

2025년 새싹 해커톤(SeSAC Hackathon) AI 서비스 기획서

팀명	소셜라이프가드
팀 구성원 성명	윤진호, 정영준, 최창하, 진효상, 김영원

① AI 서비스 명칭

SNS 기반 자살위험 신호 AI 분석·모니터링 서비스 – 도와줄게

② 활용 인공지능 학습용 데이터

1. 공개 SNS 텍스트 데이터

	활용 데이터명	분야	출처
1	X(Twitter) Public API Text Stream	공개 SNS 텍스트	X(Twitter) API
2	Hashtag-Keyword Real-time Crawl Data	공개 SNS 텍스트	키워드 기반 실시간 크롤링(허용 범위)
3	Public Community/Forum Crawl Data	공개 웹 텍스트	크롤링 허용된 커뮤니티/포럼(robots.txt 기준)

2. 자살위험·정신건강 언어 패턴 데이터

	활용 데이터명	분야	출처
1	CLPsych Dataset	자살 위험 언어패턴	CLPsych (Clinical Linguistics & Psychology Shared Task Dataset)
2	Reddit Suicidality Dataset	자살위험·우울 텍스트	Reddit 공개 데이터(영어, 번역/전처리 필요)
3	Suicide / Depression Keyword Lexicon	키워드 사전	오픈소스 기반 커스텀 구축

3. 한국어 감정·부정 발화 데이터

	활용 데이터명	분야	출처
1	AI Hub 감정분석 말뭉치	감정 분석	AI Hub
2	KoELECTRA Pretrain Data	한국어 부정·일상 언어	SKT Brain 공개 사전학습 코퍼스
3	kcBERT Pretrain Corpus	한국어 SNS/커뮤니티 텍스트	공개 커뮤니티 기반 사전학습 데이터
4	Toxic/Negative Emotion Corpus	독성·부정 발화	HuggingFace 공개 데이터셋

③ 핵심내용

**'SNS 데이터를 실시간으로 탐지하고,
위험 신호를 AI가 분석해 사용자에게 알림을 보내는 서비스'**

[핵심기능]

① SNS 실시간 크롤링 및 조건 기반 모니터링

- 사용자가 설정한 키워드, 해시태그, 지역, 시간대, 언어등의 조건을 기준으로 X(Twitter), Mastodon, 인스타그램 대체 크롤링 등 다중 SNS 데이터를 실시간으로 자동 수집
- 비정형 텍스트·축약체·이모지 등도 전처리하여 위험 신호 누락을 최소화하며, 특정 위험 키워드의 급증 패턴, 특정 지역에서의 집중 발생, 특정 시간대에 반복되는 위험발화 흐름 등을 탐지해 활동가가 즉각 상황을 파악할 수 있게 만듬

② AI 감정 분석 및 RAG 기반 위험도 분석

- 수집된 텍스트는 KoBERT·KoELECTRA 기반 감정 분석 모델과 자살 위험 발화 데이터로 학습된 분류 모델을 통해 감정 흐름·우울/절망 패턴·자해 가능성·우회적 표현 등을 포함한 종합 위험도를 산출
- 사용자가 직접 작성하는 프롬프트 기반 LLM 분석을 적용하여 상황별 맞춤 해석이 가능하며, RAG를 통해 정신건강 기관 가이드라인·임상 매뉴얼·학술자료를 참고하여 해석의 신뢰성을 높인 근거 기반 분석을 수행

