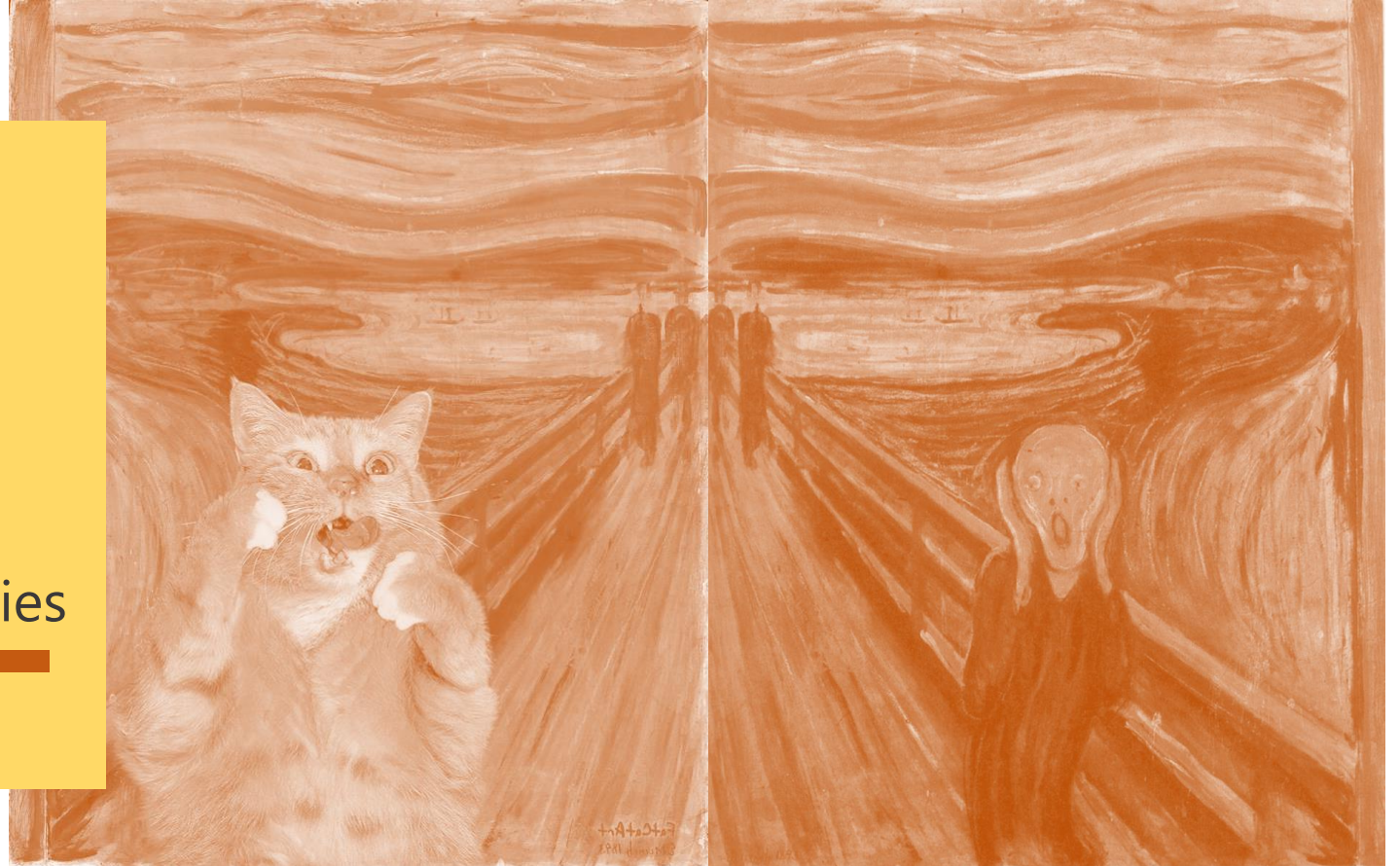


img src: <https://fatcatart.com/2016/08/scream-munch/?lang=en>

# Artificial Intelligence & Art

Understanding and predicting  
the chatter from **online** communities

---



**DSI-41 Project Group 2:** Faiz, Gabriel, Lionel

# Market trends – AI and the arts

The digital artwork\* market size is expected to reach  
**US\$ 12.12 billion**  
**by 2030,**  
from US\$ 4.09 billion in 2023, at a  
**CAGR of 16.8%** during the  
forecast period.

