

AI 음성 메모 · 자동 정리 SaaS · 풀스택 개발

## LogTalk

# AI 기반 음성 메모 & 자동 정리 서비스

— 회의·강의·대화를 녹음하면 AI가 텍스트로 변환하고 자동 정리해 주는 B2C 음성 메모 서비스 —

AI / 음성인식 STT

Next.js 15 · React 19

OAuth · 포인트 결제

도메인 AI 음성인식 · B2C SaaS · 생산성

담당 역할 기획 · UI/UX · 풀스택 개발

개발 기간 약 3개월

규모 20+ 화면 · 30+ API · 40+ 컴포넌트

Phase 1 개발 완료

라이브 운영 · logtalk.trenet.kr

## TABLE OF CONTENTS



## 목차

1	프로젝트 개요 — 서비스 · 타겟 · 작업 범위	03
2	서비스 메인 화면 — 업로드 & 변환 진입	04
3	주요 기능 — STT · AI 정리 · 결제 등 6대 기능	05
4	변환 결과 · 화자 분리 화면	06
5	기술적 주안점 — 엔지니어링 차별점	07
6	프로젝트 배경 — 문제 · 목표 · 해결	08
7	기술 스택 · 아키텍처	09
8	진행 단계 — 8단계 · 3개월	10
9	프로젝트 성과 — 정량 · 정성 · 품질	11
10	라이브 데모 · 유사 프로젝트 경험	12

## OVERVIEW

1

## 프로젝트 개요

회의·강의·대화를 녹음하면 AI가 자동으로 텍스트로 변환하고 깔끔하게 정리해 주는 **B2C 음성 메모 서비스**입니다. 업로드하면 **고정밀 STT 변환 → 화자 분리 → 자동 정리**까지 한 흐름으로 처리합니다.

## ■ 핵심 가치

- **변환을 넘어 정리까지** — 단순 받아쓰기가 아니라 안건·결정사항·핵심 개념까지 문서로 완성
- **누가 말했는지까지** — 화자 분리로 발언 주체를 구분해 그대로 읽히는 회의록 생성
- **사용한 만큼만** — 포인트 기반 과금으로 부담 없이 필요한 만큼 이용

20+

개발 화면

30+

API 엔드포인트

40+

UI 컴포넌트

## ■ 메인 타겟



## 직장인

회의를 자주 하고 빠른 기록이 필요한 실무자.



## 콘텐츠 크리에이터

녹음 기반 기록·정리가 필요한 제작자.



## 학생 · 강사 · 컨설턴트

강의·상담을 문서화하는 전문가.

## ■ 작업 범위

- **기획 · 설계** — 음성 처리 기술 조사, 기능 명세, 와이어프레임, 정보구조(IA) 설계
- **UI/UX · 프론트엔드** — 반응형 · 다크모드 디자인, Next.js 15 · React 19 · TypeScript · Tailwind
- **백엔드 · 연동** — Next.js API Routes · Express, JWT 인증, Soniox · GPT-4 · OAuth · PortOne 연동
- **배포 · 운영** — Vercel 배포, Docker, AWS ECS(예정) / 반응형 웹(PC · 태블릿 · 모바일)

