

안재웅

1999.02.03 | Seoul, South Korea
(+82) 10-7392-2367 | cubeahn185@gmail.com



문제 발굴부터 제품화까지 전 과정을 스스로 밀고 나가는 AI Product Engineer입니다. 마케팅 AI Agent 신사업을 아이템 탐색부터 맡아 개발 착수 7개월 만에 첫 본계약을 체결했고, Backend부터 Frontend, Infra, DX까지 전 영역을 설계 단계부터 담당합니다. 이전에는 디지털 치료제 팀의 개발을 리드해 v1.0 출시와 TIPS 선정을 이끌었습니다. ML/LLM을 실제 제품에 결합해온 경험을 바탕으로, 현재는 빠르면서도 안정적인 AI-Driven Development를 위한 하네스 엔지니어링 체계를 만들어 팀과 함께 쓰고 있습니다.

경력

Abear

AI Product Engineer

May. 2025 - Present

퍼포먼스 마케터를 위한 마케팅 AI Agent *Pervis* 신사업 및 기존 제품 *Windly* 개발 담당

- 신사업 *Pervis*
 - 아이템 발굴부터 제품화까지 주도
 - Finite State Machine 기반 AI Agent 설계·개발
 - Auth, 광고 플랫폼 외부 API 연동, 성과 데이터 기반 분석 기능 개발
 - AI-Driven Development 하네스 엔지니어링 시스템 구축
 - Sentry·Axiom·Langfuse 기반 에이전트 실행 관측 체계 구축
- 기존 제품 *Windly*
 - 서비스 내장 안내 챗봇 원비 개발
 - 배경 제거 서버 v2 개발 (BiRefNet-based Model)
- Tech Keywords:
 - Nest.js, Next.js, GCP, Supabase, Axiom, Sentry, Langfuse, LangGraph, Langchain, RAG, LLM

Andreia

Backend & ML Engineer, TPM

May. 2023 - Sep. 2023

May. - Aug. 2023 (Part-time),

Aug. - Sep. 2023 (Full-time)

우울증 조기 치료를 위한 감정 일기 기반 디지털 치료제 쪽다쪽다 개발 참여 및 프로젝트 개발팀 리드

- 프로젝트 개발팀 리딩 및 프로젝트 관리
- 감정 일기·온보딩·알림 등 서버 기능 개발
- LLM 기반 대화 상대 매칭 시스템 개발
- 서비스 아키텍처 설계 및 인프라 구축
- Tech Keywords:
 - Python, FastAPI, AWS, Docker, PostgreSQL, Langchain

Abear

Intern, Backend Engineer

Jan. 2023 - Aug. 2023

Jan. - Feb. 2023 (Full-time),

Mar. - Aug. 2023 (Part-time)

이커머스 사업자를 위한 B2B SaaS 솔루션 *Windly* 내 기능 개발 진행

- 상품 카테고리 추천 시스템 v1 개발 (FastText 기반 분류 모델)
- 고품질 배경 제거 기능 개발 (SAM 세그멘테이션, ONNX 경량화)
- 제휴 정산 백오피스 개발
- PG 교체 PoC 및 도입
- 데이터·트랜잭션 종류에 따른 DB 분리 및 최적화
- 개발 서버 백엔드·프론트엔드 CI/CD 구축
- Datadog 기반 프로젝트 모니터링 시스템 도입
- **Tech Keywords:**
 - Python, Django, FastAPI, AWS, MySQL, DynamoDB, Pytorch, ONNX, Github Actions

TULON

Co-founder, Full-stack Engineer

Jan. 2022 - Sep. 2022

ML 연구자를 위한 데이터 접근 권한 관리 솔루션 *TULON* 공동창업, 서비스 백엔드 설계와 개발 전담

- DDD 기반 아키텍처 설계 및 MSA 도입
- 데이터 접근 권한 관리 백오피스 개발
- 접근 요청·승인 워크플로와 권한 정책 관리, 감사 로그 개발
- 학습 격리 환경 및 데이터 반출 통제 개발
- 메인 서비스 및 학습 격리 환경에 대한 모니터링 시스템 구축
- **Tech Keywords:**
 - TypeScript, Next.js, Nest.js, PostgreSQL, RabbitMQ, Docker, AWS