\* refers to art that is created and presented in digital forms using digital technology

Eva Toorenent has been working as a freelance artist and illustrator since 2019. Late last year, she was shocked to discover that another artist had taken artwork she'd posted on Instagram and used it to "fine tune" the AI model Midjourney to produce AI art in her style. The other artist then sold the artwork to a gallery.



Eva Toorenent has worked as an artist and illustrator since 2019. Eva Toorenent

"It was hurtful," she told Insider. "My work has been stolen before, but this was a new kind of violating feeling that I haven't felt before."

Since then, she has increasingly encountered generative AI in her work, and says that the technology is being used to cut corners and costs, leaving freelance artists to pay the price.

"It's scary to see," she said. "But honestly, it feels inevitable."

<https://www.businessinsider.com/ai-taking-jobs-fears-artists-say-already-happening-2023-10>

# Current topics regarding AI and the arts

---

- **Impact on artists**
  - in terms of work and revenue
- **Authenticity / creativity**
  - AI algorithms may not truly reflect an understanding of human creativity or emotion
- **Copyright and authorship**
  - complex, and regulations are still being worked out in this very new area...
- **Ethical and privacy concerns**
  - how training data is being used, and how algorithms may perpetuate extant biases

...On the other hand, AI is here to stay with **huge potential** as in terms of market demand and as a collaborative tool

**BUT WHAT PEOPLE ARE REALLY TALKING ABOUT?**

# Problem statement and objectives

---

- To train a model using Natural Language Processing (NLP) to **accurately predict which of 2 subreddits a given post comes from**,
- How we can understand **key differences in 'pro-AI' and 'anti-AI' discourse and sentiments** to guide policymakers

**r\DefendingAIArt** vs **r\ArtistHate**



# Why this matters

---

Our model solution offers the option of **a quick ground sensing to assess online chatter** and understand which way the discourse is evolving

Useful for guiding policymaking, regardless of whether you are:

