

대규모 엔터프라이즈의 안정성부터 상용 멀티모달 AI 에이전트 까지 설계하는 테크 리더

2001년 대한민국 전자정부(G4C), 청와대, 금융감독원 시스템부터 대기업(LG전자, KB국민은행, 현대기아차), 글로벌 파트너십(일본 반다이남코)의 대규모 서비스 설계를 이끌어왔습니다. 최근에는 **Google Gemini, ElevenLabs, Web Audio API, ReportLab**을 융합하여 상용 수준의 **생성형 AI 에이전트 3종**을 성공적으로 상용화하고, 인프라 병목(CPU/Memory Full) 해결 및 클라우드(AWS/KT) SRE 안정화를 주도했습니다.

20+ 년

소프트웨어 & 아키텍처 경력

80% 절감

클라이언트 오디오 믹싱 서버 부하 경감

95%+

하이브리드 CV + LLM 데이터 정제율

27명

개발팀 총괄 디렉팅 조직 규모

ENGINEERING EXCELLENCE

🔧 심층 트러블슈팅 & 기술적 의사결정 (Case Studies)



주기적 CPU & 메모리 Full 장애 해결

소속: 에듀랩코리아 | 인프라 SRE 최적화

문제: 주기적으로 서버 CPU 및 메모리가 100%에 도달하여 서비스가 중단되는 현상 발생.

원인 분석: 비동기 파일 가공 태스크 중 커넥션 미반환 누수와 PostgreSQL Connection Pool 고갈, 프로세스 미회수 확인.

해결책 & 성과: psycopg2 커넥션 풀 재설계 및 타임아웃/컨텍스트 매니저 엄격 적용, 리소스 프로파일링 후 백그라운드 태스크 격리 분리로 서버 다운 0건 달성.



클라이언트 오디오 믹싱으로 서버 부하 80% 경감

프로젝트: VoiceMake AI 오디오 에이전트

문제: 다중 화자 음성 합성 시 서버에서 모든 오디오 조각을 병합할 경우 CPU 병목 및 대역폭 비용 급증.

기술 결정: 브라우저 내장 Web Audio API (OfflineAudioContext)를 활용하여 클라이언트 메모리 내에서 청크 믹싱 및 페이드 인/아웃 실시간 처리.

성과: 서버 연산 부하 80% 이상 절감 및 사용자 즉각 청취(체감 레이턴시 70% 단축).



OpenCV Inpaint vs PyMuPDF Redaction

프로젝트: 워크북 PDF 자동 변환 시스템

문제: 페이지 번호 삭제 시 기존 OpenCV Inpaint 방식은 배경 일러스트와 벡터 그래픽을 덧칠하여 훼손 발생.

기술 결정: PDF 내부 벡터 오브젝트와 텍스트 스트림을 직접 분리 제어하는 PyMuPDF Redaction 라이브러리 개발.

성과: 배경 벡터 그래픽 100% 무손실 보존 + 텍스트 레이어만 정밀 삭제 성공.

FEATURED ARCHITECTURE

 핵심 프로젝트 & 아키텍처 파이프라인

동화책 자동 제작 AI 에이전트 (Storybook Generation)

DOCX 원고 파싱 → Gemini 일러스트 자동 생성 → 실시간 슬라이드 에디터 → 300 DPI 출판용 PDF 렌더링

Next.js 16

React 19

Zustand

FastAPI

Google Gemini

ReportLab (300 DPI)

🔗 END-TO-END PIPELINE ARCHITECTURE

1. 대본 파싱 & 스타일 추출

python-docx 대본 파싱 + PyMuPDF 레퍼런스 PDF 스타일 분석

2. Gemini AI 일러스트

회사 보유 IP 가이드라인 프롬프트 + 고품질 삽화 자동 생성

3. 실시간 에디터 & 롤백

Next.js 16 + Zustand 기반 형상 관리 및 즉시 롤백 UI

4. 300 DPI 인쇄용 렌더링

ReportLab 엔진 기반 출판 규격 (1287x853pt) 고품상도 PDF 비동기 출력

🔑 핵심 문제 해결

수작업 삽화 제작 및 출판 편집에 소요되던 수주의 리드타임을 5분 이내로 단축. 출판 인쇄소에 그대로 발발 가능한 고품상도 규격을 백엔드 비동기 태스크(aiofiles)로 가용성 확보.

