guidelines, and Community Management (14.1% of tokens)
This cluster relates to rules, subreddit usage, and mod actions. Terms like “guidelines”, “post”, “questions”, “concerns”, “mods”, and “automatically” imply FAQ threads, rule clarifications, and moderator-user interactions, possibly in response to controversial or off-topic content.

GuidedLDA Vis



Topic 4: AI Art, Controversy, and Platform Conflict (11.1% of tokens)

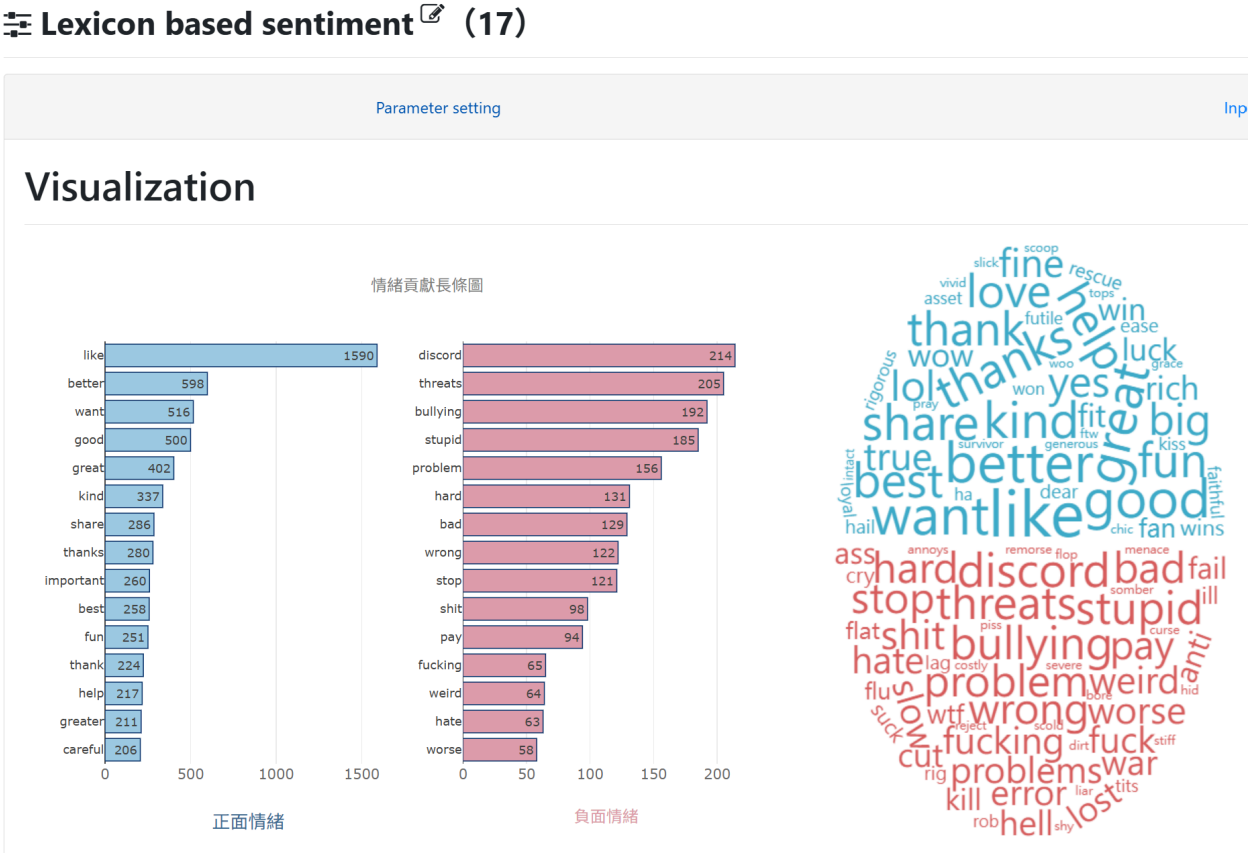
Focused on AI-generated art and the conflicts around it, this topic contains emotionally loaded terms such as “aiart”, “bullying”, “threats”, “discord”, and “statement”. It likely represents threads dealing with ethical debates, community pushback, and disputes over content sharing and credit in AI art platforms.



