

2025년 새싹 해커톤(SeSAC Hackathon) AI 서비스 기획서

팀명	vibe
팀 구성원 성명	정민성, 전민오

1. AI 서비스 명칭

plit - 시간 적응형 오디오-비주얼 하이브리드 콘텐츠 플랫폼

2. 활용 인공지능 학습용 데이터

	활용 데이터명	분야	출처
1	동영상 자막/스크립트 데이터	멀티미디어 콘텐츠 분석	Youtube, blog 등 공개 콘텐츠
2	이미지, 영상 데이터	시각자료 제공/동기화	영상 특정 구간의 프레임 캡쳐, 온라인 이미지
3	사용자 활동 데이터	사용자 시간 최적화, 게시물 추천 로직	사용자 활동을 통해 수집 (사용자 동의 기반)

3. 핵심내용

plit은 TTS 기반 음성을 중심으로 콘텐츠를 제공하는 새로운 형태의 미디어 플랫폼이다. 화면을 계속 보아야 하는 기존 영상 콘텐츠의 한계를 넘어, 콘텐츠를 듣기 중심으로 소비하고 필요에 따라 시각 정보를 선택적으로 확인할 수 있다.

또한, 1분~1시간 사이의 다양한 길이로 자동 변환하여 사용자의 상황에 맞는 최적의 소비 경험을 제공한다.

- 듣기 기반 중심 설계

영상처럼 화면에 둑이지 않으며, 이동, 운전, 운동 등의 상황에서도 소비 가능

- 선택적 시각 정보 제공

필요한 경우에만 이미지, 자료를 제공하여 이해도 집중력 향상, 일반적인 경우에는 이목을 집중시킬 수 있는 관련 이미지를 제공하여 시각적 흥미도 유발

- 시간 가변 콘텐츠 구조

짧은 시간에는 핵심 중심, 여유있는 시간에는 예시를 통해 이해도 높은 깊이있는 학습을 제공

- 크리에이터 친화적인 자동 제작 시스템

URL, 텍스트, 문서(파일) 형식만으로 콘텐츠를 제작하고 시각자료와 TTS가 자동 생성

4. 제안배경 및 목적

제안 배경

[사소한 것들의 문화학] OTT 서비스의 지배

사소한 것들의 문화학 김현경 전문 기자 입력 2025.06.06 12:35 수정 2025.06.06 23:48



HOME > Board > 보도자료
"소셜미디어, '숏폼 소비 플랫폼'으로 재정의 중"…인크로스 2025
아이엠 리포트

濡대정 | 쟁인 2025.11.01 10:44 | 댓글 0



스트리밍 서비스가 바꾼 영상 콘텐츠 소비 방식

숏폼 광고 선호, 인자 클릭 아이 구매로도

소셜미디어 내 콘텐츠 소비도 숏폼 중심으로

AI·영상적 의사결정 보조 도구로 자리매김



위 기사들을 보면 현대의 정보 전달 방식이 빠르게 다양해지고 있으며, 많은 사람들이 짧은 시간 안에 핵심 정보를 얻고자 한다는 점을 알 수 있다. 그러나 현재의 콘텐츠 소비 방식에는 다음과 같은 한계가 존재한다.

- 영상 기반은 화면 시청 필수, 멀티태스킹 어려움
- 블로그, 기사 등 텍스트 위주의 읽기 피로도 증가
- 팟캐스트는 시각자료 제한으로 인한 이해도 제한
- 공통적으로 콘텐츠 길이가 고정되어 있어 사용자의 요구 시간에 유연하게 대응하지 못함

목적

Plit은 현재의 콘텐츠 소비 방식을 벗어나 새로운 콘텐츠 소비 방식의 패러다임을 제시하고자 한다.

1. 듣기와 시각 효과가 모두 접목 된 새로운 콘텐츠 소비 구조 제공
2. 화면을 선택적으로 사용하는 시각 정보 자유 확보
3. 사용자 상황에 맞춘 시간 적응형 콘텐츠 선택
4. 이동, 운전, 운동 등 멀티태스킹 환경에서의 최적 경험 제공

5. 기술 콘텐츠 및 정보 콘텐츠의 이해도 향상

5. 세부내용

활용 데이터 및 AI 모델

plit은 텍스트 데이터, 시각 자료, 사용자 행동 데이터를 중심으로 AI 모델을 활용해 콘텐츠를 자동 생성·변환하는 구조를 갖는다.

