

2024년 1학기 UROP 참여과제 리스트

연번	교수명	연구실명	연구과제명	연구주제
1	강 유	데이터 마이닝 연구실	Large Language Model Compression	<ul style="list-style-type: none"> - 언어 모델 압축 기술은 학습이 되어 있는 언어 모델의 크기를 줄여 메모리 및 계산 필요량을 줄이는 것을 목표로 한다. - 최근 언어 모델의 성능이 비약적으로 발전하고 있으며, 그 과정에서 언어 모델의 크기 또한 거대해졌기 때문에, 높은 성능을 유지하면서 모델을 압축하는 기술이 매우 중요해지고 있다. - 대규모 언어 모델 압축(large language model compression) 기술에는 양자화, 가지치기, 그리고 지식 증류 등의 다양한 기술이 있다. - 본 연구에서는 언어 모델 압축(large language model compression) 기술을 연구하고 구현하는 것을 목표로 한다. - 본 연구에는 선형대수, 딥러닝, Python 등에 대한 지식이 요구된다.
2			Diversified Sequential Bundle Recommendation	<ul style="list-style-type: none"> - 추천 시스템은 사용자에게 개인화된 서비스를 제공하기 위해 많은 온라인 플랫폼에서 사용되는 중요한 기술이다. - 특히, 사용자의 과거 일련의 행동을 기반으로 미래의 행동을 예측하는 순차적 추천 시스템(sequential recommender systems)과, 여러 아이템을 동시에 추천하는 번들 추천은 매우 중요한 연구 주제이다. - 그러나 기존의 추천 시스템은 추천 결과의 다양성을 다루지 않기 때문에 동일한 카테고리 또는 비슷한 특성을 가진 항목을 지나치게 많이 추천하는 경향이 있다. - 이러한 다양성 부족 문제는 사용자의 경험을 한정시키고, 장기적으로 추천의 효과도 떨어트릴 수 있다. - 본 연구에서는 다양화된 순차적 번들 추천을 연구하고 구현하는 것을 목표로 한다. - 본 연구에는 선형대수, 딥러닝, Python 등에 대한 지식이 요구된다.

연번	교수명	연구실명	연구과제명	연구주제
3	권태경	인터넷 융합 및 보안	<p>Ensuring the integrity of DNS data by leveraging monitors in certificate transparency</p>	<p>1. 연구주제 일반적인 클라이언트가 서버에 연결할 때 DNS의 정보를 기반으로 해서 연결을 설정함. 그러나 현재 DNS 데이터는 man-in-the-middle attack, poisoning attack 등 여러 가지 공격에 의해 변조될 수 있음. 따라서 DNS 서버는 자신의 DNS 레코드들을 자신의 인증서(certificate)의 공개키와 연결된 개인키로 서명함. 이 때 중간에서 해당 DNS 레코드들을 모두 제거하는 downgrade attack이 가능함. 이를 방지하기 위해 클라이언트는 별도로 certificate transparency의 monitor들과 연결을 맺어 해당 서버의 인증서를 별도로 받아옴. 이를 통해 DNS 레코드가 중간에 변조 혹은 제거되지 않았는지 검사하는 과정을 추가하여 DNS 데이터들의 무결성을 보장하려고 함.</p> <p>2. 요구사항</p> <ul style="list-style-type: none"> - TCP/IP 프로토콜, DNS 프로토콜, PKI 등의 기술을 이해 - 2023년 겨울방학과 2024년 여름방학은 100% 연구주제에 참여 - 2024년 봄학기는 30% 연구주제에 참여 - 2023년 가을학기 현재 인터넷보안 수업 수강 <p>3. 기타</p> <ul style="list-style-type: none"> - 인건비 일부 지급 가능 - 연구 진행 속도에 따라 조기 종료 가능