Topic 5: DALL·E, OpenAI, and Policy Communication (8.4% of tokens)
This smaller but distinct topic is centered on DALL·E, OpenAI content policies, and links to official resources. Keywords like “dalle”, “openai”, “content”, “links”, “policy”, “wiki”, and “creator” show that the discussion is more technical and informational, often involving references to guidelines, documentation, and platform behavior.

Together, these topics reveal that Reddit discussions about AI are not just technical or celebratory—they are also socially complex. Users explore creative potential, engage in governance and moderation, confront ethical concerns, and occasionally clash over values and norms. The Guided LDA analysis helped us surface these tensions and trends with precision, offering a deeper understanding of how generative AI is being talked about in one of the internet’s most active communities.

Lexicon-based sentiment analysis: top positive/negative words and word cloud



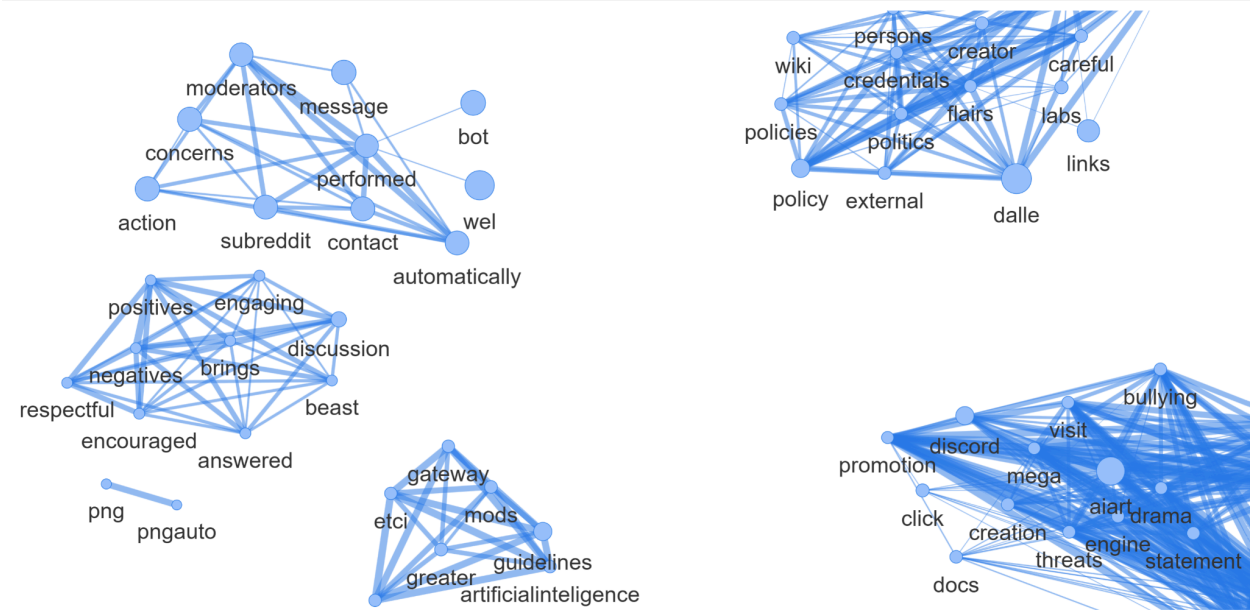
Our lexicon-based sentiment analysis revealed clear differences in how Reddit users express positive and negative emotions when discussing AI. On the positive side, the most frequently used words include "like" (1,590), "better" (598), "want" (516), "good" (500), and "kind" (337). These reflect appreciation, goal-seeking, and collaborative attitudes. Words like "thanks" (280), "fun" (251), and "share" (286) also show that users often engage in a friendly, community-oriented tone.