- 1) 텍스트 기반 데이터: 텍스트 분석, TTS모델 활용
 - 텍스트 재작성 모델(GPT): 구어체로 변환, 핵심 요약을 시간에 맞게 생성
 - TTS모델(GPT-TTS): 자연스러운 음성 변환, 문장 단위 타임스탬프 생성
 - 활용 방식: 읽기 중심 콘텐츠 -> 오디오 중심 콘텐츠로 변환
- 2) 시각 자료 데이터: 멀티모달 분석 및 동기화
 - 멀티모달 분석 모델(Gemini): 스크립트와 시각자료 연결, 적절한 이미지, 영상 프레임 추천
 - 이미지 생성 모델(GPT, Gemini): 상황에 맞는 이미지 자료를 상황에 따라 효율적인 모델을 선정하여 이미지 생성
- 3) 사용자 행동 데이터: 시간 적응형, 개인화
 - 시간 적응형 모델(자체 AI): 사용자 상황(듣기 속도, 평균 사용 길이, 사용 시간대 등)에 맞는 길이를 자동으로 추천. 실제 재생 길이 세부 조정
 - 개인화 추천 모델: 선호 콘텐츠 유형, 길이, 시각 자료 사용량 분석

서비스 아이디어 및 구조

plit은 텍스트, 오디오, 시각 자료를 결합해 듣기 중심 방식의 콘텐츠를 제공한다.

사용자는 화면을 계속 볼 필요 없이, 필요한 순간에만 화면을 확인하면 된다.

- 1) TTS 기반 오디오 중심 재생
 - 콘텐츠의 기본 흐름은 오디오(음성) 중심
 - 화면이 깨진 상태에서도 모든 내용을 이해할 수 있도록 구성
- 2) 선택적 시각 자료 표시
 - 중요한 개념, 전문 지식이 등장할 때 AI로 이해를 돋는 상세한 시각자료 제공
 - 이미지·영상은 자동 동기화되어 매 순간 흥미를 유발하는 시각자료 제공
- 3) 시간 가변 콘텐츠 제공
 - 1분~1시간 사이의 난이도 및 길이별 자동 재편집

- 일정량 학습 후에는 자동으로 사용자 맞춤형 길이의 콘텐츠 제공
- 4) 자동 제작 도구 기반 크리에이터 워크플로우
 - URL이나 텍스트 혹은 파일만 업로드하면 스크립트, 이미지, TTS가 자동 생성
 - AI가 콘텐츠 구조와 흐름까지 자동 조정
 - 5) 사용자 상태 기반 인터랙션
 - 화면 켜짐 상태에 따라 시각 자료 자동 제어(리소스 사용 제어)
 - 챕터 단위 이동, 재생 속도 조절, 북마크 기능 등 재생 편의 기능 제공

서비스 방법 및 기술 적용

전체 기술 흐름: 텍스트 입력 -> 스크립트 변환 -> TTS 생성 -> 시각 자료 매칭 -> 시간 변환 -> 개인화 추천

- 1) 콘텐츠 변환
 - 텍스트 분석 AI가 문장을 정리하고 듣기 좋은 형태로 재작성
 - TTS 모델이 자연스러운 음성으로 변환
 - 멀티모달 모델이 문장별로 필요한 시각 자료와 매칭
- 2) 시간 적응형 콘텐츠 생성
 - 전체 내용을 1분~1시간 단위로 재배열
 - 핵심만 유지하거나 예시, 설명, 이미지 비중을 조절해 시간별 내용 깊이 차별화
- 3) 크리에이터 도구 적용
 - URL, 텍스트, 문서 업로드 시 AI가 자동으로 콘텐츠 분류
 - 타임라인 기반 편집 UI 제공
 - 시각 자료, TTS, 구간 전환 시점을 자동 배치
- 4) 사용자 맞춤 기반 콘텐츠 제공
 - 사용자 패턴(길이 선호, 사용 시간대, 시청 방식)을 지속적으로 학습
 - 콘텐츠 추천, 재생 길이 조정, 시각 자료 노출 빈도를 개인화

서비스의 창의성 및 구현 가능성

창의성

- 듣기 중심 콘텐츠와 선택적 시각 자료라는 새로운 미디어 구조
- 콘텐츠 길이가 고정되지 않고 사용자 상황에 따라 자동 변환
- 영상·오디오·텍스트의 장점을 결합한 하이브리드 형태
- 크리에이터는 텍스트만으로 영상급 콘텐츠 생성 가능
- 기존의 정적 정보 전달과는 차별화 된 서비스

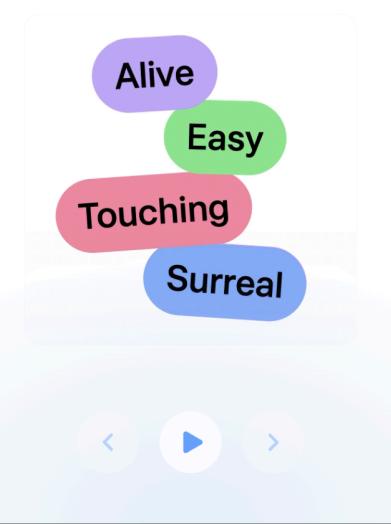
구현 가능성

- 텍스트 분석·TTS·이미지 분석 기술은 기존의 여러 API나 오픈소스 모델을 통해 구현이 가능하며 자체 개인화 AI모델과 유기적으로 연결되어 서비스에 최적화 된 AI제공 가능
- 초기 모델에서 텍스트→오디오 변환과 단순 이미지 동기화만 구현해도 서비스 가치 충분
- 새로운 미디어, 콘텐츠 구조를 제공한다는 점에서 무한한 확장성을 가짐
- 웹, 모바일, TV 등 멀티 플랫폼 확장도 기술적으로 가능함

