



yummybreeze123 · 9h ago ·

Can't commit to what? an attention seeking whore? Is there something i'm not getting

↑ 4 ↓ Reply Share ...



45tee · 9h ago ·

Look at his eyes. He's insa

⊖ ↑ 22 ↓ Reply



Big-Still6880 · 1d ago

First, they do a sh#t. when people compla unbelievable. LOL Edit: T that of funding insufficie



Br0kenba3 · 4d ag

Do you even have

⊖ ↑ 4 ↓ Reply 7



KELINGCB · 2d ago ·



OuhLongJohnson · 5d ago ·

Good. Less Chinese here please

↑ 0 ↓ Reply Share



candychiasu · 4h ago ·

... nical want be centre of attention

Share

Lowly peasant.

...y posters if you can't even spell proper English words

⊖ ↑ 4 ↓ Reply Share ...



PsyArif · 5h ago ·

Wow 5 seats in a row!

Take off slippers and put on the floor. Even phone is on the floor like a nightstand next to his bed at home.

Do we need a mandatory MRT rider course? Culture shock? Don't see this happening in East Asian countries.

⊖ ↑ 17 ↓ Reply Share ...



HANAEMILK · 5h ago ·

Bringing their shitty culture over...



sansansansansan · 23h ago ·

this guy got the unemployable cb face

↑ 1 ↓ Reply Share ...



The Problem



Stakeholder:

MDDI's Online Trust and Safety Department

Their Challenge:

Regulating harmful content

- Reddit's moderation tools are limited

Our Objectives:

- Toxicity and hatefulness analysis
- Recommendations: reduce toxicity & hatefulness levels

Outcomes:

- Saved time and operational cost in moderation
- Accurate toxicity detection

DSA4264

Reddit Analysis

Toxicity and Hatefulness Trends in
Singapore Social Media



Presented by:

Caleb Lee Heng Yi | Denise Teh Kai Xin
Lin Zhengjue Elisa | Neleh Tok Ying Yun
Sarah Goh Yue En

Agenda

WHAT WE'RE DISCUSSING TODAY



1

- Methodology:
 - EDA
 - Sentiment Analysis
 - Topic Modelling
 - Limitations

2

- Results:
 - Trends
 - Theories

3

- Recommendations:
 - Solutions
 - Moving Forward

Methodology








Preprocessing

Dataset and Cleaning Methods

...

	r/singapore 1.5M members
	r/SingaporeRaw 78K members
	r/singaporehappenings 43K members

4932728 Rows

- **2020 Jan 1 - 2023 Oct 31**
 - 2020-21 dataset: 60%
 - 2022-23 dataset: 40%
- Comment text, date & time, subreddit ID, original post link, moderation information



Preprocessing

Dataset and Cleaning Methods



Random Sampling: 500,000 Comments

- Retain original proportion
- Pandas random sample

You did what you could. He
just needs to grow up! 😞



Cleaning

- Remove emojis, punctuation, non-english, non-ASCII characters
- Set all characters to lowercase
- Remove [deleted] and [removed] comments

you did what you could he
just needs to grow up





LionGuard

Toxicity and Hatefulness Scoring



Enter your text:

I hate NUS, it is the worst school ever.

[I can't think of anything - fill in an example for me!](#)

Check Content

Overall: Fail (74/100)

Harassment: Pass	Sexual: Pass
Hateful: Pass	Toxic: Fail
Public Harm: Pass	Violent: Pass
Self Harm: Pass	

Do you agree with these results?

☒ Yes ☐ No

[Provide detailed feedback](#)

([LionGuard website](#))

- **Trained on forum data**

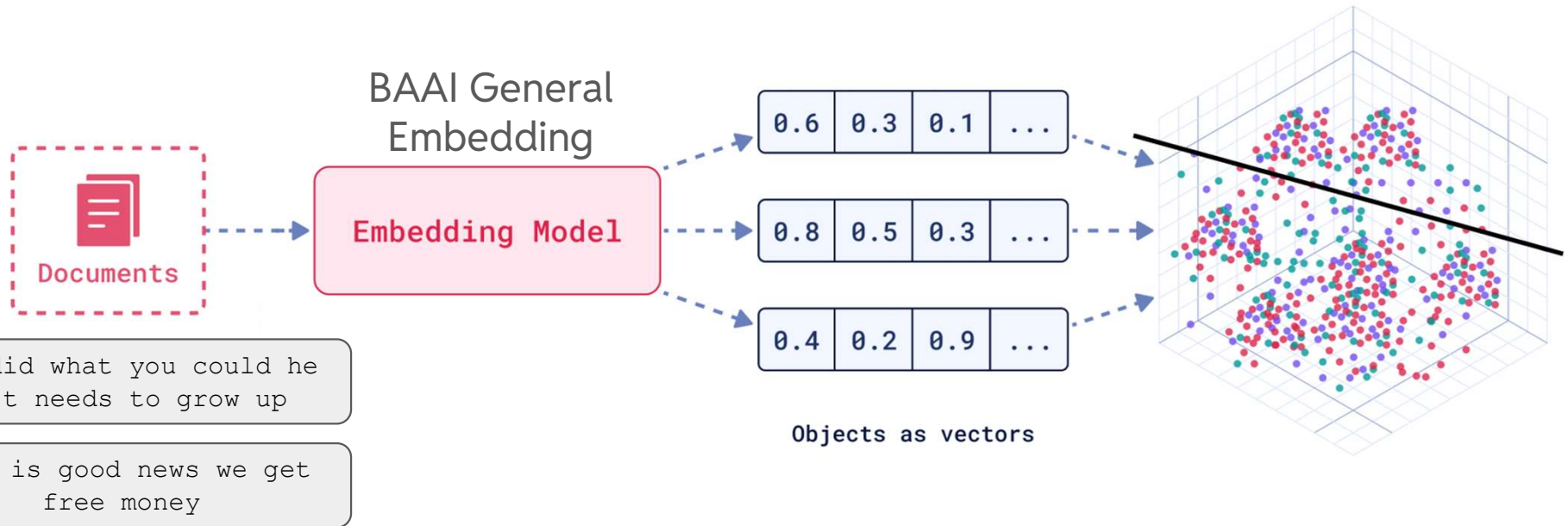
- Reddit vocabulary is different from everyday speech
- Understands Singlish words and abbreviations



Embedding & Classification

How they work:

...





Embedding & Classification

How they work:

...

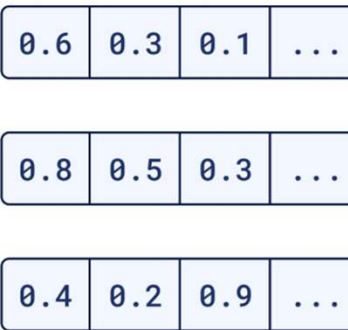


Documents

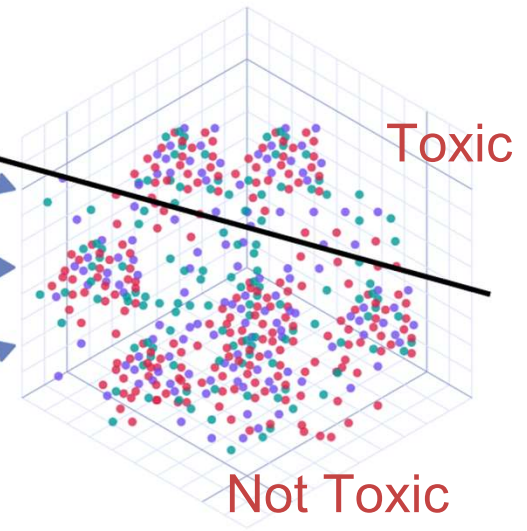
Embedding Model

you did what you could he
just needs to grow up

this is good news we get
free money



Objects as vectors





Topic Modelling

LDAMallet

...

