

AI 활용 심리치료 서비스

시공간의 제약 없이 마음을 돌보는
AI기반 심리상담 플랫폼



20252364 강한서 20251236 박주언 C조

프로젝트 배경 및 필요성



사회적 배경

현대 사회는 스트레스, 우울, 불안 등 다양한 정신 건강 문제의 증가를 겪고 있습니다. 이러한 문제들은 개인의 삶의 질을 저하시킬 뿐만 아니라 사회 전반에 걸쳐 부정적인 영향을 미치고 있습니다.



기술적 배경

최근 자연어 처리(NLP) 기술과 생성형 AI의 발전은 인간과 유사한 자연스러운 대화 구현이 가능해지면서 심리 상담 분야에서의 활용 가능성이 크게 부각되고 있습니다.



문제 인식

기존 심리 상담의 높은 비용과 시간적 부담은 많은 사람들이 상담을 주저하게 만드는 요인입니다. AI 기반 심리케어 서비스는 24시간 언제든지, 어디서든 접근 가능한 저비용의 상담 서비스를 제공할 수 있습니다.

프로젝트 목표 및 기대효과

🎯 목표



AI 챗봇 개발

사용자와의 대화를 통해 감정 상태를 정확하게 파악하고, 이에 기반하여 위로와 조언을 제공하는 AI 챗봇을 개발합니다.



감정 변화 시각화

대화 기록 및 심리 검사 결과를 바탕으로 사용자의 감정 변화를 그래프 등의 형태로 시각화하여 제공합니다.



자기 돌봄 콘텐츠 제공

스트레스 해소를 위한 명상, 음악 추천 등 간단한 자기 돌봄 콘텐츠와 함께 개인별 맞춤 개선 사항을 제안합니다.

★ 기대효과



저비용·고접근성 심리상담 서비스

기존 상담 서비스의 비용 및 시간적 제약을 극복하여 누구나 쉽게 접근할 수 있는 심리 상담 서비스를 제공합니다.



감정 표현 및 자기 이해 능력 향상

사용자가 자신의 감정을 자유롭게 표현하고, AI의 피드백을 통해 자기 이해 능력을 향상시키는 데 기여합니다.



학교·단체 멘탈케어 서비스 활용

학교나 기업 등 단체에서 구성원들의 멘탈케어를 위한 보조적인 서비스로 활용될 수 있습니다.

핵심 아이디어 및 접근 방법

본 프로젝트의 핵심 아이디어는 AI가 사용자의 감정을 이해하고 공감하며, 궁극적으로는 심리적 회복을 돕는 '디지털 상담사' 역할을 수행하는 것입니다.

1단계



사용자 입력

사용자가 자신의 감정이나 고민을 텍스트로 입력합니다.



2단계



감정 분석

AI가 입력된 텍스트를 분석하여 사용자의 현재 감정 상태를 파악합니다.



3단계



공감/조언 생성

분석된 감정을 바탕으로 공감하고 위로하며, 적절한 조언을 생성하여 제공합니다.



4단계



시각화

대화 기록 및 감정 변화를 시각적으로 보여줍니다.

적용 AI 기술 개요



자연어 처리 (NLP)

사용자의 입력 텍스트에서 감정과 의도를 분석하는 데 사용됩니다. KoBERT, HuggingFace에서 제공하는 사전 학습된 모델들을 활용하여 한국어 텍스트의 감정 분류 및 의미 분석을 수행합니다.



생성형 AI

사용자의 감정에 공감하고 적절한 위로와 조언을 제공하는 대화 응답을 생성합니다. ChatGPT API를 활용하여 자연스럽게 맥락에 맞는 대화 생성을 구현합니다.



감정 분석 모델

사용자의 텍스트에서 긍정, 부정, 중립 등 다양한 감정 상태를 식별합니다. KoBERT와 HuggingFace 모델을 기반으로 한국어 감정 데이터셋에 최적화된 모델을 구축합니다.