연번	교수명	연구실명	연구과제명	연구주제
4	김건희	시각 및 학습 연구실	컴퓨터 비전 및 자연어 처리 관련 연구	<p>Conduct research and development on various topics related to computer vision, natural language processing, and multimodal.</p> <p>One example is the collection of multimodal funny video datasets.</p> <ul style="list-style-type: none"> - As short-form funny videos on social networks are gaining popularity, it becomes demanding for AI models to understand them for better communication with humans. Different from previous datasets, we devise a pipeline to collect the videos that have both visual and verbal elements contributing to humor. - After collecting and annotating funny videos, we devise a zero-shot prompting method with state-of-the-art models to explain why the video is funny. Also, to evaluate generated explanations, we design a task that finds funny moments given videos and generated explanations. <p>Requirement:</p> <ul style="list-style-type: none"> - 기계학습, 컴퓨터 비전, 자연어 처리 관련 교과목 세 과목 이상 수강. - 해당 주제(vision and language)와 관련된 배경 지식이 있는 학생. - PyTorch 프로그래밍 경험이 있는 학생.
5	김진수	시스템 소프트웨어 및 구조 연구실	DPU(Data Processing Unit)를 이용한 스트리밍 서버 최적화	<p>최근들어 고성능 네트워크 인터페이스 카드에 프로그래밍 가능한 가속화 엔진을 추가한 DPU(Data Processing Unit) 장치를 이용하여 인공지능, 보안, 통신, 스토리지 등의 기능을 향상시키고자 하는 연구가 활발히 진행되고 있다.</p> <p>본 연구는 DPU 장치를 이용하여 비디오 스트리밍 서비스의 성능과 품질을 개선하기 위한 방안을 도출하는 것을 목표로 한다.</p> <p>Required skills:</p> <ul style="list-style-type: none"> - 컴퓨터 구조 및 운영체제에 대한 이해 - TCP/IP 프로토콜 및 소켓 인터페이스에 대한 이해 - Linux 사용 경험

연번	교수명	연구실명	연구과제명	연구주제
			[Requirements] - Linux 사용 경험, Python 코드 작성 및 실행 경험 - Pytorch, TensorFlow 등 딥러닝 프레임워크 사용 경험 - 선택: 생물학/화학적인 내용에 거부감이 없는 학생	
6	김 선	생물정보 및 생명정보 연구실	환자 특이적 Biomarker 발굴을 위한 딥러닝 모델 연구	- 기존의 Biomarker 연구들은 집단 수준의 연구가 주로 수행됨 - 본 UROP 연구에서는 집단 수준을 넘어, 환자 개별적인 Biomarker 발굴을 목표로 함 - 기본적인 Feature selection 알고리즘부터 SOTA 기계학습/딥러닝 모델을 재현함 - Transformer, GNN 등 다양한 딥러닝 구조를 통해 모델의 성능을 분석하고 개선 방안을 모색함
7			대용량 화합물 DB에 활용 가능한 화합물 검색 모델 연구	- 화합물의 유사도를 계산하는 방법은 비슷한 구조, 특징 등을 보유한 화합물을 탐색할 수 있다는 점에서 신약개발, 신소재 개발 등 다양한 분야에서 중요한 기술임 - 화합물 DB의 크기 및 탐색 대상이 되는 화합물의 복잡도에 따라, 검색에 걸리는 시간은 수 시간 ~ 수일에 걸쳐서 다양함 - 본 UROP 연구에서는 기존의 화합물 탐색 알고리즘 및 모델들의 재현하고, 한계점 및 개선점을 연구함
8			화합물 구조와 설명문 매칭을 위한 딥러닝 모델 연구	- 자연어 처리 기술의 발달로 텍스트와 다른 형태의 데이터(예: 텍스트-이미지) 간 양방향 생성에 관한 연구가 활발히 수행 중임 - 본 UROP 연구에서는 화합물의 구조와 설명문 간 매칭을 통해 화합물-텍스트 간 양방향 생성 모델을 연구함 - 기존 SOTA 모델을 재현해보고 한계점 및 개선점을 탐색함 - 이미지 분야의 멀티 모달 매칭에 관한 최신 연구들을 조사하고 적용할 예정임

