류예찬(Yeh Chan Yoo)

대한민국 | yehchanyoo.github.io/ko | linkedin.com/in/yehchan-yoo/?locale=ko-KR

학력

워싱턴대학교(시애틀 본교)(University of Washington - Seattle) *2024.09 ~ 2026.03*

통계학 석사(M.S. in Statistics - Advanced Methods and Data Analysis)

- 주요 이수 과목: 고급 머신 러닝(Advanced Machine Learning), 응용통계 캡스톤(Applied Statistics Capstone), 계산생물학(Computational Biology), 실험설계 및 분석(Design & Analysis of Experiments), 계산통계학(Statistical Computing), 응용회귀분석(Applied Regression), 통계추론(Statistical Inference)

UC 버클리(University of California, Berkeley) *2017.08 ~ 2023.12*

통계학·정치경제학 복수전공 학사, 데이터과학 부전공(B.A. in Statistics and Political Economy, Minor in Data Science)

- 주요 이수 과목: 인과추론(Causal Inference), 데이터·추론·의사결정(Data, Inference, and Decisions), 인공지능개론(Introduction to Artificial Intelligence), 현대통계예측 및 머신 러닝(Modern Statistical Prediction and Machine Learning), 데이터과학 원리 및 기법(Principles and Techniques of Data Science)

경력

앨런연구소(Allen Institute) *2025.09 ~ 2026.03*

방문학생 연구원(Visiting Student) · 미국 워싱턴주 시애틀

- 앨런연구소 내 앨런뇌과학연구소(Allen Institute for Brain Science)에서 500GB 이상의 단일세포 데이터를 이용해 PyTorch 기반 딥러닝 모듈을 포함한 베이지안 파운데이션 모델을 설계·학습·검증
- Docker 컨테이너화, Shell 자동화 스크립트, 단위/기능 테스트 작성을 통해 AWS 및 HPC 환경에서 코드 재현성과 유지보수성을 확보
- 단일세포 전사체 변이와 개인별 표현형을 연결하는 생성형 프레임워크 scVIP 모델 논문에 제2저자로 참여, 오픈소스 패키지 scvi-tools에 편입될 수 있도록 설계

  scVIP 프리프린트 논문: https://www.biorxiv.org/content/10.64898/2026.04.20.717759v2

  scVIP 모델 코드베이스: https://codeocean.allenneuraldynamics.org/capsule/4344243/tree/v3

마인드풀 컨버전(Mindful Conversion) *2024.01 ~ 2024.08*

데이터과학자(Data Scientist) · 미국 뉴욕주 뉴욕

- 마케팅 데이터 분석 스타트업 Mindful Conversion의 전담 데이터 사이언티스트로서 비즈니스 분석을 주도하고 제품·마케팅·분석 프로젝트 지원
- 예측형 검색엔진 최적화(SEO) 제품 Kixely의 데이터 파이프라인을 개발·운영, Azure 및 Google Cloud Platform(Vertex AI 포함) 환경에서 데이터 수집·처리·모델 통합 지원
- 약 10만 개의 키워드·URL 관련 데이터를 수집·분석하여 Kixely 랭킹 예측 모델의 프로토타입 개발부터 프로덕션 배포까지 단독으로 주도하면서, Kixely의 LLM 파이프라인 구축 및 평가 병행
- Python과 SQL을 활용해 10GB 이상의 마케팅 데이터에 대한 탐색적 데이터 분석(EDA)을 주도하였으며, 분석 결과를 이용해 데이터 수집 및 콘텐츠 전략의 핵심 문제를 발굴·해결하고 프로덕트 개선을 견인

- 비즈니스 및 커뮤니케이션 역량을 바탕으로 5건 이상의 데이터 기반 보고서 및 발표 자료를 작성하여 자사 및 클라이언트의 마케팅 전략 수립에 기여

