

MY website plan

Younghee Seo

KBSI(한국 기초과학지원연구원)

Younghee Seo

[image]
Aboutme

- AI engineer
- Bioinformatics engineer
- Python Programmer
- LLM Application Engineer
- Tech Communication

- 07seoy@naver.com , 07seoy@kbsi.re.kr
- 010-2565-5960
- 기술 블로그 - <https://misty-snail-dce.notion.site/blog-18f57ff58e91813880dbe50d83fbd52c?pvs=74>
- Github - <https://github.com/tjdudgml3>



교육

- 영국 Saint James High school(2007.10~2009.10)
- 봉은중학교(2009.10~2012.02)
- 휘문고등학교(2012.03~2015.02)
- 홍익대학교 컴퓨터공학과(2016.03~2022.02)
- 마이크로소프트 AI스쿨(2022.10~2023.01)
- 한양대학교 인공지능학과(석사)(2023.01~2025.02)

수강과목

- 학부: 선형대수학 A+, 확률과통계 A+, 인공지능 A0, 알고리즘분석A+,
- 대학원: 컴퓨터비전 A+, 딥러닝기초 A0, 단백질정보학 A+, 기계학습 A+, 인공지능논문연구 A+

경력

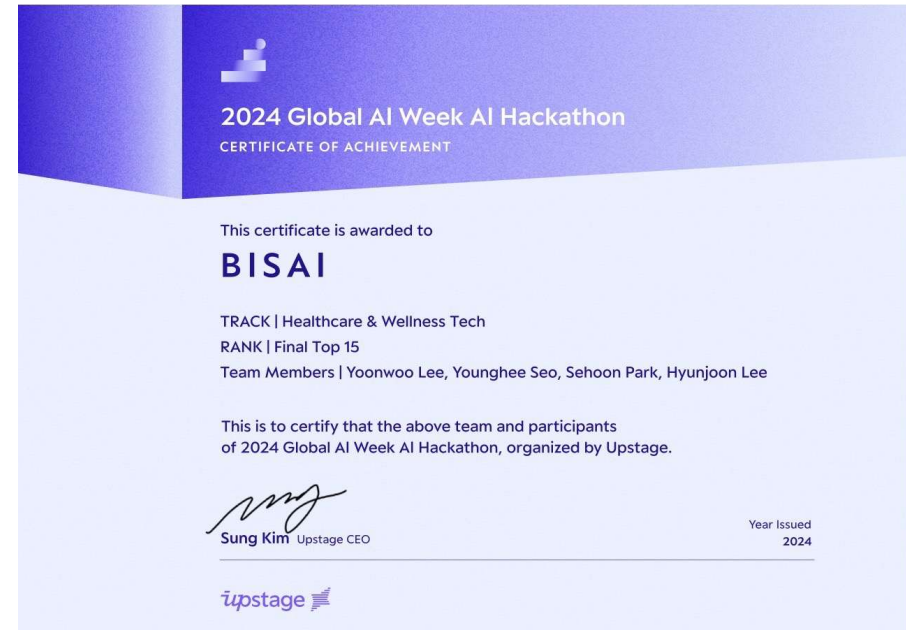
- 2025.09~ 현재: 한국기초과학지원연구원 연구원 - 단백질 분석 알고리즘 연구 및 개발, LLM을 활용한 실험자(wet-lab)을 위한 Guideline 생성(ProteoLLM), Lab Automation
- 2025.02~2025.06: 한양대학교 BISLAB 연구원- Upstage Hackathon에서 아이디어를 얻어, Library에 있는 sequence정보를 spectrum과 같이 임베딩하는 방식으로 성능을 개선하고, 미국 질량분석학회(ASMS)에서 포스터 발표
- 2020.06~2020.08: 원더풀플랫폼 인턴 - 챗봇 학습을 위한 데이터 수집, 전처리. API를 활용해 날씨 알리미 기능 추가