## ■ 서비스 카테고리

분류

AI · 음성인식

B2C SaaS

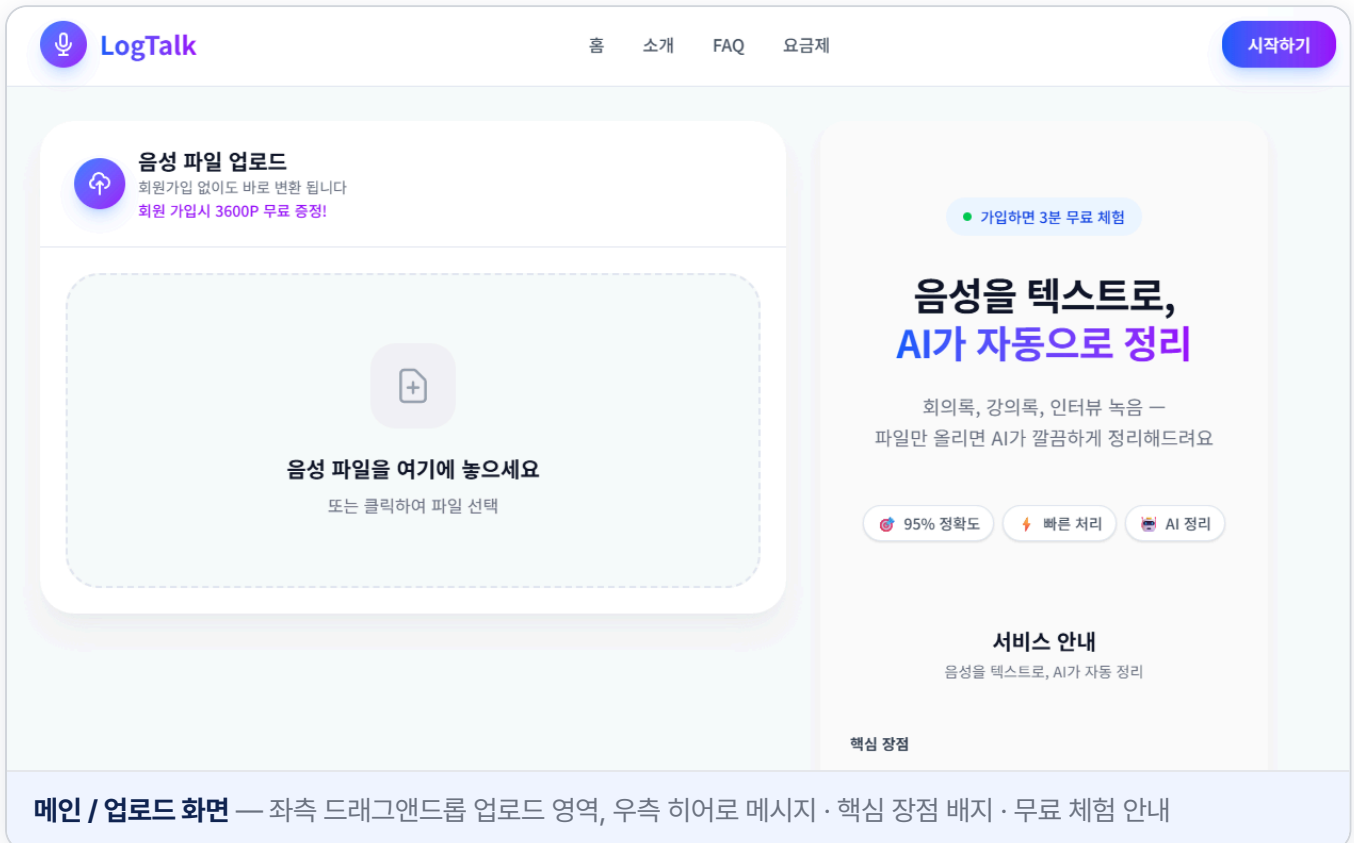
생산성 도구

STT · 화자 분리

## SERVICE MAIN SCREEN

## 2 서비스 메인 화면

처음 방문한 사용자가 복잡한 설정 없이 음성 파일을 올리는 즉시 변환을 시작할 수 있도록 설계한 진입 화면입니다. 드래그앤드롭 업로드와 가치 제안을 한 화면에 배치했습니다.



## ■ 화면 구성

- **즉시 업로드** — “음성 파일을 여기에 놓으세요” 드래그앤드롭, 회원가입 없이도 바로 변환 진입
- **가치 제안** — “음성을 텍스트로, AI가 자동으로 정리” 히어로 + 95% 정확도 · 빠른 처리 · AI 정리 배지
- **가입 혜택** — 회원 가입 시 3,600P 무료 증정, 가입하면 3분 무료 체험
- **반응형 · 다크모드** — PC · 태블릿 · 모바일 대응, 다크모드 지원