UC 버클리 교육대학(UC Berkeley School of Education) *2023.06 ~ 2023.12*

학생 연구원(Student Assistant II) · 미국 캘리포니아주 버클리

- 수학·통계 문제에 대한 200개 이상의 학생 답안 자동 채점을 위한 자연어 처리 모델을 Jupyter Notebook 및 Google Colab 환경에서 Hugging Face Transformers로 연구·개발
- PyTorch 및 CUDA 기반으로 직접 파인튜닝 및 지식 증류(distillation)한 RoBERTa 모델을 활용하여 테스트 셋 정확도 75% 달성
- 박사후연구원과 협업하여 캘리포니아주 공립 중등교육 학생 대상 통계 평가 문항 다수를 설계 및 작성

프로젝트

디지털 병리학을 위한 딥러닝: PatchCamelyon(PCam) 데이터셋 기반 종양 탐지(Deep Learning for Digital Pathology: Tumor Detection Using the PCam Dataset) *2026.01 ~ 2026.03*

- 30만 장 이상의 림프절 조직 병리 스캔 데이터(PatchCamelyon)에 이미지 인식·객체 탐지 기법을 적용한 종양 탐지 컴퓨터 비전 프로젝트 수행(워싱턴대 계산생물학 과목 팀 프로젝트)
- 로컬 환경에서 Visual Geometry Group 스타일 합성곱 신경망(convolutional neural net) 모델을 직접 설계하고, 클라우드에서 Optuna를 활용해 ResNet18 모델을 파인튜닝 및 하이퍼파라미터 최적화
- SHAP 및 GradCAM 시각화를 통해 두 모델의 성능-해석 가능성 트레이드오프를 비교 분석

※ 상세 내용(영문): https://yehchanyoo.github.io/attachments/2026/CSE_527_Final_Report___PCam.pdf

사회경제적 요인 통제하에 고객 잔고 및 신용 채무불이행이 정기예금 가입에 미치는 영향(The Impact of Client Balance and Credit Default on Term Deposit Subscription, Controlling for Socioeconomic Factors) *2025.11 ~ 2025.12*

- 포르투갈 은행 뱅킹 마케팅 데이터셋을 활용, 사회경제적 변수의 결측치를 통제한 상태에서 고객 잔고 및 신용 부도가 정기예금 가입 여부에 미치는 영향을 분석(워싱턴대 결측데이터 통계분석기법 과목 기말 프로젝트)
- 완전사례분석(CC), 베이지안 우도 추정, 다중대치법(MI) 세 가지 결측치 처리 기법을 비교 적용해 잔고 10,000유로 증가당 가입 확률 약 26% 증가, 신용 부도 시 약 38~41% 감소라는 결과 도출
- 완전사례분석이 신용 부도의 부정적 효과를 과대추정함을 규명, 결측치 처리 방법 선택에 따른 편향 가능성 제시

※ 상세 내용(영문): https://yehchanyoo.github.io/attachments/2025/BIOST_531_Final_Project.pdf

놀이 속 감정: AI와 실체적 도구를 활용한 아동의 정서 학습(Emotions at Play: Children's Emotional Learning with AI and Tangible Tools) *2024.10 ~ 2025.09*

- 워싱턴대 정보대에서 아동을 위한 실시간 감정 메시지·스티커 생성 서비스의 프론트엔드·백엔드를 아우르는 Android 앱 및 웹 앱 설계·개발
- OpenAI 및 Google 멀티모달 모델 연동을 위한 백엔드 API 통합 레이어 구현 — 프롬프트 엔지니어링, 응답 파싱, 실시간 콘텐츠 안전성 탐지·필터링 로직 설계

※ 이 프로젝트는 다른 프로젝트들과 달리 통계학과 전공 수업과 관련되어 있지 않지만, 일반 인공지능 업무와 관련되어 있음.