연번	교수명	연구실명	연구과제명	연구주제
9	박근수	컴퓨터 이론 및 응용 연구실	NP-hard 그래프 문제를 위한 실용적인 알고리즘	NP-hard 그래프 문제들을 대규모 그래프에서 풀 수 있는 세계 최고 성능의 알고리즘들을 개발하고, 이를 기반으로 NP-hard 그래프 문제들에 특화된 공개 SW 프레임워크를 개발한다. 주요 성능지표인 알고리즘 수행시간에 있어서 현존 최고 성능의 알고리즘 대비 개선을 100% 이상을 달성한다. 이를 통해 국제 산업계와 학계를 선도하는 기술을 확보하고, NP-hard 그래프 문제에 대한 연구 및 응용 SW 개발의 오픈 생태계를 구축한다.
10	박재식	시각 및 기하지능 연구실	효율적인 딥러닝 생성 모델 제작	대표적인 딥러닝 기반 영상 생성 모델은 probabilistic denoising process를 수백번에 걸쳐 반복적으로 거치면서 clean data를 생성한다. 반복적인 inference의 계산 부담을 낮추기 위해 최근에는 distillation기법등이 제안되고 있다. 본 연구에서는 기존의 생성 모델을 새롭게 디자인하여, 한번의 network inference로 probabilistic denoising process를 모사하는 새로운 기법에 대해 탐구한다. 본 연구를 위해서는 딥러닝 및 머신러닝에 대한 이해와 consistency model에 대한 기법을 필요로 한다.
11			광역시차 래디언스 필드 기법 연구	사진으로부터 공간 및 모습을 복원하고, 원본의 사진과 유사하게 영상을 렌더링하는 너프(NeRF) 기술은 높은 활용성과 유용성으로 인해 최근들어 크게 각광을 받고 있다. 본 연구주제에서는 현재 너프 기술들의 한계점으로 지적되는 광역 (wide baseline) 복원 기술에 대해 탐구하고 새로운 해결 기법을 제시하는 것을 수행한다. 본 연구를 위해서 딥러닝, 머신러닝, 너프 기술의 이해를 필요로 한다.

연번	교수명	연구실명	연구과제명	연구주제
12	이상구	Intelligent Data Systems Lab.	한국어 능력 강화를 위한 Llama fine-tuning	Requirements - Linux 사용 경험, Python 코드 작성 및 실행 경험 - Open-source LLM(Large Language Model)에 대한 관심 연구 주제 - 2023년 2월 LLaMA의 공개 이후, 이를 기반으로 한 LLM 연구들이 활발하게 진행되고 있음 - 그러나 LLaMA의 경우 90% 정도의 학습 데이터가 영어로 되어있어, 한국어 등 소수 언어에 직접 활용하기 어려움 - 데이터 수집 및 모델 Fine-tuning을 통해 한국어 성능을 높이고자 함 목표 - Llama 모델 한국어 학습을 진행하여 한국어에도 잘 동작하는 대화형 LLM 모델을 만드는 것이 목표 기타 - 참여 집중도에 따라 인건비 지급여부 결정
13			실용적 RAG(Retrieval-Augmented Generation) 활용 방법 연구	Requirements - Linux 사용 경험, Python 코드 작성 및 실행 경험 - LLM(Large Language Model)에 대한 관심 연구 주제 - 자연어 처리 분야에서 RAG는 뛰어난 혁신을 보여주고 있는 영역 - LLM의 능력을 확장하여 사용자의 질문이나 요청에 대해 정보가 풍부한 답변을 제공하도록 도움 - RAG는 데이터 검색의 정확도를 높이고 텍스트 생성의 품질을 향상시키는 것이 중요함 - 일반 문서에 대한 연구는 활발히 진행되고 있으나, 상품정보와 같은 정형데이터와의 연계에 대한 탐구는 찾아보기 힘들 목표 - 상품 정보를 검색하고 사용자의 질문을 답변하는 RAG 시스템을 개발함으로써 실용적 LLM 활용 방법을 연구하는 것을 목표로함 기타 - 참여 집중도에 따라 인건비 지급여부 결정

