

강민지 (Stella Kang) - AI/NLP 인턴십 포트폴리오

mkang17@vt.edu | 블랙스버그, 버지니아 (Blacksburg, VA, USA) | +1-540-230-8328

LinkedIn: <https://linkedin.com/in/stellak-VT> | GitHub: @stellak-vt

소개 (Introduction)

버지니아 공대(Virginia Tech)에서 응용이산수학(Applied Discrete Mathematics)을 전공하고, 컴퓨터과학(Computer Science)을 부전공하고 있는 학부 3학년 재학생 강민지입니다. 대학 1학년 이후 진로를 깊이 고민하던 중, AI 챗봇이 우울증과 불안 증상 개선에 도움을 준다는 임상 연구를 접하며 큰 영감을 받았습니다. 이 경험을 계기로, AI 기술이 사람의 감정을 이해하고 공감하며 정신 건강 회복을 돕는 도구가 될 수 있다는 가능성에 매료되었습니다. 제 궁극적인 목표는 정신질환을 보다 부담 없이 진단하고 회복할 수 있도록 돕는 AI 기반 치료 챗봇을 연구·개발하는 것입니다. 수학적 사고력과 프로그래밍 역량을 바탕으로 실제 문제 해결에 기여하는 AI 연구에 진심으로 참여하고자 하며, 학기 중 100% 리모트로 진행되는 무급 인턴십 또는 연구 협업 형태의 참여를 희망합니다.

[프로젝트] 스킨케어 제품 추천 시스템

기간: 2025년 3월 ~ 진행 중

기술: Java, Eclipse, VS Code

- 사용자 피부 타입과 선호도를 기반으로 스킨케어 제품을 추천하는 규칙 기반 시스템 설계
- CLI 기반 프로토타입 구현 완료 (텍스트 입력을 통한 조건 필터링)
- 제품 속성, 사용 대상, 제형 등의 데이터를 기반으로 추천 로직 개발 중
- 추후 콘텐츠 기반 추천 및 GUI 버전으로 확장 예정

Developed a rule-based product recommendation system based on user attributes such as skin type and preferences. Currently working on extending it into a content-based filtering system.

GitHub: [SimpleRecommenderSystem](#)

[리서치] 자연어처리 연구 보조

소속: Virginia Tech University Libraries

기간: 2025년 6월 - 8월

도구: PyTorch, Hugging Face Transformers, Jupyter, Kaggle

- 2025년 가을학기 NLP 리서치 참여 전, 사전 멘토링 프로그램 참여
- Hugging Face Transformers 튜토리얼 실습 및 감정 분석 예제 학습
- PyTorch 모델 학습/평가 파이프라인 구조 이해

Reviewed and practiced fine-tuning Hugging Face Transformer models for text classification and sentiment analysis.

[리서치] Submodular Optimization 기반 이미지 분류 연구

소속: Virginia Tech 응용대수 연구실 / NSF REU 프로그램

기간: 2025년 5월 - 8월

기술: Python, PyTorch, NumPy, Matplotlib, Overleaf

- 대표 데이터셋 추출을 위한 서브모듈러 최적화 기법 분석 및 구현
- CRAIG(Coreset Selection for Efficient and Robust Training) 알고리즘 기반 코어셋 선택 성능 비교 실험 수행
- stochastic greedy vs. regular greedy 방식 비교, loss 그래프 시각화
- PyTorch 기반 실험 코드를 구성하고, NumPy 및 Matplotlib을 활용하여 성능 결과를 분석
- Overleaf를 활용해 논문 기반 개념 및 실험 해설 자료 정리 (비전공자도 이해 가능하도록 작성)

Conducted research on subset selection via submodular optimization to improve training efficiency in image classification tasks.

참고자료:

[1] K. Killamsetty, X. Zhao, F. Chen, and R. Iyer, "Retrieve: Coreset selection for efficient and robust semi-supervised learning," 2021. [Online]. Available:

<https://arxiv.org/abs/2106.07760>

[2] B. Mirzasoleiman, J. Bilmes, and J. Leskovec, "Coresets for data-efficient training of machine learning models," 2020. [Online]. Available:

<https://arxiv.org/abs/1906.01827>

[교육활동] 수학 아웃리치 TA

소속: Virginia Tech 응용대수 연구실

기간: 2025년 6월 - 7월

- 중학생부터 대학생까지 다양한 수준의 학생들을 대상으로 수학 및 컴퓨팅 개념 강의 및 실습 진행
- 이진수, ASCII, Lagrange 보간, Reed-Solomon 코드 등 추상적 개념을 실제 컴퓨터과학 및 암호 응용 사례와 연결하여 설명

세부 활동:

- COS Camp - Math TA (2025년 7월): 중학교 1학년, 고등학교 1학년 학생 약 100명을 대상으로 이진수와 ASCII 원리 강의 및 활동 지원 2회
- Life Science Camp - Math TA (2025년 6월): 고등학교 1~2학년 학생 48명을 대상으로 Lagrange 보간과 Reed-Solomon 코드 실습 지원
- JMU REU Program 참여 보조 (2025년 6월): JMU REU 프로그램 참가자들과 함께 유한체 위 다항식 관련 연구 주제에 참여

[워크숍] OCR & AI 문자 인식 실습

장소: Virginia Tech Library

기간: 2025년 5월

- 저해상도 문서 이미지의 OCR 적용 및 전처리 실습
- LLaMA 기반 AI 모델을 활용한 텍스트 추론 시도

Applied OCR and LLM tools (LLaMA) on noisy scanned documents using Python.

기술 역량

- 언어: Python, Java
- 프레임워크/라이브러리: PyTorch, NumPy, Matplotlib, Hugging Face Transformers
- 개발 환경: Jupyter, VS Code, Eclipse, Git (기초 수준)
- 언어능력: 한국어(모국어), 영어(상급)

마무리 안내 (Final Note)

이 포트폴리오는 현재까지의 프로젝트 및 리서치 준비 과정을 바탕으로 작성되었으며, 향후 연구 결과 및 구현물이 완성되는 대로 계속 확장해 나갈 예정입니다. 자연어처리 분야에서의 실무와 협업 경험을 통해 더 깊이 있는 성장을 이루고 싶습니다.