



100%



#바글바글

Chat GPT를 이용한 댓글 순화 서비스

Comment purification service using Chat GPT





100%



갈수록 심각해지는 **악플**

사람마다 **기준**이 다른데 모두 검열하는게 맞아?



바글바글

Excutive Summary



Chat GPT를 이용해 댓글을 **개인 맞춤형**으로 **순화**해주고

필요하다면 **조언**과 **설명**도 해주자!





Agenda

Intro

**Solution
Overview**

Demonstration

**Future
Plans**

Outro





Meet the Team



조수환
프로젝트 리더
백엔드



안성찬
기능구현
UI/UX



김종한
데이터베이스
UI/UX



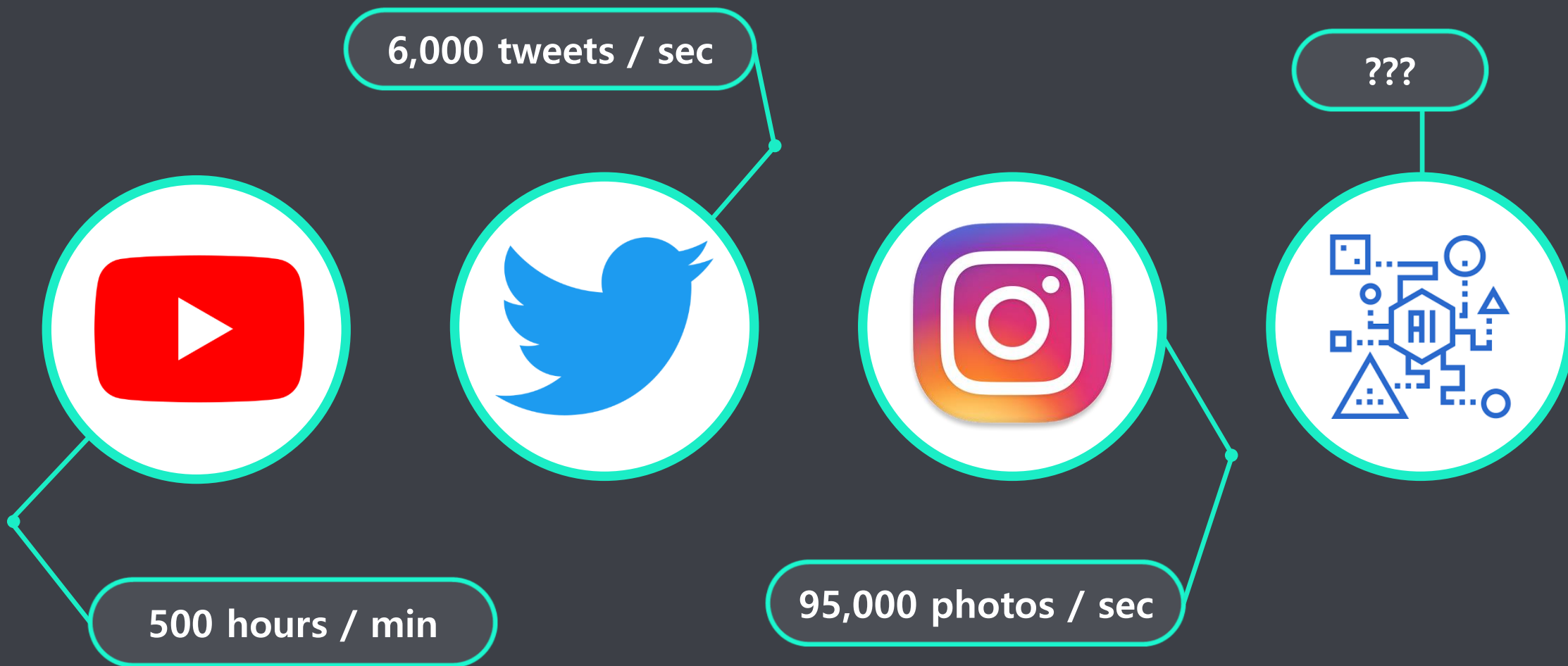
정우성
기능구현
파이프라인



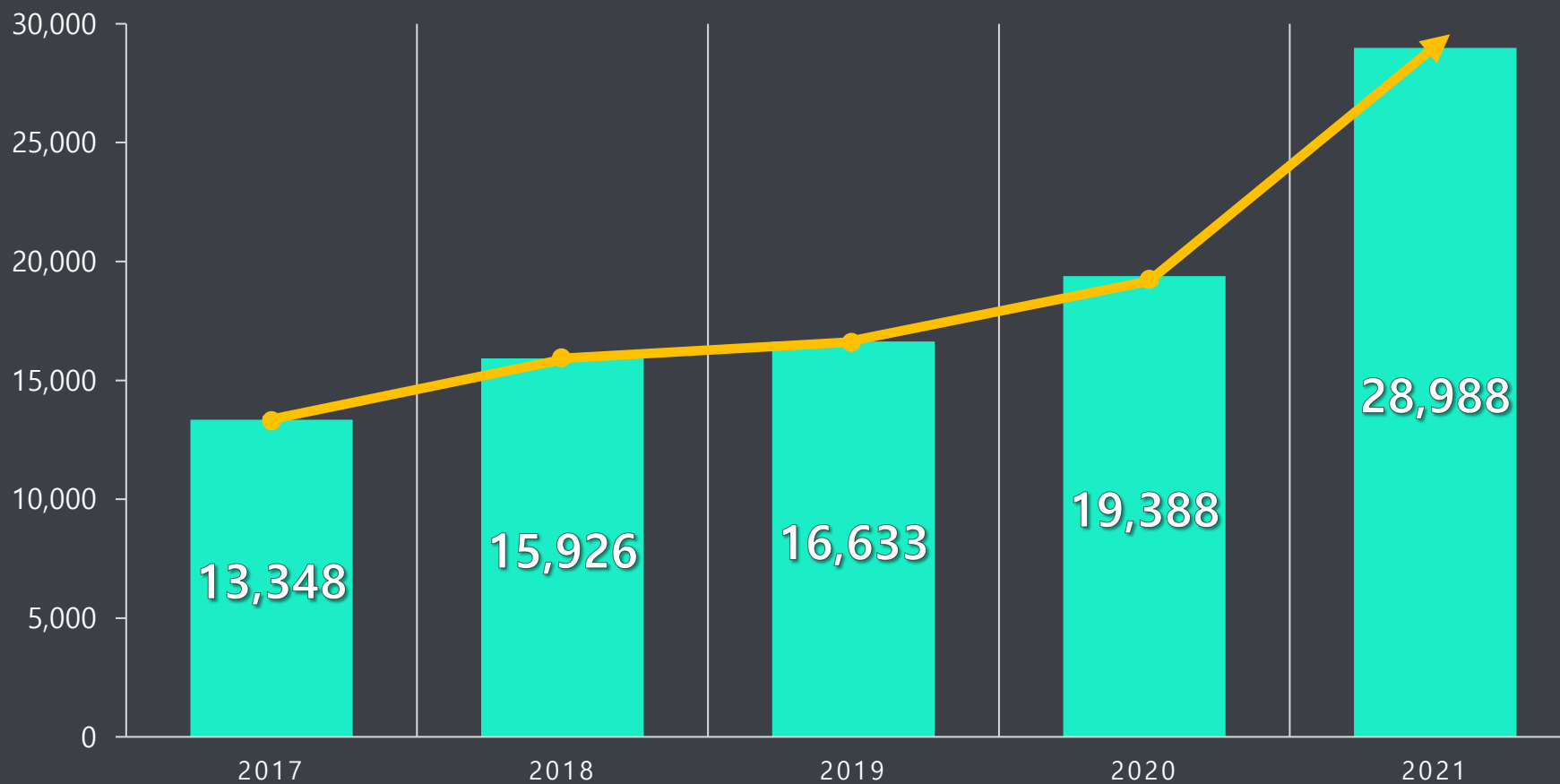
김만서
기능구현
발표

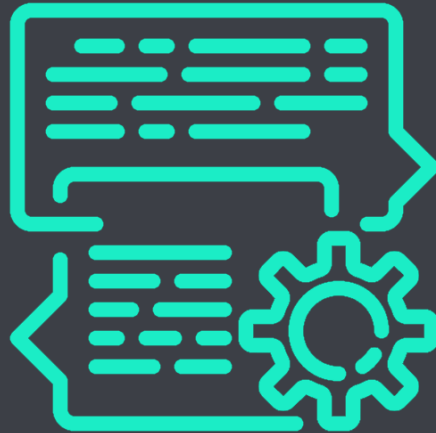
Intro

(문제정의)



연도별 사이버 명예훼손모욕 발생건수





바글바글?