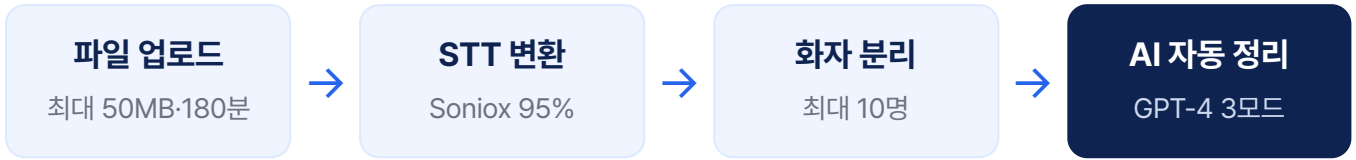
## ■ 레이아웃 · 내비게이션

- **상단 내비게이션** — 홈 · 소개 · FAQ · 요금제 메뉴와 우측 ‘시작하기’ CTA로 핵심 동선 고정
- **2단 구성** — 좌측 ‘작업(업로드)’, 우측 ‘가치 제안·서비스 안내’로 한눈에 이해되는 분할

FEATURES

### 3 주요 기능

■ 음성 → 정리 핵심 파이프라인



■ 6대 기능

<p></p> <p><b>고정밀 음성 인식(STT)</b></p> <p>Soniox API · 한국어 95% 정확도. 6개 오디오 형식 지원, 5초 폴링으로 실시간 진행 상태 확인.</p>	<p></p> <p><b>AI 텍스트 자동 정리</b></p> <p>OpenAI GPT-4 · 회의록(안건·결정·액션) / 강의록(챗터·키워드) / 대화록(흐름·요약) 3모드.</p>	<p></p> <p><b>대용량 파일 분할</b></p> <p>ffmpeg.wasm 브라우저 분할 · 60분 청크 Codec Copy, 여러 청크 병렬 업로드로 처리 가속.</p>
<p></p> <p><b>OAuth 소셜 로그인</b></p> <p>Google OAuth 2.0(Passport.js), Kakao(심사대기). JWT 15분 + Refresh 7일, 로그인 유지.</p>	<p></p> <p><b>포인트 결제 시스템</b></p> <p>PortOne 위젯 · 카드 · 카카오페이 · 네이버페이 · 계좌이체. 서버사이드 포인트 관리.</p>	<p></p> <p><b>변환 내역 관리</b></p> <p>변환 히스토리 저장, 태그 · 메모 추가, 검색 · 필터링, 포인트 사용 내역 조회.</p>