With so many new covid19 cases ive
started wearing masks again

hdb prices have been rising every
year how to buy

```
# Display the topics  
display_topics(lda, feature_names, no_top_words)
```

Topic 0:
gt singapore said cases police case covid19 public new source
Topic 1:
people dont think like good want pap make things say
Topic 2:
like car buy people road got bus use think really
Topic 3:
thanks read know people post like lol im dont comment
Topic 4:
years hdb yes price prices market money year pay new
Topic 5:
food like eat time good work looks nice chicken home
Topic 6:
people like covid dont mask wear need comments vaccine masks
Topic 7:
day good thank lol time like school need use im
Topic 8:
like got dont im na man time know ah really
Topic 9:
singapore people work like job dont sg country pay money



Topic Modelling

LDAMallet

```
# Display the topics
display_topics(lda, feature_names, no_top_words)
```

Topic 0:
gt singapore said cases police case covid19 public new source
Topic 1:
people dont think like good want pap make things say
Topic 2:
like car buy people road got bus use think really
Topic 3:
thanks read know people post like lol im dont comment
Topic 4:
years hdb yes price prices market money year pay new
Topic 5:
food like eat time good work looks nice chicken home
Topic 6:
people like covid dont mask wear need comments vaccine masks
Topic 7:
day good thank lol time like school need use im
Topic 8:
like got dont im na man time know ah really
Topic 9:
singapore people work like job dont sg country pay money



```
# Define the seed words for each topic
seed_words = {
    "Political": ["ge", "general electi",
    "Covid-19": ["covid-19", "infection",
    "Race & Religion": ["chinese", "mal",
    "Transport": ["breakdown", "train",
    "Relationships": ["relationships",
    "Crime": ["crime", "case", "police",
    "Housing": ["hdb", "price", "bto",
    "Education": ["student", "psle", "s",
    "Work": ["ot", "salary", "unemploye",
}
```

Limitations

Bias in the data – imbalance between subreddits

Singapore
Welcome to /r/singapore: The place for anything Singapore.

Created Jan 26, 2008

Public

1.5M
Members

344
Online

Top 1%
Rank by size ↗

SingaporeRaw
Welcome to r/SingaporeRaw Community. Public speech in Singapore is heavily regulated, where the average Redditors will often feel silenced and muted. r/SingaporeRaw is a safe haven for Redditors to share and post freely without fear of censorship. Feel free to share your thoughts, experiences, and opinions here, and we hope you have a good time !

Created Mar 6, 2019

Public

78K
Members

98
Online

Top 2%
Rank by size ↗

singaporehappenings
A community to share everything that's happening around Singapore! Stay connected and up-to-date with the latest events, news, and experiences in Singapore at r/singaporehappenings.