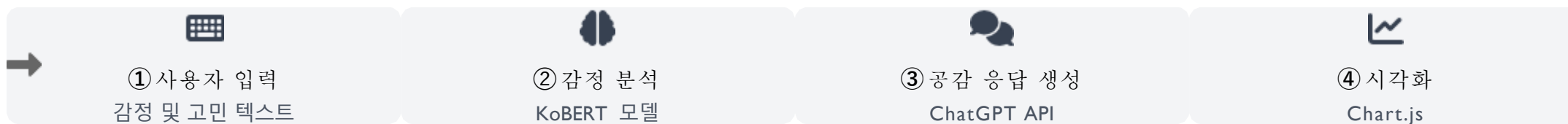


시각화

사용자의 감정 변화 추이 및 심리 검사 결과를 직관적인 그래프 형태로 제공합니다. Chart.js 라이브러리를 활용하여 웹 기반의 동적인 그래프를 구현합니다.

모델 구조 및 알고리즘 설계

데이터 흐름



감정 분석

KoBERT 모델을 활용하여 사용자의 텍스트에서 감정 상태를 분석합니다.

주요 알고리즘: 감정 분류(**Softmax**)



대화 생성

ChatGPT API를 활용하여 감정에 공감하고 적절한 조언을 생성합니다.

주요 알고리즘: 조건부 텍스트 생성 (프롬프트 기반)



시각화

Chart.js 라이브러리를 활용하여 감정 변화 추이를 직관적으로 보여줍니다.

주요 기능: 감정 상태 변화 추이 그래프

데이터 수집 및 윤리적 고려사항



데이터 출처

- ✓ 공개 한국어 감정 데이터셋(AI Hub 등) 활용 직
- ✓ 접 작성 대화: 실제 상담 시나리오 반영 다양한
- ✓ 감정 상태에 대한 균형 잡힌 데이터 확보

전처리 과정

- 🔻 불필요한 문자 제거 및 이모티콘 정리
- 🚫 욕설 필터링: 부적절한 언어 차단
- 📌 감정 레이블 부여: 학습 데이터셋 형식으로 변환



윤리적 고려 사항



개인정보 보호

사용자의 개인정보 및 상담 내용은 철저히 익명화하고, 상담 후에는 폐기하는 원칙을 준수합니다.



익명성 보장











모든 사용자 데이터는 익명으로 처리되어 개인 식별이 불가능하도록 합니다.



자해 관련 문장 처리

사용자가 자해 또는 극단적인 선택을 암시하는 문장을 입력할 경우, 즉시 전문가의 도움을 받을 수 있도록 안내하는 시스템을 구축합니다.

구현 도구 및 개발 환경


개발 언어	개발 도구	AI 프레임워크 및 API
<div></> Python: AI 모델 개발 및 웹 서비스 백엔드 구현</div> <div></> HTML/CSS/JS: 프론트엔드 인터페이스 구현</div> <div></> SQL: 데이터베이스 관리 및 쿼리</div>	<div> Google Colab: AI 모델 학습 및 실험</div> <div> VS Code: 코드 작성 및 프로젝트 관리</div> <div> PostgreSQL: 사용자 데이터 관리</div>	<div> PyTorch: 딥러닝 모델 구축 및 학습</div> <div> ChatGPT API: 대화 생성 기능 구현</div> <div> Chart.js: 감정 데이터 시각화</div>
<div>개발 환경 구성도</div> <div> 로컬 개발 환경 →  클라우드 학습 환경 →  서버 배포 환경 →  사용자 접근 환경</div>		

프로젝트 일정표

📅 주차별 추진 계획

<div>1주차</div> <div><ul style="list-style-type: none">✓ 프로젝트 주제 선택 및 조사✓ 프로젝트 특징 및 목표</div> <div>2주차</div> <div><ul style="list-style-type: none">✓ 기획서 작성 및 역할 분담 최✓ 종 기획서 및 개발 계획 완성</div>	<div>3주차</div> <div><ul style="list-style-type: none">✓ 공개 분석 데이터셋 정도</div> <div>4주차</div> <div><ul style="list-style-type: none">✓ 감정 분석 모델 학습 및 ChatGPT API 캐스트 시작✓ 감정 평가 + 모델링 작성</div>	<div>5주차</div> <div><ul style="list-style-type: none">✓ 감정 분석 + 대화 설정✓ 감정(Chart.js) 및 UI✓ 통합 봇 MVP</div> <div>6주차</div> <div><ul style="list-style-type: none">✓ 사용자 테스트 및 이에 대한 응답을 작성하고 발표✓ 최종 시연 및 의견</div>
--	---	---