프로젝트

Pervis

@ Abear

Nov. 2025 - Present

퍼포먼스 마케터를 위한 AI Agent

- **Background:** 광고 세팅은 퍼포먼스 마케터 업무의 대부분을 차지하면서도 피로도도가 가장 높은 구간. 이 구간의 일괄 세팅에서 시작해 광고 세팅 → 성과 분석 → 전략 수립 사이클 전체를 자동화하는 것이 목표이나, 입력으로 받는 시트는 양식이 정해져 있지 않고 암묵지로 차 있음. 이를 전부 자율적으로 해석하되 정해진 워크플로를 벗어나서는 안 돼, 워크플로 안에서 최대한의 자율성을 갖는 FSM 기반 AI Agent로 설계
- **Approach:**
 - 워크플로를 **13개 스테이지**로 명시하고 전이는 전진 전용으로 고정 - 경로는 잠그되 각 스테이지 안에서의 판단은 에이전트에 맡김
 - 스테이지마다 쓸 수 있는 도구를 한정해, 지금 단계에서 할 수 없는 일은 시도조차 못 하게 구성
 - LLM 호출 앞에 결정적 게이트를 세워 취소·주제 이탈·교정 요청을 모델에 닿기 전에 처리 - 가장 비싼 단계에 도달하는 요청 수 자체를 축소
 - 집행 직전 검수·승인 스테이지의 전이 조건 자체에 사람의 승인을 걸어, 승인 없이는 다음 단계로 넘어가지 못하게 구성
 - 구조로 막히지 않는 실패는 런타임 가드레일로 차단 - 같은 도구 반복·평풍·스테이지 회귀·질문 과다에 더해 턴·비용 상한까지 감지해 중단. 카운터를 체크포인트에 남겨 요청이 끊기거나 재시작해도 유지
 - 도구 결과는 전량을 모델에 넘기지 않고 체크포인트에 적재한 뒤 요약만 반환 - 스테이지가 늘어도 모델이 읽어야 할 양은 함께 늘지 않도록 구성
 - 플랫폼 차이는 파이프라인 템플릿으로 선언해, 엔진 하나 위에 광고 플랫폼 다섯 곳을 탑재
- **Result:**
 - 경로만 고정하고 그 안은 열어둔 구조 - 스테이지 순서는 잠겨 있지만 어떤 도구를 어떤 순서로 쓸지는 에이전트가 정함. 금지 사항을 프롬프트로 나열하는 대신, 지금 할 수 없는 일은 도구 목록에서 아예 빠짐
 - 양식이 정해지지 않은 시트를 그대로 입력으로 받음 - 구조 파악·컬럼 의미 매핑·셀 값 해석·셀 간 추론을 층으로 나누고, 규칙이 확실하는 구간은 규칙이 끝내고 모호한 구간만 모델로 넘김
 - 같은 입력에 매번 다른 경로를 밟는 모델을 구조가 감쌌 - 전이는 게이트가, 실행은 가드레일이 잡아 해매는 실행이 실패가 아니라 중단으로 끝남
 - 스테이지를 13개까지 늘리고도 모델이 한 번에 읽는 양은 늘지 않음
 - 플랫폼별 차이를 템플릿 선언으로만 흡수해, 엔진 하나 위에 광고 플랫폼 **5곳**(Meta·Google Ads·TikTok·Kakao·Naver)과 에이전트 도구 **217개**를 탑재
- **Tech Keywords**
 - Nest.js, Next.js, GCP, Supabase, Axiom, Sentry, Langfuse, LLM