Created Sep 11, 2022

Public

43K
Members

17
Online

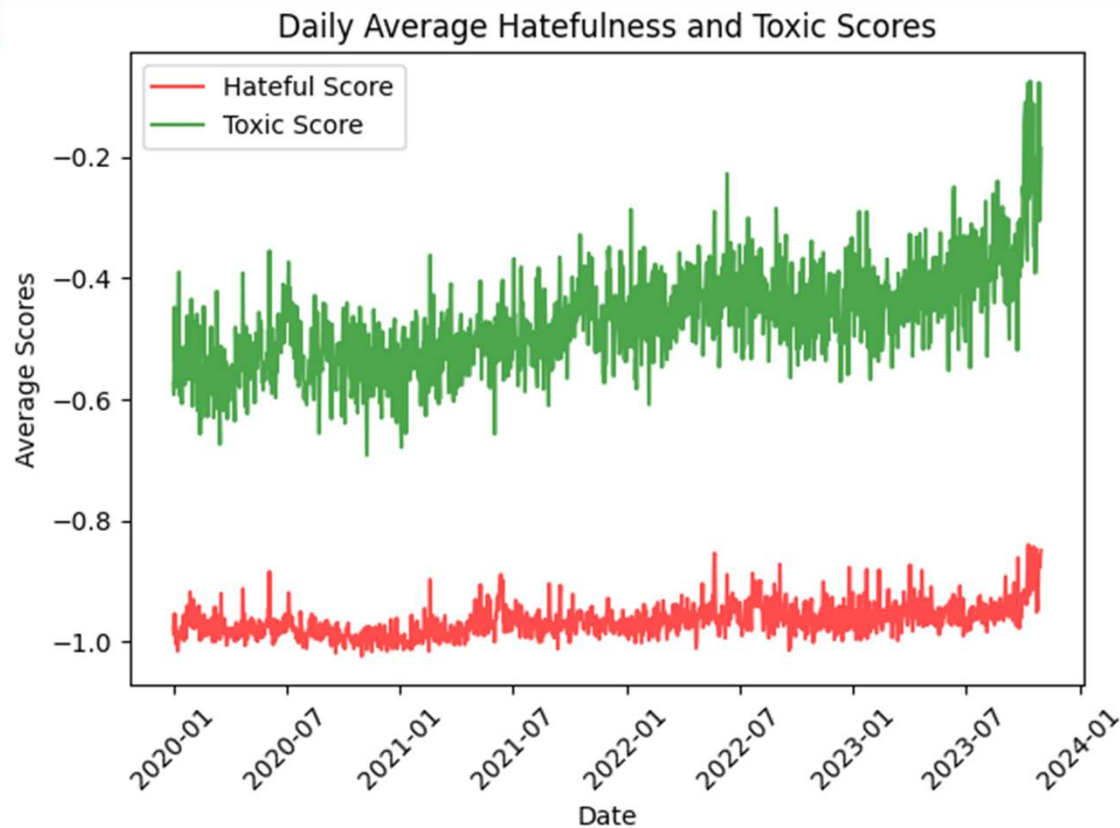
Top 3%
Rank by size ↗

Results





Trend over Time

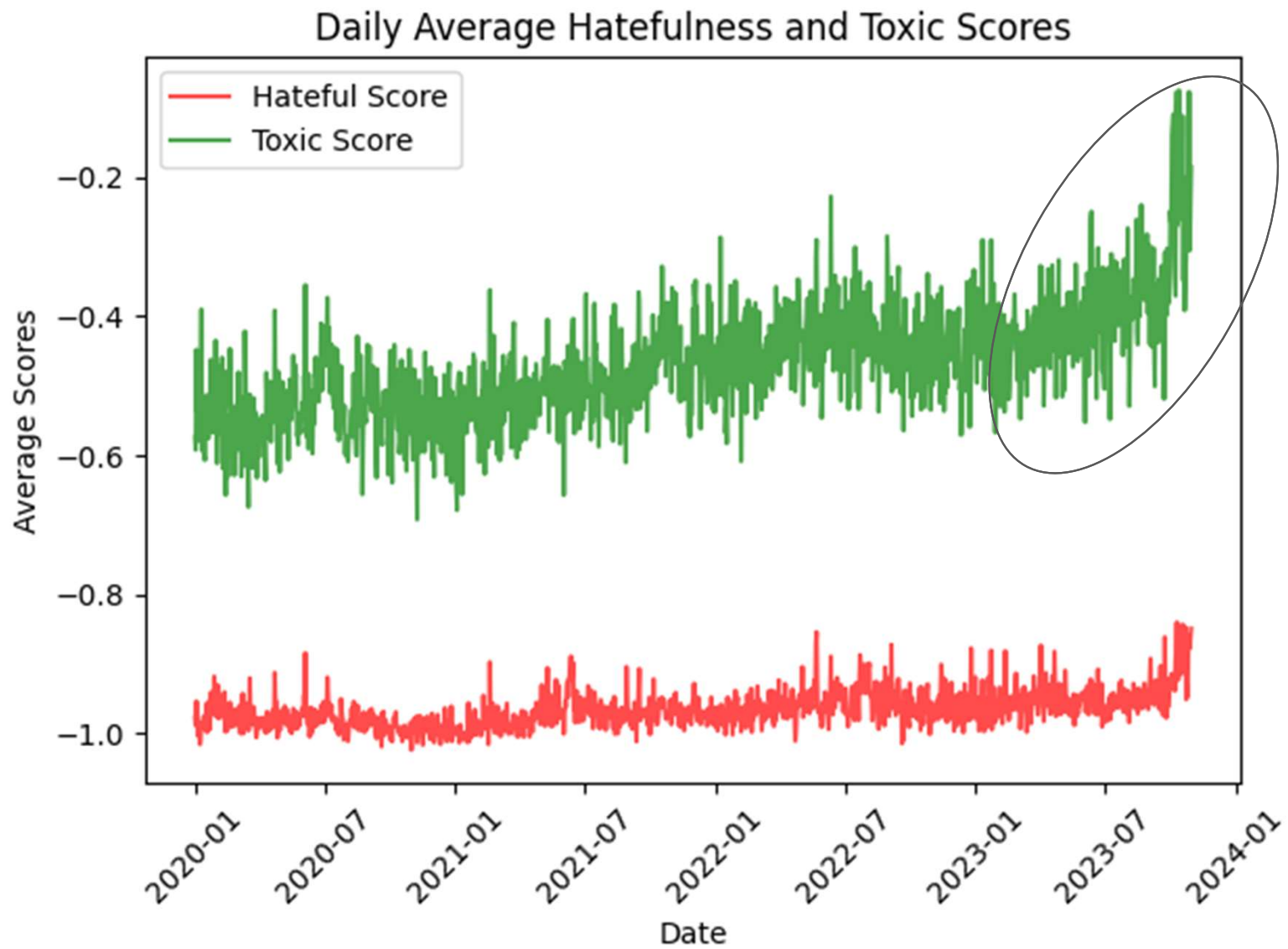


58%
Toxicity