수상, 자격증

- Upstage Global AI Hackathon – Top15 certificate

[image]
업스테이지_cert

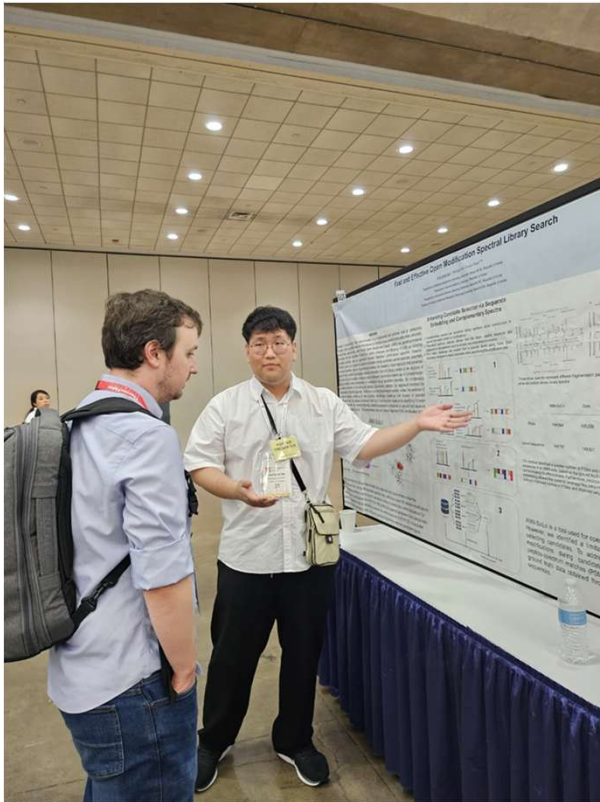
- Microsoft Azure AI900
- Microsoft Azure DP900