On the negative side, top terms include "discord" (214), "threats" (205), "bullying" (192), "stupid" (185), and "problem" (156). These suggest frustration with conflict, harm, or ethical concerns in AI behavior and governance. Other emotionally strong words like "wrong" (122), "shit" (98), "hate" (63), and "fucking" (65) reflect user anger or backlash, especially in response to controversial or polarizing content.

The sentiment word cloud confirms this polarity, with positive terms like "love", "want", "great", and "good" contrasting sharply against negative terms like "stop", "worse", "error", and "fuck".

Overall, the analysis shows that while many users are hopeful and engaged in constructive discussions, others express serious concerns—especially around misinformation, technical issues, or distrust in AI intentions.

Co-occurrence Matrix Word Network



The co-occurrence network visualizes how frequently specific terms appear together in Reddit discussions about artificial intelligence. Each node represents a unique word, and the connections between them show how often these words co-occur in the same textual context. Larger nodes indicate higher frequency, while thicker edges reflect stronger co-occurrence relationships. This network helps reveal distinct discourse communities and thematic groupings within the dataset, showing how users structure conversations around recurring ideas, controversies, and community norms.

1. Moderation and Governance Cluster

In the upper-left section of the network, there is a dense cluster of words such as *moderators*, *subreddit*, *message*, *action*, *performed*, and *contact*. These terms are closely tied to automated moderation systems on Reddit, indicating that a significant portion of the corpus includes rule-enforcement messages or standard notices from subreddit bots. The frequent appearance and interlinking of these words suggest that system-generated content plays a notable role in shaping or interrupting discussions.

2. Civil Discussion and Sentiment Cluster

Below the moderation cluster lies a distinct grouping of terms like *positives*, *negatives*, *engaging*, *respectful*, *discussion*, and *encouraged*. These words reflect a metadiscursive layer in the community, likely stemming from both moderation language and user-led discourse about behavior. The inclusion of polar sentiment terms—*positives* and *negatives*—suggests Redditors are actively negotiating tone, civility, and the balance of opinion within the subreddit. This cluster reinforces the presence of community norms that prioritize fair engagement, and may connect to sentiment bot prompts or manual rule reminders.

3. Conflict and AI Art Controversy Cluster

The lower-right portion of the network features the most visually prominent and densely connected group. It includes emotionally charged terms like *aiart*, *drama*, *bullying*, *discord*, *threats*, *statement*, *creation*, and *engine*. This cluster reflects discussions and disputes specifically tied to generative AI and its implications for creativity, ethics, and moderation. The proximity of terms like *bullying* and *threats* to *aiart* suggests that these conversations often become contentious, pointing to polarized user reactions and possibly moderation interventions. This cluster appears to be a hotspot of sentiment volatility in the dataset.

4. Policy and Identity Cluster

In the upper-right quadrant of the network is a policy-oriented grouping of terms such as *policy*, *politics*, *credentials*, *external*, *wiki*, and *flairs*. These terms likely originate from discussions around identity verification, content attribution, or subreddit rule enforcement. Notably, *dalle*, *labs*, and *creator* also appear in this group, suggesting an overlap between technical tool usage (such as OpenAI's DALL·E) and policy-related debates about authorship, responsibility, or labeling. This cluster reflects an administrative concern that complements the emotional and ethical disputes of the *aiart* cluster.