Harness Engineering

@ Abear

Nov. 2025 - Present

AI-Driven Development를 빠르면서도 안정적으로 만드는 하네스 시스템 구축

- **Background:** AI 에이전트 기반 개발은 구현 속도를 크게 높이지만, 산출물의 품질과 일관성이 보장되지 않으면 검수·재작업 비용이 속도 이득을 상쇄. 개발 속도와 안정성을 동시에 확보할 수 있는 체계가 필요
- **Approach:**
 - 검증 게이트를 베이스라인 방식으로 설계 - 복잡도·순환 의존·마이그레이션 적합성을 기준선 파일로 고정해 기존 부채는 통과시키되 신규 위반만 차단
 - 편집·커밋·컴팩션·종료 등 8개 시점에 혹 26개를 걸어 에이전트가 규칙을 우회할 수 없게 구성
 - 도메인 규칙 문서 17종과 브랜치별 세션 상태 파일로 에이전트가 읽을 컨텍스트를 구조화해 컨텍스트 압축 전후로 보존
 - 역할을 기획·실행·도구 3계층 19개 에이전트로 분리하고, 워크트리와 락으로 병렬 작업을 격리
 - 스펙과 코드의 드리프트를 야간 워크플로로 감시하고, 에이전트 폭주는 서킷 브레이커로 차단
 - 운영 이슈와 버그 대응을 컨텍스트 수집·원인 파악·수정·배포까지 한 파이프라인으로 묶어 자동화
- **Result:**
 - 업계 통상 공수 산정으로 1개월 규모인 에픽 급 기능을 5일 만에 개발 완료
 - 코드 작성·PR 리뷰·품질 유지·개발 스펙과 정책 통일까지 개발 과정 전반을 이 체계가 맡아, 서비스 기획과 구현 자체에만 집중할 수 있게 됨
 - 운영 이슈와 버그를 컨텍스트 수집부터 수정·배포까지 자동으로 처리해 운영 안정성 개선
 - 테스트 1,757개·62.8만 줄로 프로덕션 코드의 1.12배, 커밋 6,257건 중 revert 38건(0.6%) - 속도를 올린 대가로 품질이 깎이지 않았다는 근거
 - 검증 스크립트 135개, 혹 26개, 에이전트 19개 규모로 팀 전체가 상시 사용 중
 - 이 체계 위에서 만든 Pervis가 개발 착수 2개월 내 PoC 고객사, 5개월 내 유료 PoC, 7개월 내 첫 본계약에 도달 - 확보한 속도가 사업 일정으로 그대로 옮겨짐
- **Tech Keywords:**
 - Claude Code, Codex, MCP, TypeScript, Zod, Vitest, Jest, Husky, Github Actions

원비

@ Abear

Aug. 2025 - Sep. 2025

Windly에 내장된 안내 챗봇

- **Background:** 신규 유저의 약 60%가 제품 기본 사용법과 해외구매대행 도메인 지식을 모두 CS 문의로 해결해 내부 CX팀 리소스가 과도하게 소모되던 상황. 반복되는 단순 질의응답만이라도 Windly 사용법과 도메인 지식을 함께 갖춘 자동 응답으로 걷어낼 필요
- **Approach:**
 - 답변 근거인 노선 문서는 이미지·영상이 섞이고 문서끼리 상대 경로 링크로 얹힌 구조 - 내보낸 묶음을 통째로 받아 링크와 첨부을 풀어 옮기고 원본 주소는 출처로 남기는 반입 경로를 별도 구축
 - 제품과 정책이 계속 바뀌는 도메인이라 모델에 지식을 급지 않고 문서를 고치면 답변이 따라 바뀌는 RAG 구조 채택 - 제목 계층으로 나눈 뒤 1000자·200자 겹침으로 청킹해 pgvector에 적재
 - 질문에서 마켓플레이스(쿠팡·네이버·11번가·타오바오·1688 등)와 요금제 등급을 LLM으로 먼저 추출해, 검색 전에 후보 문서를 좁히는 메타데이터 필터 단계를 전치
 - LangGraph로 ReAct 에이전트 그래프 11종을 구현하고 요청 단위로 교체 가능하게 만들어, 같은 질문에 검색 전략을 나란히 붙여 비교
 - 응답 10초 예산을 맞추는 쪽으로 결론 - 재귀 루프가 있는 변형 대신 단일 패스 그래프를 기본값으로 고정하고 top-k를 5로 축소
 - 문서 관리 백오피스와 서비스 내장 챗봇을 함께 개발해 운영까지 아우르는 구조 설계
- **Result:**
 - 노선 문서 450여 개를 근거로 답하는 서비스 내장 챗봇 구축
 - 한 번 물어보고 마는 도구가 아니라 답이 나올 때까지 이어 묻는 경로가 됨 - 출시 첫 달(2025.08-09) 258명이 410건의 대화, 한 대화당 평균 2.6회 질문
 - 대화 로그에서 응답 시간·토큰 사용량까지 뽑아내는 모니터링 추출 경로를 만들어 운영 지표를 직접 확인
 - 신규 기능 출시와 가이드 문서 주입을 하나의 프로세스로 묶어, 제품이 바뀌어도 챗봇 지식이 뒤쳐지지 않도록 운영 체계 정착
- **Tech Keywords:**
 - LangGraph, RAG, FastAPI, pgvector, Next.js