서비스의 예상 UI/UX

미디어 재생 UI

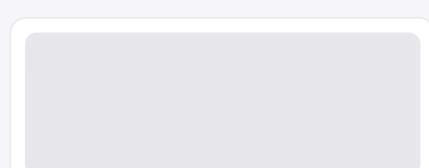
- TTS소리를 텍스트 애니메이션으로 표현하여 시각적인 흥미 유발
- 상황에 따라 적절한 이미지, GIF, 자료를 제시하여 흥미 유발과 함께 이해도 향상
- 섹션별 구분으로 앞/뒤로 넘겨가며 소비 가능
- 시간 최적화 내용 구성으로 사용자의 목적에 따라 정보 및 맞춤 자료 제공



콘텐츠 목록

긴 강의를 듣기 좋은 길이로 재구성한 폴리 목록입니다.

[전체](#) [학습](#) [기술](#) [인터뷰](#) [즐겨찾기](#)



콘텐츠 목록 UI

- 카테고리별 정렬
- 맞춤형 플레이리스트 형식으로 여러개의 콘텐츠를 제한된 시간 내에 재구성하여 소비 가능
- 커뮤니티를 통해 사용자간의 커뮤니케이션을



활성화하고 전문 지식을 공유할 수 있음

시연 영상: <https://youtu.be/W0KC95DEXc4>

6. 기대효과

사회적 효과

- 1) 이동, 운동, 운전 등 상황에서의 정보 접근성 확대
 - 기존 영상, 텍스트 기반 콘텐츠는 이동 과정에서 소비하기가 어려움
 - plit은 듣기 및 선택적 보기 중심으로 이동, 운동, 운전등의 상황에서도 소비 가능
 - 현대인들의 짧은 빈시간까지 효율적으로 활용할 수 있음
- 2) 정보 접근성 및 포용성 향상
 - 시각 장애인도 오디오 중심으로 동일한 콘텐츠 소비 가능
 - 텍스트 기반 자료를 음성, 시각 자료로 재구성하여 정보 격차 감소
 - 화면을 보기 어려운 상황에서도 정보 소비 가능
- 3) 콘텐츠 소비 학습시간 최적화
 - 1분~1시간까지 시간 여유가 없는 사람들에게까지 맞춤형 정보 제공
 - 짧은 시간에는 핵심 요약, 여유 있는 시간에는 심화 설명을 통해 학습 효율 상승 및 정보 과부하 방지
- 4) 지식/정보 콘텐츠의 대중화
 - 크리에이터가 텍스트 만으로 콘텐츠 제작이 가능하여 다양한 분야의 전문가가 영상 촬영 및 편집 능력이 없어도 자료 제작 가능
 - 전문 지식이 더 쉽게 오디오, 시각 혼합 자료로 재구성되어 전문지식의 대중화 기여

5) 새로운 미디어 소비 문화 형성

- 보는 콘텐츠에서 듣고 필요에 따라 시각적인 자료까지 볼 수 있는 콘텐츠로 변화
- 짧은 시간 단위의 콘텐츠 소비 패턴 확장되며 빠르고 정확한 정보 전달 문화 강화
- 오디오와 비주얼이 결합된 새로운 형태의 콘텐츠 시장 형성

경제적 효과

1) 미디어 제작 비용 절감

- 영상 촬영/편집 없이 텍스트와 내/외부 자료만으로 콘텐츠 생산 가능(진입장벽 하락)
- 이미지 생성, 적합한 이미지 제공 기능으로 자료조사 및 디자인 비용 절감

2) 콘텐츠 소비량 증가

- 1분~1시간 사이로 원하는 정보를 얻을 수 있음
- 다양한 시간대에 적합한 콘텐츠 공급

3) B2B 시장 진입 가능성

- 교육 플랫폼, 기관과 학교 등 정보 제공이 필요한 곳을 대상으로 맞춤형 서비스 제공
- 텍스트, 오디오 시각과 솔루션으로 SaaS 비즈니스 확장 가능

4) 구독 및 프리미엄 확장성

- 기본형 오디오 콘텐츠 제공, 프리미엄형 고품질 TTS와 시각자료 등으로 B2C 수익 구조 구축
- 시간 적응형 콘텐츠로 재생량 상승 -> 재생시간 및 구독 전환율 상승

5) AI기반 재해석 콘텐츠 시장 창출

- 기존의 콘텐츠를 재구성하여 2차 콘텐츠 제공 가능
- 크리에이터 수익 모델 다각화
 - > 원작자와의 수익 배분
 - > 자신의 창작물을 다양한 형태로 재배포

6) 글로벌 시장 확장성

- TTS, LLM 기반 서비스로 다국어 지원이 용이함
- 녹음, 촬영이 필요없는 텍스트 기반 생성 방식은 전 세계 어디서나 동일한 방식으로 이용 가능

* 상세 설명을 위해 도표, 스케치 등 별도파일 추가 가능

* 제출한 기획서는 온라인 예선 심사 전 구체화하여 깃허브(GitHub)에 필수로 게시