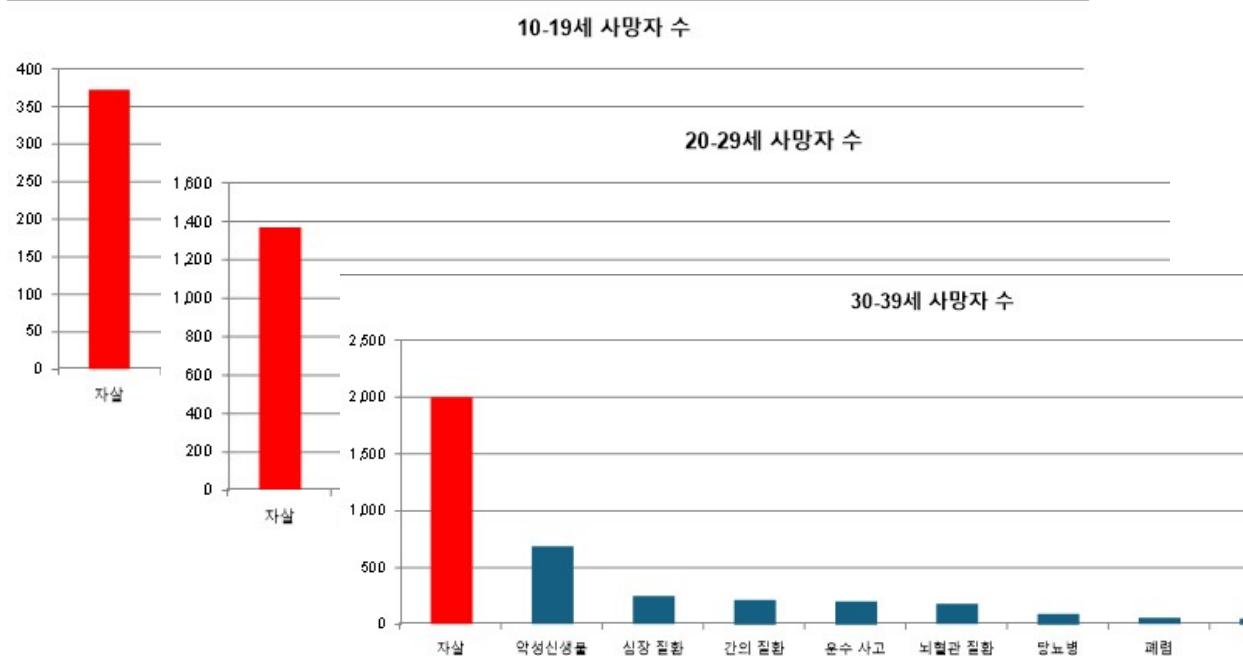
③ 대시보드·알림 기능 및 대응 가이드 제공

- 분석된 모든 결과는 실시간 대시보드에 High/Medium/Low 위험도, 시간대별 감정 흐름, 최근 급증한 키워드, 지역 기반 위험 발화 히트맵 형태로 시각화
- High 위험도가 감지될 경우 텔레그램·슬랙·카카오톡 등으로 즉시 알림이 전달되며, 게시글을 클릭하면 원문·감정 그래프·AI 분석 요약·위험 요인·관련 상담기관까지 확인 가능
- 사용자 상황에 맞춰 상담센터 연락처, 행동 매뉴얼, 대화 가이드 등 즉시 활용 가능한 대응 지침을 함께 제공하여 활동가의 현장 대응력을 강화

④ 제안배경 및 목적

[제안배경]

1. 청년·서울 중심 자살 문제의 심각성



10~30대 사망사유 1위 자살 (국가통계포털 KOSIS, '2024년 연령별 자살 사망자 수 분포')

- 서울시 자살 사망자 수 **2,234명** (국가통계포털 KOSIS, '2024년 시도 분기별 자살 사망자 수 분포')
- 온라인 자살 유발 정보 모니터링·AI 기반 위험군 탐지를 주요 정책 방향으로 채택
- 청년 자살률 2023~2024년 3년간 상승세

2. 새벽 2시, 침묵 속에 울리는 알림음 하나가 소중하다



자살 예방 운동가 현장 스케치 유튜브 : https://youtu.be/x5WVYNwS5Xw?si=Xlc1rZeA1kYh_4QY

심각한 사회문제인 자살 문제를 해결하고자 SNS 피드 속에는 아님을 바라는 마음으로 스크롤하는 누군가가 있습니다. 이 사람은 바로 당신의 눈에 보이지 않는 곳에서 "살고 싶다"는 메시지 하나에도 귀 기울이고, "끝내고 싶다"는 글 하나에도 가슴이 떨리는 자살예방운동가입니다.

■ 화면 너머 누군가가 보내는 한 줄의 글

"다른 사람들에겐 아무렇지 않아 보여도...제가 없으면 다 나아질까요?"