쑥닥쑥닥

@ Andreia

May. 2023 - Sep. 2023

우울증 조기 치료를 위한 감정 일기 기반 디지털 치료제

- **Background:** 조기 개입은 기록이 쌓이는 것만으로 성립하지 않고 다음 행동으로 이어져야 하나, 일기는 정형 필드가 없는 자유 서술 텍스트라 그 자체로는 추천의 입력이 되지 못하는 형태. 본문에서 정서 신호를 뽑아내고 이를 다른 factor와 묶어 사람을 연결하는 2단계가 필요
 - **Approach:**
 - 감정 라벨 세트 고정 - LLM이 그 안에서만 고르게 해, 같은 정서를 매번 다른 말로 적는 자유 서술을 추천이 그대로 읽을 수 있는 값으로 변환
 - 일기 본문에서 해당 라벨을 추출해 사용자별로 축적
 - **neural CF 기반 추천 시스템 개발** - 추출한 감정 라벨을 데모그래픽 데이터·추출 키워드와 함께 factor로 투입해, 규칙이 아니라 상호작용에서 학습하도록 구성
 - 추출과 매칭은 배치로 처리해 일기 저장 경로에서 분리
 - 추출 결과와 추천 품질을 지표로 검증
 - **Result:**
 - 쑥닥쑥닥 v1.0 출시
 - 핵심 기술인 감정 추출·추천 시스템으로 **TIPS 선정**, 지원금 5억 원 유치
 - **Tech Keywords:**
 - Python, FastAPI, AWS, Docker, PostgreSQL, Langchain
-

ML 연구자를 위한 데이터 접근 권한 관리 솔루션

- **Background:** ML 연구자가 보안이 요구되는 데이터를 다룰 때 접근 권한을 체계적으로 관리할 수단이 부재. 게다가 의료 데이터는 보안이 기능이 아니라 관문이라, 병원 심사를 통과하기 전에는 도입 논의 자체가 열리지 않는 시장
- **Approach:**
 - DDD 기반 아키텍처 구상 및 MSA 도입
 - 데이터 접근 권한 관리 백오피스 웹 서비스 개발
 - 접근 요청·승인 워크플로, 권한 정책 관리, 감사 로그 등 서버 기능 개발
 - 학습 격리 환경 구축 - VM 프로비저닝·회수 자동화, 환경 이미지 템플릿 관리, 네트워크 격리 및 방화벽 정책
 - 데이터 반출 통제 - 반출 경로 차단, 결과물 반출 심사 플로우, 세션 활동 기록 및 모니터링
 - 병원 계정 체계 SSO 연동 및 VM 리소스 쿼터 관리
- **Result:**
 - 민감 데이터 접근을 요청·승인·감사 기록까지 통제하는 플랫폼과 학습 격리 환경 구축
 - 개발 막바지에 국내 4대 병원 중 2곳을 도입 예정처로 확보
- **Tech Keywords:**
 - TypeScript, Next.js, Nest.js, PostgreSQL, RabbitMQ, Docker, AWS

