

AI · 안전 신뢰성 전공

Convergence Major of Trustworthy AI

(데이터사이언스학+소프트웨어학+지능정보보호학+경영학+교양학부 융합)

교 육 목 표

1. 인간 중심 AI 활용에 대한 융합적 지식과 통찰 함양 [지]
2. 책임성과 윤리 의식을 갖춘 AI 리더 양성 [덕]
3. AI 안전신뢰성 기술 실무 역량과 현장 적용 능력 강화 [술]

학 과 소 개

AI안전·신뢰성 전공은 인간 중심의 안전하고 신뢰할 수 있는 AI 활용 체계를 설계·검증·운영할 수 있는 융합형 전문 인재를 양성하는 전공이다.

본 전공은 AI 기술 개발이나 활용을 넘어, AI 시스템의 오류·편향·보안·윤리·법·거버넌스 문제를 통합적으로 이해하고 해결할 수 있는 교육 모델을 지향한다. 특히 AI 안전성(Safety), 신뢰성(Reliability), 책임성(Accountability), 설명가능성(Explainability) 등의 AI 품질 문제를 교육과정의 주요 축으로 설정하고, AI 기술의 사회적 책임과 공공적 가치에 대한 이해를 바탕으로 현대 사회가 직면한 다양한 문제를 해결할 수 있는 역량을 함양한다.

또한 본 전공은 이러한 이론적·윤리적 이해가 실제 문제 해결 역량으로 확장될 수 있도록 프로젝트 기반 학습(Project-Based Learning), 산학협력형 실습, 인간-AI 협력 중심 교육 등을 운영한다. 이를 통해 학생들은 산업과 사회 현장에서 요구되는 AI 품질 관리 및 책임 있는 AI 활용 역량을 체계적으로 습득하고, AI 기반 문제 해결을 선도하는 실천형 여성 리더로 성장할 수 있다.

교 과 과 정

학 년	학 기	학수번호	이수 구분	교 과 목 명	학 점	시 간	과목 구분	비 고
1	1	TA00001	전필 복필	AI프로그래밍기초 Fundamentals of AI Programming	3	3	이론 실습	
	1	TA00002	전선	AI안전신뢰성개론 Introduction to AI Safety and Reliability	3	3	이론	
	2	DT00006	전선	데이터사이언스기초수학 Elementary Mathematics for Data Science	3	3	이론	· 데이터사이언스학과
	2	DT00007	전선	데이터전처리 Data Preprocessing	3	3	이론 실습	· 데이터사이언스학과
2	1	TA00003	전필 복필	AI윤리와법규 AI Ethics and Law	3	3	이론	
	1	TA00004	전필 복필	Python기반LLM기초 Introduction to Large Language Models	3	3	이론	
	1	DT00010	전선	데이터베이스 Database	3	3	이론 실습	· 데이터사이언스학과
	1	TA00005	전선	Vibe코딩기반RAG구축 RAG System Design and Implementation	3	3	이론 실습	

학 년	학 기	학수번호	이수 구분	교 과 목 명	학 점	시 간	과목 구분	비 고
2	2	CS13024	전선	인공지능과정보보호 Artificial Intelligence and Information Security	3	3	이론 실습	· 사이버보안전공
	2	TA00006	전선	휴먼-AI인터랙션디자인 Human-AI Interaction Design	3	3	이론 실습	
	2	TA00007	전선	AI거버넌스와위험관리 AI Governance and Risk Management	3	3	이론	
3	1	DT00022	전선	AI모델최적화 AI Model Optimization	3	3	이론	· 데이터사이언스학과
	1	TA00008	전선	AI와소셜이노베이션실습 AI for Social Innovation	3	3	실습	
	2	TA00009	전필 복필	AI보안설계및신뢰성평가 Secure and Reliable AI System Design	3	3	이론 실습	
	2	DT00031	전선	빅데이터플랫폼 Big Data Platform	3	3	이론 실습	· 데이터사이언스학과
	2	TA00010	전선	설명가능한AI(XAI) Explainable AI	3	3	이론 실습	
	2	TA00011	전선	글로벌AI인증및표준 Global Standards and Governance for Artificial Intelligence	3	3	이론	
4	1	DT00023	전선	데이터공학 Software Engineering	3	3	이론 실습	· 데이터사이언스학과
	1	TA00012	전선	MLOps및AI운영관리 MLOps and AI Operations	3	3	이론 실습	
	1	TA00013	전선	AI리스크평가및손실관리 AI Risk Assessment and Loss Management	3	3	이론 실습	
	2	TA00014	전필 복필	AI안정신뢰성캡스톤디자인 Capstone Design for Safe and Reliable AI Solutions	3	3	실습	
총 63학점 (전필 15학점, 전선 48학점) / (복필 15학점)								