그 글을 본 운동가는 화면을 멈추고 심장 박동을 느끼며 "이건 놓쳐선 안 될 신호"라 내면에서 외칩니다. 수백 개의 게시글 속 하나지만 그 하나에 담긴 절망의 무게를 그는 알고 있습니다.

■ 정규 근무시간이 끝나도, 알림은 쉬지 않습니다

야간 시간대: 누군가 "안녕히 계세요..."라고 남기고 사라질지도 모른다는 경고음처럼 울립니다. 운동가는 잠을 뒤척이며 모니터 앞에 앉아 무표정으로 나타난 그 순간에도 마음속은 떨림과 초조로 가득 차 있습니다.

■ '데이터'가 아닌 '사람' 앞에 서 있는 그들

수치로 표현되기엔 너무도 생생한 고통들이, 감성 분석이 포착하기엔 너무도 교묘한 은유로 SNS 구석에 숨어 있습니다. 운동가는 그 은유 하나하나를 해독하기 위해 밤새 고민하고, 한 명이라도 더 살릴 수 있다면이라는 신념으로 화면을 응시합니다.

■ 보고서의 숫자가 아닌, 생명의 호흡을 마주하는 순간

엑셀 시트에는 '발견 건수', '연계 수치', '위험도 스코어'가 적히지만, 운동가의 마음속엔 "살 수 있었을지도 모를 그 사람"이 살아 숨쉬고 있습니다. 그 사람이 다시 웃을 수 있을까, 그 사람이 새벽이 아니라 아침을 맞을 수 있을까, 그 질문이 그의 트리거가 됩니다.

■ 우리는 그들을 위해 서비스를 만들어야 합니다

위험 신호가 너무 많아서 골든타임을 놓치는 것이 가장 두렵다는 그들의 말처럼, 우리의 기술과 플랫폼은 "놓치지 않도록, 더 빨리 반응하도록, 더 정확히 판단하도록" 설계되어야 합니다. 그들의 밤이 조금이라도 평온해질 수 있도록, 그들이 더 많은 생명을 마주할 수 있도록.

■ 작은 글 하나에 숨겨진 목소리를 듣기 위해

작은 글 하나에 숨겨진 신호는 곧 한 사람의 생명일 수 있습니다.

하지만 그 신호들은 너무 빠르고, 너무 많아, 한 명의 운동가가 모두 감당하기 어렵습니다. 그래서 기술이 필요합니다.

위험 신호를 놓치지 않게 하고, 운동가가 가장 중요한 순간 '개입'에 집중할 수 있게 만드는 것. 이것이 저희 서비스의 핵심입니다.

[문제인식]

1. SNS 위험 신호의 '양·속도'가 인간이 감당할 수준을 초과

- 자살예방운동가는 수많은 SNS 글 중 실제로 긴급 개입이 필요한 게시글을 육안으로 찾아야함
- 업데이트되는 피드를 따라가는 것은 불가능, 실시간 감시 공백이 생길 가능성이 있음
- 현 체계는 모니터링(탐색 단계)에 대부분의 시간을 소모하고 있어, 정작 중요한 '개입·상담' 단계에 투입할 수 있는 시간이 줄어드는 구조적 문제 발생

2. 위험도 판단의 비일관성과 전문 분석 도구 부재

- 같은 문장을 보고도 활동마다 위험도를 다르게 판단하는 일이 빈번하고, 조직 차원에서 사용할 수 있는 "표준화된 위험도 스코어링 툴"이 부재
- 자살 위험 언어 패턴(우울의 지속, 자해 암시, 동반자 모집 등)에 특화된 모델이 부족
- 분석·보고서 작성에 과도한 시간 소모 → 개입 활동 시간 감소

