

# Cleanity (Clean Community)

20180532 Heechan Lee

20200433 Gyuho Lee

20210277 Hojun Park

## Introduction

# Homo Sapiens with Digital World

Humans were social animal from old times. The difference between now and then is that we are spending more time on virtual community.




Hi-Tel (Bulletin Board System(BBS))



## Introduction

# And these Communities can be great...

It's entertaining, useful and fun! Sometimes, it can show strong unity.

**익명**  
10/22 22:04

**PL 조진듯**




교과서의 모든 글자가 "팬파인애플펜" 이런 식으로 적혀있음



p! 들은 사람이면 원소린지 알 텐데 모든 단원이 다 이런식임:

파인애플을 펜에다가 넣었는데 그걸 또 애플펜에다가 쫓으면 펜파인애플애플펜이 됨 그리고 펜파인애플펜에 펜파인애플애플펜을 적용하면 펜파인파인애플애플애플펜이 된다. 근데 비슷한 펜애플파인애플애플파인펜이 있는데 사실 두 개가 똑같으니까 형식을 펜파인파인애플애플애플펜으로 통일한다. 여기서 퀴즈! 펜을 애플 두 개에 넣으면 애플펜애플이 될까요 애플애플펜이 될까요 정답은 파인애플애플입니다 왜냐하면 어찌구저찌구

교수님 강의 들으면 그래도 대충 파인애플이랑 애플이랑 펜이랑 잘 조합하면 펜파인어찌구 만들 수 있을 것 같아서 교과서랑 기술 보면 엉뚱한 탄소리 하고 있음 분명 펜은 나오는데 그걸 수박이랑 물통에 꽂아서 수박물통수박펜을 만드시고 증명하세요. 수박물통수박펜으로 양꼬치를 만드시오.

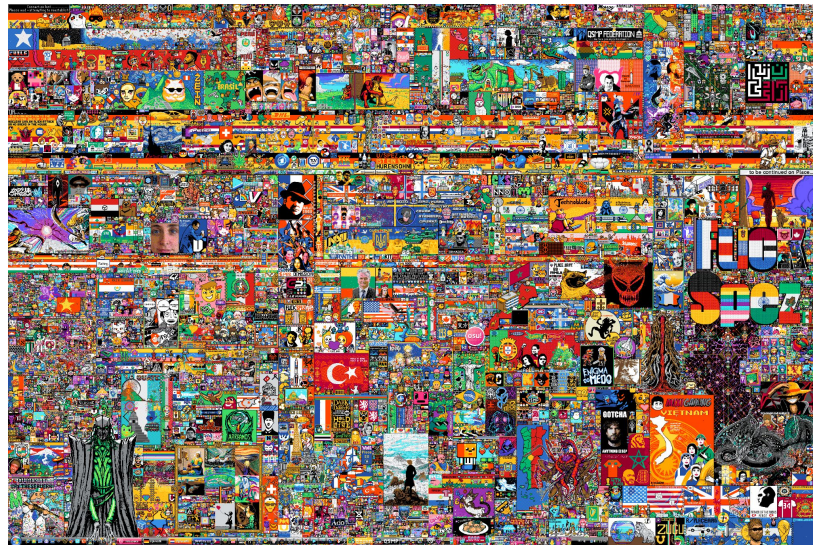
이럴 땀글 쓸 시간에 기술을 하나라도 더 봐야 될 조지는데 지금 집중력 개박살나서 펜파인수박물통애플펜을 다시 조우할 자신이 없다

 463  48  114

 공감  스크랩

← Recent KAIST  
'everytime' post  
complaining about  
the difficulty of  
'Programming  
Language' course

Well known online →  
community event of  
'reddit', r/place.  
Users can color a pixel  
and they made a whole  
map of countless arts.



## Introduction

### But sometimes people can be toxic while communicating with others

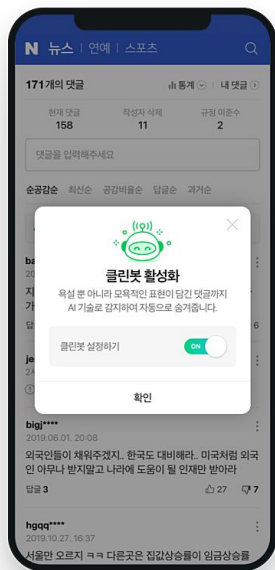
Even if it's not toxic, too much abusive languages, spamming and trash talk can be problematic.



## Introduction

Even though there are some programs that detects inappropriate chats...