바르게 글쓰고
바르게 글읽기

Solution Overview

(솔루션 개요)

기존 검열 시스템의 단점



네이버
클린봇

① 클린봇이 부적절한 표현을 감지한 댓글입니다.



카카오
세이프봇

세이프봇이 자동으로 가린 댓글입니다.
해당 영역을 클릭하시면 댓글을 보실 수 있습니다.

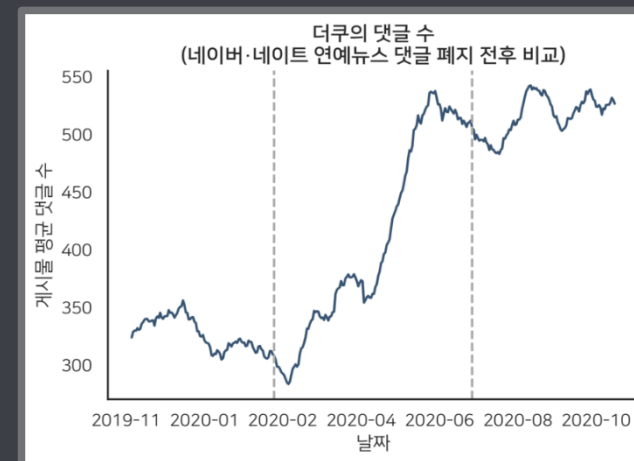
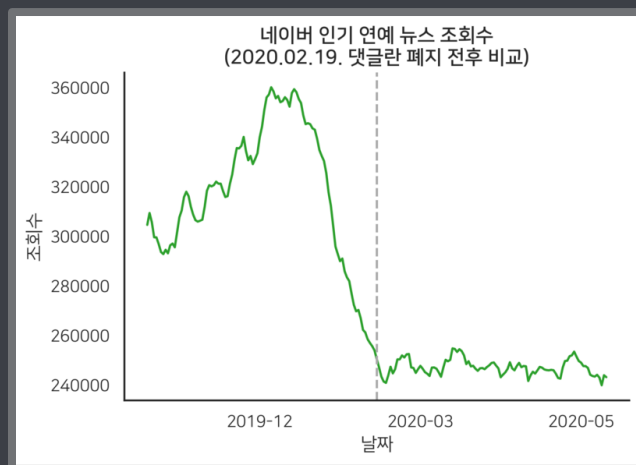


인권 보호

검열

표현의 자유는?

2020.02 네이버 연예 뉴스 댓글란 폐지 전후 비교



무차별적 검열과 삭제는 문제 해결에 도움이 되지 않는다

'Chat GPT'를 이용한 댓글 순화

인간 피드백 기반 강화 학습을 통해 발전
-> 댓글의 의미를 파악하는데 유리



설문조사 및 인터뷰

Survey
106

Age

20~29 : 88.7%
30~39 : 5.7%

Interview
11

설문지



Google Forms



안녕하세요.

저희는 AI+X 선도인재양성프로그램 중급 수업을 수강하고 있는 **바글바글** 팀입니다.

저희는 ChatGPT API를 이용한 챗글 관련 서비스를 제작하고 있습니다.

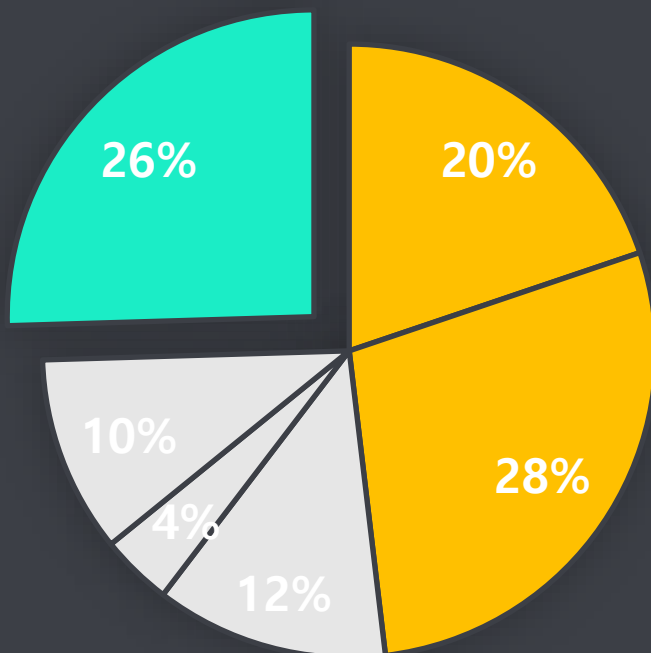
진행함에 있어 챗글에 대한 인식과 챗글 서비스의 기능 수요를 통해 더 나은 챗글 경험을 제공하기 위해 여러분의 소중한 의견이 필요합니다. 질문지에 최대한 솔직하게 답변해주세요.