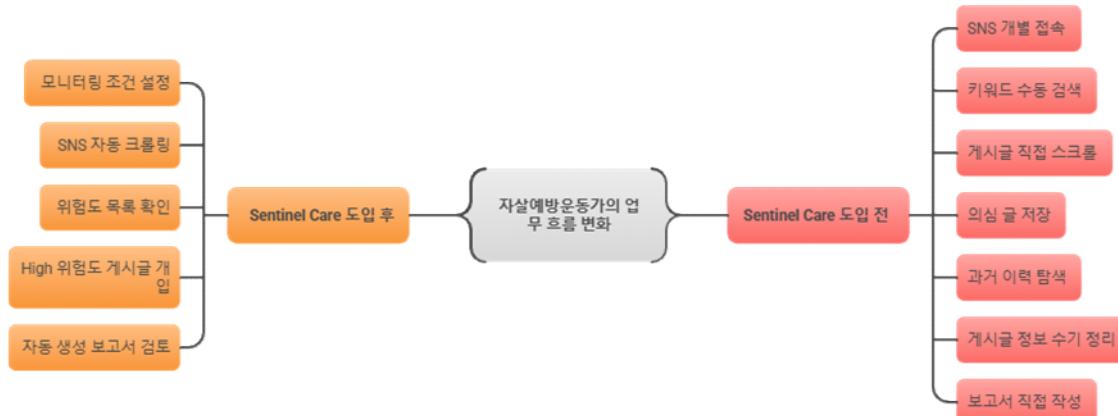
[목적]

1. SNS 기반 자살 위험 신호의 조기·정밀 탐지 (분석 정확도 향상)

- 사람의 눈으로 놓치기 쉬운 미묘한 표현, 장기간에 걸친 감정 변화, 특정 시간·지역에서의 위험도 급증을 AI 모델과 RAG 기반 분석으로 정밀하게 포착
- 기존 수작업으로 소요되던 모니터링·선별·기초 분석·리포트 초안 작성 단계를 Sentinel Care가 자동화함으로써, 활동가는 개입·상담·현장 출동에 더 많은 시간을 쓸 수 있음
- "문서·분석에 쓰이던 시간을 줄여 골든타임 대응 속도를 높이고, 현장 개입 횟수와 질을 높이는 것"이며, 이를 통해 실제 생명을 지키는 결과로 이어지게 하는 것이 1차 목표