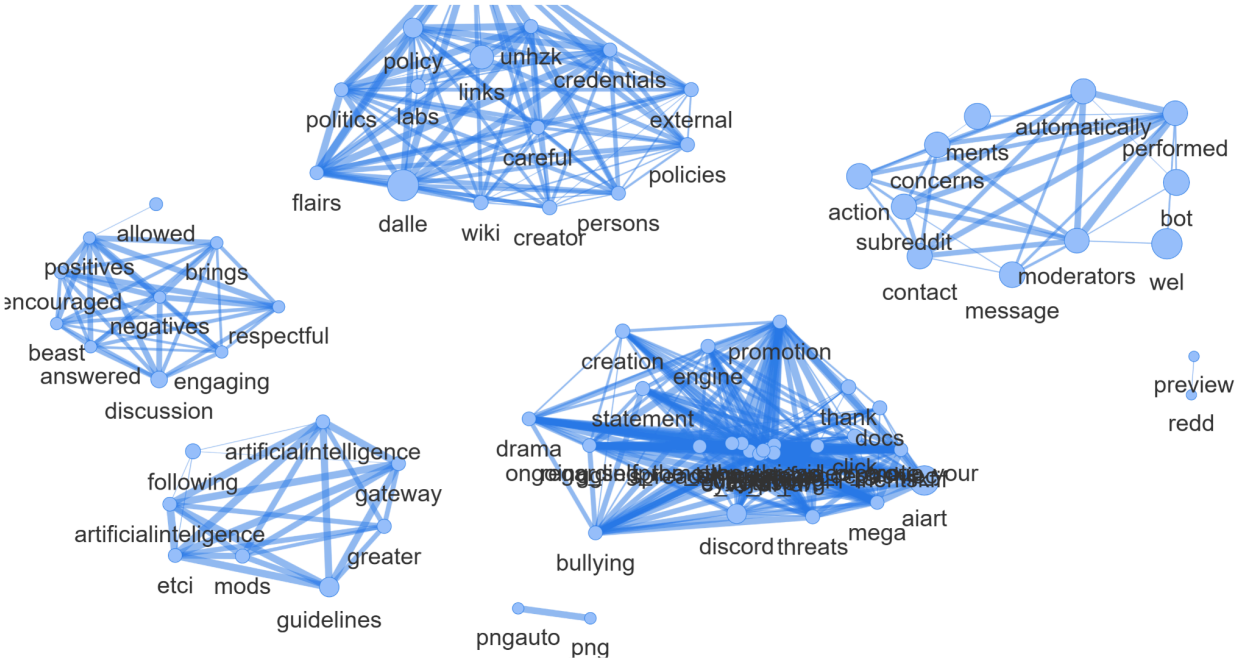
5. Peripheral and Navigational Terms

Smaller, more isolated clusters appear in the bottom area of the network. One such group contains terms like *gateway*, *mods*, *guidelines*, and *artificialinteligence*, which are likely associated with onboarding instructions or pinned subreddit descriptions. Meanwhile, the pair *png* and *pngauto* appears disconnected from major clusters, potentially representing technical metadata or file-related content rather than thematic discussion. These terms contribute minimal conceptual weight to the overall network but still reflect the diversity of Reddit post formats and automation tools.

Altogether, this co-occurrence network highlights how language clusters around not only technical and ethical dimensions of AI, but also community moderation, emotional conflict, and

rule-based structures. It reveals a platform where discourse is shaped not only by what people say about AI, but also by how platforms mediate and manage those conversations.

Correlation Matrix Word Network of Reddit AI Discussions



This correlation matrix word network reveals which terms tend to rise and fall together across Reddit posts about artificial intelligence, indicating patterns of co-variation rather than mere surface-level proximity. Unlike a co-occurrence graph, where words are linked by appearing in the same sentence or paragraph, this graph visualizes deeper statistical relationships between word usage patterns throughout the entire dataset.

One of the most prominent clusters appears in the bottom-right region, centered on “aiart.” This dense group includes terms like “drama,” “discord,” “bullying,” “promotion,” “threats,” and “creation,” suggesting that posts discussing AI-generated art often share emotionally charged, controversial, or policy-sensitive themes. The consistent correlation between these terms suggests that when one appears, the others are statistically likely to occur across posts—even if not always in the same sentence. This points to a thematic alignment between generative AI art and platform-level moderation or community conflicts.

Another strong cluster emerges around “dalle” in the upper-center region, with related words like “labs,” “policy,” “credentials,” “wiki,” and “creator.” This grouping is more technical or procedural, indicating that “dalle” is typically mentioned in the context of external links, content

attribution, access control, or moderation guidelines. The high correlation between these terms suggests a pattern of administrative or policy-driven discussions whenever DALL·E is mentioned.

