

Juntaek's Profile

Information

- **Name** 오준택
- **Birth** 1994.01.26
- **Blog** <https://eunmastudio.tistory.com/>
- **Web** <https://eunmastudio.netlify.app/>
- **Github** <https://github.com/Oh-JunTaek>



About Me

안녕하세요! '**Comfort-zone**'에서 벗어나 새로운 도전을 하는 '**AI/DATA Engineer**' 오준택입니다.

- **Effective Communicator** : 명확하고 효과적인 의사소통으로 팀워크의 중요성을 최우선합니다.
- **Curious Learner** : 새로운 기술을 배우고 데이터 분석 트렌드를 탐색하는데 적극적입니다.
- **Collaborative Growth** : 개인적인 성장을 넘어, 함께 공유하며 발전하는 문화를 추구합니다.

Timeline

- **2025.03 – 2025.06** : 인포뱅크 (키움증권 팟봇 프로젝트 – 프리랜서)
- **2024.07 – 2024.12** : 카카오테크 부트캠프 – 생성형 AI 1기 수료
- **2024.07 – 2024.08** : Data Science 2024 코칭스터디
- **2024.05 – 2024.06** : Let's AI 2024 (데이터 분석 기초 및 시각화)
- **2024.01 – 2024.02** : AI 엔지니어 기초 다지기 (네이버커넥트)
- **2019.03 - 2023.02** : 보육교사
- **2013.03 – 2019.02** : 경북대학교 아동학과 졸업 (부전공: 국어국문학)

Tools & Skills

- **Language** #Python #Kotlin
- **AI skills** #RAG #LangChain #LLM #Prompt Engineering
- **Data** #SQL #MongoDB
- **Tools & Environment** #Github #VSC # Notion

Education & Activities

- 키움증권 사내 챗봇 [2025.03 – 2025.06] (<http://chatist.ai/>)
 - DATA 정제
 - RAG 챗봇에 기반한 raw 데이터 정제
 - QA
 - 실제 성능 테스트를 통해 2차 정제 및 RAG 로직 개선
 - 봇빌더 기반 시나리오 디버깅
 - 봇 테스트 환경 내 대화 시나리오 논리 흐름 검토 및 단위별 분기 처리 검증
- 카카오테크 부트캠프 [2024.07 – 2024.12]
 - RAG 기반 챗봇 프로젝트 (<https://github.com/kakaotech-bootcamp-11>)
 - 데이터 크롤링 - 데이터 수집 및 전처리
 - Python 기반 자동 분류 파이프 라인 구축 - 문서 단위 임베딩 처리 및 벡터 DB 구축
 - RAG 로직 구현 - BM25 + FAISS 하이브리드 구조, 연속질의 응답 흐름 개선
 - LLM 기반 이력서 자동 생성 프로젝트 (<https://github.com/KTB-Sixmen>)
 - 사용자 유형 기반 데이터 항목 설계 및 대용량 병렬 처리 로직 구현
 - 프롬프트 설계 최적화 및 API 연동 기반 LLM 응답 품질 개선
 - 팀 내 백엔드/프론트 협업을 통한 전체 서비스 흐름 이해도 향상
- Data Science 2024 코칭스터디 (네이버 부스트코스) [2024.07 – 2024.08]
 - 데이터 분석 기초
 - 데이터 분석 - Python 기초 및 pandas 활용
 - 공공데이터를 활용한 실습 - 결측치, 추출, 요약(seaborn), 시각화(matplotlib)
- Let's AI 2024 (네이버 부스트코스) [2024.05 – 2024.06]
 - 머신러닝
 - 머신러닝의 개념 및 모델의 구조 이해 - Pandas와 Numpy 라이브러리 활용
 - 데이터 시각화(matplotlib)
 - 딥러닝 기초
 - 딥러닝 학습 방법 - 신경망의 구조, 역전파 알고리즘, 경사하강법(기초)
- AI 엔지니어 기초 다지기 (네이버커넥트) [2024.01 – 2024.02]
 - 데이터 활용
 - 정형 데이터의 분류 및 전처리 실습
 - 캐글 데이터를 활용한 분석 및 실습
 - 머신러닝 기초
 - 분류 알고리즘 실습 - 의사결정트리, 랜덤포레스트
 - 성능 평가 지표 기준 모델 선택 및 개선 - F1-score, Accuracy
 - 이미지 모델 분류 실습
 - CNN - 개/고양이 분류를 위한 비정형 데이터 처리 (Scikit-learn, TensorFlow(Keras))

