

BACKEND / AI ENGINEER



검증 가능한 AI를 서비스 판단으로 연결합니다

김성민

Java/Spring 기반 금융 서비스 백엔드에서 시작해 시계열 ML, LLM, RAG, AI Agent 검증으로 영역을 확장해왔습니다.

모델 출력을 그대로 신뢰하기보다 데이터·정책·검증 계층을 분리하고 실제 서비스 의사결정으로 연결합니다.

좋은 결과뿐 아니라 실패·편향·재현성의 한계도 기록해 다음 설계의 근거로 사용합니다.

ENGINEERING FOCUS

01

Backend Systems

금융 데이터 수집·배치·API·운영 흐름을 하나의 서비스로 연결합니다.

02

AI Decision Layer

모델 예측과 서비스 결정을 분리하고 정책 계층에서 책임을 명확히 합니다.

03

Verification

백테스트·독립 실행·교차 검토로 결과의 조건과 한계를 함께 확인합니다.

SELECTED LINKS

GitHub

github.com/kimseongmin0301

Career Detail

kimseongmin.co.kr/career

Live Signals

kimseongmin.co.kr/signals

Current Interest

Financial AI · AI Agent · RAG · Human-AI Workflow · AI Service Engineering

AI ENGINEERING RESEARCH

AI를 어디까지 믿고 어디에서 사람의 판단을 남길 것인가

세 연구는 모델 순위를 정하는 실험이 아니라 AI를 개발 워크플로에 배치하는 기준을 찾는 과정입니다.

01 AI Coding Agent Benchmark

Question 업무 성격에 따라 적합한 Coding Agent는 달라지는가?

Tested 동일 Repository·요구사항·baseline에서 Codex와 Claude를 3개 Task로 독립 실행하고 교차 평가했습니다.

Result 조사·전략은 Claude, 버그 수정은 실질 동급, 계획·기능 구현은 Codex 우세였습니다.

Learned 하나의 Agent 순위보다 작업 성격에 따른 역할 배치가 중요했습니다.

[OPEN GITHUB ↗](#)

02 AI PR Review Agent

Question 테스트가 통과한 PR도 AI가 의미적으로 검증할 수 있는가?

Tested 고정 PR 5개를 독립 Reviewer 세션에서 diff·관련 코드·pytest 결과로 검토했습니다.

Result 기대 판정 5/5 일치. 테스트가 놓친 로직 오류·동작 회귀를 탐지하고 설계 판단은 HUMAN_REVIEW로 분리했습니다.

Learned Test Pass는 Correctness와 같지 않으며 최종 Merge는 사람의 책임입니다.

[OPEN GITHUB ↗](#)

03 Human-AI Review Benchmark

Question Human과 AI는 어떤 지점에서 다르게 판단하는가?

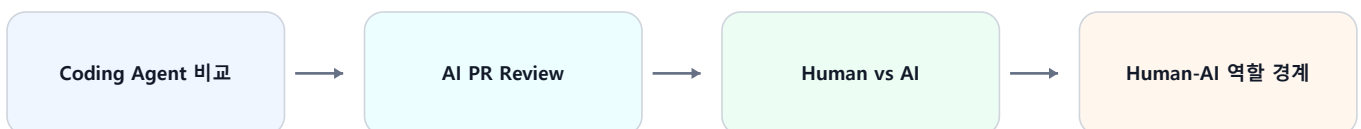
Tested 8개 고정 Case에서 Human Only와 AI Only를 비교했습니다. Human + AI 정량 조건은 중단했습니다.

Result 사전 판정 일치도 Human 6/8, AI 7/8. AI는 넓은 검증, Human은 맥락과 최종 판단에서 강점을 보였습니다.

Learned 협업의 우월성을 입증한 것이 아니라 역할 경계를 찾기 위한 다음 질문을 만들었습니다.

[OPEN GITHUB ↗](#)

RESEARCH EVOLUTION



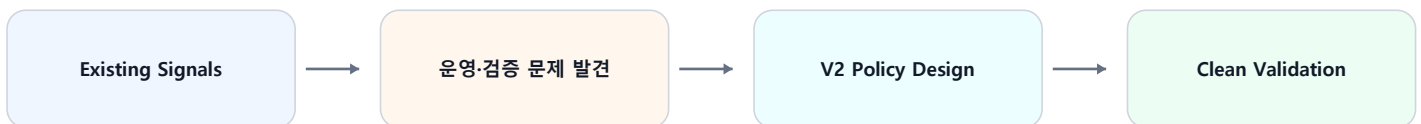
연구별 실행 환경과 모델이 다르므로 결과는 절대적인 모델 순위가 아니라 각 실험 조건 안의 관찰로 해석합니다.