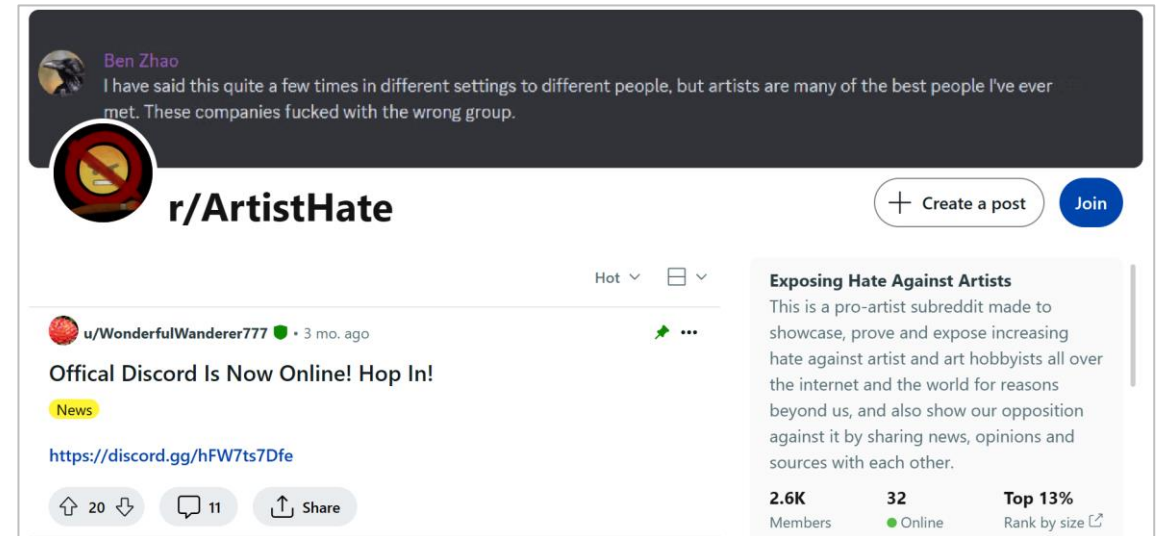
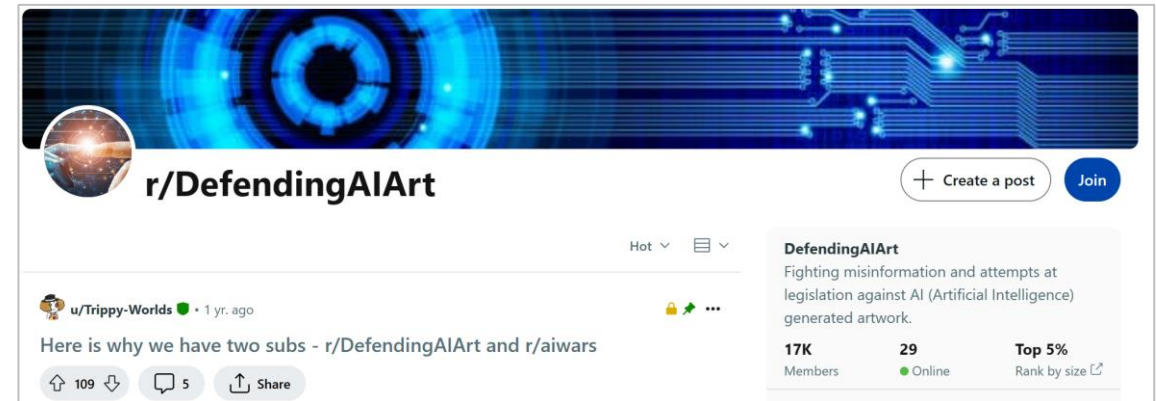
- furthering **related industry and AI tech initiatives** (e.g. iMDA),
- **developing the local arts scene** and **engaging with artists/art institutions/the general public** (e.g. NAC)
- establishing relevant **guidelines and regulatory frameworks** to keep up with the pace of AI development (e.g. PDPC, IPOS)



# The subreddits

**Both subreddits were created in 2023:** recency and comparability in time periods

- **r/DefendingAIArt:**  
i.e. the 'pro' camp. Describes itself as a "space for pro-AI activism" and is aimed at "fighting misinformation and attempts at legislation against AI generated artwork"
- **r/ArtistHate:**  
i.e. the 'anti' camp. Named as such for Redditors to discuss what is being observed as "increasing hate against artist and art hobbyists" when their interests should be better protected amidst AI