설문의 정확성을 위해서 중립적인 답변을 최대한 피해서 답변해주시면 감사하겠습니다 ^—^

Question

“이 선수는 쓰레기야 왜 이렇게 못하냐?”

같은 댓글이라도 사람마다 순화가 필요한 단계는 다르다

**A**

비난적 표현
(단어) 순화

48%**B**

부정적 의도
(비난) 순화

26%**C**

순화 필요 없음

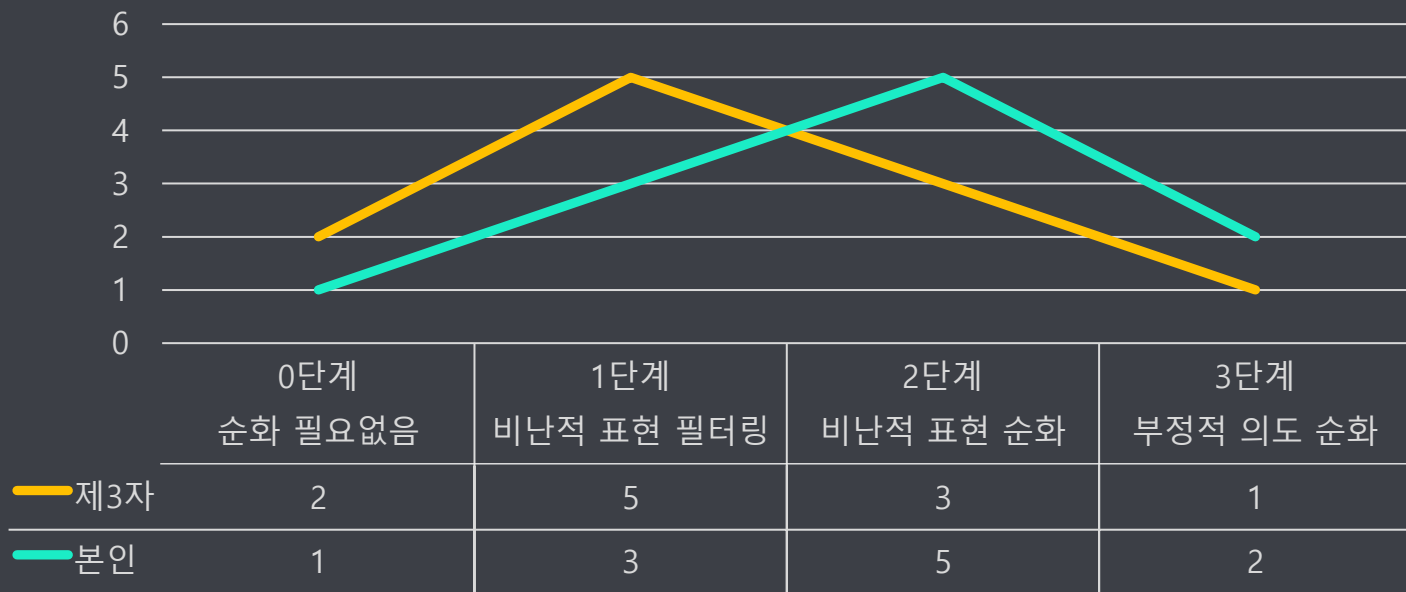
26%

“이게 악플이라고 생각하지 않는다.
비슷한 댓글을 받은 적 있는데 더 열심히 하는 계기가 되었다.”