🔧 주요 기술 의사결정

- 프롬프트 인젝션 및 IP 이탈 방지를 위한 시스템 프롬프트 엔지니어링 설계.
- psycopg2 Connection Pool 튜닝 및 AWS S3 자동 백업 연동.

VoiceMake — AI 음성 & 오디오 생성 에이전트

Gemini 2.5 Flash 구조화 출력 → ElevenLabs TTS → 브라우저 Web Audio API 클라이언트 믹싱

Next.js 16

Web Audio API

ElevenLabs API

Gemini 2.5 Flash

Node.js VM

AUDIO PROCESSING PIPELINE

1. LLM 화자 구조화 분석

Gemini 2.5 Flash
Structured
Outputs로 대사/나
레이션/감정 분리

2. 멀티 화자 음성 합성

ElevenLabs API 연
동으로 캐릭터별 차별
화된 보이스 합성 청
크 생성

3. Web Audio API 믹싱

브라우저
OfflineAudioContext
상에서 무음 간격 보
정 및 Fade 클라이언
트 렌더링

4. Node VM MP3 인코딩

Node.js vm 모듈로
브라우저 전용
lamejs를 백엔드 격
리 바인딩 고속 압축

💡 핵심 성과

대본 화자 매핑 자동화로 제작 공수 90% 이상 단축. 서버 오디오 트랜스코딩 연산 부하 80% 경감.

🔒 보안 및 세션 아키텍처

외부 솔루션 없이 HMAC-SHA256 서명 기반 8시간 유효 쿠키 세션 자체 설계 및 PostgreSQL 세션 복구 이력 적재.

워크북 PDF → 인터랙티브 웹 교재 자동 변환 GUI 툴

OpenCV 윤곽선 검출 + Gemini Vision 하이브리드 파이프라인 & Prompt Ops 검수 시스템

Python 3.11

OpenCV

PyMuPDF (fitz)

Gemini Vision

CustomTkinter

🔧 기술 구현 내용

- **하이브리드 CV + Vision AI**: HSV 색 공간 마스킹 및 Contour 검출로 스티커 투명 PNG 분리 + Gemini Vision으로 노이즈 파편 필터링 및 앵커 텍스트 매핑.
- **무손실 PDF Redaction**: 벡터 일러스트 보존하며 페이지 번호만 정밀 제거 (데이터 정제율 95%+).

🔥 데스크톱 GUI & Prompt Ops

CustomTkinter + DnD 드래그앤드롭 데스크톱 앱 구축. 내장 로컬 웹서버(http.server) 및 관리자가 "22 페이지 제외해줘"라고 자연어로 지시하면 AI가 메타데이터를 수정해 재빌드하는 Prompt Ops 피드백 루프 탑재.

Cloud & SRE

클라우드 서버 구성 및 시스템 유지보수

AWS
BastionKT
CloudDocker /
Nginx

AWS Bastion 보안망

VPC 내부 격리 서브넷 환경 및 Bastion Host 경유 접근 통제 구축으로 보안 감사 요건 충족.

하이브리드 백업 체계

KT클라우드와 AWS 멀티 클라우드 연동, S3 자동 적재 및 사내 백업 이중화 파이프라인.

인프라 안정화 & 튜닝

Nginx 역방향 프록시, Docker 가상화 및 주기적 장애 원인 프로파일링을 통한 고가용성 보장.

대규모 엔터프라이즈 포털 & IoT 시계열 파이프라인

TICK
StackApache
NiFiTekken 7
API

포항시 전기버스 TICK 모니터링

Telegraf, InfluxDB, Chronograf, Kapacitor 기반으로 LTE 환경 배터리 센서 데이터 실시간 스트리밍 및 OCPP 2.0 규격 충족.

LG전자 포털 & Apache NiFi ETL

개발 리더(DL), Spring Boot 백엔드 및 Apache NiFi를 통한 Oracle to MySQL 대용량 데이터마이닝 파이프라인 구축.