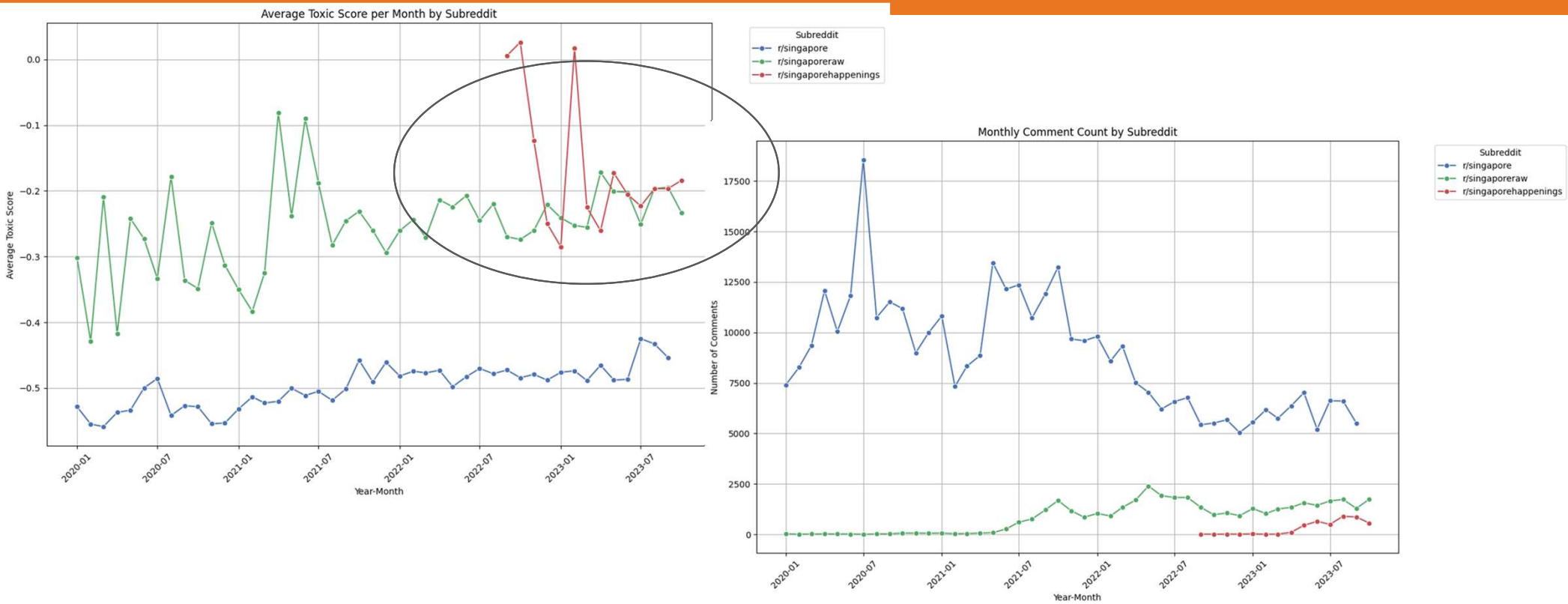
8%
Hatefulness





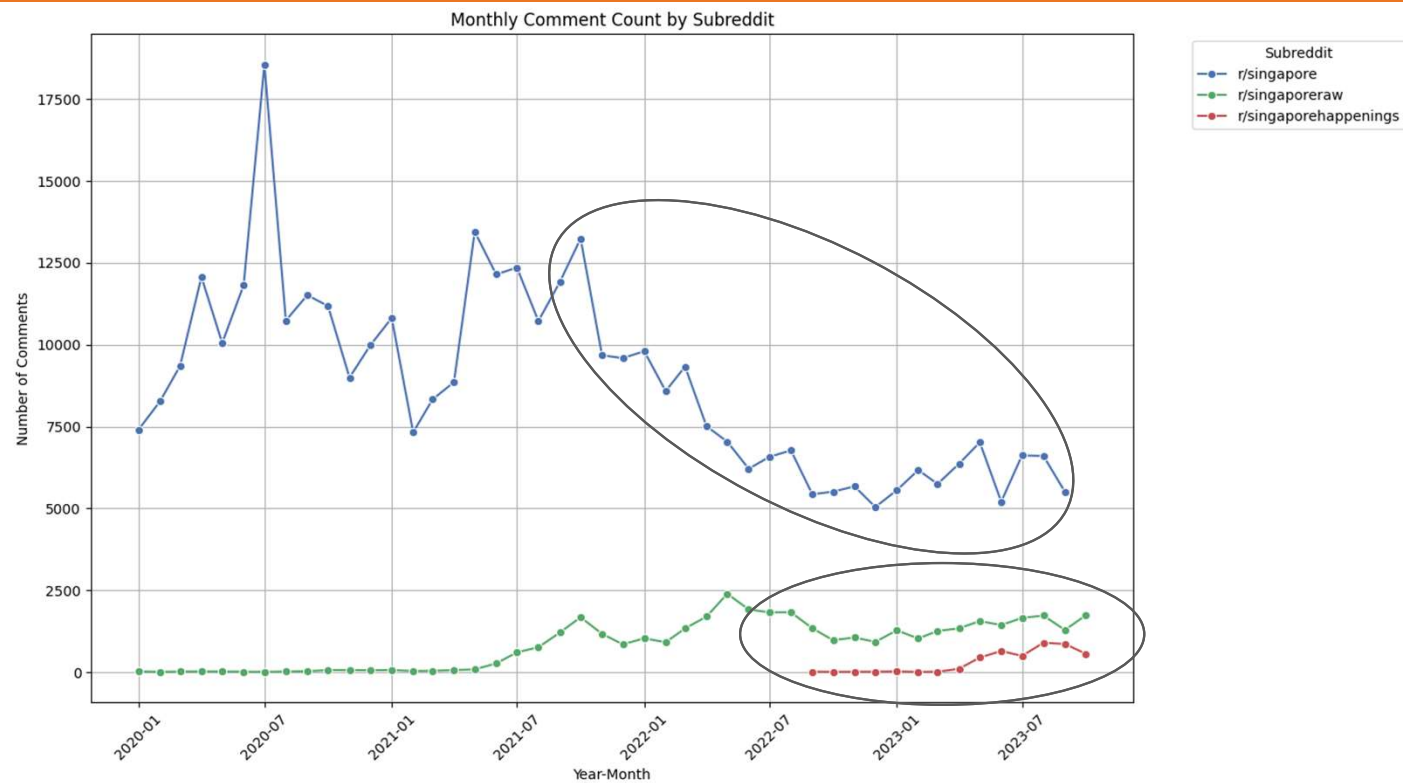


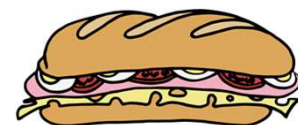
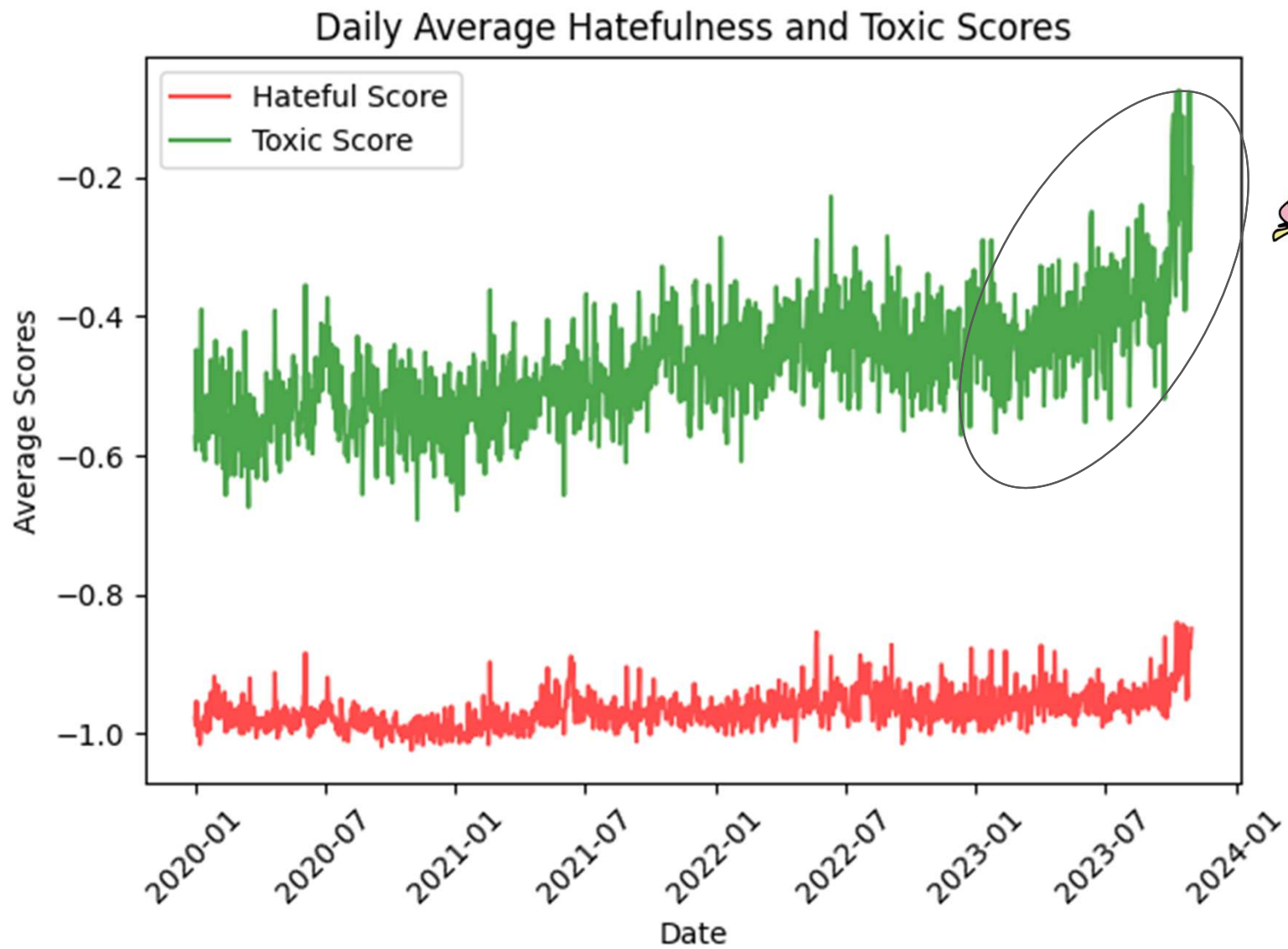
Trend Across Subreddits

