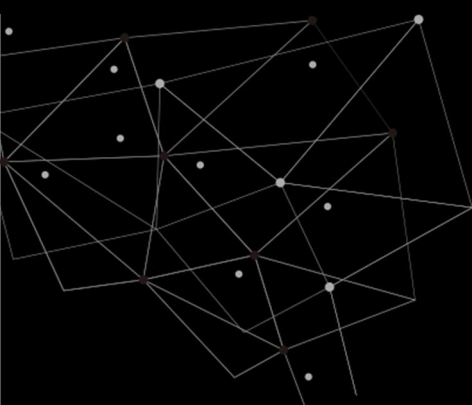
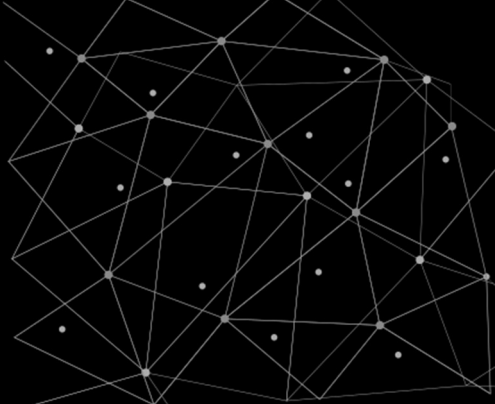


2026 오픈소스 AI 활용 인재 양성 특화형 과정

온프레미스 경량 LLM 에이전트 개발·검증

Project Guide



1

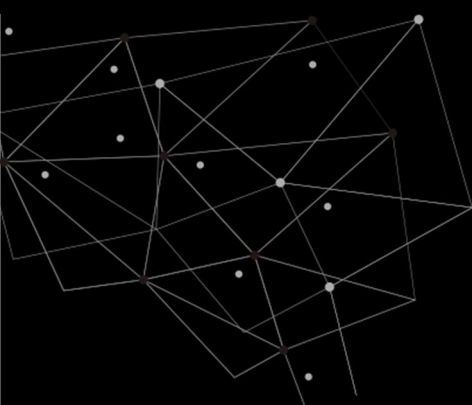
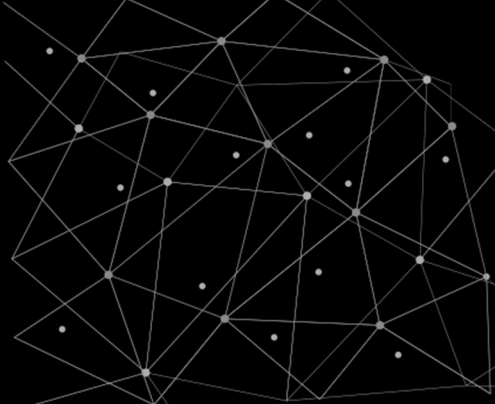
과정 개요

활용 기술 / 난이도 / 활용 언어 등



① 과정 개요

- ✓ **과정명** : 온프레미스 경량 LLM 에이전트 개발·검증(Potato LLM)
- ✓ **난이도** : ★★★★★☆ (대상 : 시니어·현업 전문가 권장)
- ✓ **활용 언어** : Python
- ✓ **활용 기술** : 경량 LLM(Qwen·Gemma·Llama), 양자화(GGUF/AWQ/GPTQ), Tool-Calling Agent, llama.cpp·Ollama, 오픈소스 벤치마크



2

과정 소개

과정 상세 소개



2

과정 소개



프로젝트 세부 내용

[프로젝트 소개]

- 외부 AI API·고가 GPU 없이 저사양 PC(CPU)에서도 동작하는 경량 LLM에 에이전트를 결합해, 실제 업무를 수행하는 AI를 직접 만들고 그 성능을 오픈소스 벤치마크로 검증·개선합니다.

[프로젝트 목표]

- 저사양·온프레미스 환경에서 동작하는 경량 LLM 에이전트 구축
- 실제 업무 수행 능력을 재는 오픈소스 벤치마크(PotatoBench)로 검증·개선
- 경량 AI의 실무 적용 가능성과 한계를 정량적으로 확인

[최종 산출물]

- Baseline Potato LLM · PotatoBench · Potato Scoreboard (오픈소스 공개)

2

과정 소개



프로젝트 세부 내용

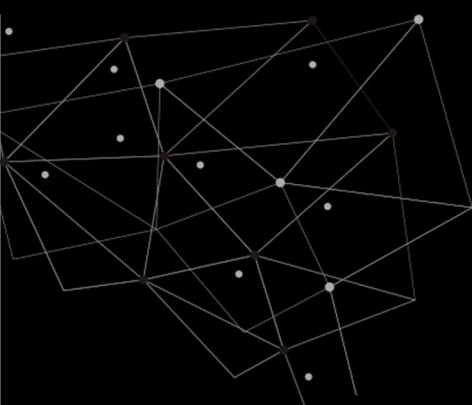
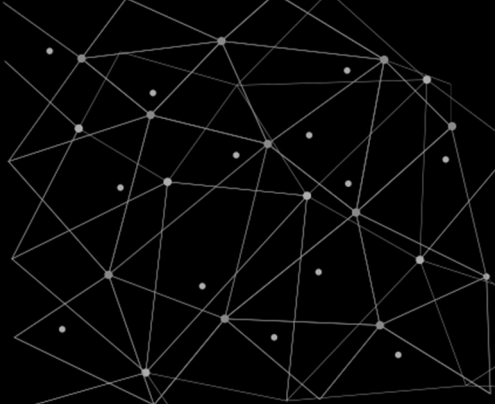
[운영 방안]