팀 역할분담

 강한서

 기획

 데이터

 모형

 박주언

 UI/UX

발표 담당

위험 요소 및 대응 방안

기술적 위험

- ⚠ 감성 분석 모델의 정확도 저하
 - 대응: 철저한 데이터 전처리, KoBERT 모델 하이퍼파라미터 최적화
- ⚠ 생성형 AI 응답 품질 및 일관성 부족
 - 대응: 감성 상태를 프롬프트에 명확히 포함, 감정별 기본 응답 틀 설계
- ⚠ 데이터 품질 문제 (중복, 오타, 잘못된 라벨링)
 - 대응: 중복 제거, 직접 검증, 데이터 증강으로 감성 불균형 해소

프로젝트 관리 위험

- ⚠ 짧은 기간(6주)으로 인한 일정 지연 가능성
 - 대응: 주차별 세부 목표 설정, 주기적 진행상황 점검, 모델링-UI 병행
- ⚠ 협업 과정의 코드 충돌 및 작업 중복
 - 대응: GitHub를 통한 버전 관리, 명확한 역할 분담, 정기적 코드 리뷰

윤리적 위험

- ⚠ 자해, 폭력 등 위험 발언에 대한 부적절한 AI 반응
 - 대응: 위험 단어 필터링, 전문가 상담 안내 시스템 구현
- ⚠ 개인정보 및 민감 데이터 보호
 - 대응: 민감 대화 내용 즉시 폐기, 익명화 처리 원칙 적용

성능 평가 지표 설정

데이터셋 분리

감정 데이터셋 중 일부(예: 20%)를 테스트 세트로 따로 분리해 사용

정량적 평가 지표

- ✓ 정확도(Accuracy): 전체 데이터 중 정확히 분류된 비율
- ✓ 정밀도(Precision): 특정 감정으로 분류한 항목 중 실제 해당 감정인 비율
- ✓ 재현율(Recall): 실제 특정 감정인 항목 중 올바르게 분류된 비율
- ✓ F1-score: 정밀도와 재현율의 조화평균


감정별 평가

감정별(기쁨, 슬픔, 분노, 불안, 중립 등)로 위 지표를 계산해 평균 성능 산출


목표 성능: **F1-score 기준 0.80 이상** 을 목표로 설정

대화 품질 평가

 BLEU Score: 생성된 응답의 자연스러움 정량 평가

 Human Evaluation: 실제 사용자 평가(1-5점)
공감도 관련성 유용성 자연스러움 등 평가

 BLEU Score 계산 후, 사람 평가 점수와 함께 종합 점수를 평균화

 Human Evaluation 점수가 **4.0 이상** 일 경우 "공감형 대화 가능" 수준으로 판단

사용자 테스트 또는 설문 계획

평가 대상

- 무작위 30~50명 (스트레스·우울감 경험자 중심)
- 실험군: AI 심리치료 챗봇 이용
- 비교군: 기존 심리상담센터 또는 일반 자가 대처 방법 이용

조사 방식

- 사전/사후 설문조사 (4주 간격)
- 심층 인터뷰 (일부 참여자 10명 대상)
- 자동 로그 데이터 분석 (대화 길이, 응답 시간, 감정 단어 비율 등)

조사 항목

사용성(UX)

- 인터페이스 직관성, 대화 자연스러움, 감정 공감 정도

심리적 변화

- 우울/불안 척도(PHQ-9, GAD-7 등)
- 자기효능감(Self-efficacy)

만족도

- 전반적인 만족도(1~5점 척도)
- "다시 사용하고 싶다" 여부

개방형 피드백

- 불편했던 점, 추가되었으면 하는 기능 등

결과 분석 및 개선 방안 제시

결과 분석

- 우울·불안 지표 변화 평균 비교 (사전 대비 사후 개선율)
- 사용성 평가 평균 점수, 피드백 내용 분류
- 로그 데이터에서 AI 응답 시간/정확도/공감 문장 비율 분석