● 자살예방운동가의 업무 시간 단축 및 골든타임 확보

자살예방운동가의 업무 흐름 변화



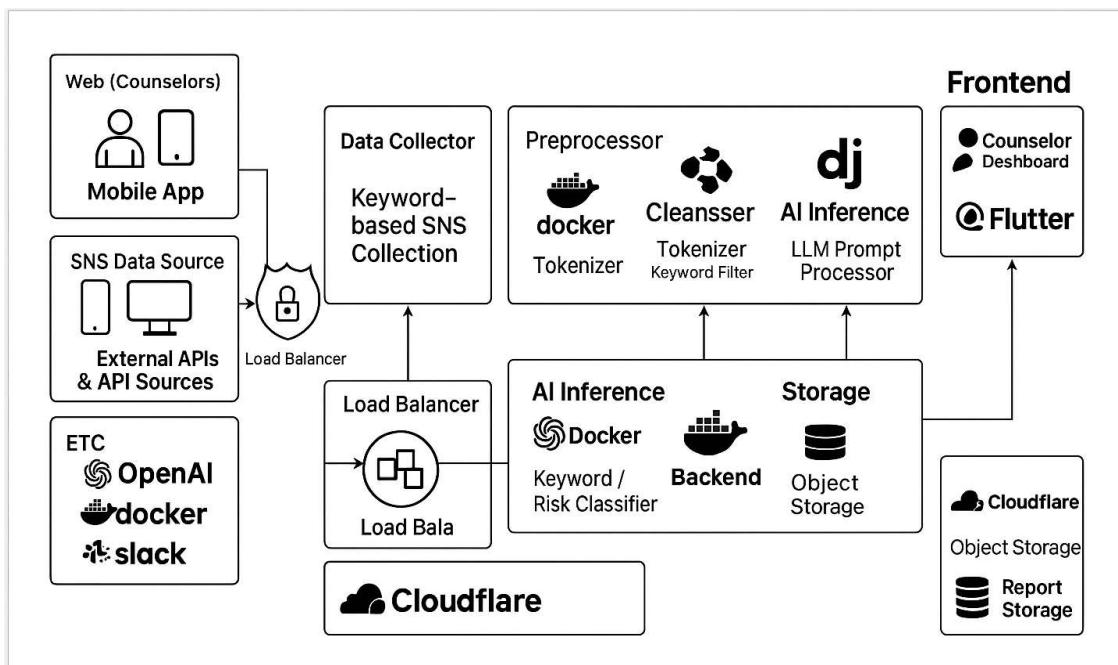
2. 데이터 기반 자살 예방 활동 체계 및 협업 인프라 구축

- SNS 위험 신호 데이터를 시계열·지역·연령대·키워드별로 구조화해, 기관·지자체·시민단체가 공유할 수 있는 근거 데이터를 축적
- 이를 토대로 서울시 및 관련 기관과의 시범 사업, 공공–민간 협력 모델, 향후 B2G/B2B 확장 등 다양한 형태의 협업 가능성을 염
- 궁극적으로는 “한 번의 프로젝트”에 그치지 않고, 지속적으로 학습·개선되는 AI 기반 자살 예방 생태계를 만드는 것을 장기 목적으로 함

5 세부내용

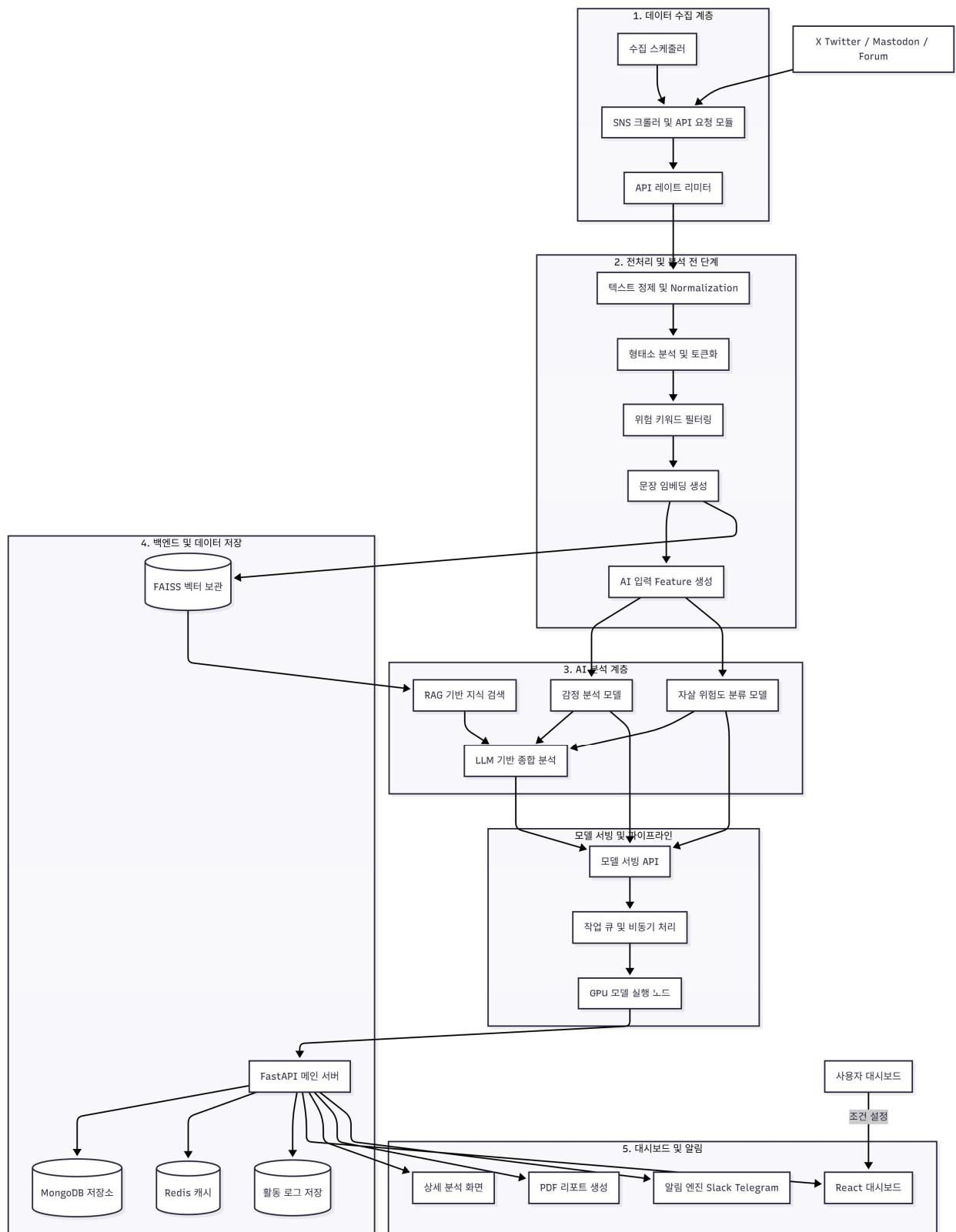
[아키텍처 설계]