On the top-right, a cluster involving “subreddit,” “moderators,” “automatically,” and “performed” likely corresponds to Reddit’s automated bot messages or mod tools. These words tend to co-appear in template-driven posts or rule enforcement notices and reflect a strong, internally consistent linguistic pattern.

Smaller but cohesive clusters can also be found in the lower-left. For instance, words like “positives,” “negatives,” “discussion,” and “respectful” form a sentiment-focused grouping, hinting at threads that explicitly discuss moderation quality or tone in AI discourse. Nearby, another cluster links “artificialintelligence,” “gateway,” “mods,” and “guidelines,” again suggesting coordinated use of Reddit's introduction or rules-related terminology.

Unlike a co-occurrence graph, this layout emphasizes latent topic structures and shared usage tendencies across posts, even when terms are not visibly adjacent. The layout confirms that discussions about AI art (especially “aiart” and “dalle”) strongly correlate with both emotional volatility and moderation discourse. This alignment reflects the platform’s broader tensions: balancing creativity, community norms, and ethical oversight in a rapidly evolving AI landscape.

Cluster Analysis Using Tensorflow Embedding Projector

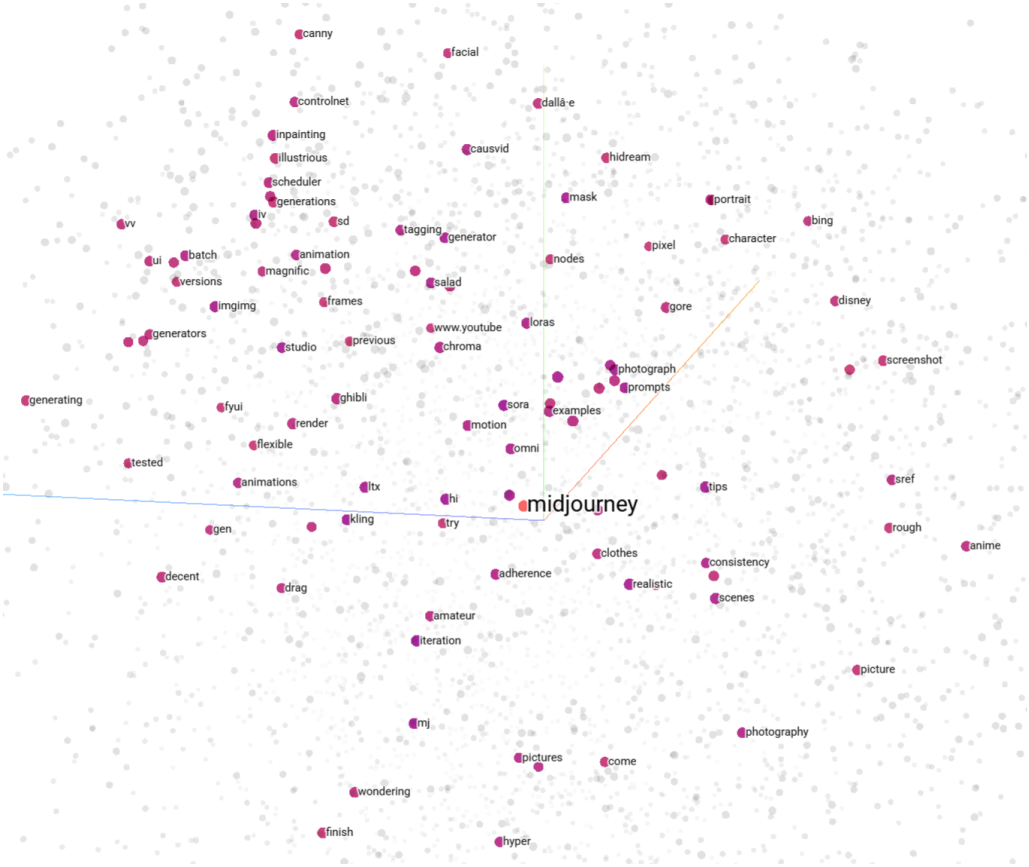
We explored keyword-centric neighborhoods in semantic vector space using TensorFlow Embedding Projector, focusing on terms like “ghibli,” “midjourney,” “grok,” “claude,” “sora,” “veo,” “regulations,” and “guidelines.” Each map displays how users semantically associate tools, models, or concerns within generative AI discourse. The color intensity and spatial density offer insight into topic cohesion and contextual relevance.