1학년 교과내용

TA00001 전필 AI프로그래밍기초

이 교과목은 인공지능 개발에 필요한 기초 프로그래밍 지식을 다루며, 특히 Python 문법과 핵심 프로그래밍 구조, 그리고 주요 라이브러리의 실용적 활용에 중점을 둔다. 학생들은 코드를 작성하고 읽고 관리하는 방법을 익히는 동시에, 객체지향 프로그래밍(OOP)의 개념과 이를 통해 모듈화되고 재사용 가능한 소프트웨어를 설계하는 원리를 학습한다. 또한 알고리즘적 사고, 코드의 가독성, 계산 효율성의 중요성을 이해하고, 이를 바탕으로 기본적인 문제 해결 절차를 구조적으로 구현하는 능력을 기른다. 아울러 예외 처리와 관련 기법을 실습하여 코드의 견고성과 안정성을 높이는 방법도 함께 익힌다. 기초 프로그래밍 역량과 실습을 결합함으로써, 이 교과목은 이후 인공지능 모델링, 데이터 처리, 소프트웨어 개발에 필요한 심화 학습의 기반을 마련하도록 설계된다.

TA00002 전선 AI안전신뢰성개론

이 교과목은 인공지능 안전성과 신뢰성의 핵심 개념을 기초 수준에서 소개하며, 특히 두 개념의 차이와 상호 연관성, 그리고 인공지능 시스템의 설계와 운영에서 갖는 중요성을 중점적으로 다룬다.

DT00006 전선 데이터사이언스기초수학

데이터사이언스 전공을 학습하는데 필요한 수학적 기초를 습득한다. 구체적으로 확률과 통계, 최적대수, 최적화, 정보이론의 기초적인 내용을 학습한다. 이러한 수학적 기초를 통해 데이터 분석, 머신러닝, 인공지능 등 다양한 분야에서 데이터 분석, 예측 모델 제작, 데이터의 패턴과 통찰을 얻는데 필요한 수학적 기초를 탄탄히 할 수 있다.

DT00007 전선 데이터전처리

데이터 전처리는 데이터를 분석하기 전, 데이터를 정리하고 준비하는 중요한 과정이다. 이 과정은 데이터 정제, 어노테이션, 통합, 임베딩으로 구성되어 있으며, 각 단계에서 어떤 방법으로 데이터를 처리하는 지에 대해 학습한다. 특히, 데이터 어노테이션은 기업이 제시한 주제로 실습을 진행한다.

2학년 교과내용

TA00003 전필 AI윤리와법규

이 교과목은 인공지능의 개발과 활용을 규율하는 윤리 원칙과 법적 체계를 체계적으로 소개한다. 학생들은 공정성, 투명성, 책임성, 설명가능성과 같은 핵심 규범 개념을 학습하고, 개인정보 보호와 자동화된 의사결정 규제를 포함한 국내외 인공지능 관련 법·제도와 규제를 함께 검토한다. 또한 인공지능 서비스의 기획, 개발, 운영 전 과정에서 발생할 수 있는 윤리적·법적 쟁점을 식별하고, 컴플라이언스와 리스크 관리의 관점에서 적절한 판단을 내리는 역량을 기르는 데 중점을 둔다. 나아가 기술, 윤리, 법의 상호작용을 종합적으로 탐구함으로써 책임 있는 인공지능에 대한 통합적 이해를 형성하고, 인간 중심의 인공지능 시스템 설계와 거버넌스를 위한 기준을 제안할 수 있는 능력을 함양한다.