연번	교수명	연구실명	연구과제명	연구주제
14	이상구	Intelligent Data Systems Lab.	안정적 서비스를 위한 sLLM 최적화 연구	Requirements <ul style="list-style-type: none"> - Linux 사용 경험, Python 코드 작성 및 실행 경험 - Open-source LLM(Large Language Model)에 대한 관심 연구 주제 <ul style="list-style-type: none"> - 최근 LLM을 활용한 다양한 어플리케이션들이 주목받고 있음 - 주어진 입력에 대한 콘텐츠를 생성하는 작업은 핵심 기반이 되는 기능이나, sLLM의 경우 상대적으로 할루시네이션, 동일 단어 생성 문제, 생성된 글의 다양성 및 참신성 등의 문제가 있어 실용성이 떨어짐 - Prompt engineering / Fine-tuning 기술들을 활용하여 이를 해결하고자 함 목표 <ul style="list-style-type: none"> - 모델 fine-tuning 및 prompt engineering을 통해 상품 정보 등 주어진 입력에 대해 타이틀이나 설명문과 같은 콘텐츠를 제대로 생성할 수 있는 sLLM 모델을 만드는 것이 목표 기타 <ul style="list-style-type: none"> - 참여 집중도에 따라 인건비 지급여부 결정
15			특정 스타일을 반영할 수 있는 이미지 생성 모델 연구	Requirements <ul style="list-style-type: none"> - Linux 사용 경험, Python 코드 작성 및 실행 경험 - Diffusion model에 대한 관심 연구 주제 <ul style="list-style-type: none"> - 이미지 생성 AI의 대표적인 기술은 GAN(Generative Adversarial Network), diffusion model이 있음 - 이 중 Text-to-Image 생성이 가능한 diffusion model은 성능이 좋아 관심도가 높으나 유저가 생성된 이미지의 스타일을 유지하는 것은 난제임 - 본 연구에서는 특정 스타일을 유지하는 diffusion model을 재학습하는 방법에 대해 연구함 목표 <ul style="list-style-type: none"> - 실제 브랜드 상품의 이미지를 재학습하여 diffusion model에서 특정 스타일을 유지하는 방법을 연구하는 것을 목표로함 기타 <ul style="list-style-type: none"> - 참여 집중도에 따라 인건비 지급여부 결정