Rec-R1 재현 연구(Replication of Rec-R1) *2025.04 ~ 2025.06*

- 대형 언어 모델과 추천 시스템을 결합한 강화학습 기반 프레임워크 Rec-R1 프리프린트 논문의 결과를 재현 및 비판적으로 검토(워싱턴대 고급 머신 러닝 과목 팀 프로젝트)
- 벡터 데이터베이스(FAISS) 및 실험 추적 도구(Weights & Biases)를 활용해 추천·상품 검색 성능을 평가

※ 상세 내용(영문): https://yehchanyoo.github.io/attachments/2025/CSE_493S_599S_Final_Project.pdf
※ 코드: https://github.com/yehchanyoo/Rec-R1_magic

결측 데이터 조건별 대치법 평가

(Evaluation of Imputation Methods Under Different Missing Data Conditions) *2025.03 ~ 2025.06*

- 2023년 미국 지역사회조사(ACS) 뉴욕주 마이크로데이터를 활용, MCAR/MAR/MNAR 세 가지 결측 메커니즘 하에서 5가지 대치법(무대치·평균·랜덤·최근접 이웃·회귀)의 성능을 시뮬레이션 비교(워싱턴대 표본조사기법 과목 기말 프로젝트)
- 임대료(RNTP) 및 부동산 가치(VALP) 변수에 대해 대치법별 평균·분위수 추정 정확도를 정량 평가하고, MNAR 조건에서 랜덤 대치가 분위수 추정에 가장 우수함을 규명

※ 상세 내용(영문): https://yehchanyoo.github.io/attachments/2025/stat529_final_project_yehchan_yoo.pdf

※ 포스터(영문): https://yehchanyoo.github.io/attachments/2025/stat529_poster_final.pdf

「401(k) 가입자격 변화가 저축에 미치는 영향」 재현 및 개선 연구(Replication and Improvement on "How do 401(k)s A↵ect Saving? Evidence from Change in 401(k) Eligibility")

2022.10 ~ 2022.12

- 1996 SIPP 원자료를 직접 클리닝해 Gelber(2011) 논문의 이중차분법 분석을 재현, 401(k) 가입자격 취득이 401(k)·IRA 저축에 미치는 인과효과(계수 약 1.03, 0.46)를 확인(UC버클리 인과추론 과목 팀 프로젝트)
- 아웃라이어 윈저화·성향점수 층화·1년차 자산누적 등 3가지 강건성 검정을 재현해 원 논문 결과의 한계(특히 성향점수 층화 시 신뢰구간이 0을 포함) 규명
- 성향점수매칭 후에도 남은 공변량 불균형 문제를 이중강건추정량(doubly robust estimator)으로 보완, 401(k) 저축 효과는 유의하나 IRA 저축 효과는 유의하지 않음을 재확인

※ 상세 내용(영문): https://yehchanyoo.github.io/attachments/2022/group04_YY_XZ.pdf

공모전

창발 이노베이션 해커톤 2026(Changbal Innovation Hack 2026)

2026.06

최고 음성 AI 활용상(Best Voice Award)

• 음성 기반 자동화 저널링 및 임상 문서화 웹 플랫폼 Cadence 공동 개발

• BizCrush의 음성 전사(Transcription) API를 도입, 실시간 오디오를 구조화된 요약 및 실행 목표로 전환하는 파이프라인 구축

• 실시간 교차 언어 번역 기능을 성공적으로 구현하며 기술적 활용도를 인정받아 Best Voice Award 수상

※ 프로젝트 링크: https://devpost.com/software/cadence-ewfcan

UC 버클리 사회적 가치 창출 데이터톤(Datathon for Social Good)

2022.11

도시과학(Urban Sciences) 트랙 2위

• 음성• 캘리포니아주 카운티별 대중교통 통계와 1인당 주택 수의 상관관계 분석

• 선형 회귀(Linear Regression) 모델링 및 데이터 시각화 기법을 활용하여, 대중교통 인프라와 주거 밀도 간의 유의미한 패턴 도출

• 도시과학 분야에서의 우수한 데이터 분석 역량을 인정받아 전체 2위 수상

※ 프로젝트 링크: http://bit.ly/41aml23

기술 스택(Skill set)

- 프로그래밍 언어: **Python, R, SQL**, Git, Shell, C/C++, JavaScript(React, Node.js), Kotlin
- 개발 환경 및 플랫폼: Docker, AWS, Google Cloud Platform(Vertex AI 포함), Microsoft Azure, Jupyter Notebook, RStudio, GitHub, Android Studio