외주 프로젝트

2024.09 - 12	NeuroMoment · Slyric	수면 유도 음원을 생성해 재생하는 알람 앱 - 한 곡(기본 7시간) 생성 50분 → 약 4.5분 . Flutter iOS/Android 앱과 백엔드 단독 개발
2024.04 - 06	Abear · Windly	LLM 기반 번역 텍스트 정제 파이프라인을 약 100만 개 상품의 옵션명 · 상품명 · 태그에 적용
2024.04 - 06	Abear · Windly	상품 카테고리 추천 v2 - FastText 분류기를 LLM 분류로 대체해 마켓당 70분 재학습 제거 , 정확도 90% → 92%
2023.03 - 07	TebahSoft	멀티모달 감정 분석 - 문장 부호 복원 연구의 융합 구조를 재사용해 음성을 추가 모달로 결합, Ekman 6종 + 중립 7종 분류에서 텍스트 단일 모델 대비 F1 0.71 → 0.81
2023.03 - 07	TebahSoft	상담 요약 및 화자 분리 - pyannote 화자 분리와 Whisper STT의 서로 다른 타임라인을 화자 전환 지점에서 병합. 음성 입력만으로 화자별 기록이 나와 기록 작성 5분 내 완료 (DER 10%)
2023.02 - 04	삶의질연구소 · 사브작	부모 · 자녀 일기 · 커머스 플랫폼의 백엔드를 단독 설계 · 개발, 출시 완료
2022.09 - 12	TebahSoft (산학)	멀티모달 문장 부호 복원 연구 - UniPunc을 한국어로 이식(KoBERT + wav2vec2). 음성이 없는 샘플엔 학습된 virtual embedding을 주입해 추론 시 음성 없이도 동작 . KoBERT 단일 모달 대비 F1 +15포인트 , 기업 이관 완료

기술 & 외국어

Programming Languages	<i>Typescript, Python, Javascript, Go, C/C++</i>
ML/DL	<i>Pytorch, Langchain, ONNX, BentoML, Streamlit</i>
Web	<i>React, Next.js, HTML5, CSS</i>
Server	<i>Nest.js, FastAPI, Django, Fiber, Flask, Express</i>
Cross Platform App	<i>Flutter, React-Native</i>
Database	<i>PostgreSQL, Supabase, MySQL, MongoDB, DynamoDB, Firebase</i>
Cloud Providers	<i>AWS, GCP</i>
Observability	<i>Sentry, Langfuse, Datadog, Axiom, New Relic</i>
DevOps & Infra	<i>Docker, Github Actions, Jenkins, RabbitMQ</i>
Collaboration	<i>Github, Slack, Notion, Jira, Confluence</i>
Languages	Korean (Native), English (Business proficiency), Chinese (Conversational)

학력

KAIST (한국과학기술원)	- 전공: 전기및전자공학부 (School of Electrical Engineering) - 복수전공: 전산학부 (School of Computing) - 부전공: 수리과학과 (Department of Mathematical Sciences)
Mar. 2018 - Feb. 2026 졸업	B.S. · GPA 3.6/4.3

Mar. 2015 - Feb. 2018

졸업

기타

KATUSA

Battalion S1 Human Resource Specialist (42A)

2ID, 2CAB, 602D ASB

- 400명 이상의 병력 인사 관리 및 행정 업무 지원

Oct. 2023 - Apr. 2025

Sr. KATUSA

Camp Humphreys

- 대대 소속 KATUSA 교육, 인솔, 관리 담당

수상

May 2024

2024 육군 창업경진대회

우수상 (3위) · LLM 기반 야전 기술 교범 챗봇 *TecChat*
- 도전 K-Startup 범부처 대회 진출

May 2024

2024 공군 창업경진대회

창의상 (120팀 중 30위) · 적대적 오디오로 음성 데이터
학습을 막는 *MyVoice*

Dec 2020

E*5 KAIST
(학생창업지원프로그램)

Finalist (Top 10) · 프로젝트 포트폴리오 기반 개발자
채용 플랫폼 *TurninPoint*