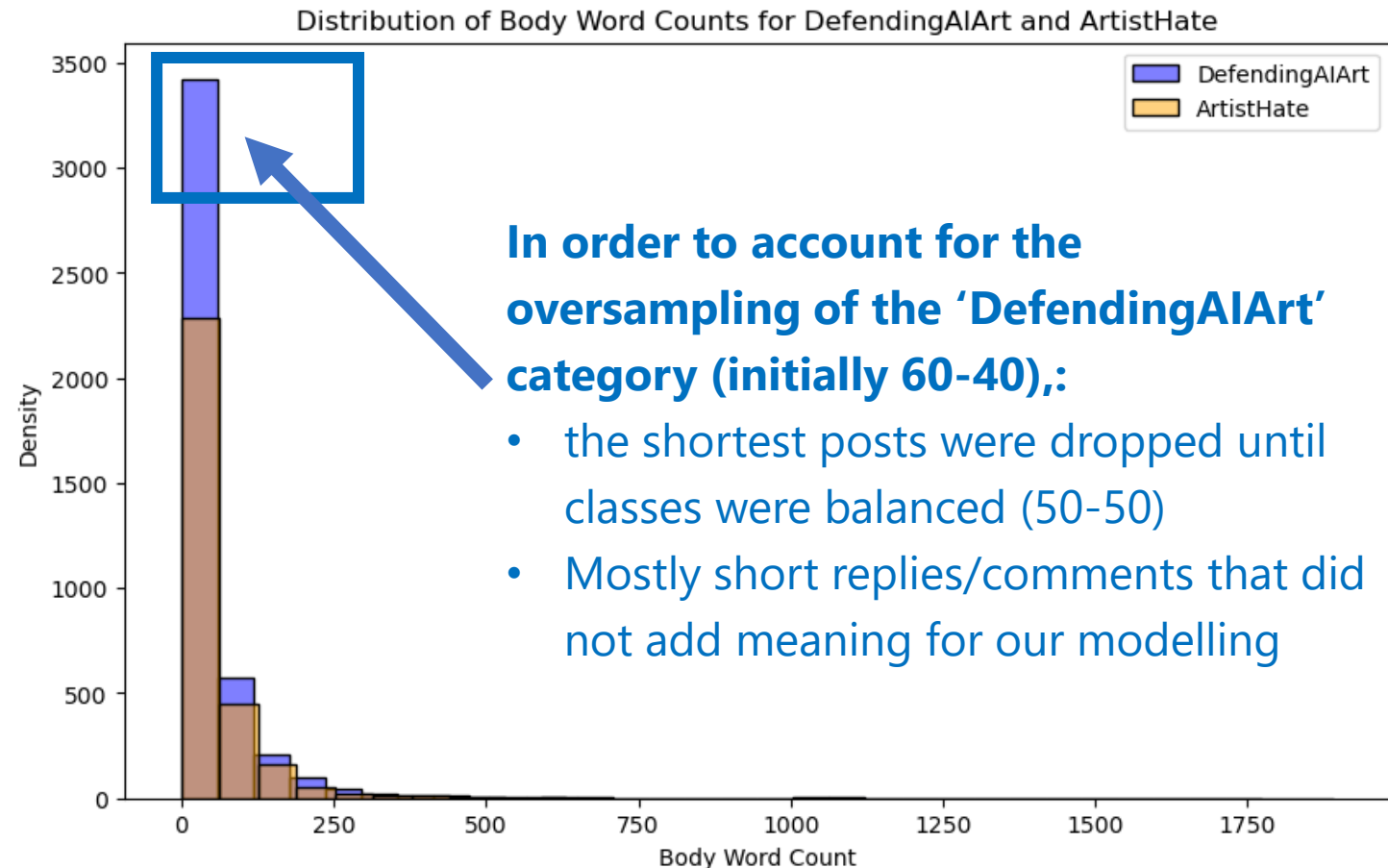
# Preprocessing of data

---

- Scraping of posts via PRAW to access Reddit API
- Text cleaning and pre-processing to capture essence of words
  - Removing deleted / bot-generated posts
  - Rendering emoticons and emojis as text descriptions to capture possible sentiment
  - Removing urls and gif links

# Preprocessing of data

- Scraping of posts via PRAW to access Reddit API
- Text cleaning and pre-processing to capture essence of words
  - Removing deleted / bot-generated posts
  - Rendering emoticons and emojis as text descriptions to capture possible sentiment
  - Removing urls and gif links
- Balancing of classes (50-50) to better evaluate predictions for either camp
- Resulting in a **total of ~n=6,000 usable rows of data** consisting of original posts and comments





# Sentiment analysis

## What the scores mean for both subreddits on average:

- Overall **somewhat positive** sentiments (compound scores >0.05)
- **Balance of subjective opinions and objective discussions** (subjectivity scores around 0.5)

## Insights:

- Encouraging as it suggests the current discourse is not strongly polarising or emotional
- However the overall positive sentiment inferred has its shortcomings as humour or sarcasm may not be picked up by the current algorithm

	Compound score* for sentiment	Subjectivity Score^
DefendingAIArt	0.07	0.49
ArtistHate	0.11	0.46

\* Based on VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis tool

^ Based on TextBlob (Python NLP library)