■ 지원 형식 · 사양

**지원 형식**

MP3   WAV   M4A   AAC   OGG   FLAC

---

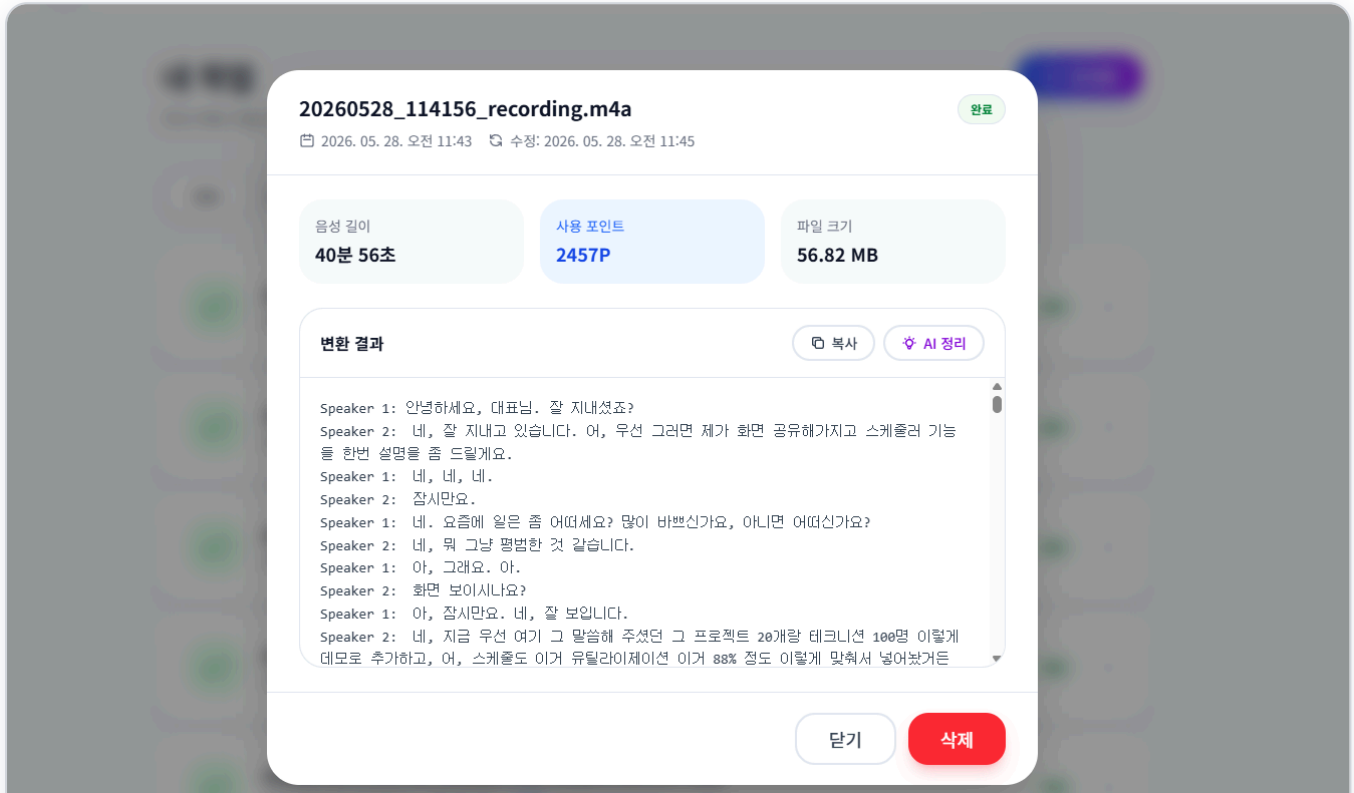
**처리 사양**

최대 50MB · 180분   화자 최대 10명   정리 3모드   결제수단 4종   TXT · PDF 내보내기

## TRANSCRIPTION RESULT

## 4 변환 결과 · 화자 분리

변환이 끝나면 화자별로 구분된 전체 텍스트와 음성 메타데이터를 한 화면에서 확인하고, 버튼 한 번으로 AI 정리 · 복사 · 내보내기까지 이어집니다.



**변환 결과 화면** — 음성 길이 · 사용 포인트 · 파일 크기 메타데이터 + Speaker 1/Speaker 2 화자 분리 텍스트 + 복사 · AI 정리 액션

## ■ 화면 구성

- **화자 분리 출력** — “Speaker 1 / Speaker 2 ...” 형태로 발언 주체를 자동 구분해 그대로 읽히는 대화록
- **음성 메타데이터** — 음성 길이(40분 56초) · 사용 포인트(2,457P) · 파일 크기(56.82MB) 표기
- **원클릭 후처리** — ‘복사’와 ‘AI 정리’ 버튼으로 변환 결과를 즉시 정리 · 재사용
- **작업 상태 추적** — 완료 배지와 생성 · 수정 시각으로 변환 이력 관리

## ENGINEERING HIGHLIGHTS

## 5 기술적 주안점

단순 API 연동을 넘어, 서버 비용 · 보안 · 안정성을 좌우하는 핵심 지점을 직접 설계 · 해결했습니다.

## 1 브라우저 기반 파일 처리 — 서버 부하 50% 감소

ffmpeg.wasm으로 대용량 음성을 브라우저에서 직접 분할. 재인코딩 없이 Codec Copy 방식으로 처리해 서버 처리량을 절반으로 절감.

## 2 화자 분리(Speaker Diarization)

Soniox Speaker Diarization으로 화자별 발언을 자동 구분, 최대 10명까지 "Speaker 1 · Speaker 2 ..." 형태로 분리 출력.