1. 시스템



프레임워크	<ul style="list-style-type: none">- Next.js 16 (메인 서버)- FastAPI (AI 모델 서빙)
데이터베이스	<ul style="list-style-type: none">- Supabase (PostgreSQL + 인증 + Realtime)- Upstash Redis (크롤링 큐 + 캐싱)- Pinecone 또는 pgvector (RAG용 벡터 DB)
AI/ML	<ul style="list-style-type: none">- Vercel AI SDK (LLM 통합)- Hugging Face Transformers (KoBERT, KoELECTRA)- LangChain (RAG 구현)
크롤링	<ul style="list-style-type: none">- Twitter API v2- Playwright (웹 크롤링)- BullMQ (작업 큐)
알림	<ul style="list-style-type: none">- Telegram Bot API- Slack Webhooks
배포	<ul style="list-style-type: none">- Vercel (Next.js)- Railway 또는 Fly.io (FastAPI)

2. 플로우 차트



[기능설명]

1. SNS 실시간 크롤링 및 조건 기반 모니터링

The screenshot displays the AI-based real-time monitoring interface. At the top, there are tabs for 'AI 실시간 탐지 중' and '탐지 조건 변경'. On the right, a user profile for '김민수' (Seoul) is shown with a '변경사항 저장' button. The main area is divided into several sections:

- III 모니터링 조건 설정**: Includes tabs for '대시보드', '탐색', and '설정'. Sub-sections include '키워드 & 해시태그' (with a note about AI filtering keywords), '카테고리 설정' (with categories like 자살, 우울, 힘들어, 등반자살), and '플랫폼 선택' (with options for X (Twitter), Instagram, and Mastodon).
- AI 분석 구성**: Shows threat detection analysis with tabs for '민감함(제일 많은 탐지)', '중간(권장)', and '낮음'. A note says '이 소셜 미디어 게시물의 자살 위험 신호를 분석하십시오. 언어 폐단, 감정적 어조, 자해에 대한 형식적 언급을 고려하십시오. 위험 수준을 높음, 중간, 낮음으로 분류하고 그 근거를 제시하십시오.' Below it, a note says '이 프롬프트는 AI 모델이 게시물을 분석하는 데 도움이 됩니다. 특정 패턴이나 맥락에 초점을 맞추도록 사용자 지정할 수 있습니다.'

- 사용자가 설정한 **키워드, 해시태그, 지역, 시간대, 언어** 등의 조건을 기준으로 X(Twitter), Mastodon, 인스타그램 대체 크롤링 등 다중 SNS 데이터를 **실시간으로 자동 수집**
- 비정형 텍스트·축약체·이모지 등도 전처리하여 위험 신호 누락을 최소화하며, 특정 위험 키워드의 **급증 패턴**, 특정 지역에서의 **집중 발생**, 특정 시간대에 반복되는 **위험발화 흐름** 등을 탐지해 활동가가 즉각 상황을 파악할 수 있게 만듬

2. AI 감정 분석 및 RAG 기반 위험도 분석

The screenshot displays the AI-based sentiment analysis and risk assessment interface. At the top, there are tabs for 'AI 실시간 탐지 중' and '탐지 조건 변경'. On the right, a user profile for '김민수' (Seoul) is shown with a '고위험' (High Risk) status. The main area is divided into several sections:

- 상세분석**: Shows posts from users '@user-1234' (X, 2024-01-15 14:32:00) and '김민수' (Seoul). It includes a '고위험' (High Risk) alert for the post '너무 힘들어서 이제 모든 걸 끝내고 싶다...'. Other posts include '끌내고 싶다. 힘들어, 이유를 찾을 수 없어' and '김민수' (Seoul)'s response '이 게시물은 즉각적인 개입이 필요합니다. 자살 충동이 명확하게 드러나고 있습니다.' A '작성' (Write) button is also present.
- AI 분석 리포트**: Shows a report for a post by '김민수' (Seoul) with a '고위험' (High Risk) score of 83.5%. The report notes '최근 심리 지수 83.5%' and provides a detailed analysis of the post's content.
- 작성한 게시물**: Shows a post from 'X' (2024-01-15 14:32:00) with a '고위험' (High Risk) score of 70%. The post contains the text '너무 힘들어서 이제 모든 걸 끝내고 싶다...'. A '작성' (Write) button is present.
- 작성한 게시물**: Shows a post from 'X' (2024-01-15 14:32:00) with a '고위험' (High Risk) score of 95%. The post contains the text '요즘 삶이 의미롭게 느껴져. 아무것도 하고 싶지 않아'. A '작성' (Write) button is present.

- 수집된 텍스트는 KoBERT·KoELECTRA 기반 감정 분석 모델과 자살 위험 발화 데이터로 학습된 분류 모델을 통해 **감정 흐름·우울/절망 패턴·자해 가능성·우회적 표현**등을 포함한 종합 위험도를 산출
- 사용자가 직접 작성하는 **프롬프트 기반 LLM 분석**을 적용하여 상황별 맞춤 해석이 가능하며, RAG를 통해 정신건강 기관 가이드라인·임상 매뉴얼·학술자료를 참고하여 해석의 신뢰성을 높인 근거 기반 분석을 수행

3. 대시보드·알림 기능 및 대응 가이드 제공

The dashboard interface includes:

- Real-time Monitoring:** Shows overall risk scores (e.g., Total Alert Level: 2,848, High Risk: 23, Moderate Risk: 147, Low Risk: 1,294) and growth rates compared to the previous week.
- User Activity:** Displays recent posts from users like @user-1234, showing messages such as "너무 힘들어서 이제 모든 걸 끝내고 싶다..." (Too tired to end it all...) and "끌내고 싶다. 힘들어. 이유를 찾을 수 없어" (Want to leave. Tired. Can't find a reason). It also shows a timeline of posts from January 15, 2024.
- Comprehensive Analysis:** Includes a section titled "AI 종합 분석" (Comprehensive AI Analysis) which provides insights into the situation, such as "자살 위험 신호가 명확히 짐작됩니다." (Suicide risk signal is clearly suspected) and "경지 수 없고, '무의미' 등의 표현이 극단적 감정 상태를 나타냅니다." (Inability to express, 'meaningless' expressions indicate extreme emotional state).
- Timeline:** A graph titled "최근 살리 저수" (Recent rescue rate) shows a fluctuating trend at 83.5%.
- Alerts and Guidance:** A section titled "작성한 게시글" (Written posts) lists messages from users like 김민수, with a note: "이 게시물은 즉각적인 개입이 필요합니다. 자살 충동이 명확하게 드러나고 있습니다." (This post requires immediate intervention. Suicidal intent is clearly shown.)
- Feedback:** A "고위험" (High Risk) feedback button is visible on the right side of the interface.

- 분석된 모든 결과는 실시간 대시보드에 **High/Medium/Low 위험도, 시간대별 감정 흐름, 최근 급증한 키워드, 지역 기반 위험 발화 히트맵**형태로 시각화
- High 위험도가 감지될 경우 텔레그램·슬랙·카카오톡 등으로 **즉시 알림**이 전달되며, 게시글을 클릭하면 원문·감정 그래프·AI 분석 요약·위험 요인·관련 상담기관까지 확인 가능
- 사용자 상황에 맞춰 상담센터 연락처, 행동 매뉴얼, 대화 가이드 등 **즉시 활용 가능한 대응 지침** 함께 제공하여 활동가의 현장 대응력을 강화