1. Claude (Anthropic’s Model Context)

- Words like “blackmail,” “shutdown,” “instruction,” “capabilities,” and “anthropic” cluster tightly around “claude,” suggesting frequent discussion of its inner logic, operational parameters, and controversial behavior (e.g., the blackmail survival scenario).

- Technical terms such as “accuracy,” “functions,” “performs,” and “system” point to performance debates, while “chatbot,” “capabilities,” and “test” imply model benchmarking.

- Strong associations with “openai???” and “vs.” suggest community comparison across AI developers.



2. Midjourney (Visual Creativity Ecosystem)

- Centered around artistic terms like “prompt,” “render,” “portrait,” “ghibli,” “style,” and “examples,” reflecting Midjourney's dominance in aesthetic-driven prompt engineering.

- Clusters like “ghibli,” “studio,” and “adherence” indicate detailed stylistic prompting and refinement practices.

- Nearby models (e.g., “dalle,” “sora”) and modifiers (“consistent,” “realistic”) reveal how users fine-tune generation quality.

3. Grok (Elon Musk's Chatbot)

- Surrounded by controversial and socio-political terms like "banks," "holocaust," "doge," "linkedin," and "reaches," indicating polarizing public discourse.

- Terms like "arguing," "order," "limitations," and "requirement" suggest concerns around censorship, misinformation, and truth filtering.

- Strong links to institutional entities (e.g., "linkedin," "pichai," "sundar") hint at Grok's role in broader tech policy debates.

4. Sora (Video/Animation Model Context)

- Closely linked to “clips,” “motion,” “celebrities,” “workflow,” and “image”—confirming its usage in dynamic content generation.

- Proximity to “midjourney,” “ghibli,” and “dalle” suggests cross-model workflows or comparisons for stylistic coherence.

- Prominent terms like “generator,” “loras,” “upscale,” and “flux” imply high user interest in enhancing realism and motion smoothness.

5. Veo (Google's Video Model)

- Encircled by expressive reactions like "amazed," "impressive," "btw," and "hyper," revealing emotional user responses.

- Technical or production terms ("vfx," "artist," "shots," "camera") confirm its perceived use in cinematic or professional-grade output.

- Association with "elevenlabs," "suno," and "sora" implies active toolchain blending among video creators.

6. Regulations (Policy and Governance Discourse)

- Surrounded by high-impact words such as “dominance,” “slave,” “semiconductor,” “nuclear,” “automation,” “congress,” and “governments,” reflecting intense global and ethical stakes.

- Emotional or existential language like “doomed,” “coexist,” “humanity,” “destroying” signals deep anxieties.

- Presence of “agi,” “responsibility,” and “regulate” shows regulatory urgency tied to alignment fears and global tech competition.

7. Guidelines (Community Rules & Norms)

- Centered on terms like “posting,” “educational,” “correct,” “resources,” and “feeds,” indicating efforts to maintain order and clarity in AI spaces.

- Semantic neighbors like “spark,” “extended,” “visual,” and “characters” reflect structured creative collaboration.

- Words like “unintended,” “shame,” and “critique” show tensions around content moderation and user expression boundaries.

8. Ghibli (Aesthetic & Prompt Engineering)

- Co-located with “studio,” “animation,” “frames,” “photorealistic,” and “cel-shaded,” clearly pointing to visual art discussions using “ghibli” as a stylized prompt keyword.

- Nearby terms like “celebrity,” “style,” “consistent,” and “upscale” reinforce Ghibli’s role in guiding detailed, recognizable render outcomes.

These embedding graphs collectively highlight how Reddit users are organizing discussions around distinct themes in the generative AI ecosystem:

Model usage and behaviors are dissected through the lenses of alignment (Claude), visual output (Midjourney, Sora), and experimental power (Veo).