일본 반다이남코 Tekken 7 연동

일본 본사 개발진과 직접 통신 아키텍처 협업. 바나패스포트 전적 연동 및 파이트머니 기반 인게임 아이템 웹 커스터마 이징 구현.

CAREER HISTORY

🏆 20+년 커리어 타임라인 (Professional Experience)

에듀랩코리아 | Lead AI Systems Architect & DevOps

최근

영유아 교육용 AI 에이전트 3종(동화책 생성기, VoiceMake, 워크북 변환기) 아키텍처 설계 및 풀스택 개발 총괄. KT클라우드/AWS 인프라 구성, Bastion 보안 환경 및 주기적 시스템 장애(CPU/Memory Full) 원인 해결.

프리랜서 테크 리더 | Principal Software Engineer

2025.01 ~ 2025.08

- (주)엘에스엠트론 (마이파머스앱): FastAPI + RAG + LLM 활용 데이터 가공 파이프라인 및 Flutter 하이브리드 앱 연동.
- 화재 시험 성적서 웹 관리 시스템: Next.js + WebSocket 기반 실시간 동시 수정 권한 제어 구현.
- 알림모아 (푸시알림 관리 앱): Flutter 안드로이드 앱 개발 및 Python Flask + AI STT/TTS 백엔드 연동.

(주)포스팻 | 개발총괄 (이사/본부장)

2021.07 ~ 2025.01

- LG전자 사내 포털: 개발 리더(DL), Spring Boot 백엔드 및 Apache NiFi 기반 대용량 ETL 파이프라인 구축.
- AI 이미지 생성 서비스: Stable Diffusion 기반 백엔드(FastAPI) 및 프론트엔드 구축.
- KB국민은행 서여의도 본부: 자리 배치 시스템 설계 및 Docker Compose 기반 CI/CD 자동 배포 체계 수립.
- 현대기아차 디자인연구소: 지식공유/브레인스토밍 시스템 오픈소스 아키텍처 구축.

- 포항시 전기버스 배터리 모니터링: TICK Stack (Telegraf, InfluxDB, Chronograf, Kapacitor) 및 MQTT 실시간 시계열 파이프라인 구축.
- 서산 태양광 EMS & EV 충전기: OCPP 2.0 규격 총족 전기차 충전기-서버 간 WebSocket 양방향 통신 시스템 구축.

- 일본 반다이남코 Tekken 7 온라인 서비스 연동 시스템 아키텍처 설계 및 개발 총괄.
- 신발 디자이너 글로벌 오픈 플랫폼, 모바일 게임 퍼블리싱 포털 AWS 인프라 설계/운영.

- 드림어스: 대한민국 전자정부(G4C), 청와대 대표 홈페이지, 금융감독원 리뉴얼, 국민카드 여행 패밀리 사이트 구축.
- CCR: 대작 MMORPG "RF온라인", "포트리스2 블루" 글로벌 운영 지원 및 백오피스 개발.
- GSP인터렉티브: 온라인 TPS "헤썬" 게임 포털 및 런처/패쳐 개발 총괄.
- KTH / 노리공방: 영국 축구게임(FMO) 웹 컨버전 및 중국 마켓 액션 RPG 게임 서버 개발 총괄.

LEADERSHIP & PHILOSOPHY

기술 리더십 & 엔지니어링 철학

27명 조직 총괄 & 프로세스 정립

개발 총괄 및 기술 책임자로서 27명 규모의 개발팀을 리드하며, GitLab Registry, Redmine, Slack을 통한 애자일 스프린트와 표준화된 개발 프로세스를 구축했습니다.

글로벌 테크 파트너십

일본 반다이남코(Bandai Namco), 중국 현지 퍼블리셔, 홍콩 외주사 등 다양한 국가의 개발진과 기술 스펙을 직접 협의하고 완벽한 품질로 납품한 글로벌 커뮤니케이션 역량을 보유하고 있습니다.

비즈니스 임팩트 우선주의

화려한 신기술 도입에만 머무르지 않고, 서버 비용 80% 절감, 수작업 공수 90% 단축 등 실제 비즈니스 지표와 운영 효율성을 극대화하는 실용적 아키텍처를 추구합니다.