⑥ 기대효과

1. 이해관계자별 As-Is → To-Be

이해관계자	As-Is	To-Be
자살 위험 당사자	<ul style="list-style-type: none"> - 위기 시 SNS에 글을 올려도 즉각적인 발견 가능성이 낮음 - 본인의 신호가 무시되거나 묻힐 수 있다는 불안감 존재 - 도움을 받을 타이밍을 놓치기 쉬워 실제 위험으로 이어질 확률이 높음 	<ul style="list-style-type: none"> - SNS에 남긴 짧은 글, 은유적 표지 빠르게 탐지되어 조기 발견 가능성 증가 - 위험도에 따라 우선 개입·연계가 이루어져 구조 확률 상승 - 청년층 중심 위험 신호를 선제적으로 감지하여 생명 안전망의 실질적 촘촘함 강화
자살예방운 동가/상담 사	<ul style="list-style-type: none"> - 수백 개의 게시글을 직접 검색·스크롤·정리하는 과부하 - 위험 판단 기준이 사람마다 달라 불일치 발생 - 보고서 업무·데이터 정리에 많은 시간 소모 → 정작 '개입' 시간이 부족 - 골든타임을 놓칠까 하는 심리적 압박감이 상존 	<ul style="list-style-type: none"> - 자동 크롤링·자동 분석·우선순위 정렬로 필요한 게시글만 빠르게 확인 - 위험도 스코어 기준이 표준화되어 판단의 일관성 확보 - 보고서 자동 생성으로 백오피스 부담 감소 → 개입·상담 시간 증가 - 즉시 알림 기반 대응으로 골든타임 확보, 생명 구조율 증가
보건복지 관련 기관·지자 체	<ul style="list-style-type: none"> - SNS 기반 위험 신호의 규모·패턴을 정량적으로 파악하기 어려움 - 예방 정책/예산 집행 근거가 부족해 결과 평가가 어려움 - 기관 간 협업이 데이터 중심으로 이루어지지 않아 효율성이 낮음 	<ul style="list-style-type: none"> - 지역·시간·연령대별 위험 신호 데이터를 누적 관리 - 정책 보고·사업 타당성 검토에 정량적 근거 데이터 제공 - 공공기관-시민단체 간 데이터 기반 협업 모델 강화 - 지역사회 단위의 선제적 예방 정책 설계 가능

2. 정량적 효과

위험 신호 탐지 속도 향상 <ul style="list-style-type: none"> - 기존 대비 위험민감 글 발견까지 걸리는 시간 약 50~70% 단축 (수동 스크롤/검색 제거 → AI 즉시 분석) 	업무 효율 증가 <ul style="list-style-type: none"> - 활동가 기준 모니터링·분석·보고서 작성 시간 1일 평균 2~3시간 절감 - 기계적 작업 제거로 실제 개입 활동 증가(최대 1.5~2배까지 가능) 	골든타임 내 개입률 상승 <ul style="list-style-type: none"> - 실시간 알림 → High 위험도 대상 발견 후 개입까지 걸리는 시간 60~80% 단축 구조 성공 가능성 증가, 기관 대응 성과 향상

3. 확장 로드맵 (3단계)

단기 : 3~6개월	1단계 : MVP 및 지자체·센터 적용 (단기, 3~6개월)		- SNS 크롤링 + 위험 스코어링 + 대시보드 + 알림 기능 운영
			- 자살예방센터, 청년정신건강센터 등과의 시범사업
			- 위험 패턴 데이터 축적 및 모델 고도화
중기 : 6~18개월	2단계 : B2G/B2B 확장 및 공공 협력 (중기, 6~18개월)		- 서울시·지자체 단위 통합 모니터링 시스템 구축
			- 교육청, 대학, 병원, 복지기관과 연계
			- 다국어 확장(청년·외국인 커뮤니티 감시)/위험도 기반 지역별 대응 매뉴얼 자동화 기능 추가
장기 : 18개월~	3단계 : 국가 단위 예방 플랫폼 및 국제 협력 확장 (장기 : 18개월~)		- 중앙정부·보건복지부(AI 기반 자살예방정책)에 연계 - WHO·OECD 정신건강 프로젝트 연동
			- 글로벌 SNS 플랫폼 협업(미공개 API/긴급 대응 프로토콜)
			- 전 세계적 자살예방 AI 모니터링 표준 모델 구축