연번	교수명	연구실명	연구과제명	연구주제
16	이재진	천둥 연구실	대규모 언어 모델 추론 시스템 최적화	<p>대규모 언어 모델 (Large Language Model)이 다양한 분야에서 높은 성과를 보여주면서 수요가 폭발적으로 증가하고 있다. 이에 따라 GPU를 효율적으로 사용해 모델의 추론 시간을 줄이는 것이 중요해졌다.</p> <p>본 연구에서는 대표적인 대규모 언어 모델인 LLaMa의 추론 시스템을 다수의 GPU가 장착된 환경에서 최적화하는 것을 목표로 한다. 본 연구에 참여하는 학생은 기존 시스템의 병목 지점을 파악하고 이를 다양한 방식으로 해결해 볼 것이다. 모델의 품질을 떨어뜨리지 않는 선에서 가능한 다양한 최적화 방식을 공부하고 직접 적용해 보는 과정에서 딥 러닝 시스템 최적화 전반에 대한 경험을 쌓을 수 있을 것이다.</p> <p>Linux 사용에 익숙해야 하고, Python, C++ 코드를 읽고 개발할 수 있어야 하며, PyTorch 등의 딥 러닝 프레임워크 사용 경험이 있으면 좋다. 관련 경험이 없으면 배워가면서 연구를 진행하게 된다.</p>
17	이창건	실시간 유비쿼터스 시스템 연구실	디지털 일상활동 데이터 기반 청소년 마음건강 진단·케어 AI 모델링	<ul style="list-style-type: none"> • 목표: 디지털 일상활동 데이터 기반 청소년 마음건강 진단 AI 모델 개발 - 청소년들의 디지털 일상활동 (학습활동, SNS 교류, 인터넷/유튜브 브라우징) 속에서 언어/행동데이터를 수집하고, 개인별 자가 심리검사를 진행함. 수집된 데이터에서 마음건강을 진단할 수 있는 중요 feature들을 찾아내고, 그것을 기반으로 청소년 마음건강 진단 및 개인 맞춤형 케어 AI를 모델링 • 무엇을 하는지? - 디지털 일상활동 (학습활동, SNS교류, 인터넷/유튜브 브라우징) 데이터 분석 및 visualization - Multi-modal sensor data에서 주요 feature extraction을 통한 Digital phenotyping - 청소년 마음건강 진단 및 개인 맞춤형 케어 AI 모델링 • Requirements: - Python(Pytorch, Tensorflow), 웹(React.js)/모바일(Swift, Kotlin) 경험 - 인공지능 관련 강의 수강생
18			UAV시뮬레이터 개발 및 Fault Tolerance 시스템 연구	<ul style="list-style-type: none"> - 오픈소스 UAV(Unmanned Aerial Vehicle) 시뮬레이터 개발 및 최적화 - Fault Tolerance 시스템 연구 - Fault Tolerance 이론의 UAV 모델 응용 <p>위와 같은 연구를 진행합니다. 고성능 개인 PC 제공하며, C++ 및 Python 개발 경험이나 CPS, 시뮬레이션 관련 강의 수강생 우대합니다.</p>

연번	교수명	연구실명	연구과제명	연구주제
19	장병탁	바이오지능연구실	실세계 데이터 기반 인공지능 연구 및 서비스 로봇 응용	<p>연구주제</p> <ul style="list-style-type: none"> - 로봇이 실내외 환경에서 지각 가능한 다양한 센서 정보를 이용하여 학습하고 추론하는 기계학습 모델 연구 - 새로운 문제를 해결하는 자기주도 인공지능 기술 개발 - 스스로 불확실성을 자각하며 질문하면서 성장하는 에이전트 기술 개발 - 이동형 로봇의 서비스 수행 시나리오 설계 및 데모 프로그램 개발 <p>최소 참여 요건</p> <ul style="list-style-type: none"> - Python 프로그래밍 가능자 - TensorFlow, PyTorch 등 기계학습 솔루션 경험이 있거나, 없더라도 '성실'하게 공부할 수 있는 학생 <p>권장 요건</p> <p>ROS(Robot OS) 사용 경험, SLAM 개발 경험, 신경망 기반 응용프로그램 개발 경험, 멀티모달 데이터 처리 경험 및 강화학습 연구 경험</p> <p>※신청자의 과제 수행 내용은 담당 교수 및 멘토 연구원들과의 면담을 통해 구체화</p>
20	주한별	비주얼 컴퓨팅 연구실	Building a multi-view system for high-quality digital human reconstruction	<p>연구주제:</p> <ul style="list-style-type: none"> - 멀티 카메라 시스템을 구축하여, 영상으로부터 high-quality 3D digital humans (face, body, hand motion) 을 복원하는 연구를 진행한다. - UROP 학생은 dataset processing, open source repo building, system building, 3D reconstruction 등의 다양한 모듈 중 한 파트를 담당하게 될 예정이다. - 좋은 연구 결과는 top-tier 학회에 제출할 예정이며 contribution에 따라 authorship을 가질 수 있다. <p>Requirement:</p> <ul style="list-style-type: none"> - Strong coding skill (C++ & Python) in Linux environment (required) - Prerequisite courses: linear algebra, and at least one course in AI (e.g., Computer Vision, Machine Learning, Deep Learning)