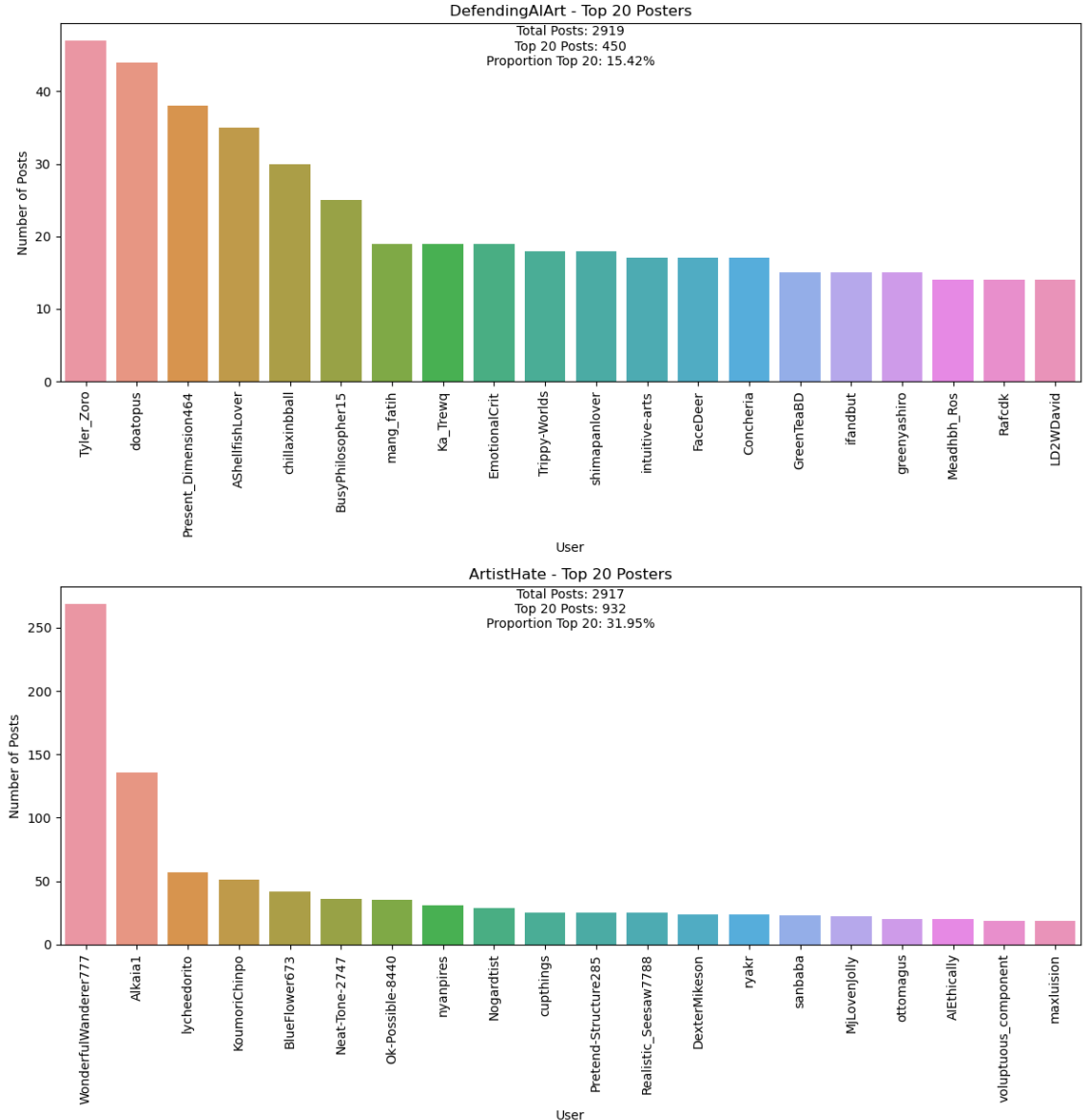
# Other observations:

## Redditor base

The nature of debate and conversations may be dominated by a select few in the 'anti-AI'-camp:

- **r\DefendingAIArt**: almost **twice the no. of unique users** (1096) compared to ArtistHate (603)
- **r\ArtistHate**: much **higher proportion of posts** (32.0%) **contributed by the top 20 users** compared to DefendingAIArt (15.4%)

Top 20 Posters and Proportion of Posts by Top 20 Users



# Other observations:

## Single-word counts

Big overlap when looking at single words  
(even after removing stop-words)

### Top 20 words by frequency from r\DefendingAIArt

- |           |            |
|-----------|------------|
| 1. ai     | 11. use    |
| 2. art    | 12. want   |
| 3. people | 13. image  |
| 4. artist | 14. time   |
| 5. just   | 15. way    |
| 6. like   | 16. good   |
| 7. make   | 17. know   |
| 8. work   | 18. really |
| 9. think  | 19. human  |
| 10. thing | 20. say    |