FINANCIAL AI · PRODUCTION TO RESEARCH

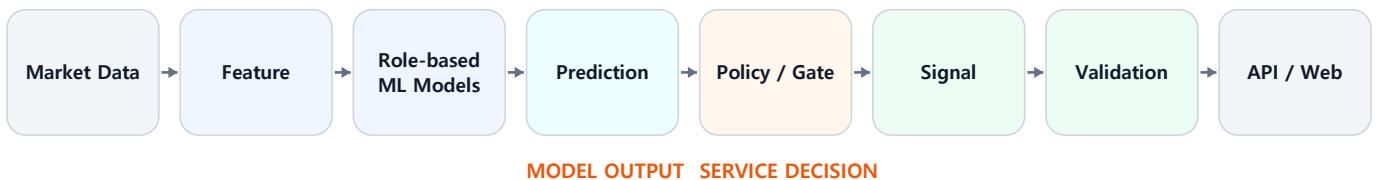
운영 결과에서 검증 절차를 다시 설계한 금융 AI 시스템

기존 Signals 서비스를 운영하며 발견한 선택 편향과 재현성 문제를 바탕으로 **Model Output**과 **Service Decision**을 분리한 V2 정책 구조를 설계했습니다.

FROM OPERATION TO V2



ARCHITECTURE



PROBLEM → DECISION

Problem

Test-Live 결과가 정책 선택에 일부 영향을 준 Selection Bias를 발견했습니다.

Decision

모델과 정책을 분리하고 Validation-only Selection, Embargo, One-shot Test Evaluation으로 절차를 다시 고쳤습니다.

VERIFICATION → LEARNING

Time split

Train 2012~2023 → Validation 2024 → Test 2025 → Forward 2026~

Learning

높은 성과 수치보다 정책 선택 과정과 재현 가능한 전진 검증이 중요했습니다. Immutable Forward Ledger는 운영 적용 전입니다.

2025 CLEAN TEST · SIGNAL-BASED HISTORICAL SIMULATION

155

Signals

67.10%

Win Rate

+9.19%

Avg. Return

+7.82%

Median

-4.53%

MDD

거래비용·세금·슬리피지·실제 체결·자본 제약·동시 보유 한도를 반영하지 않은 신호 단위 시뮬레이션입니다. 2026 구간에서는 MDD 확대도 관찰되어 낮은 위험을 구조적 특성으로 주장하지 않습니다.



GitHub · AI Stock Signal System

github.com/kimseongmin0301/ai-stock-signal-sys...

Live · Existing Signals

kimseongmin.co.kr/signals

RAG · DATA DESIGN BEFORE RETRIEVAL

검색 품질을 embedding 이전의 데이터 설계부터 다뤘습니다

개인 이력서 RAG에서는 더 큰 모델보다 무엇을 제거하고 어떤 의미 단위로 검색할 것인지를 먼저 정의했습니다.

PIPELINE



DATA DESIGN

PII before embedding

이름·연락처·주소·개인 보상·개인 URL을 검색 데이터에서 제거했습니다.

Semantic chunking

고정 글자 수가 아닌 경력·프로젝트·문제 해결 단위로 분리했습니다.

Context-independent

각 청크가 단독 검색되어도 대상과 맥락을 이해할 수 있게 작성했습니다.

RETRIEVAL DESIGN

Metadata and provenance

category·period·company·role·technology·source를 보존합니다.

Scope-aware retrieval

회사 경력과 개인 프로젝트를 구분하고 넓은 질문은 관련 범위를 다양하게 검색합니다.

Final validation

필수 필드·중복 ID·이메일·전화번호·URL·연봉 패턴을 다시 검사했습니다.

CURRENT PRODUCTION SNAPSHOT

20

Resume chunks

49

Total chunks

Gemma

Embedding

JSON

Index

현재는 구조를 이해하고 검증하기 위해 Ollama embeddinggemma와 로컬 JSON 벡터 인덱스를 사용합니다. Elasticsearch 전환은 계획 단계이며 이미 도입한 기술로 표현하지 않습니다.

Live · Resume RAGkimseongmin.co.kr/

ENGINEERING VIEW

프로젝트는 달라도 판단 기준은 일관됩니다

AI를 실제 시스템에 적용할 때 반복해서 확인한 다섯 가지 원칙입니다.

01

Model Output ≠ Service Decision

모델 예측은 정책·위험·시장 조건을 거쳐 서비스 신호가 됩니다.

02

Test Pass ≠ Correctness

통과한 테스트가 요구사항·계약·회귀를 모두 보장하지는 않습니다.

03

AI Output ≠ Final Judgment

AI는 근거를 확장하고 프로젝트 맥락과 최종 책임은 사람이 가집니다.

04

Performance ≠ Reproducibility

높은 수치보다 선택 편향·검증 절차·전진 재현성을 함께 확인합니다.

05

Retrieval Starts Before Embedding

RAG 검색 품질은 모델보다 먼저 개인정보 정책과 청크 설계에서 시작합니다.

CURRENT INTEREST

Financial AI

AI Agent

RAG

Human-AI Workflow

AI Service Engineering

● **GitHub**
github.com/kimseongmin0301

● **Career Detail**
kimseongmin.co.kr/career

Problem → Decision → Verification → Result