Career

- 인포뱅크 (프리랜서) – 키움증권 챗봇 프로젝트 참여 [2025.03 ~ 2025.06] (<http://chatist.ai/>)
 - 생성형 AI 부트캠프에서의 RAG 챗봇 프로젝트 경험을 바탕으로 실무에 빠르게 적용
 - 도메인 특화 데이터 정제: 금융 분야의 복잡한 raw 데이터(FAQ, 금융 상품 설명 등)를 LLM 임베딩 최적화를 위해 문맥 단위로 분절 및 노이즈 제거, 토큰수 정규화하여 벡터화 효율 향상
 - QA 및 RAG 로직 개선 제안: 챗봇의 질의응답 품질 점검을 위한 QA 시나리오 수립 및 수행, 응답 오류 패턴 분석을 통한 재정제 및 RAG 세팅 최적화 제안
 - Bot Builder 디버깅 및 테스트: 대화 시나리오 기반 테스트 수행, 봇빌더 상의 로직 결함 및 UX 상 허점 발견 후 디버깅 수행

Cover Letter

- 지원 동기 및 관심

생성형 AI 프로젝트를 진행하며, 데이터가 모델 성능과 사용자 경험에 얼마나 직접적인 영향을 주는지 실감할 수 있었습니다. 챗봇이 사용자 발화 의도에 맞는 데이터 선정 적중률이 만족스럽지 못하는 이슈가 있었습니다. RAG에 사용되는 알고리즘의 문제인 것인지, LLM 모델 선정의 문제인 것인지, 양상을 둔 기법 비율의 문제인지 원인을 찾지 못해 다양한 시도를 해보았는데, 데이터 정제를 새롭게 한 이후 챗봇의 응답 성능이 **77%**에서 **86%**까지 약 **9%p** 상승하는 경험을 토대로, 자연스럽게 데이터 처리 및 품질 관리의 중요성에 주목하게 되었습니다.

데이터 처리 역량을 증가시키기 위해 SQLD를 취득했고, 키움증권의 챗봇 프로젝트에 참여하여 실무 경험을 쌓았습니다.

저는 새로운 기술을 빠르게 습득하고, 실전에서 유연하게 적용하는 역량, 그리고 AI 서비스 프로젝트 참여 경험과 책임감을 바탕으로, 데이터 엔지니어로서 의미 있는 기여를 할 수 있다고 확신합니다. 특히 AI 프로젝트 참여 경험은 데이터를 다양한 상황에서 주도적으로 다루는 데 큰 강점이 될 것입니다.

- 직무역량

- **RAG** 기반 서비스 개발 역량
 - 처음 실무 프로젝트에 투입될 때 예정에 없던 인력이 필요해지면서 문서화된 업무 지침이나 제대로 된 인수인계가 없는 어려움이 있었습니다. 하지만 저는 부트캠프에서 RAG 챗봇 프로젝트를 진행하면서 데이터 수집 단계부터 RAG 로직 구현을 위한 검색 알고리즘 및 벡터 db 구축, 마지막으로 QA 까지 진행하며 성능 개선 과정 전부를 진행한 경험이 있습니다. 이러한 경험을 통해 실무 프로젝트에서 데이터 정제 역할을 담당하였을 때에도 스스로 해야 할 일을

파악하고 진행하여 기대 이상의 역량을 보여주고, 타 팀원의 RAG 구현과정에 도움을 줄 수 있었습니다. 그리고 그 과정을 통해 RAG 챗봇의 전체적인 워크 플로우에 대한 높은 이해도를 보여주어 지식구축팀에도 도움을 주고 프롬프트 엔지니어 설계 개선 및 RAG 로직 성능 향상에도 참여하여 모두에게 필요한 역할을 수행할 수 있었습니다.

- QA 및 성능 개선 경험

- 프로젝트 내 직접 LLM 응답의 품질을 정량/정성 분석하며 개선 작업을 수행하면서 프롬프트 엔지니어링 기술이 향상되었습니다. 실무 프로젝트에서는 봇빌더를 활용하여 시나리오 단위 테스트도 경험하였으며 이 과정에서 협업자도 찾아내지 못한 치명적인 이슈도 찾아내면서 팀 전체의 오류 검증 및 디버깅 업무를 추가적으로 받을 수 있었습니다.

- SQL 및 Python, 생성형AI 활용 역량

- 데이터 품질에 대한 관심과 실무 연계성을 높이기 위해 **SQLD** 자격증을 취득하여 SQL 활용 능력을 향상시켰습니다. 프로젝트 목표에 맞는 이해도를 기반으로 데이터를 자동으로 수집, 정제, 분류할 수 있는 데이터 파이프라인을 구축하고 LLM을 활용하여 자동정제로직을 구축하여 시간 및 휴먼리소스를 절감한 효율적인 업무진행을 할 수 있습니다.