## 3 서버사이드 보안 — 포인트 조작 차단

포인트는 서버에서만 관리하고 JWT에 포함하지 않아 클라이언트 측 조작을 원천 차단.

## 4 한글 인코딩 처리

TextDecoder + encodeURIComponent 조합으로 JWT 내 한글을 완벽 지원, 깨짐 없는 토큰 처리.

## 5 결제 멱등성(Idempotency)

orderId 기반 중복 결제 방지 처리로 동일 주문의 이중 결제 이슈를 원천 차단.

## 6 인앱 브라우저 대응

카카오톡 · 인스타그램 인앱 브라우저를 감지해 OAuth가 정상 동작하는 외부 브라우저로 리다이렉트.

## 7 성능 최적화

Next.js SSR/SSG, 이미지 최적화, 코드 스플리팅으로 초기 로드와 렌더링 성능 확보.

## ■ 안정성 · 호환성

- 비동기 폴링 처리 — 5초 폴링으로 대용량 변환의 진행 상태를 실시간 추적, 타임아웃 회피
- 광범위한 코덱 호환 — MP3 · WAV · M4A · AAC · OGG · FLAC 6종 입력을 일관되게 처리
- 크로스브라우저 검증 — 주요 브라우저 테스트로 ffmpeg.wasm · OAuth 동작 안정성 확보

## ■ 한눈에 보는 효과

## 효과 지표

서버 부하 50% ↓

화자 최대 10명

JWT 한글 지원

중복 결제 차단

인앱 OAuth 대응

## BACKGROUND

## 6

## 프로젝트 배경

기존 STT 도구는 “받아쓰기”에서 멈춰 정리는 사용자 몫이었습니다. LogTalk은 변환부터 정리·과금 까지를 하나의 흐름으로 묶어 이 빈틈을 메우는 데서 출발했습니다.

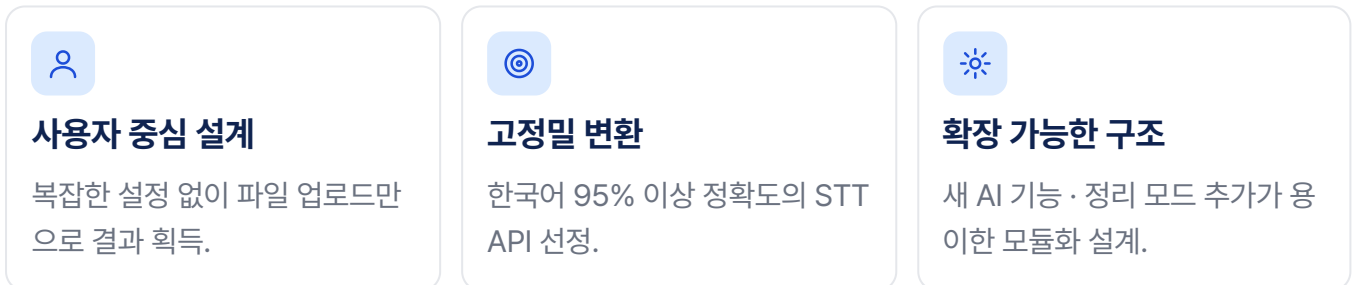
## ■ 문제점

- 수기 정리의 부담 — 회의·강의 후 직접 정리하는 데 많은 시간 소요
- 변환만 하는 STT — 기존 서비스는 단순 텍스트 변환만, 정리는 사용자 몫
- 대용량 처리 한계 — 큰 음성 파일에서 서버 부하 및 타임아웃 이슈
- 낮은 가독성 — 화자 구분 없이 텍스트만 나열되어 읽기 어려움

