



김루비

Cloud & AI & Full Stack Engineer

✉️ dev.rubykim@gmail.com

📍 서울, 대한민국

🌐 <https://dev-rubykim.vercel.app>

⤷ [ruby-kim](#)

Education

세종대학교

소프트웨어 학사 • 3.5 / 4.5

2017.03 - 2022.02

Technical Skills

Aa Language

Python3, C, C++, C#, HTML, CSS, JavaScript

☁️ Cloud & Infra

Docker, Kubernetes, AWS, GCP, Azure, NCP, Terraform

⌚ Database

MySQL, MongoDB, Redis, PostgreSQL, AWS DynamoDB, Turso

🌐 Frontend

React.js, Next.js, Vue.js, Astro

⚙️ Backend

Flask, Django, FastAPI, Express.js

🧠 AI / LLM

Gemini, OpenAI, Prompt Engineering, Claude, RAG Pipeline, Grok

Certifications

Professional Cloud Architect Certification

May 2025

Google

Credential ID:

4b0c2a11b47041cfbc0f0bea5f33a29b

AWS Certified Cloud Practitioner

Sep 2024

Amazon Web Services (AWS)

Credential ID:

e0f4ecdb69f64249a2dac97ab5daf721

Languages

한국어

원어민

영어

업무상 자유로운 의사소통 가능

Summary

- LLM 기반 서비스 및 AI SaaS 설계: Gemini, GPT API를 활용하여 다국어 번역 파이프라인(Globalization SaaS) 및 글로벌 문화 가이드(CultureExplorer)를 직접 기획하고 구축
- 금융/의료급 안정성 인프라 설계: 4년 이상의 경력을 통해 FDA 승인 의료 플랫폼의 글로벌 인프라 아키텍처 및 Zero-Trust 보안 체계 구축
- 비용 및 성능 최적화 전문: AWS Lambda Adapter 및 Cloudflare Edge(Workers/R2)를 활용하여 상시 가동 비용을 절감하고 글로벌 저지연 서비스 구현
- 데이터 파이프라인 및 구조화: 비정형 텍스트 데이터의 실시간 수집, 정제 및 RAG 파이프라인 설계 역량 보유

Major AI & Cloud SaaS Projects

Gemini Globalization SaaS

Lead Architect

2026.01 - 현재

- LLM 파이프라인: Gemini API를 활용하여 기술 블로그 포스트의 마크다운 구조를 유지하며 다국어 자동 번역 수행
- 멀티 테넌시 설계: AWS Cognito 인증과 DynamoDB 메타데이터를 분리하여 테넌트별 독립 도메인(CNAME) 및 AdSense ID 동적 주입 로직 구현
- 인프라 최적화: AWS Lambda Adapter를 도입하여 상시 가동 비용을 \$0에 가깝게 유지하며 상품화 단계 진입

CultureExplorer

Lead Developer

2025.08

- AI 데이터 큐레이션: Gemini/GPT/Claude 멀티 모델을 활용해 20개국 비즈니스 에티켓 데이터를 수집 및 구조화하는 자동화 엔진 구축
- 에지 컴퓨팅: Cloudflare Workers 및 R2 기반의 저비용 에지 서빙 환경 구축 및 중동 지역 대응을 위한 RTL(Right-to-Left) 레이아웃 최적화

<https://culture.evoa.one>

Core Work Experience

Sibel Health

원격근무

Cloud Engineer II

2021.08 - 현재

1. 글로벌 임상 데이터 플랫폼 (Discovery Hub) 아키텍처 고도화

- 서버리스 인프라 구축: AWS Lambda 기반 고가용성 아키텍처 설계로 대규모 가변 트래픽 대응 및 관리 오버헤드 최소화
- 데이터 무결성 확보: AWS SQS 기반 비동기 파이프라인 도입, 실시간 고용량 텔레메트리 데이터 전송 성공률 100% 달성
- 글로벌 거버넌스 수립: CloudFormation(IaC) 기반 멀티 리전(US, EU, APAC) 자동 배포 체계 구축, 서비스 확장 속도 70% 단축 및 데이터 주권 준수
- 유지보수 최적화: 공통 로직을 Lambda Layer로 모듈화하여 코드 중복 제거 및 시스템 유지보수 효율성 제고

2. FDA 승인 실시간 의료 플랫폼 (Anne Hub) 보안 및 성능 최적화

- Zero-Trust 보안 아키텍처: HiveMQ mTLS 인증 및 서버 측 TLS 로테이션 시스템 구축, FDA 승인 클리니컬급 보안성 확보
- 네트워크 지연 개선: AWS Global Accelerator 및 애니캐스트 IP 라우팅 최적화로 글로벌 네트워크 지연(Latency) 150ms 이상 단축
- 비동기 부하 제어: FastAPI 및 Python AsyncIO 기반 백프레셔(Backpressure) 로직 구현, 트래픽 폭증 시 시스템 가용성 유지