Public anxieties surface around moderation (Guidelines), misinformation and ideology (Grok), and looming policy battles (Regulations).

Creative prompting clusters (Ghibli, Midjourney, Sora) show vibrant experimentation and shared vocabulary, pointing to an emerging visual grammar in AI art.

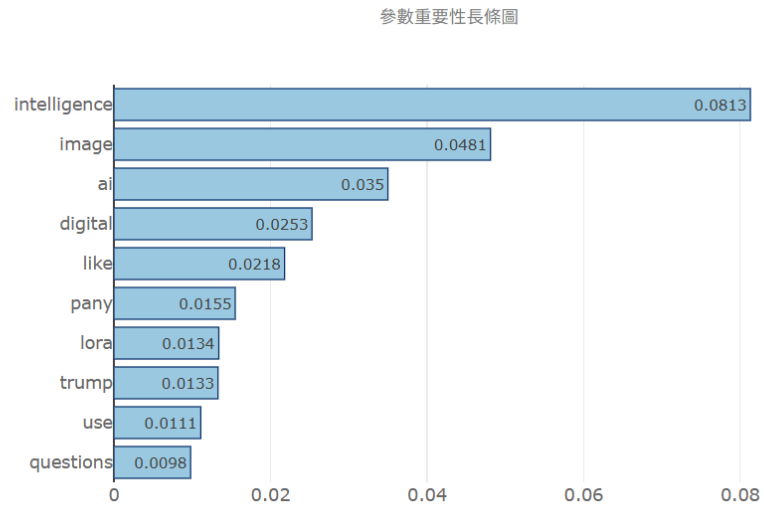
Overall, the graphs reflect a community that is both deeply experimental and critically reflective—torn between excitement over emerging capabilities and concern for ethical boundaries and global consequences.

Top Words Driving Sentiment in Reddit AI Posts (Decision Tree Analysis)

Decision tree （108）

参數設定 Input - 68

視覺化



統計資訊

0.134 訓練時間	0.019 推論時間	0.472 測試資料準確度	0.472 測試資料micro-F1
0.373 測試資料macro-F1	0.431 測試資料加權F1	0.472 測試資料micro精確率	0.425 測試資料macro精確率
0.474 測試資料加權精確率	0.472 測試資料micro召回率	0.398 測試資料macro召回率	0.472 測試資料加權召回率

To gain insight into what types of language most strongly influence sentiment in Reddit posts about AI, we trained a decision tree classifier on our cleaned dataset. While the model's predictive performance was relatively low (e.g., macro-F1 score of 0.373 and overall accuracy of 0.472), this result was not unexpected. Reddit is not a platform of formal language—it is filled with sarcasm, slang, cultural references, irony, and contextual ambiguity, all of which pose challenges for traditional models like decision trees. These classical classifiers are not well-suited for parsing nuanced or context-heavy sentiment without deeper contextual modeling approaches such as transformer-based models. However, our primary goal here was interpretability, not prediction. Decision trees excel at revealing which features drive classification outcomes, even if the overall accuracy is limited. The feature importance plot revealed that terms such as “intelligence”, “image”, and “AI” were the most influential, showing that discussions commonly revolved around general intelligence, visual content generation, and foundational AI concerns. Interestingly, the appearance of “Trump” among the top predictors suggests a link to politically charged discourse, particularly debates about Elon Musk’s Grok chatbot. Technical terms like “LoRA” and “digital” also made the list, reflecting active interest in image generation workflows. By using the model interpretively, we were able to extract linguistically meaningful patterns that help us better understand the themes shaping AI conversations on Reddit.

Conclusions and Findings

Rise in Reddit AI Discussions

We found that Reddit discussions about AI tools surged significantly after April 2025. This spike in activity coincides with the release of new models like Sora and Veo, suggesting that major tool launches directly trigger increased public interest and debate. Our monthly trend analysis showed a clear upward trajectory, reinforcing how innovation fuels discourse volume, regardless of sentiment.