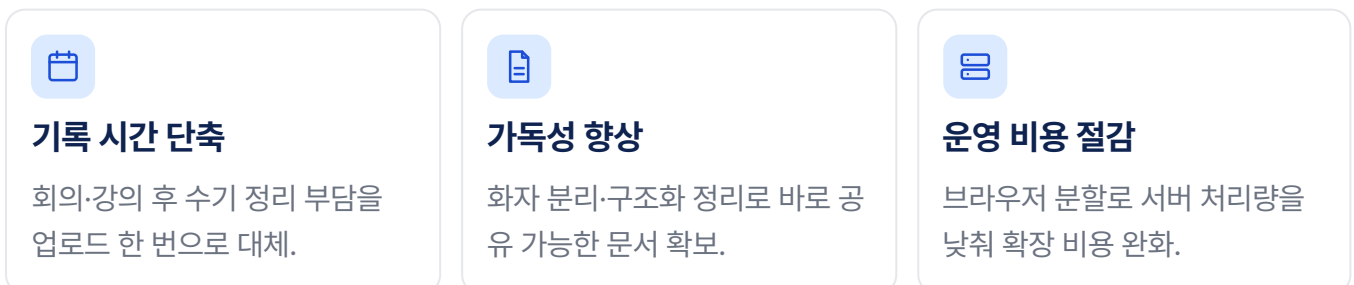
## ■ 목표 → 해결 방안



## ■ 핵심 설계 원칙



## ■ 기대 효과



TECH STACK & ARCHITECTURE

# 7 기술 스택 · 아키텍처

## FRONTEND

- Next.js 15.5.4
- React 19
- TypeScript 5.9
- Tailwind CSS 4.1
- Radix UI
- ffmpeg.wasm 0.12
- React Markdown

## BACKEND

- Next.js API Routes
- Express.js 5.1
- Passport.js
- jsonwebtoken
- bcryptjs
- multer

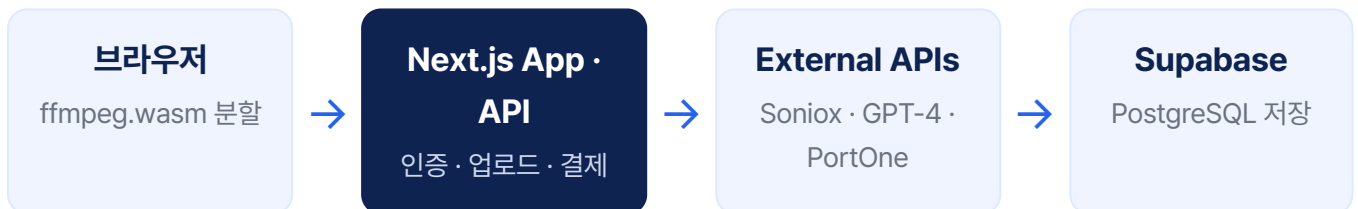
## DATABASE / INFRA

- Supabase (PostgreSQL)
- Vercel
- Docker
- AWS ECS (예정)

## EXTERNAL APIS

- Soniox STT
- OpenAI GPT-4
- Google OAuth 2.0
- Kakao OAuth
- PortOne 결제

### ■ 처리 아키텍처



### ■ 기술 선택 근거

<p><b>👤</b></p> <p><b>Next.js 풀스택</b></p> <p>프론트 · API를 한 코드베이스로 일원화, SSR/SSG로 초기 로드 최적화.</p>	<p><b>🗄️</b></p> <p><b>Supabase</b></p> <p>PostgreSQL 기반 데이터 영속화와 인증 · 스토리지 통합 운영.</p>	<p><b>☀️</b></p> <p><b>ffmpeg.wasm</b></p> <p>클라이언트 분할로 서버 부하 절감, 재인코딩 없는 초고속 처리.</p>
---	--	---

\* 음성 분할은 클라이언트(ffmpeg.wasm)에서 수행해 서버 부하를 낮추고, 변환 · 정리 · 결제 결과는 Supabase에 영속 저장.