어학

- TOEIC – 855
- Opic – IH

발표



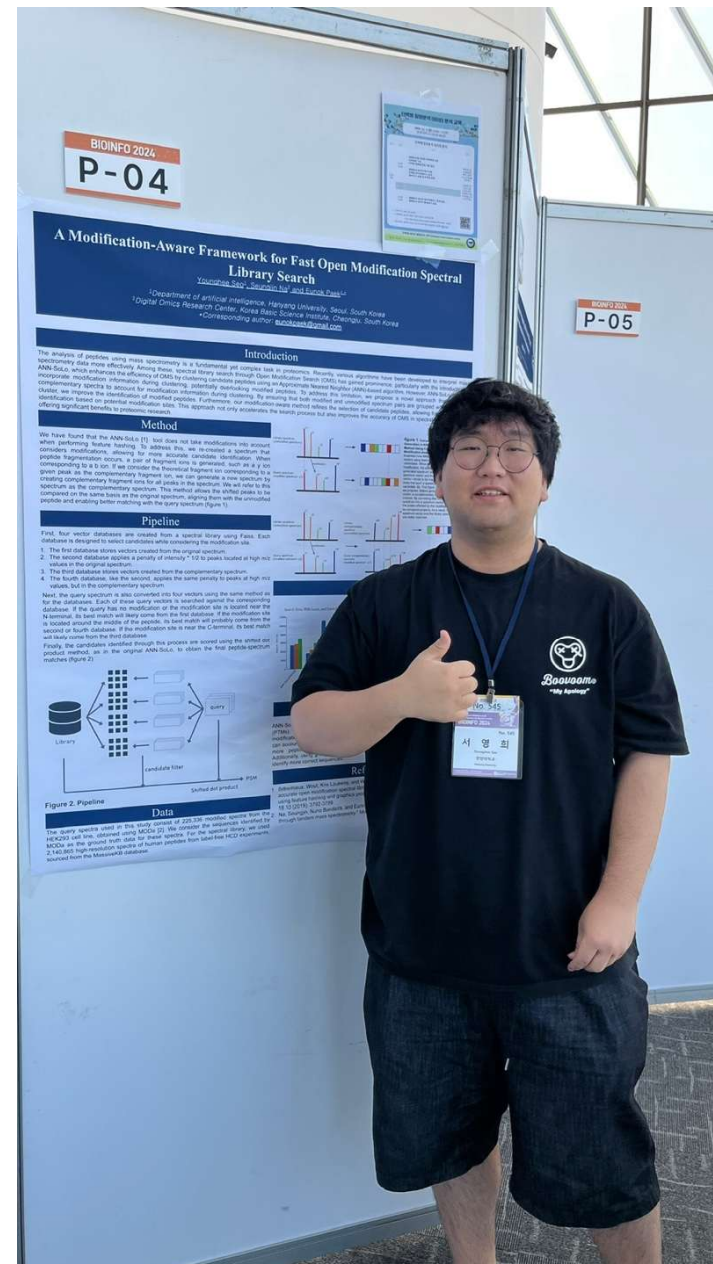
2025.06. – USA Baltimore,
ASMS포스터발표

[image]
ASMS발표

발표

2024.10 – 한국 경주, 생
명정보학회 포스터발표

[image]
생명정보학회

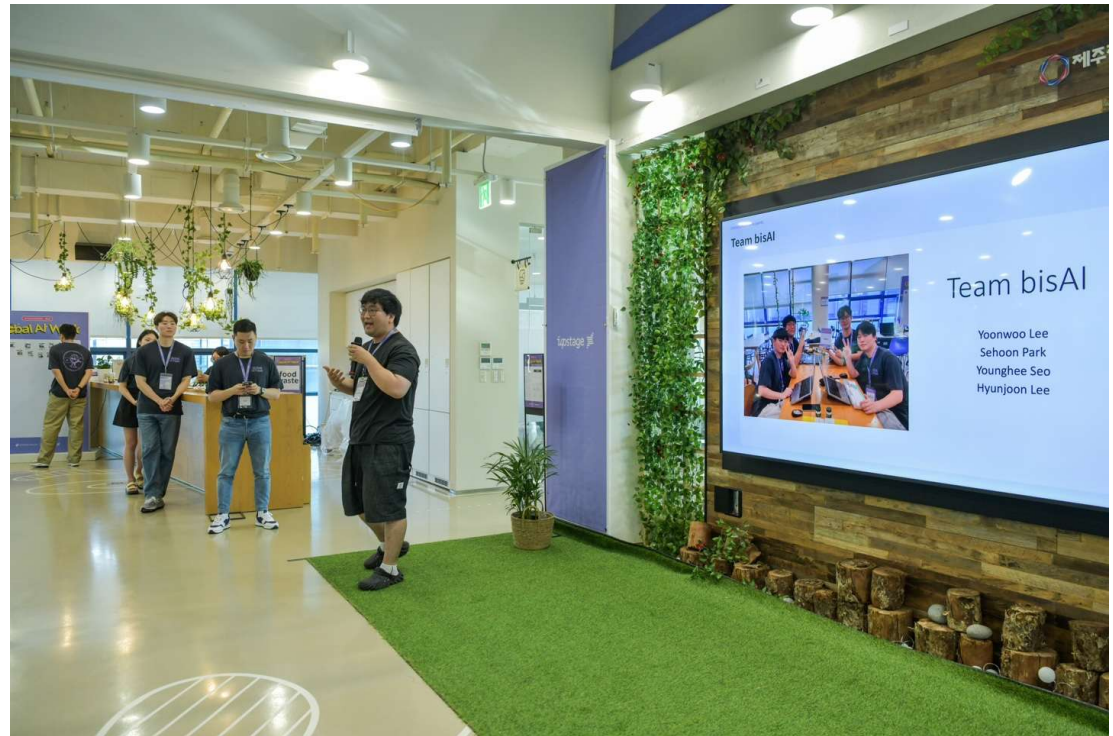


발표

2024. 10 – 한국 제주도,
Upstage Global AI
Hackathon 프로젝트 발표

<https://www.youtube.com/watch?v=8Zi2z13o0AU>
1:22:00 부터

[image]
업스테이지발표



Projects

- ANN-Complex
- ProteoLLM
- NutriPharmAI
- ImagePicker
- Country Zombie game

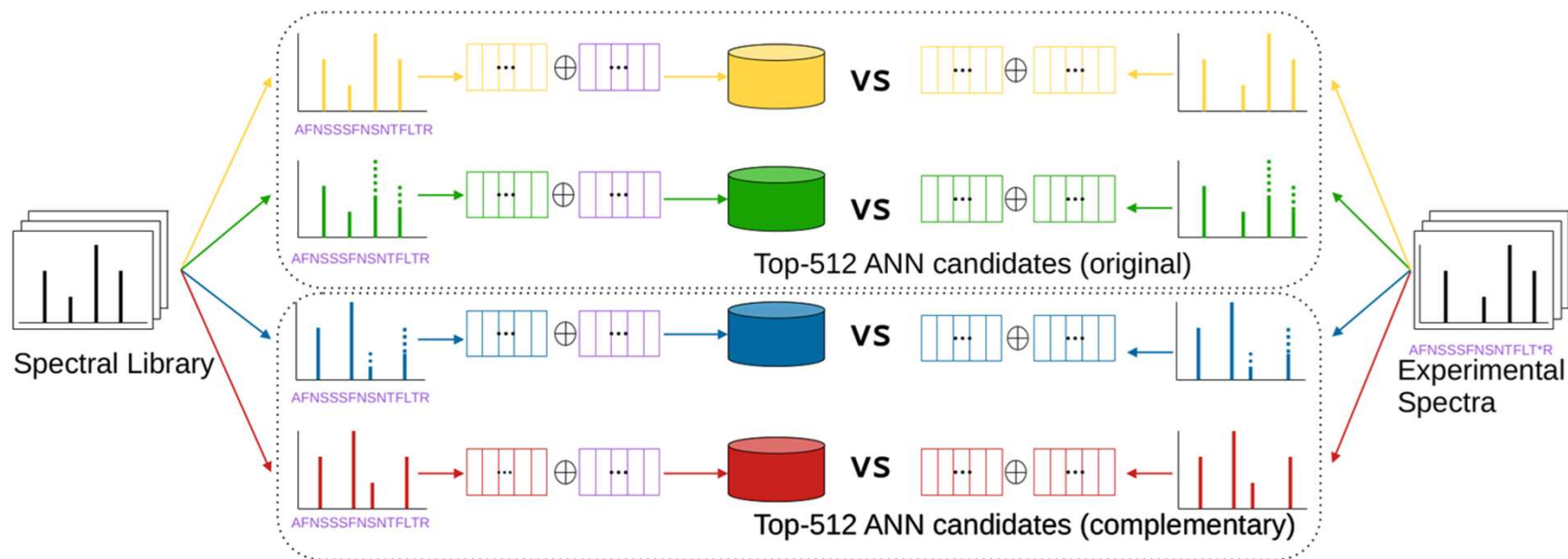
ANN-Complex

- https://github.com/tjdudgml3/fast_open_mod_search
- ASMS포스터 발표
- 한양대학교 석사 졸업연구