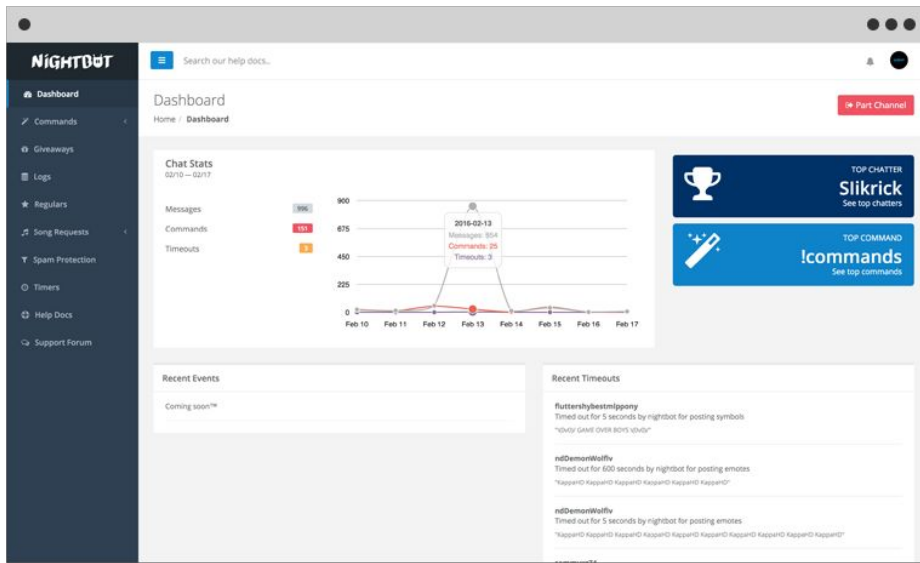
It's not that effective. (Only good at catching spams, ambiguous standard, etc.)



클린봇 설정 레이아웃



가사 하단 댓글리스트



Youtube/Twitch NightBot

## Introduction

### Even though there are some programs that detects inappropriate chats...

It's not that effective. (Only good at catching spams, ambiguous standard, etc.)

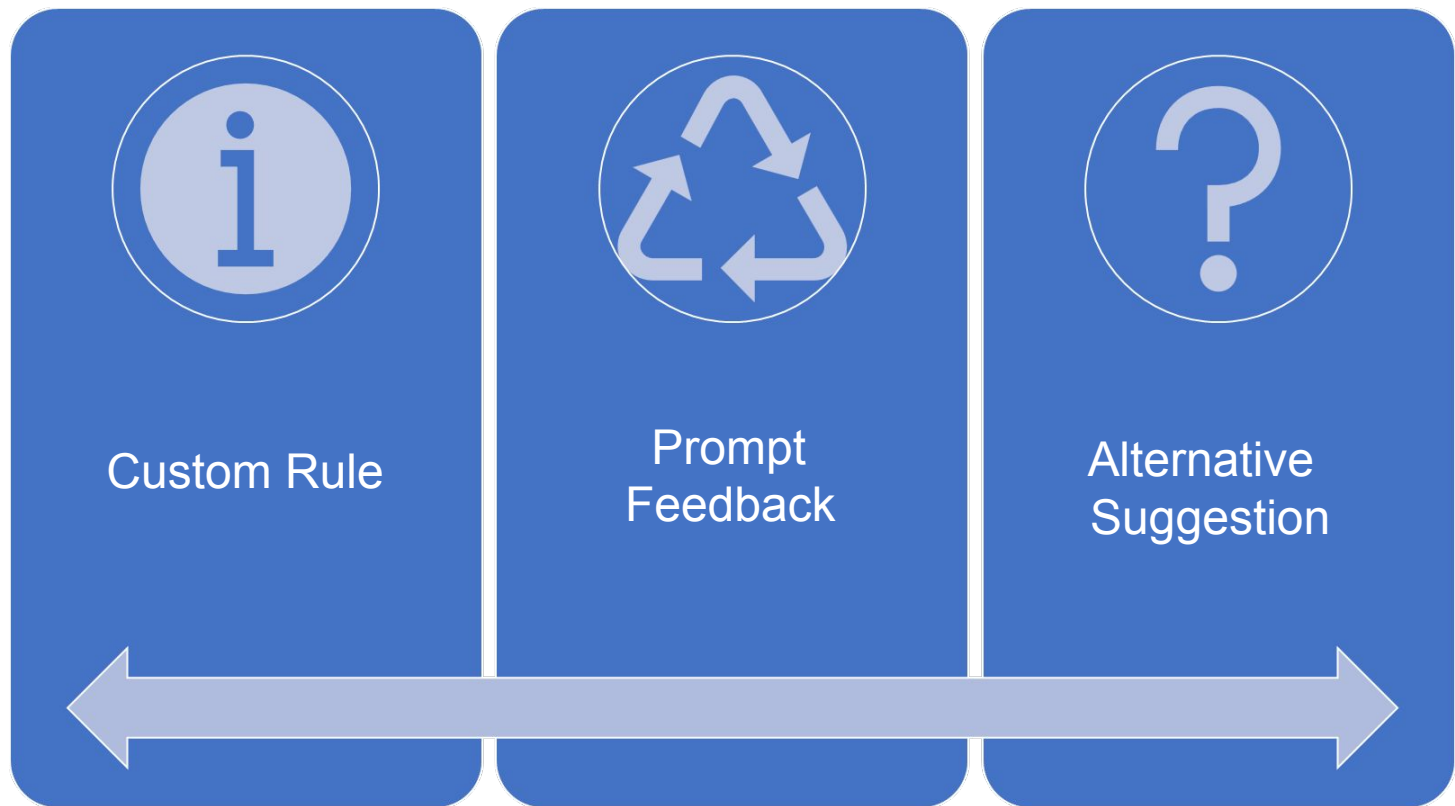
We all have an experience about...