PROCESS

# 8 진행 단계



## 1 기획 · 설계 (3주)

STT API 비교 분석, 경쟁 서비스 벤치마킹, 기능 명세 · 정보구조(IA) · 와이어프레임 · DB 스키마, 디자인 시스템 · 반응형 · 다크모드

## 2 프론트엔드 개발 (4주)

컴포넌트 개발, 페이지 구현, ffmpeg.wasm 브라우저 파일 처리 로직

## 3 백엔드 · 결제 연동 (4주)

API 개발, OAuth 연동, 외부 API 통합, PortOne 결제 위젯 · 웹훅 · 포인트 시스템

## 4 테스트 · 배포 (1주+1일)

기능 · 크로스브라우저 테스트, 버그 수정, Vercel 배포 · 도메인 연결 · 운영 환경 구성

### ■ 진행 방식

#### 커뮤니케이션

주 2회 정기 미팅 및 진행 상황 리포트 제공.

#### 일정 관리

마일스톤 기반 일정 관리, 주요 기능별 데모 제공.

#### 품질 관리

TypeScript 기반 코드 품질 확보, 코드 리뷰.

\* 총 예상 기간 약 3개월 (1인 개발, Phase 1 기준).

RESULTS

## 9 프로젝트 성과

### ■ 정량 성과 (규모 기준 · 예상)

**3개월**

개발 기간

1인 개발 · Phase 1

**20+**

개발 페이지

메인·대시보드·결제·내역

**30+**

API 엔드포인트

인증·음성처리·결제·관리

**40+**

UI 컴포넌트

재사용 가능한 컴포넌트

**50%**

서버 처리량 절감

ffmpeg.wasm 브라우저 분할

### ■ 정성 성과

- 서버 부하 감소 — ffmpeg.wasm 브라우저 분할로 서버 처리량 50% 절감
- 안정적인 인증 — JWT + Refresh Token으로 토큰 탈취 시에도 피해 최소화
- 결제 안정성 — 멍등성 처리로 중복 결제 이슈 원천 차단
- 확장 가능한 아키텍처 — 새로운 AI 기능 · 정리 모드 추가 용이
- 유지보수 용이성 — TypeScript 엄격 모드, 관심사 분리로 코드 품질 확보

### ■ 품질 지표



#### 보안

서버사이드 포인트 관리, JWT + Refresh Token, 결제 멍등성으로 조작 · 중복 차단.



#### 운영 안정성

Vercel 자동 배포, 비동기 폴링 기반 변환 상태 추적, Docker 컨테이너화.

### ■ 향후 계획 (Roadmap)

- Kakao 로그인 정식 오픈 — 개발 완료 후 심사 대기 중인 카카오 OAuth 활성화
- 백엔드 배포 확장 — AWS ECS 기반 백엔드 배포로 처리량 · 가용성 강화
- AI 기능 확장 — 신규 정리 모드 추가 및 요약 · 키워드 추출 고도화

LIVE DEMO & EXPERIENCE

# 10 라이브 데모 · 유사 경험

## ■ 라이브 데모

 **서비스 바로가기**





<https://logtalk.trenet.kr>

- 회원가입 후 **무료 포인트**로 음성 변환 체험 가능 (가입 시 3,600P 증정)
- Google 계정으로 **간편 로그인** 지원
- 샘플 음성 파일로 **STT 변환 + AI 정리** 기능 테스트 가능

## ■ 데모 체험 흐름



## ■ 유사 프로젝트 경험

 <p><b>AI/ML API 연동</b> OpenAI · 음성인식 API 등 AI 모델 연동 경험.</p>	 <p><b>B2C SaaS 플랫폼</b> 사용자 대상 SaaS 서비스 기획 · 개발 경험.</p>
 <p><b>OAuth · 결제 구축</b> 소셜 로그인 및 결제 시스템 구축 경험 다수.</p>	 <p><b>대용량 파일 처리</b> 대용량 파일 분할 · 스트리밍 처리 경험.</p>

## ■ 활용 시나리오

 <p><b>회의록 자동화</b> 녹음만으로 안건·결정사항·액션 아이템 정리.</p>	 <p><b>강의 · 세미나 노트</b> 챕터·핵심 개념·키워드로 구조화 된 강의록.</p>	 <p><b>인터뷰 · 상담 정리</b> 화자 분리로 발언 주체가 명확한 대화록.</p>
--	--	--