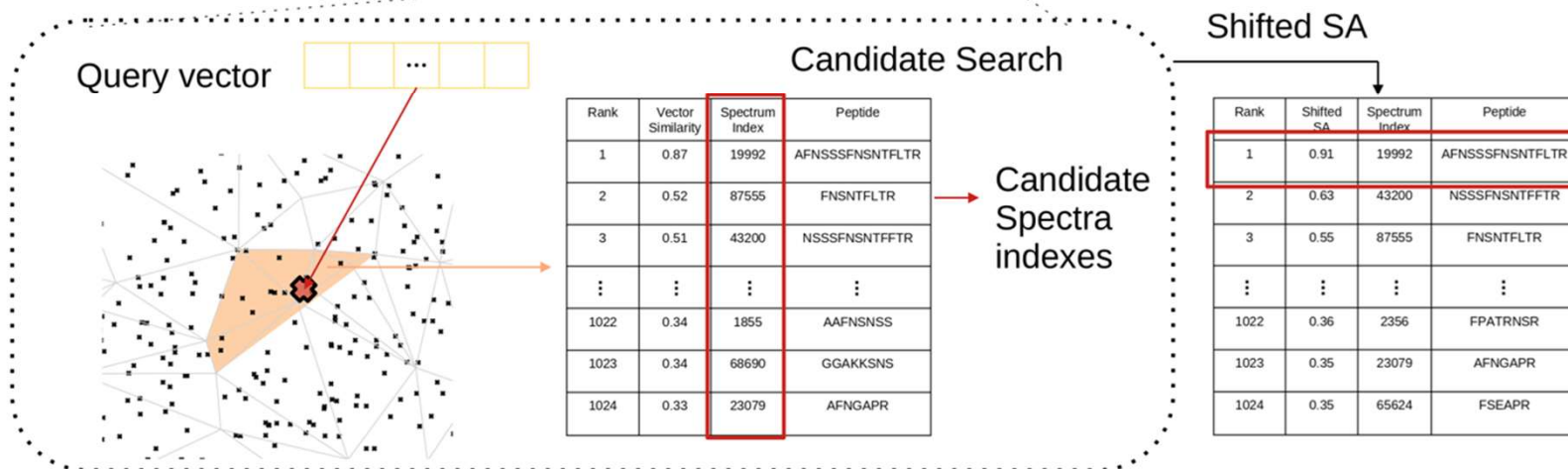
설명: 본 연구는 MS/MS 스펙트럼 기반의 단백질 동정 과정에서 발생할 수 있는 정보 손실 문제를 해결하고, 잠재적인 post-translational modification (PTM)의 탐지 민감도를 높이기 위한 알고리즘을 개발하였습니다. 특히 실험 스펙트럼에서 누락될 수 있는 이온 신호를 보완하기 위해, 이론적으로 생성 가능한 이온 위치를 계산하여 새로운 피크를 생성하고 이를 벡터화하는 complementary spectrum generation 방법을 제안함으로써 정보의 충실도를 향상시켰습니다. 기존 방법, complementary 방법, 두 방법의 가중치를 결합한 방식 등 총 4가지 전처리 방식을 각각 적용하여 4개의 데이터베이스를 구성하고 검색을 수행하였으며 이를 통해 정확도를 높이는 동시에 기존 방식에서 놓치기 쉬운 변형 펩타이드까지 포착할 수 있도록 설계하였습니다. Upstage Global AI Hackathon에서 아이디어를 받아 Sequence를 Spectrum과 같이 임베딩하여 성능을 5퍼센트 이상 향상시켰으며 이 연구 결과는 2025년 미국 질량분석학회(ASMS)에서 포스터 발표를 통해 기술적 성과와 실용 가능성을 인정받았습니다.

#Bioinformatics, Machine Learning, Vector Search, Data Analysis, Faiss, Python

Images for ANN-Complex



[image]
Paper_figure



ProteoLLM

- 한국기초과학지원 연구원에서 실험자(wet-lab)과 분석가(dry-lab)의 의사 소통에서 발생하는 문제가 빈번하여 고안한 프로젝트
- 한국기초과학지원연구원
- 설명: 실험자(Wet-lab)와 분석가(Dry-lab) 간의 데이터 소통 비효율을 해결하기 위해, Google ADK 기반의 '실험실 맞춤형 멀티에이전트 시스템'을 구축했습니다. 이 시스템은 Router, Search, Evaluator 에이전트가 유기적으로 협업하여 맥락 없는 데이터 전달 문제를 해결합니다. 특히 Faiss와 2-Stage Search를 도입해 논문 검색 속도를 획기적으로 단축했고, Gemini 3와 Flash 모델을 이원화하여 비용 효율을 최적화했습니다. 또한, JSON 통신과 Self-Evaluation Loop를 설계해 데이터의 무결성을 확보함으로써, 연구소 내 협업 생산성을 실질적으로 향상시키고 있습니다.