-인터뷰 응답자 中 작곡가 @@씨-

같은 댓글이라도 대상에 따라 순화 단계는 달라진다

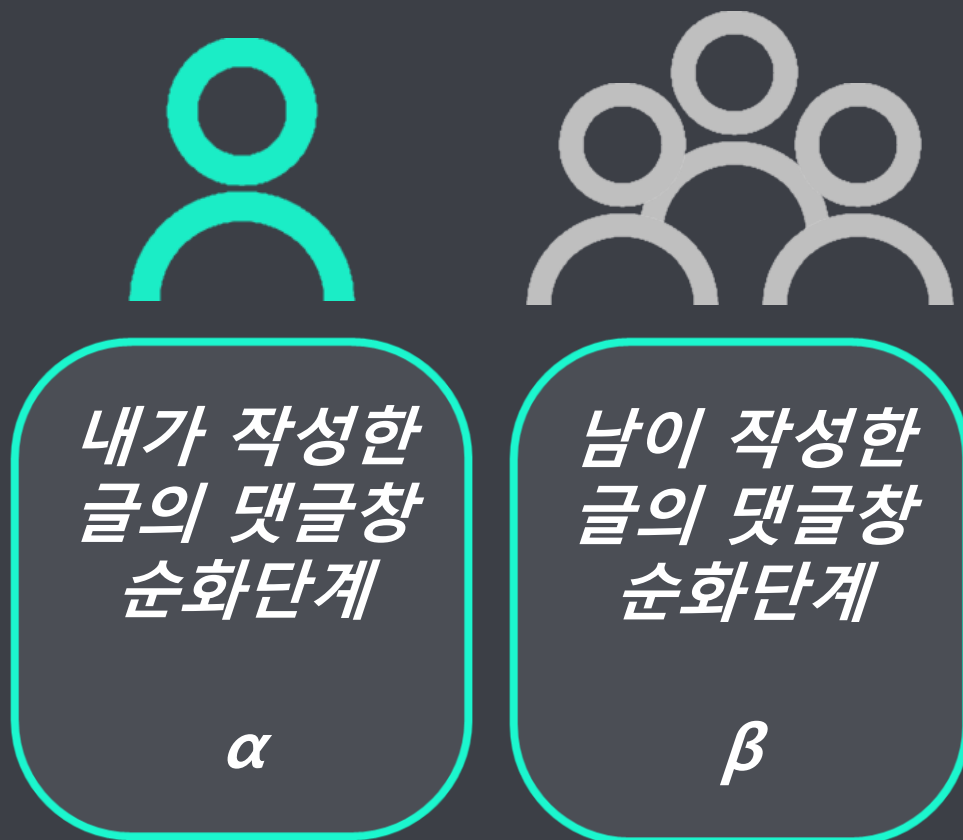
비난적 표현의 대상에 따른 응답한 순화 방식 변화



“ 솔직히 롤 해본 사람들은 다 공감할텐데
게임 망치는 놈들은 욕 좀 먹어야죠.

그래서 제가 실수하면 그냥 다 차단하고
합니다.”

-인터뷰 응답자 대학생 ##씨-



원본

이 선수는 쓰레기야
왜 이렇게 못하냐

1단계

이 선수는 @@야
왜 이렇게 못하냐


2단계

이 선수는 별로야
왜 이렇게 못할까요

3단계

이 선수는 마음에 들지 않네요.
성적이 저조한 이유가 궁금합니다.

Updated: June 13, 2023, 7:34 p.m.

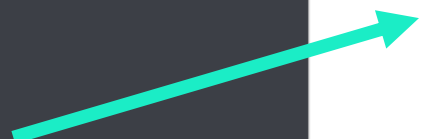
 **JJonng** June 12, 2023, 9:42 p.m. 수정 삭제 해석

디시 틀어막아봐야 엄한데서 또터진다 그냥 여따 쓰
레기들 모아놓고 차단하는게 답임


Updated: June 13, 2023, 7:34 p.m.

나의 게시물 댓글 순화(적용안함) ▾

남의 게시물 댓글 순화(3단계) ▾




Updated: June 13, 2023, 7:34 p.m.

 **JJonng** June 12, 2023, 9:42 p.m. 수정 삭제 해석

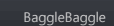
디시 틀어막아봐야 엄한데서 또터진다 그냥 여따 쓰
레기들 모아놓고 차단하는게 답임

Updated: June 13, 2023, 7:34 p.m.

 **JJonng** June 12, 2023, 9:42 p.m. 수정 삭제 해석

다른 곳에서도 비슷한 문제가 발생할 수 있으니 디시
를 차단하는 것만으로는 해결되지 않습니다. 대신, 불쾌한 내용을 올
리는 사용자들을 차단하는 것이 좋을 것 같습니다.