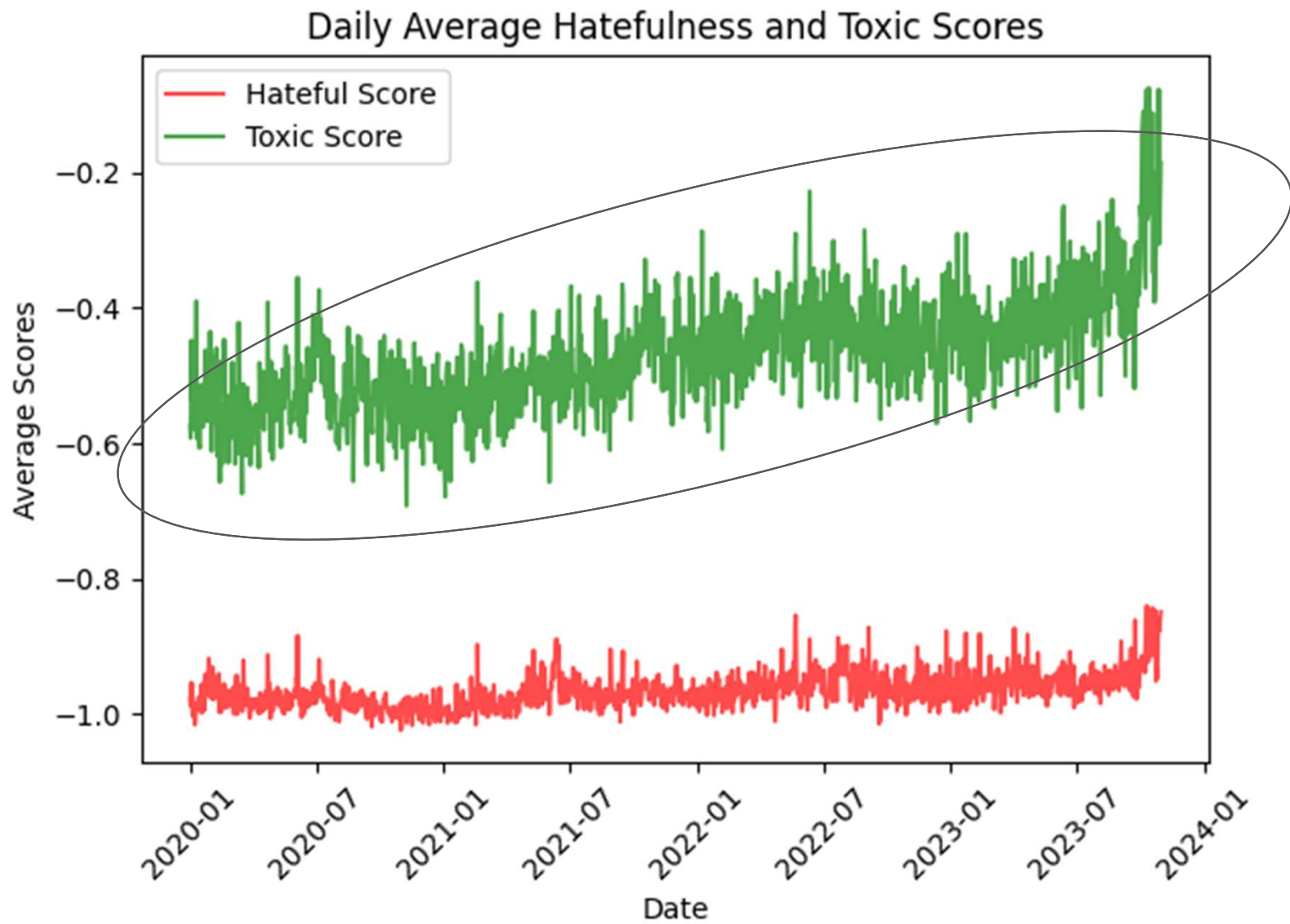




Trend Across Subreddits



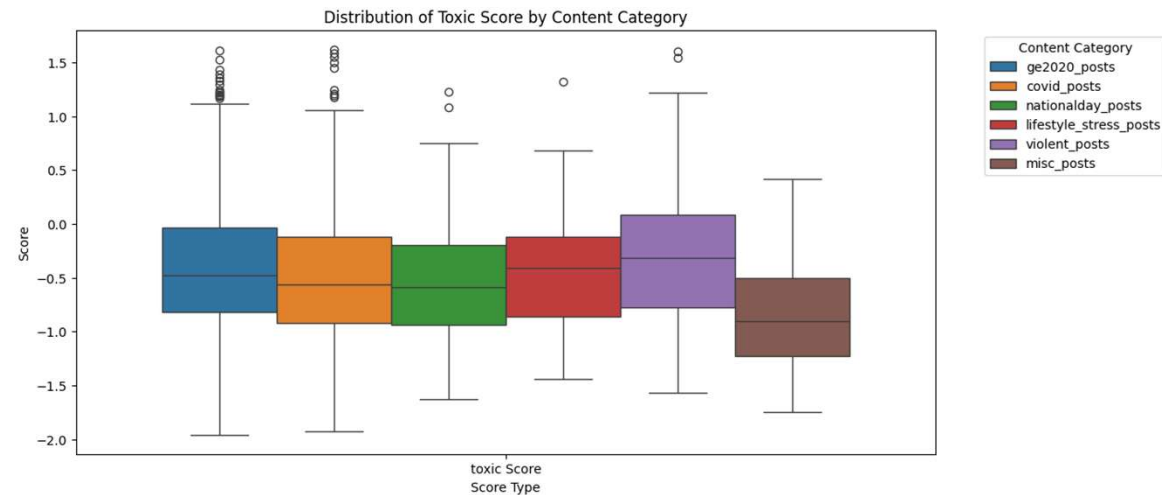
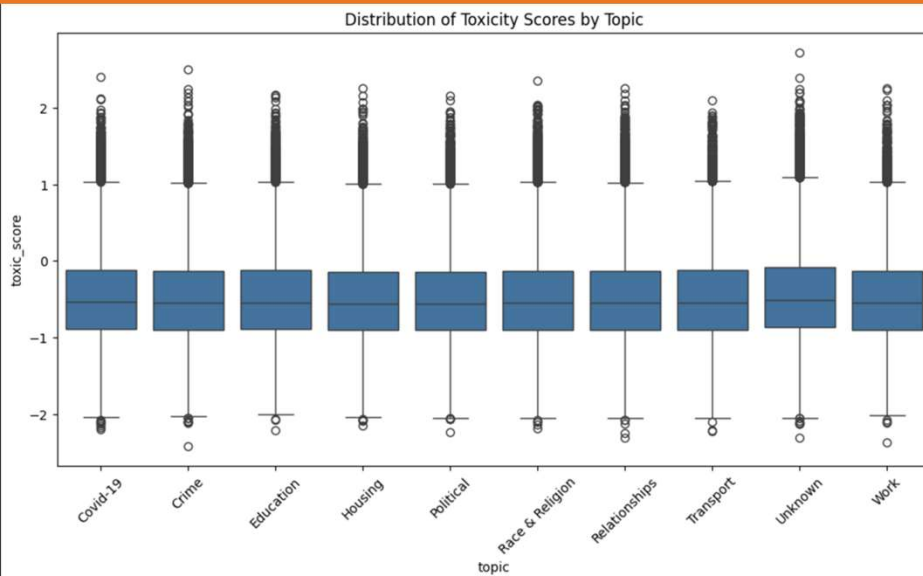






Trend Across Topics

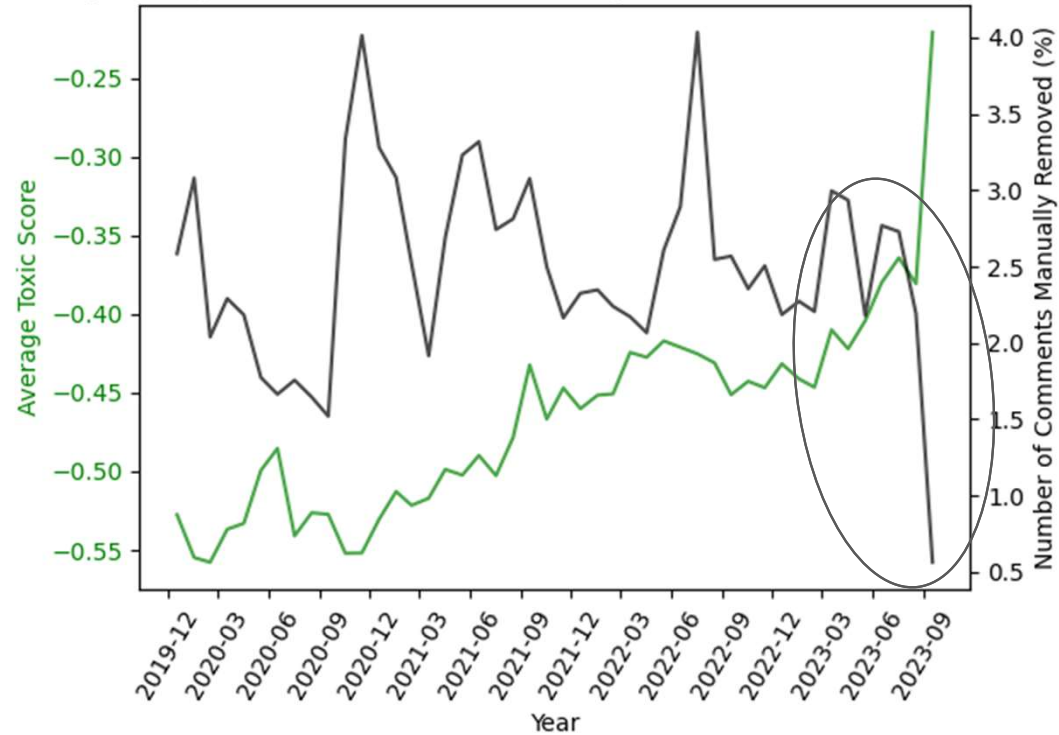
...