### Top 20 words by frequency from r\ArtistHate

- |           |            |
|-----------|------------|
| 1. ai     | 11. image  |
| 2. people | 12. want   |
| 3. art    | 13. time   |
| 4. just   | 14. use    |
| 5. like   | 15. way    |
| 6. artist | 16. human  |
| 7. work   | 17. know   |
| 8. make   | 18. really |
| 9. thing  | 19. good   |
| 10. think | 20. model  |

# Other observations:

## 2-word counts

Still several common words, suggesting they are discussing **similar** AI-specific issues – simply with different perspectives.

### Top 20 words by frequency from r\DefendingAIArt

- |                 |                     |
|-----------------|---------------------|
| 1. ai art       | 1. make art         |
| 2. anti ai      | 2. like ai          |
| 3. use ai       | 3. stable diffusion |
| 4. using ai     | 4. people like      |
| 5. ai artist    | 5. ai people        |
| 6. ai generated | 6. fan art          |
| 7. look like    | 7. art ai           |
| 8. feel like    | 8. fair use         |
| 9. just like    | 9. lot people       |
| 10. ai tool     | 10. pro ai          |

### Top 20 words by frequency from r\ArtistHate

- |                      |                  |
|----------------------|------------------|
| 1. ai art            | 1. using ai      |
| 2. ai bros           | 2. ai image      |
| 3. ai generated      | 3. art just      |
| 4. generative ai     | 4. lot people    |
| 5. look like         | 5. ai bro        |
| 6. ai artist         | 6. thing like    |
| 7. feel like         | 7. just want     |
| 8. just like         | 8. training data |
| 9. use ai            | 9. people want   |
| 10. stable diffusion | 10. ai just      |

# Other observations:

## 2-word counts

Still several common words, suggesting they are discussing **similar** AI-specific issues – simply with different perspectives.

### Top 20 words by frequency from r\DefendingAIArt

- |                     |  |
|---------------------|--|
| 1. ai art           | 11. make art                             |
| 2. anti ai          | 12. <b>Comments about AI-generated</b>   |
| 3. use ai           | 13. <b>output... is it 'really' art?</b> |
| 4. using ai         | 14. people like                          |
| 5. ai artist        | 15. ai people                            |
| 6. ai generated     | 16. fan art                              |
| <b>7. look like</b> | 17. art ai                               |
| <b>8. feel like</b> | 18. fair use                             |
| <b>9. just like</b> | 19. lot people                           |
| 10. ai tool         | 20. pro ai                               |

### Top 20 words by frequency from r\ArtistHate

- |                      |                   |
|----------------------|-------------------|
| 1. ai art            | 11. using ai      |
| 2. ai generated      | 12. ai image      |
| 3. ai generated      | 13. art just      |
| 4. generative ai     | 14. lot people    |
| <b>5. look like</b>  | 15. ai bro        |
| 6. ai artist         | 16. thing like    |
| <b>7. feel like</b>  | 17. just want     |
| <b>8. just like</b>  | 18. training data |
| 9. use ai            | 19. people want   |
| 10. stable diffusion | 20. ai just       |



# Other observations:

## 2-word counts

Still several common words, suggesting they are discussing **similar** AI-specific issues – simply with different perspectives.

### Top 20 words by frequency from r\DefendingAIArt

- |                 |                      |
|-----------------|----------------------|
| 1. ai art       | 11. make art         |
| 2. anti ai      | 12. like ai          |
| 3. use ai       | 13. stable diffusion |
| 4. using ai     | 14. people like      |
| 5. ai artist    | 15. ai people        |
| 6. ai generated | 16. ai image         |
| 7. look like    | 17. art just         |
| 8. feel like    | 18. lot people       |
| 9. just like    | 19. ai bro           |
| 10. ai tool     | 20. thing like       |