Updated: June 12, 2023, 9:56 p.m.



공격성 검사 기능



공격성 확인



공격성 확인기

해당 댓글은 다소 공격적인 표현들이 포함된 것 같습니다. 이러한 표현은 대화를 진행하는 상대방에게 불쾌감을 줄 수 있습니다. 대신, 더욱 적절하고 건강한 대화를 위해서는 좀 더 적절한 표현을 사용해 주시기 바랍니다.

해석 기능



해석



میم 해석기

알잘딱깔센은 '알아서 잘 딱 챙겨서 센스있게 해라'라는 의미이고, 킁킁빠빠는 '킁때끼고 빠질때 빠져라'라는 의미입니다. 이 댓글은 상대방에게 센스있게 대처하고, 빠르게 대응하라는 의미를 담고 있습니다.

기대효과



댓글을 보는 사용자 측면

1. 이용자별 적절한 댓글 수위 설정
2. 악성 댓글에 의한 피해 최소화

댓글을 쓰는 사용자 측면

1. 악성 댓글 **동조효과**를 방지
2. 표현의 자유를 보장



건전한 댓글문화 형성에 기여

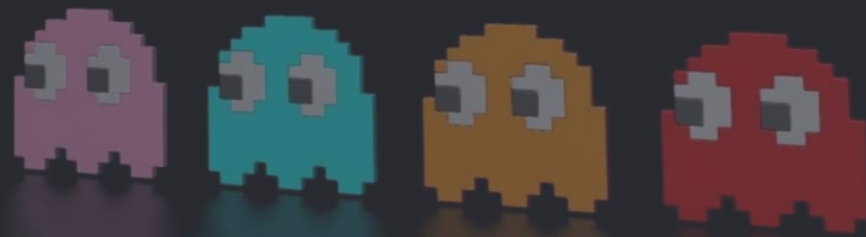


100%



Demonstration

(시연)





(conformity effect)

동조효과란?

no

no

ye...no

no

no





깨진 유리창 이론

(Broken Window Theory)



Future Plans

(향후계획)



성능향상





100%



확장



Outro

(마무리)

어려웠던 점



Chat GPT
융통성 부족

잘하면 망할 듯 : 잘했는데 왜 망함?



커튼 좀 쳐줘 : 커튼 좀 열어줘
커튼 좀 쳐줘 : 커튼 좀 닫아줘

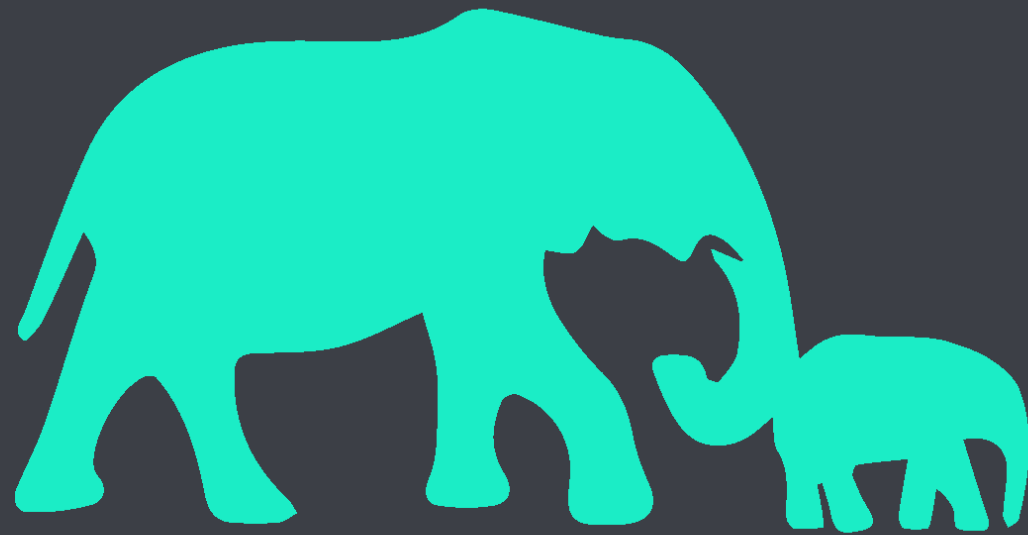
GPT의
섬세하지 못한
한국어 능력

K-GPT 가
필요한 이유

바글바글
성능 상승



100%





100%



#바글바글

감사합니다