개선 방안

대화 품질 향상

- 감정 어조(공감 문장) 자동 강화 알고리즘 적용
- 오답/무응답 사례 자동 검출 및 재학습

UX 개선

- 심리 상태를 색상·아이콘으로 시각화
- 사용자의 감정 변화 기록('감정 다이어리' 기능) 추가

맞춤형 피드백

- 스트레스 요인(학업/대인관계 등)에 따른 응답 유형 세분화
- 사용자 프로필 기반 개인화 추천

기대효과 요약



정신건강 서비스 시장 내 AI 기반 상담 플랫폼 확장성 확보

- ✓ 전문가 상담 접근이 어려운 대학생 및 청년층 대상 초기 진입으로 시장 저변 확대
- ✓ 상담 서비스 비용 절감으로 더 많은 사용자에게 접근성 제공



심리상담 및 AI 융합 분야 학습 자료 제공

- ✓ 한국어 감정 분석 데이터셋 및 모델 성능 개선 기여
- ✓ AI 심리상담 윤리 지침 및 모범 사례 수립을 위한 기초 자료 제공



저비용·비대면 심리 지원 서비스 기반 마련

- ✓ 심리상담 문화 확산과 정신건강 관리에 대한 인식 개선
- ✓ 시공간 제약 없는 24/7 지원으로 위기 상황 대응 지원
- ✓ 사회적 취약계층이나 심리지원 사각지대에 있는 대상자들의 접근성 향상

후속 연구 및 서비스 확장 방향



상용화 방향

- 대학 상담센터, 지자체 정신건강센터에 API 형태로 제공
- '상냥이', '마음이' 처럼 모바일 앱 통합 서비스화



고도화 방향

- 멀티모달 AI 적용 (음성 감정·표정 인식 포함)
- 대화형 LLM(Large Language Model) 기반 맞춤 치료 시나리오 자동 생성



지속적 발전

- 사용자 피드백 기반 주기적 모델 재학습
- 윤리적 AI(데이터 보호·비판적 발언 필터링) 체계 강화
- 장기 추적 연구(6개월~1년 단위)로 치료 지속 효과 검증

👁 장기적 비전

AI 심리상담 기술의 발전과 임상 검증이 축적됨에 따라, 종합적 정신건강 관리 플랫폼으로 확장하여 예방-진단-치료-관리의 전 과정을 지원하는 통합 솔루션으로 발전할 것으로 기대됩니다.

한계점 및 개선 가능성

! 현재 한계점

- 전문가와 연계를 필요로 하는 보조적 역할
- 복잡한 심리적 문제 해결에 제한적 기능
- 데이터 기반 응답의 한계와 윤리적 고려사항

★ 주요 장점 및 가치

- 심리상담을 향한 심리적 장벽 낮추기
- 상담에 대한 접근성 향상 및 사회적 편견 해소
- 시공간 제약 없이 가볍게 심리상담 가능
- 익명성 보장을 통한 솔직한 감정 표현 환경 제공



전문적 상담의 보완재




심리상담 진입점





개인정보 보호 강화

참고 문헌 및 자료


데이터셋 관련

 한국어 자연어 처리 데이터셋 목록 (LittleFox NLP 데이터셋)

 AI Hub 감정분석 데이터셋

 한국어 대화 코퍼스

AI 심리치료 관련 연구


 Journal of Medical Internet Research (2023): "Effects of an Artificial Intelligence Platform for Behavioral Interventions on Depression and Anxiety Symptoms"

- AI 지원 치료 그룹에서 우울 감소율 34% vs 전통치료 20%

 JMIR Mental Health (2018): "Using Psychological Artificial Intelligence (Tess) to Relieve Symptoms of Depression and Anxiety"

- AI 챗봇이 불안·우울 완화에 통계적으로 유의한 효과

기타 참고자료

 국내 대학생 정신건강 실태조사 보고서

 인공지능 심리상담 윤리 가이드라인