#Multi Agent System, Multi-stage search, Self-evaluation Loop

ProteoLLM

[image]
ProteoLLM_web

Configuration

Google API Key

Agent Observability

[LLM_RESPONSE] Model
Response received

[AGENT_END] Completed
Agent: literature

[AGENT_START] Starting Agent:
evidence_qa

[LLM_REQUEST] Requesting
gemini-2.0-flash

[LLM_RESPONSE] Model
Response received

[AGENT_END] Completed
Agent: evidence_qa

[AGENT_START] Starting Agent:
analyst_advisor

[LLM_REQUEST] Requesting
gemini-3-pro-preview

[LLM_RESPONSE] Model
Response received

[AGENT_END] Completed
Agent: analyst_advisor

[AGENT_START] Starting Agent:
explainer

[LLM_REQUEST] Requesting
gemini-2.0-flash

[LLM_RESPONSE] Model
Response received

[AGENT_END] Completed
Agent: explainer

Bioinformatics Multi-Agent System

Powered by Google ADK Framework and Gemini 3.0, 2.0

sumo가 관찰 안되고 tnf만 관찰되네

알겠습니다. SUMO가 관찰되지 않고 TNF만 관찰되는 상황에 대한 분석 및 해결 방안을 설명해 드리겠습니다.

문제: SUMOylation 신호가 일시적인 특성, 특이적인 농축 부족 또는 부적절한 시료 준비로 인해 소실되었을 가능성이 있습니다. 반면 TNF는 (처리 물질이거나 주요 염증 마커일 가능성이 높음) 풍부하게 존재합니다.

분석 계획: SUMOylation 미검출 원인 조사

1. 현재 데이터 평가 (진단 단계)

- 가설: SUMOylation은 양이 적고 비정상적입니다. 농축 없이는 일반적인 LC-MS/MS 분석에서 변형되지 않은 단백질(예: TNF)이 주를 이룹니다. 또한 세포 용해 시 내인성 SUMO 프로테아제(SENPs)가 빠르게 단백질에서 SUMO를 제거할 수 있습니다 (SENPs 억제제가 없을 경우).
- 조치:
 - 단백질 vs. PTM 확인: "SUMO가 관찰되지 않음"이 SUMO 단백질 자체(SUMO1/2/3)를 의미하는지, 아니면 SUMOylated 펩타이드를 의미하는지 확인합니다.
 - 특정 잔류물 재검색: 트립신을 사용한 경우 표준 검색 엔진은 포유류 SUMO의 트립신 잔류물이 크고 분기되어 있기 때문에 SUMO 부위를 식별하지 못하는 경우가 많습니다.
 - 참고: 특정 효소(예: WaLP)를 사용하지 않은 경우 SUMO C-말단에 해당하는 특정 질량 이동을 검색합니다 [1].
 - TNF 평가: TNF가 외인성 처리 물질(배지/상층액에서 발견되거나 수용체에 결합)인지, 내인적으로 발현되는지 확인합니다. TNF 존재는 MS 실험이 작동했음을 확인하여 장비 전체 실패를 배제합니다.