**mentions of generative AI-tools 'Stable Diffusion' (text to image algorithm)**

### Top 20 words by frequency from r\ArtistHate

- |                      |                   |
|----------------------|-------------------|
| 1. ai art            | 11. using ai      |
| 2. ai bros           | 12. ai image      |
| 3. ai generated      | 13. art just      |
| 4. Stable Diffusion  | 14. lot people    |
| 5. ai artist         | 15. ai bro        |
| 6. ai generated      | 16. thing like    |
| 7. look like         | 17. just want     |
| 8. just like         | 18. training data |
| 9. use ai            | 19. people want   |
| 10. stable diffusion | 20. ai just       |

# Other observations:

## 2-word counts

Some indication of unique arguments being put forth by opposing camps

### Top 20 words by frequency from r\DefendingAIArt

- |                 |                      |
|-----------------|----------------------|
| 1. ai art       | 11. make art         |
| 2. anti ai      | 12. like ai          |
| 3. use ai       | 13. stable diffusion |
| 4. using ai     | 14. people like      |
| 5. ai artist    | 15. ai people        |
| 6. ai generated | 16. fan art          |
| 7. look like    | 17. art ai           |
| 8. feel like    | <b>18. fair use</b>  |
| 9. just like    | 19. lot people       |
| 10. ai tool     | 20. pro ai           |

### Top 20 words by frequency from r\ArtistHate

- |                      |                          |
|----------------------|--------------------------|
| 1. ai art            | 11. using ai             |
| 2. ai bros           | 12. ai image             |
| 3. ai generated      | 13. art just             |
| 4. generative ai     | 14. lot people           |
| 5. look like         | 15. ai bro               |
| 6. ai artist         | 16. thing like           |
| 7. feel like         | 17. just want            |
| 8. just like         | <b>18. training data</b> |
| 9. use ai            | 19. people want          |
| 10. stable diffusion | 20. ai just              |

# Actual modelling – models tested

	Training accuracy	Testing accuracy	Precision	Recall	F1 score	Computation time
Multinomial Naive Bayes	0.83	0.68	0.67	0.73	0.70	1 x
Logistic Regression	0.96	0.71	0.73	0.67	0.70	2 x
K-Nearest Neighbors	1.0	0.55	0.70	0.17	0.28	2 x
Random Forest	0.93	0.65	0.63	0.71	0.67	Up to 12 x

# Actual modelling – final model

---

	Training accuracy	Testing accuracy	Precision	Recall	F1 score	Computation time
<b>Multinomial Naive Bayes</b>	<b>0.83</b>	<b>0.68</b>	<b>0.67</b>	<b>0.73</b>	<b>0.70</b>	<b>1 x</b>

- **Able to get predictions correct almost 70% of the time overall**
- Generalisable to unseen data
- Computationally **fastest (and cheapest)**
- **Simplicity** in implement – can handle discrete data (text) and word counts
- Highly **scalable** – can easily read high volume of posts

---

**DEMO time**



# Consideration for next steps

---

## Could we further strengthen the predictive accuracy of the model?

- Issue-wise, similar vocabulary being used to describe flip sides of the same coin on either camp → can make it challenging for models to distinguish
- Data limited to posts from 2023 onwards, due to recency of topic
- Our predictive accuracy of close to 70% is already **significantly better than the baseline by twenty percent!**

## Possible next steps: funding for additional research to

- collect data across more social media platforms
- supplement with other data collection methods (e.g. focus groups with artists and other individuals)



# Recommendations

## 1. More initiatives in these areas:

- **Extend our research** to other social media platforms, or other methods such as in-depth interviews/ focus groups/ surveys with artists and other individuals
- Initiatives to **promote human artist-AI system collaboration**, not replacement
- **Broader educational efforts** to raise awareness about how AI impacts artists and other professionals such as curators, art-related tech developers etc.

## 2. Review potential concerns, e.g. review regulatory/ ethical guidelines (Code of Practice):

- Clarity around **IP ownership** of AI-generated artworks to ensure artists are protected
- Safeguard individuals' **privacy and use of training data** fed into AI models
- Encourage **ethical data sourcing practices** by artists and developers
- Requiring **disclosure** in the use of AI for artwork

## 3. More public engagement

- Industry stakeholders: **artists, technology developers, art institutions**
- Involve **the public** in decision-making processes

---

# Q&A