3. 엔지니어링 운영 자동화 및 가시성(Observability) 강화

- Deterministic 빌드 파이프라인: Docker 기반 빌드 환경 표준화로 AWS Lambda 내 C-extension 라이브러리 의존성 문제 해결 및 배포 안정성 확보
- 인프라 자동화 CLI 개발: AWS SDK 기반 클라우드 리소스 추출 및 QR 코드 생성 도구 개발, 현장 기기 설정 리소스 90% 단축(오류율 0%)
- 분산 추적 시스템 구축: CloudWatch 및 X-Ray 연동을 통한 서비스 트래픽 흐름 시각화 및 장애 인지 시간(MTTD) 단축

Profiles

 ruby-kim

 Portfolio

Core Work Experience

솔루게이트 (Solute)

Research Assistant

서울

2020.03 – 2020.06

- NLP 전처리 알고리즘 개발: VOC(고객의 소리) 데이터 분석을 위한 맞춤형 철자 교정 모델을 개발하여 데이터 정제 효율성 향상
- STT 성능 고도화: 음성 인식(Speech-to-Text) 정확도 향상을 위한 데이터셋 구축 및 인식 엔진 튜닝 보조

Technical Projects

사찰 행정 통합 관리 및 디지털 전환(DX) 플랫폼

2026.01

1인 외주

- 프로젝트 개요: 수작업 기반의 사찰 행정 시스템을 클라우드로 이전하여 30일 이상의 업무를 실시간으로 자동화
- 기술적 의사결정 및 해결:
 - 비용 효율적 아키텍처: AWS Lambda Adapter를 도입하여 Next.js 앱을 서버 리스로 배포. 상시 가동 서버 대비 운영 비용 80% 이상 절감
 - 정밀 인쇄 엔진 구축: 클라이언트 사이드 렌더링 오차를 해결하기 위해 SSR(Server-Side Rendering) 기반 레이아웃 엔진을 구축하여 전문 인쇄 규격(A4) 대응
 - 세션 보안 정책: 민감 데이터 보호를 위해 Next-Auth 기반 인증과 15분 미 활동 시 자동 로그아웃 메커니즘 적용
 - 독자 로직 구현: 외부 API 의존성 제거를 위해 음력-양력 변환 알고리즘 직접 구현 및 데이터 구조화(DynamoDB)

Boba: 백오피스 구축

2021.12

1인 외주

- 프로젝트 개요: 무인 프린팅 솔루션의 행정 업무 자동화를 위한 풀스택 백오피스 구축
- 주요 성과 및 해결 방안:
 - 급격한 학습 곡선 돌파: Next.js 도입 3일 만에 아키텍처를 분석, 2주 내 실 서비스 수준의 애플리케이션 납품으로 기술적 민첩성 입증
 - 풀스택 아키텍처 설계: Next.js API Routes와 MySQL을 연동하여 기기 라이선스 모니터링 및 사용자 인증/로그 트래킹 로직 구현
 - 운영 가시성 확보: MUI 기반의 직관적인 UI를 설계하여 복잡한 관리 로직을 도식화하고 운영진의 업무 숙달 시간 단축
 - 문서 기반 협업 최적화: Notion을 통한 DTO 설계 및 API 요구사항 명세화로 소통 오류를 차단하며 개발 정확도 100% 달성

PressCheck: AI 기반 뉴스 데이터 인사이트 플랫폼

2020.09 - 2020.11

졸업 프로젝트 / 기술 리드 / 백엔드 & AI & 데이터 엔지니어링

- 실시간 데이터 수집 파이프라인 설계: BeautifulSoup4 및 Request를 활용하여 다수의 뉴스 포털로부터 실시간 데이터를 수집하는 분산 크롤링 엔진 구축
 - 성과: 시간당 수천 건의 뉴스 기기를 안정적으로 수집하고, 데이터 중복 제거 로직을 통해 저장 효율성 최적화
- NoSQL 기반 데이터 레이크(Data Lake) 구축: 비정형 텍스트 데이터의 유연한 처리를 위해 MongoDB를 활용하여 데이터 저장소 구축
 - 성과: 대용량 텍스트 데이터에 대한 빠른 인덱싱 및 검색 성능 확보
- NLP 모델 서빙 및 API 설계: KoBERT 및 FastText 기반의 요약/유사도 분석 모델을 백엔드에 연동하고, 이를 서비스하기 위한 RESTful API 인터페이스 설계
 - 성과: 뉴스 기사 간의 의미론적 유사도 분석을 통해 관련 뉴스 그룹화 (Clustering) 및 인사이트 자동화 구현
- 기술 리딩 및 성능 최적화: 데이터 크롤링부터 전처리, AI 추론 결과값 전송까지 이어지는 전체 데이터 파이프라인의 병목 현상을 진단하고 아키텍처 개선 리드