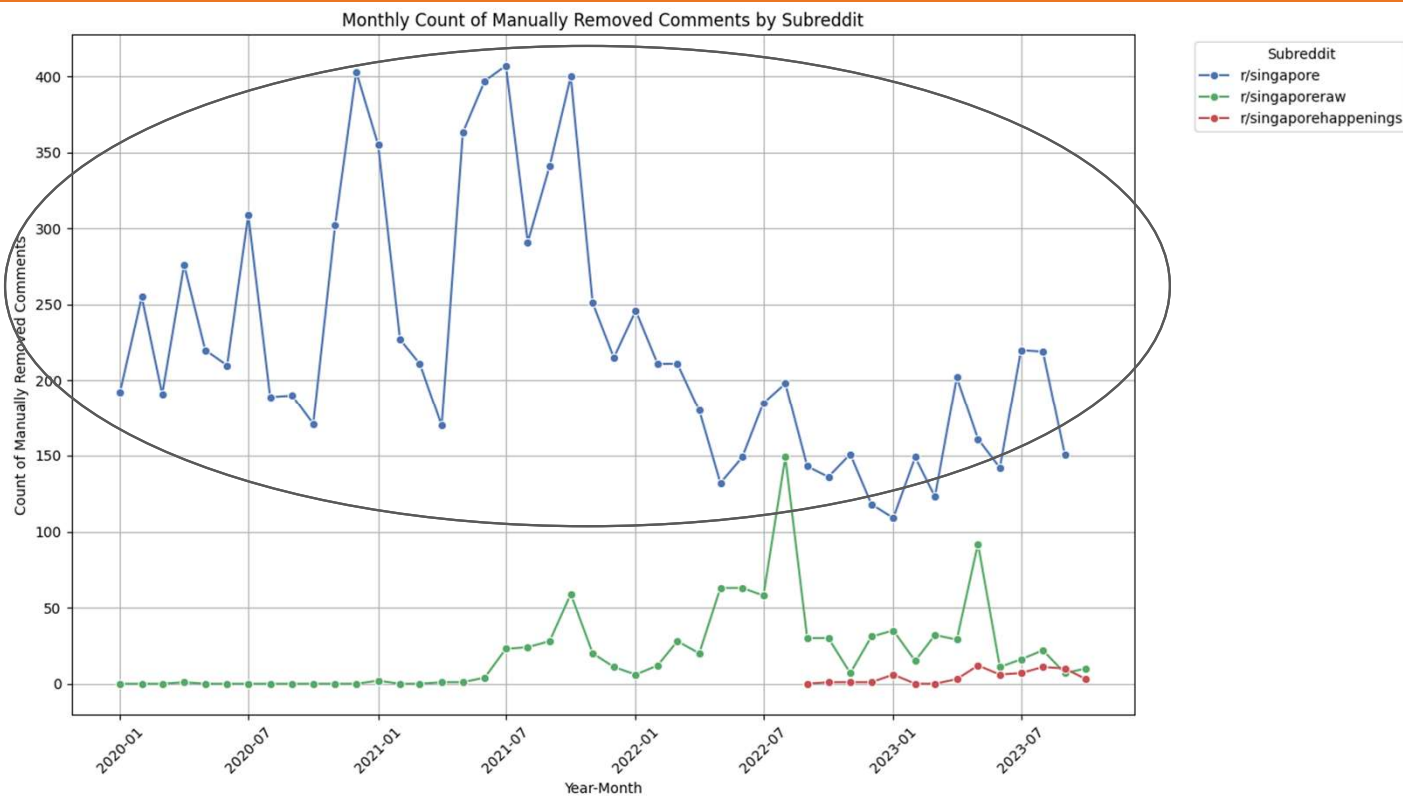
Trend of Moderation

Monthly Average Toxic Scores and Percentage of Comments Manually Removed

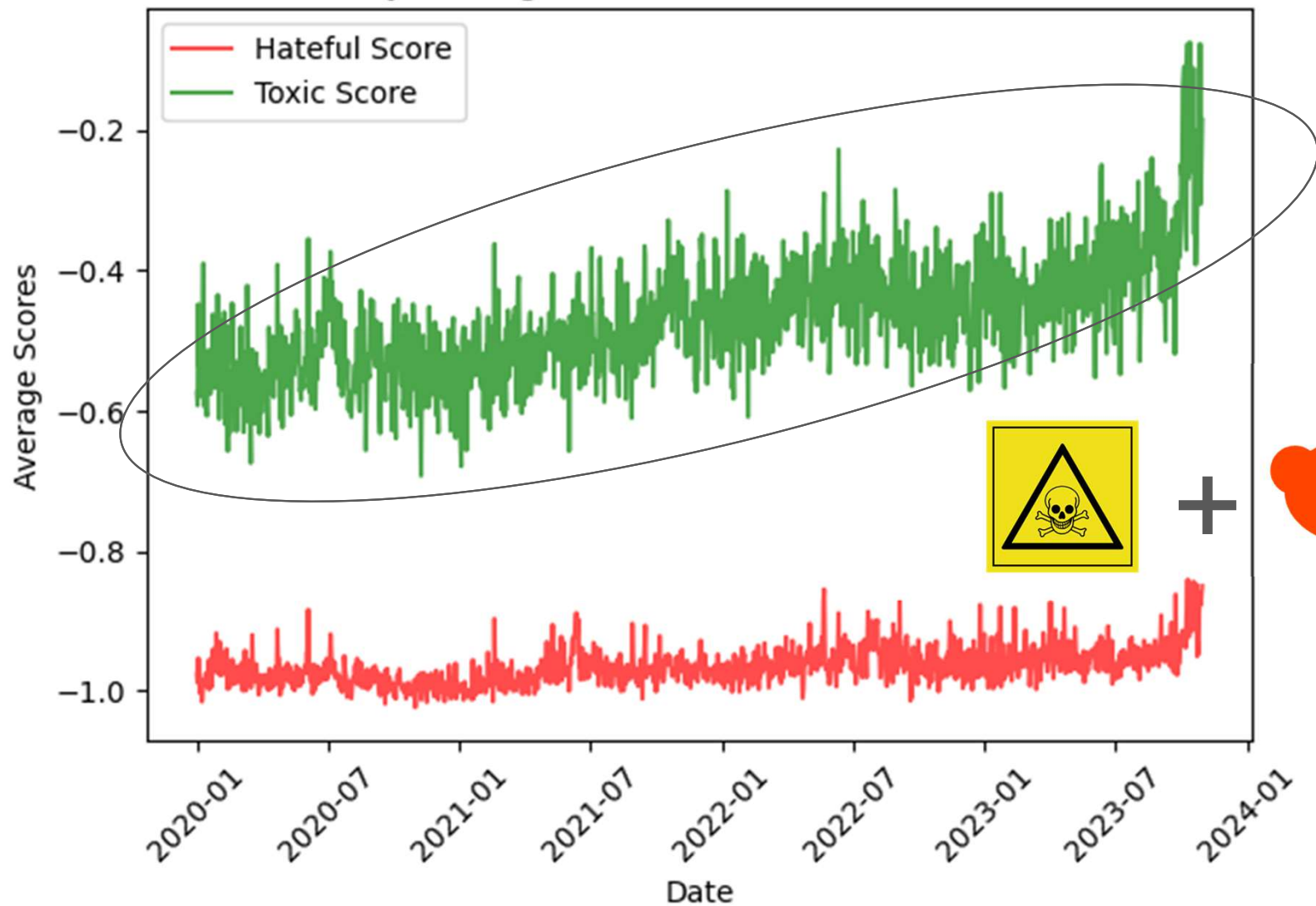




Trend of Moderation



Daily Average Hatefulness and Toxic Scores





What does this actually mean?

Our 2 Theories that explain these trends

2 Theories for 2 Types of Subreddits

r/Singapore	r/SingaporeHappenings & r/SingaporeRaw
Average toxicity increased.	Average toxicity generally higher, no recent increase.
Number of comments decreasing overtime.	Number of comments increasing overtime
Moderation has been decreasing overtime.	Moderation generally lower than r/Singapore to being with.

1

Limited moderation capacity due to Reddit's API Pricing changes

2

Rising popularity of inherently toxic subreddits



Theory 1: r/Singapore

Limited moderation capacity due to Reddit's API Pricing changes



April 19, 2023

Reddit announced API
will no longer be free

- Introduced a paid premium access point for third parties who require additional capabilities
- To prevent AI Companies from getting free data



Theory 1: r/Singapore

Limited moderation capacity due to Reddit's API Pricing changes

...



July 1 2023

\$0.24 per 1K API calls

- Intense backlash from Reddit's unpaid content moderators
- Key concern being the loss of third-party apps and moderation tools due to new costs



Theory 1: r/Singapore

Limited moderation capacity due to Reddit's API Pricing changes



- r/Singapore concerned about their mod tools



June 12 - 14, 2023

8,000+ subreddits participated in a blackout as protest



r/singapore • 1 yr. ago
mildfull MOD pang gang lo



On June 12, r/Singapore will be going dark in protest against Reddit's API changes which will kill 3rd party apps

NEW UPDATE

What's going on?

A recent Reddit policy change threatens to kill many beloved third-party mobile apps, making a great many quality-of-life features not seen in the official mobile app **permanently inaccessible** to users.



Theory 1: r/Singapore

Limited moderation capacity due to Reddit's API Pricing changes



r/singapore • 1 yr. ago

KeythKatz MOD East side best side

Blackout Reflection: How has /r/singapore going dark affected you?

Discussion

We're back! [r/singapore](#) was inaccessible for 48 hours from 12 to 14 June as part of the site-wide blackout protesting against imminent API changes. If this is the first time you're hearing about this (we know there are a lot of you), [/r/techsupport has a page with all resources](#), and [ELI5 has put up a succinct summary of the situation](#).

While we as moderators are not directly affected by the API changes for bots as our subscribers have generally been a good bunch, it has caused us to shelve plans to introduce our own bot that would have resulted in quicker and more effective moderation, and most of us use 3rd party clients, the lack of which after 30 June would result in less moderation. We also rely heavily on [r/toolbox](#) and [r/enhancement](#) for effective moderation, and while they should be unaffected for now, Reddit is sending a signal that third parties are not welcome and we are concerned for their future viability. We are also appalled at how Reddit has treated 3rd party developers and volunteers whom they rely on to keep Reddit running. They have repeatedly [slandered the developer of /r/apolloapp](#) and [doubled down in the recent AMA](#). Over the years, Reddit has promised an improvement to moderation tools on the native apps, yet has launched little that mods actually want.