2. 시료 준비 및 농축 전략 (다음 단계)

- 농축 필수: 내인성 SUMOylation은 농축 없이는 검출하기 어렵습니다.
 - 제한 방법: SUMO 잔류물 모티프 형태(예: 특정 돌연변이체 또는 화학적 처리를 사용하는 경우 K-ε-GG 형태) 또는 에피토프 태그된 SUMO(His-SUMO) 풀다운 후 Ni-NTA 농축을 사용합니다 [2, 3].
- 소화 효소 최적화:
 - 트립신은 스펙트럼을 복잡하게 만드는 큰 SUMO 잔류물을 남깁니다.
 - 권장 사항: wild-type α-lytic protease (WaLP) 사용을 고려하십시오. WaLP는 SUMOylated 단백질을 소화하여 더 작고 뚜렷한 SUMO 잔류물(diglycyl-lysine, KGG)을 가진 펩타이드를 생성하므로 MS를 통해 더 나은 식별이 가능합니다 [1].
- 탈SUMO화 효소 억제:
 - SENPs를 비가역적으로 억제하고 SUMO 사슬을 보존하기 위해 용해 완충액에 **NEM (N-Ethylmaleimide)**을 엄격하게 포함해야 합니다 [2].

3. 생물학적 해석

- TNF 맥락: TNF 신호 전달은 SUMOylation에 의해 크게 조절되는 경로의 NF-κB를 활성화합니다 (예: NEMO/IKKγ). TNF만 보이는 경우 하위 캐스케이드가 활성화될 수 있지만 PTM이 누락되었을 수 있습니다.

실험자에게 묻는 질문

SUMO 미검출 문제를 해결하려면 다음 메타데이터가 필요합니다.

- 시료 준비 (중요):
 - SUMO 특이적 농축 방법(예: SUMO 항체를 사용한 IP, His-tag 풀다운 또는 COFRADIC)을 사용했습니까? 아니면 "전체 용해물" 분석입니까?

Ask about bioinformatics (e.g., "Find targets for breast cancer")