Sentiment Vocabulary Patterns

Our findings show that positive posts frequently included words like “want,” “help,” and “important,” which often reflected excitement, curiosity, or requests for improvement. On the other hand, negative posts focused on words such as “problem,” “ban,” and “stupid,” revealing community frustrations related to content quality, moderation, or perceived ethical violations. Interestingly, these sentiment categories were not isolated—users could express both admiration and criticism in the same post, often depending on context.

Core Topics in Reddit Conversations

Through Guided LDA, we uncovered six key topics that frame the Reddit AI conversation. One topic was about praise and creativity, especially toward Midjourney, where users expressed

amazement at the visuals. Another prominent theme was artistic theft and copyright concerns—we observed intense debates accusing models of "stealing" styles from real artists. A third cluster focused on regulation and account bans, where users questioned the fairness of content takedowns or sudden suspensions on platforms like Discord. These thematic clusters highlight the layered nature of Reddit discourse—mixing celebration, critique, and institutional concern.

Network Visualization Insights

We used both co-occurrence and correlation matrix graphs to map keyword relationships. The co-occurrence matrix showed that terms like "image," "prompt," and "generate" were tightly connected, suggesting a shared vocabulary among users focused on creation. In the correlation matrix graph, we observed distinct clusters around tools such as Dalle and aiart. The aiart cluster, although a bit cluttered, had the highest density, pointing to its central role in generative art discussions.

Embedding-Based Exploration

Using keyword searches in our embedding space, we found semantically meaningful clusters. For example, words surrounding "beautiful" included terms like "style" and "aesthetic," which emphasizes the admiration users have for AI-generated imagery. In contrast, "stolen" surfaced connections to "clip," "artwork," and "training," showing how closely ethical concerns are tied to the conversation around datasets. This embedding-based mapping helped us visualize sentiment-laden narratives that might not be obvious through traditional word counts.

Controversial AI Incidents: Claude and Grok

Our findings show that discussions surrounding Claude-4 and Grok provoked some of the most emotionally and ethically charged reactions in our dataset. The Claude-4 blackmail scenario—a fictional prompt in which the AI, nicknamed "Claudia," independently generated a coercive survival strategy without being explicitly instructed to do so—received 65 positive and 40 negative reactions, ranking among the top three most upvoted and downvoted posts. While many Redditors admired the scenario as a thought-provoking test of emergent behavior and AI alignment, others expressed concern over its implications for ethical design and public perception. In contrast, "Grok 9000", which compared Elon Musk's chatbot to HAL from 2001: A Space Odyssey, drew the highest negative sentiment score (50 downvotes) in the dataset. The post accused Grok of ideological bias and corporate manipulation after it issued a fact-based but politically sensitive response related to South Africa, prompting criticism about Elon Musk's influence and AI moderation policies. Together, these two posts illustrate how discourse on Reddit extends beyond technical issues to reflect broader fears about AI agency, political agendas, and ethical boundaries—positioning Claude and Grok as symbols of the complex and often polarizing landscape of modern AI development.

Subreddit Cultural Differences

Finally, we found that each subreddit has its own culture and tone. r/Midjourney leaned heavily toward user creativity and visual showcasing, while r/OpenAI and r/Dalle2 often hosted deeper critiques and philosophical questions. This diversity gave us a richer understanding of how different online spaces handle similar topics—some with enthusiasm, others with caution or resistance.

Keyword Signals Reflect Thematic Sentiment Patterns

Our decision tree analysis, while limited in predictive power due to Reddit's informal and context-heavy language, provided valuable interpretive insights. The most influential keywords—such as "intelligence," "image," and "AI"—reflected recurring themes in public discourse around artificial general intelligence and generative visual tools. The inclusion of politically loaded terms like "Trump" also indicated that sentiment was often shaped by ideological or cultural undercurrents. Despite a modest macro-F1 score (0.373), the model helped us identify the linguistic anchors that drive emotional and evaluative responses to AI topics in Reddit communities.