- 월 2회 오프라인 밋업 + Slack·GitHub 온라인 리뷰 병행

[운영 계획]

- 1주차: 시작 & 방향 설정 - 무엇을 왜 만드는지 정의, 평가 기준(PotatoBench) 설계
- 2주차: v0.1 + 첫 측정 - 소규모 평가셋과 baseline 에이전트로 첫 실행·측정
- 3주차: 약점 보완 - 실패 사례 기반 Prompt·도구 형식·평가 기준 개선
- 4주차: 경량화 - 모델 압축(양자화) 후 속도·메모리·성능 변화 실험·분석
- 5주차: 도구 사용 개선 - 에이전트가 도구를 더 정확히 사용하도록 개선
- 6주차: 실행 최적화 - 저사양 PC에서 더 빠르고 가볍게 동작
- 7주차: 정리 & 공개 - 결과·점수판 정리, 오픈소스 공개 및 실무 적용 방향 논의

※ 세부 일정 및 내용은 상황에 따라 변경될 수 있음



3

멘토 소개

담당 멘토 소개



3

멘토 소개



최세영 멘토

✓ 주요 경력

- Brain Crew — AI Research Engineer
(문서처리 플랫폼 Docurator : LangGraph 에이전트·RAG)
- GSC 구매처등록 자동화 — OCR+LLM(FastAPI), 정확도 95.6%
- Omnicore AI Labs — Lead AI Engineer (Co-founder)
- Upstage — AI Ambassador (정치 답리서치, Ko-FreshQA 벤치 1위)

✓ 주요 활동

- 오픈소스 Lang2SQL 기여 (에이전트 아키텍처·품질 평가 벤치마크)
- 해커톤 멘토/심사 — ICT INNOHACK, 부산대 × Upstage AI Agent
- 실무 강의 — 부산테크노파크·부산정보산업진흥원
- 기술서 번역 — 『Docker Deep Dive』, 『The Kubernetes Book』

3

멘토 소개



강민석 멘토

✓ 주요 경력

- AI Engineer — RAG·AI Agent·문서 처리 자동화 설계·구현·배포
- 사내 보고서 작성 자동화 — 장기 실행 Agent Harness 설계, 평가 60→94점
- 대기업 제조사 RAG 고도화 — 데이터 대시보드·RAG 추론 인터페이스 개발
- 라이프로그 개인화 AI — GraphDB→RDB RAG 전환, 토큰·처리시간 절감
- 금융사 문서 처리 PoC — FastAPI·Celery 비동기 전처리·온프레임 배포

✓ 주요 활동

- Upstage AI Ambassador 2기 — Solar Pro 3 Agent·기술 블로그 제작
- AgentRef, langgraph-chat-ui 등 LangGraph 오픈소스 프로젝트 개발
- PNU Chat — 부산대 학생지원 챗봇, Hybrid Retrieval·RRF 성능 개선
- 2025 부산대 x Upstage AI Agent 해커톤 대상
- FIN Challenge 본선 입상, DIVE 2025 BNK 금융지주 트랙 2위

3

멘토 소개



황재성 멘토

✓ 주요 경력

- SKT AX Infra 구축 사업 — 개발
- GS칼텍스 AX 사업 — PM
- LG전자 실시간 데이터 처리 시스템 구축
- AI 문서 처리 시스템 개발 (HSAD·IBKC·SKEP)
- 야놀자 관광 생성형 AI 챗봇 개발
- 풀무원 식품 도메인 RAG 구축
- 블루레시피 식품 AI 모델 개발

✓ 주요 활동

- 당뇨병 특화 LLM 연구 (JMIR 투고, 제1저자)
- 지식그래프 기반 음식 추천 연구 (HCI 발표, 제1저자)
- BERT 기반 식단 추천 연구 (IEEE 게재, 제2저자)
- 대스타 해결사 플랫폼 최우수상(1위) 수상

3

멘토 소개



김태한 멘토

✓ 주요 경력

- AI Engineer - RAG·AI Agent·문서 처리 자동화 설계·구현·배포
- SKT 세무·복지 자동화 멀티 에이전트 - OCR·LLM·전자결재 연동
- GSC 의료비 청구 자동화 - OCR+LLM(FastAPI), 정확도 94.3%·재현성 99.3%
- LG전자 라이프로그 백엔드 - Agent API·Batch, PostgreSQL+pgvector
- 퇴직연금 멀티 에이전트 - Agentic RAG, A2A SDK 기반 통신
- AI 스마트 재고관리 - ReAct·Reflexion, Webshop 0.60→0.84
- HTP 심리분석 - CI/CD·GCP, 이미지 6.29→2.1GB(Multi Stage Build)
- 쓰리디포커스 2.5년 - Photogrammetry 3D·디지털트윈 DCONSAFE

✓ 주요 활동

- 부스트캠프 AI Tech 7기(CV) - 영수증 OCR F1+31.19%·스케치 분류 1위
- 가짜연구소 10기 Agent is All You Need - LangChain·LangGraph·A2A
- HuggingFace Agent Course 수료
- 연세대 의료영상 AI 연구실 - 국제 논문(Sensors, 2021) 게재
- 메타버스 개발자 경진대회 3등, CONVR 2022 3rd Prize 등 수상