NutriPharmAI

- <https://github.com/sehooni/NutriPharmAI>

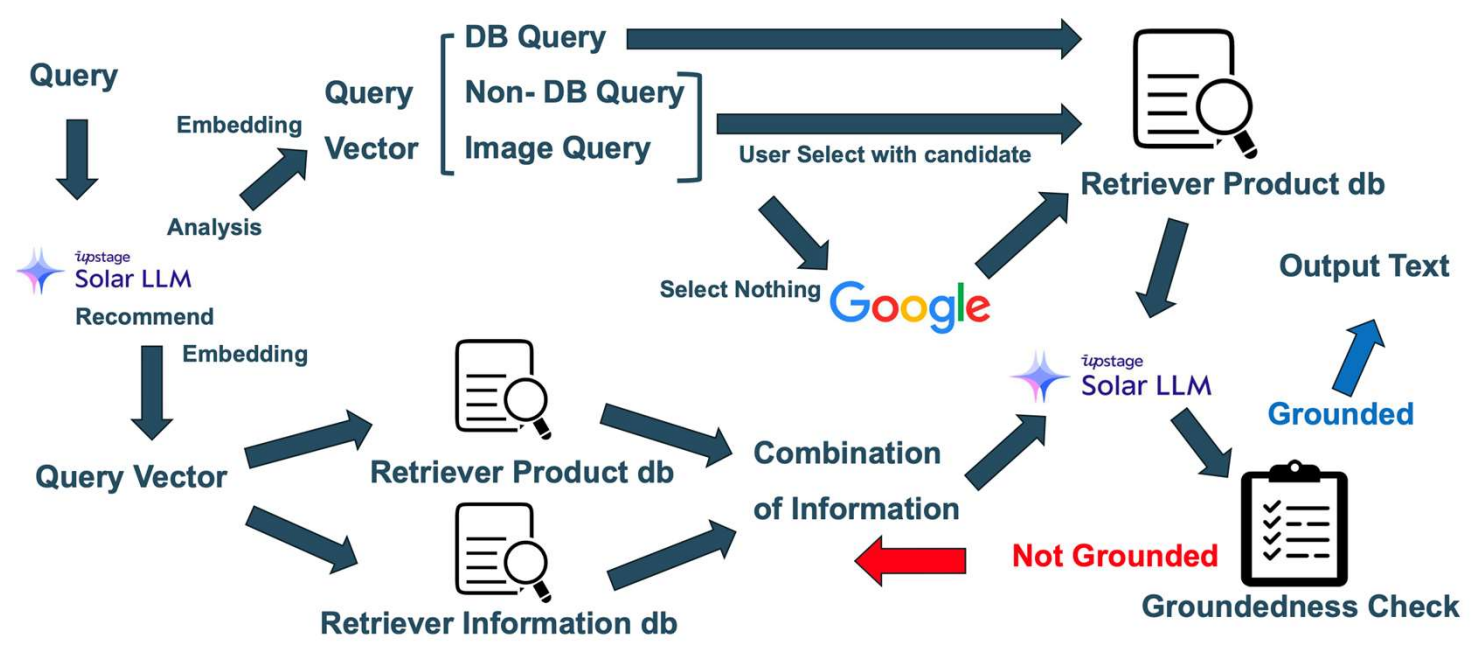
- Upstage Global Hackathon

- Top 15 Certificate

- 설명: Upstage LLM 해커톤에서 영양제 분석 웹서비스를 개발했습니다. 우선 사용자 편의를 위해 OCR 기술을 도입하여, 텍스트뿐만 아니라 제품 사진만으로도 성분을 분석할 수 있도록 노력했습니다. 또한, 단순 RAG의 한계인 할루시네이션을 해결하고자 멀티 에이전트 협업 시스템을 구축했습니다. 에이전트가 사용자 의도를 파악해 내부 DB뿐만 아니라 Google Search Engine를 통한 웹 검색까지 활용함으로써, DB에 없는 제품까지 커버하는 확장성을 확보했습니다. 또한 생성된 답변이 검색된 문서와 일치하는지 검증(Groundedness Check)하고, 근거가 부족하면 재 검색을 수행하는 루프를 통해 신뢰도 높은 정보를 제공했습니다. 해커톤 당시 닥터 임베딩을 무차단으로 변경하여 성능을 비약적으로, 이 아이디어를 석사 연구의 스펙트럼 임베딩과 아미노산 서열 정보를 같이 임베딩하는 기술로 발전시켜 알고리즘 성능을 5% 이상 향상시켰고, 그 결과 미국 질량분석학회(ASMS) 발표 기회를 얻었습니다.

#Multi Agent System, RAG, Web-crawling, LLM, OCR, Google Search Engine, Groundedness Check

NutriPharmAI



[image]
NutriPharmAI_figure

[image]
NutriPharmAI_web

2. **개이거가 상태** 특정 지변이나 안목 불요 종이 경우 영양제 선척이 영한을 미칠 수 있습니다. 예를 들어 가 지화이 안는 경우 미크씨수 선척에 조영이 필요합니다.

ImagePicker

- 홍익대학교 졸업프로젝트
- https://github.com/tjdudgml3/image_picker
- End-to-End AI project
- 설명: 학부 졸업 프로젝트로 스마트폰 사진첩을 자동으로 정리해주는 AI 애플리케이션을 개발. 실제 사용 환경인 모바일에 최적화하기 위해, 웹 기반 시연 계획을 앱으로 전환하여 개발. ResNet 모델을 학습시켜 텐서플로우 라이트를 통해 모바일 환경에 경량화하여 이식했으며, 리사이클러뷰(RecyclerView) UI를 구현. 데이터 수집부터 모델링, 앱 개발까지의 End-to-End 프로젝트.
- #App Development, Computer Vision, ResNet, Web-crawling, Tensorflow, Python, Java

ImagePicker

시연영상 -

<https://www.youtube.com/shorts/HqqoxfbQWzM>

[image]
Hongik_app



Country Zombie Game

- 학부 프로젝트
- Unity Project
- 게임개발
- 설명: 유니티를 이용한 게임개발.
- https://github.com/tjdudgml3/country_zombie_game
- # Unity, C#, Game programming

Country Zombie Game

[image]
Country_zombie