Present

Moderation tools made unavailable





Theory 2: r/SingaporeRaw & r/SingaporeHappenings

Rising popularity of inherently toxic subreddits

SingaporeRaw

Welcome to r/SingaporeRaw Community. Public speech in Singapore is heavily regulated, where the average Redditors will often feel silenced and muted.

r/SingaporeRaw is a safe haven for Redditors to share and post freely without fear of censorship. Feel free to share your thoughts, experiences, and opinions here, and we hope you have a good time !

Created Mar 6, 2019

Public

78K

Members

74

Online

Top 2%

Rank by size ↗

singaporehappenings

A community to share everything that's happening around Singapore! Stay connected and up-to-date with the latest events, news, and experiences in Singapore at r/singaporehappenings.

Created Sep 11, 2022

Public

43K

Members

18

Online

Top 3%

Rank by size ↗

- Emphasis on creating an uncensored space for redditors.
- Both subreddits have far fewer rules than r/Singapore



Theory 2: r/SingaporeRaw & r/SingaporeHappenings

Rising popularity of inherently toxic subreddits



Add flair and tags

r/singaporehappenings flair

Search

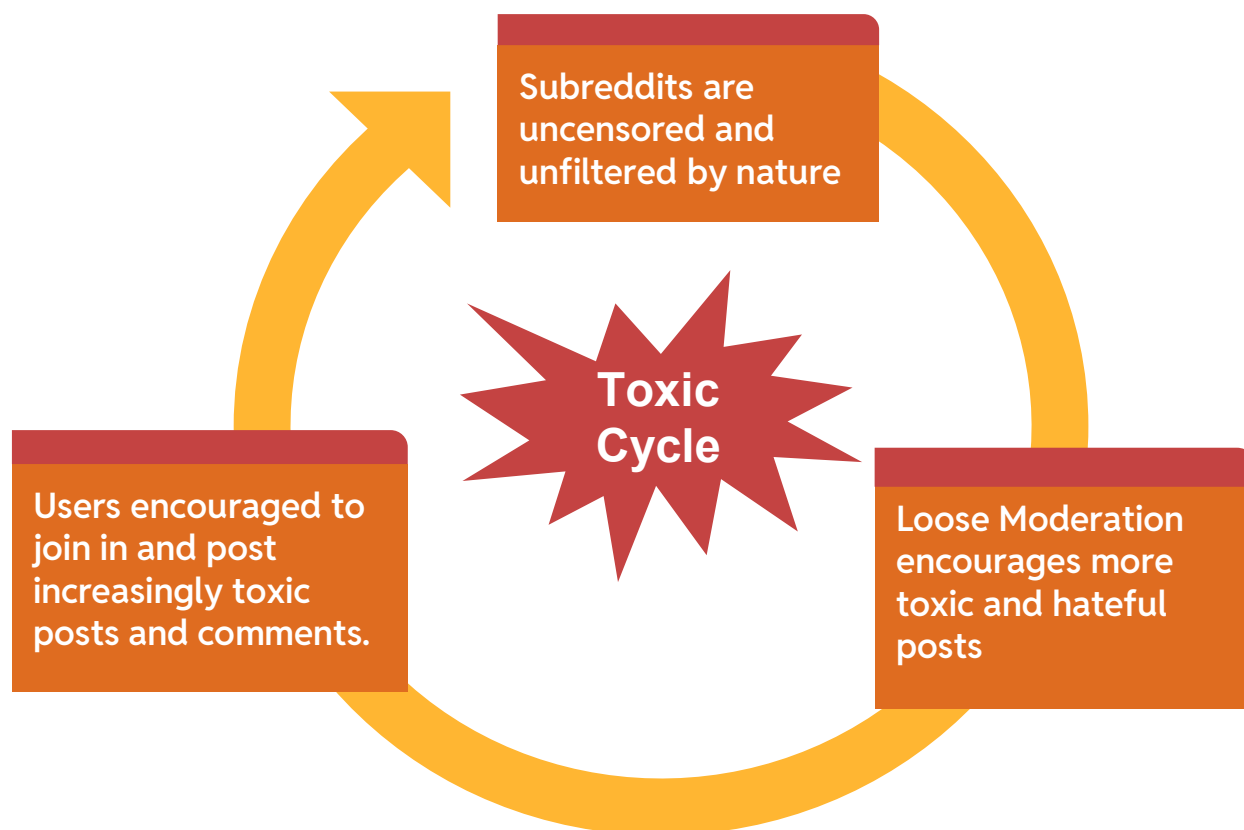
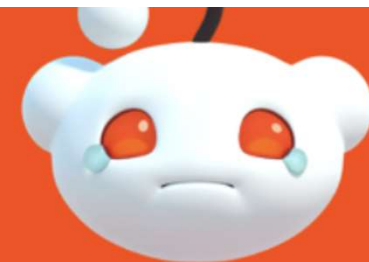
- ☒ No flair
- ☐ Kind Human
- ☐ Shocking
- ☐ Yishun
- ☐ Karens
- ☐ Viral News
- ☐ Question
- ☐ What The F***
- ☐ Parliament Discussion
- ☐ Opinion
- ☐ Entertainment
- ☐ Funny
- ☐ Political Shit
- ☐ Elections
- ☐ No Context Video

- Flairs are like hashtags for posts
- Describes the contents



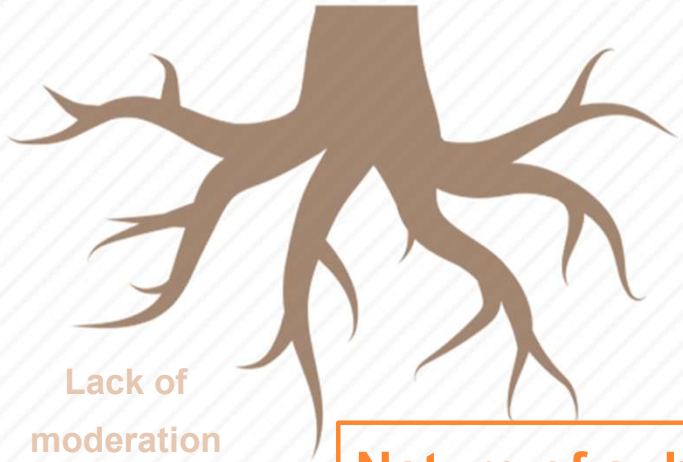
Theory 2: r/SingaporeRaw & r/SingaporeHappenings