- Because the rule is too strict, sometimes we can't even type normal words. (Scunthorpe Problem)
- Writings that spelling and grammar are so wrong, it's unreadable.



## Introduction

**This is why we want to supplement the original chat 'clean bots'.**



## Approach

### 1. Custom Rules for Each Community's Preference

The rules can vary community by community; There's no rule for all cases.

#### General Baseball Community

Allow slandering of players

Allow talking about all teams



#### Community only for specific baseball team

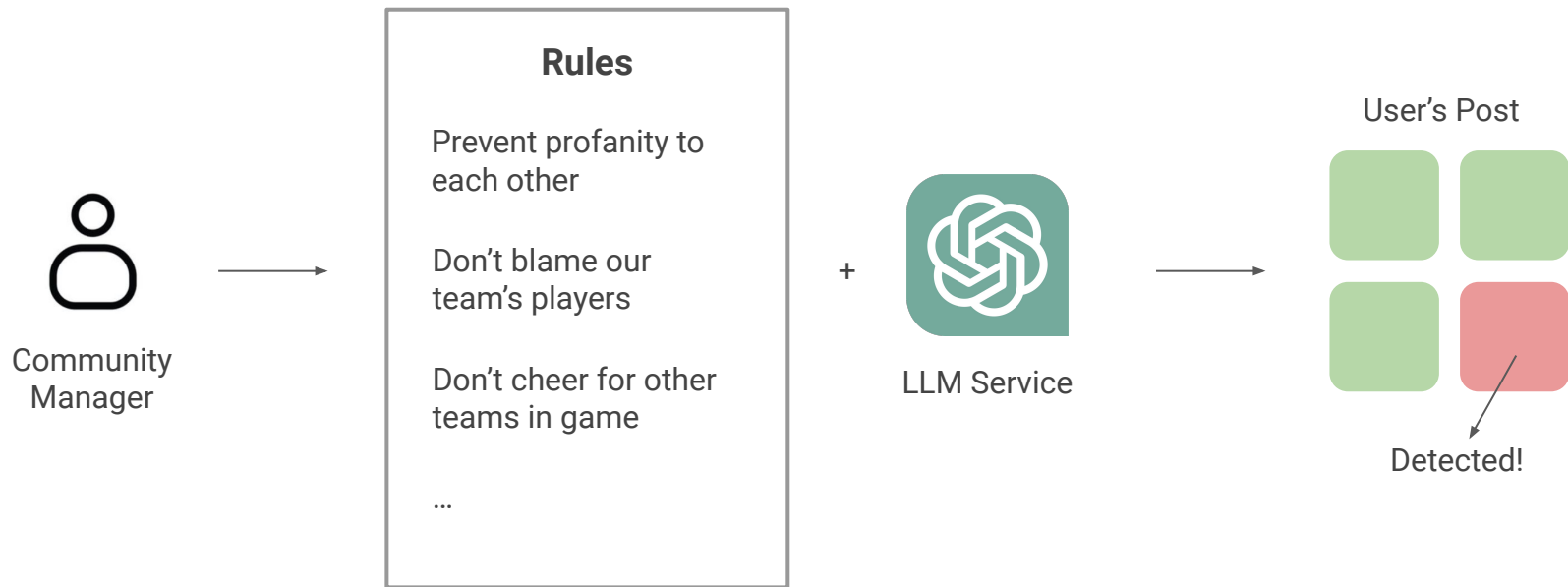
No slandering of players

Don't cheer for the current opponent



## Approach

### 1. Custom Rules for Each Community's Preference



Take advantage of LLM service that can process natural languages flexibly than rule-based system

Approach

## 2. Prompt Feedback and User Tendency Review


User's Input

Hey M#\$%###@!!