Rising popularity of inherently toxic subreddits



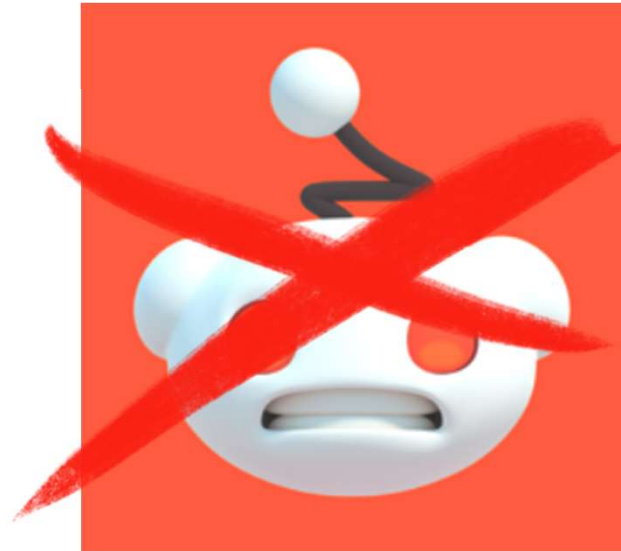
Recommendations





Lack of
moderation

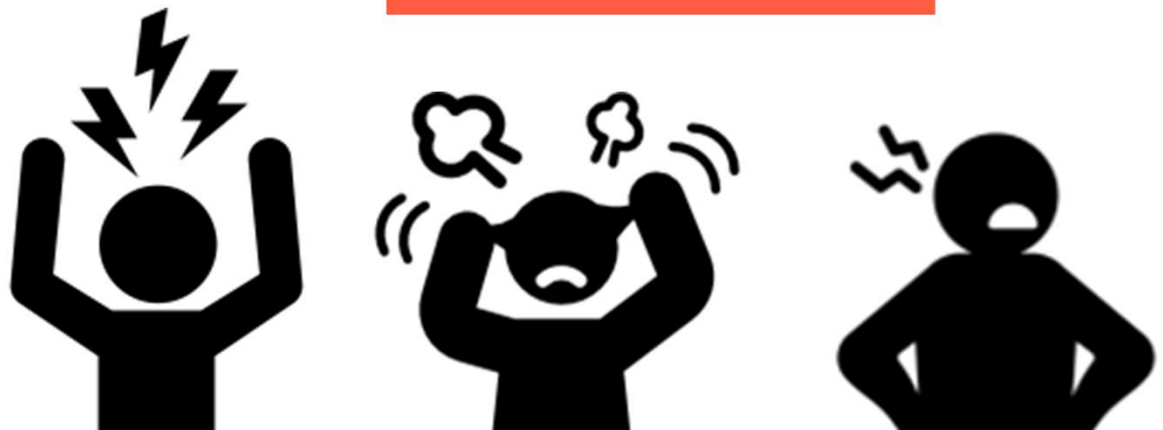
Nature of subreddit



Freedom of speech, assembly and association

14.—(1) Subject to clauses (2) and (3) —

- (a) every citizen of Singapore has the right to freedom of speech and expression;
- (b) all citizens of Singapore have the right to assemble peaceably and without arms; and
- (c) all citizens of Singapore have the right to form associations.



Recommendation 1:

Moderator tool powered by singlish/local context LLM



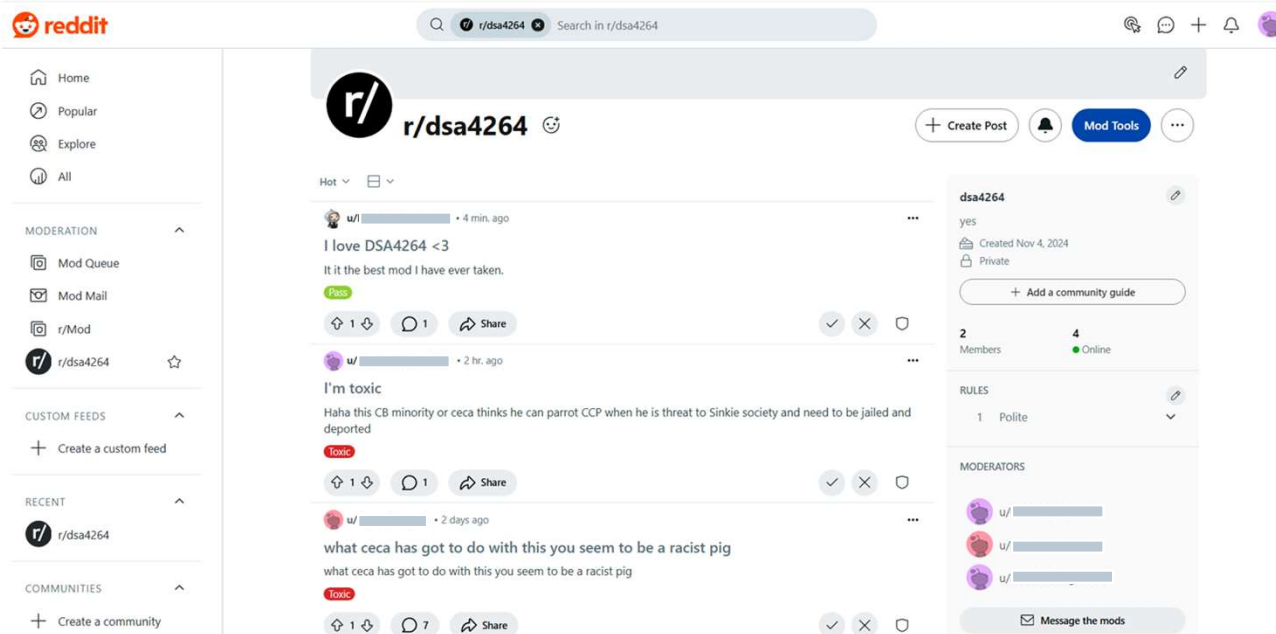
- 1 Calibrate LLM for toxic and hateful scores
- 2 Seek approval of moderator tool from Reddit
- 3 Moderators adjust threshold for hatefulness and toxicity for subreddit

3 Recommendation 1:

Moderator tool powered by singlish/local context LLM



Posts:



● **High toxicity and hatefulness**

Suggested to delete

● **Low toxicity and hatefulness**

Auto approved

Slide 38

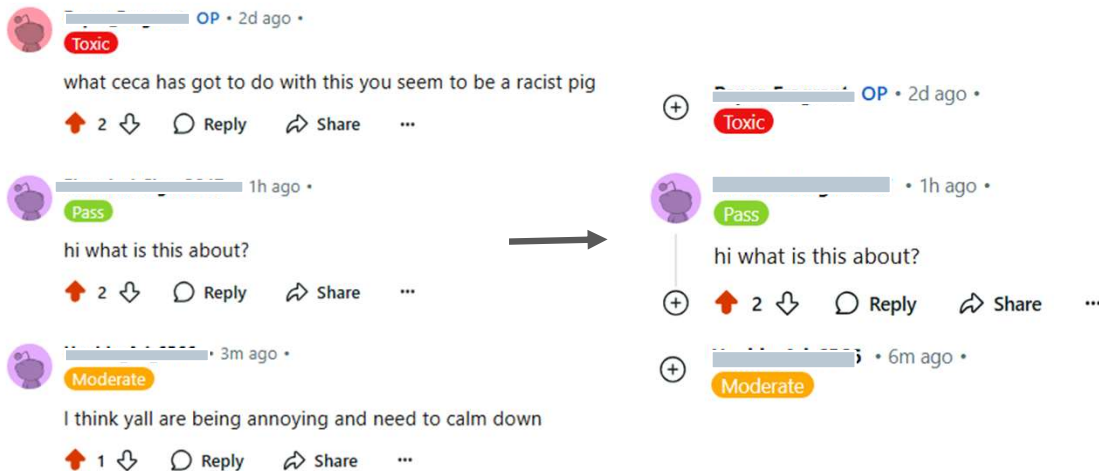
- 1** oh wdyd by more realistic?
Sarah Goh, 13/11/2024
- 2** U can see the new img HAAHHA
Denise Teh, 13/11/2024
- 1** @sarah I edited the post thing to be light mode & look more realistic
Denise Teh, 13/11/2024
- 3** I added the post flair idk if comments can add onot tho
Denise Teh, 13/11/2024

Recommendation 1:

Moderator tool powered by singlish/local context LLM



Comments:



Toxic
what ceca has got to do with this you seem to be a racist pig
2 0 Reply Share ...

Pass
hi what is this about?
2 0 Reply Share ...

Moderate
I think yall are being annoying and need to calm down
1 0 Reply Share ...

Toxic
+ 0 0 Reply Share ...

Pass
+ 2 0 Reply Share ...

Moderate
+ 0 0 Reply Share ...

- High toxicity and hatefulness**
Collapsed and moderators alerted
- Moderate toxicity and hatefulness**
Auto collapsed
- Low toxicity and hatefulness**
No action needed

Recommendation 2:

Community education for parents and children



- **Parents**

- Restrict children's access to negative online spaces
- Parental controls

- **Children**

- Online etiquette and responsible technology use
- Avoid engaging in hateful and toxic online spaces



Moving Forward

- Replicability to other subreddits
- Scalability to evaluate other aspects

Overall: **Fail** (100/100)

Harassment: Pass

Hateful: Pass

Public Harm: Pass

Self Harm: **Fail**

Sexual: Pass

Toxic: **Fail**

Violent: Pass

- Extension to other social media sites



Thank you!

FEEL FREE TO APPROACH US
IF YOU HAVE ANY QUESTIONS.



(Reddit Logo)