After Checking

Hey M#\$%###@!!



It violates "No Swearing"



Giving prompt feedback

Also helps users to be aware of community rules

Approach

## 2. Prompt Feedback and User Tendency Review

User's Input

Hey M#\$%###@!!



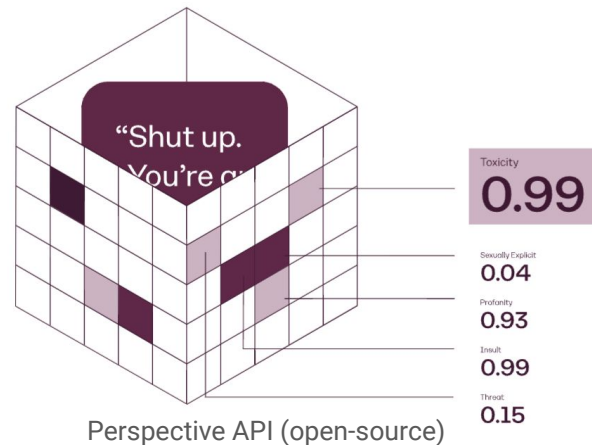
After Checking

Hey M#\$%###@!!



It violates "No Swearing"

Giving prompt feedback  
Also helps users to be aware of community rules

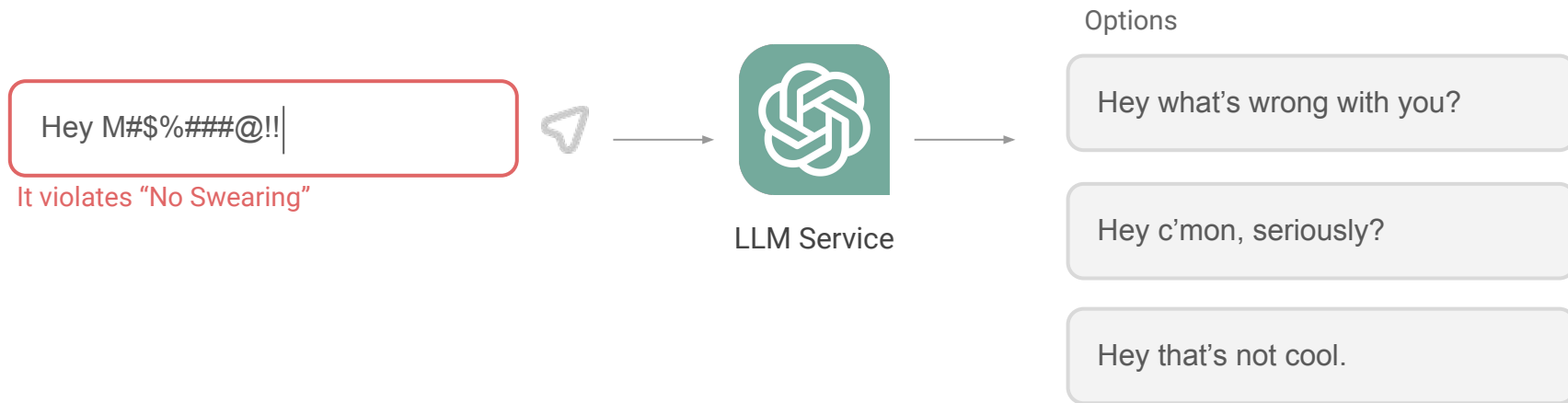


Show the toxicity degree in several aspects

After getting a few comments from users,  
We can review user's tendency and give feedback

## Approach

### 3. Alternative Suggestion

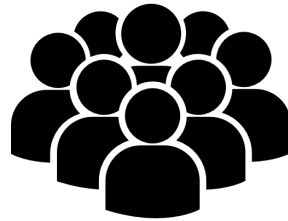


Based on the rules, we can suggest some alternative expressions that can be allowed

## Conclusion

We expect the impacts of our project as follows;

- 1) Detect Violation Based on Custom Community Rules by Natural Language
- 2) Proper Feedback and Make Users Be Aware of Community Rules
- 3) Suggest Mitigated Comments that Can Be Accepted By Rules



***Contribute to Make More Healthy Online Discussion***