TA00004 전필 Python기반LLM기초

이 교과목은 Python 기반 환경에서 대규모 언어모델(LLM)의 기본 개념과 실제 활용 방법을 기초적으로 다룬다. 학생들은 토큰화, 임베딩, 추론 과정 등을 포함한 LLM의 기본 구조와 작동 흐름을 학습하고, 프롬프트 설계와 모델 상호작용에 대한 실습을 통해 활용 역량을 기른다. 또한 LLM이 생성한 응답에 내재할 수 있는 불확실성, 오류, 편향의 기술적 원인을 살펴봄으로써, 출력 결과를 그대로 수용하는 것이 아니라 비판적으로 해석하는 능력을 함양한다. 아울러 인간의 판단과 의사결정을 지원하는 도구로서 LLM을 안전하게 활용하기 위한 기본적인 통제 개념과 책임 있는 사용 원칙도 함께 탐구한다. 기술적 이해, 실습적 경험, 비판적 성찰을 결합함으로써, 이 교과목은 LLM을 정보에 근거해 책임 있게 활용할 수 있는 기초 역량을 기르는 데 목적이 있다.

DT00010 전선 데이터베이스

데이터베이스의 기본 이론과 실습을 배운다. 데이터를 조회, 삽입, 수정, 삭제하는 SQL을 배우고 실습을 통해 데이터베이스와 상호작용하는 방법을 익힌다. 데이터베이스 작업을 하나의 단위로 묶어 원자성, 일관성, 고립성, 지속성을 보장하는 방법인 트랜잭션 관리를 배운다. 데이터 중복을 줄이고 무결성을 보장하기 위해 데이터를 여러 테이블로 분리하는 기법인 정규화를 학습한다. 실제 데이터베이스 시스템에서 사용하는 실습 프로그램을 통해 데이터베이스 설계와 운영 기술을 실무에 적용할 수 있는 법을 학습한다.

TA00005 전선 Vibe코딩기반RAG구축

이 교과목은 자연어 기반 지시 워크플로우, 이른바 바이브 코딩(vibe coding)을 활용하여 검색증강생성(RAG) 시스템을 설계하고 구현하는 방법을 실습 중심으로 소개한다. 학생들은 데이터 수집, 임베딩, 검색, 응답 생성에 이르는 RAG의 전체 파이프라인을 이해하고 구축하는 과정을 학습하며, 특히 실제 적용을 위한 시스템 설계 역량을 기르는 데 중점을 둔다. 또한 검색 정확도, 문맥 적합성, 응답 신뢰성을 핵심 기준으로 삼아

RAG 시스템의 성능을 평가하고 개선하는 능력을 함양한다. 아울러 기술적 구현을 넘어, 인공지능을 보조적 도구로 효과적으로 활용하는 과정에서 인간의 의도와 판단이 수행하는 역할도 함께 탐구한다. 시스템적 사고, 성능 분석, 인간-AI 협력 관점을 통합함으로써, 이 교과목은 신뢰할 수 있는 지식기반 인공지능 서비스를 설계하고 운영할 수 있는 기초 역량을 기르는 데 목적이 있다.

CS13024 전선 인공지능과정보보호

AI 시대의 정보보호는 단순히 공격을 막는 것을 넘어, 인공지능이 범죄에 악용되는 시나리오(해킹 코딩 자동화 등)에 대응하고 역으로 AI를 활용해 사이버 공격을 선제적으로 차단하는 공방형 기술을 요구한다. 본 강좌에서는 실무 중심의 문제 해결 프로세스를 기반으로 공개 데이터셋 등 실제 보안 데이터셋을 활용해 문제를 정의하고 해결하는 전 과정을 직접 수행한다. AI를 활용하여 보안 문제의 특성에 맞는 데이터 구조화 및 모델 최적화와 같은 고도화된 문제 해결 능력을 함양한다. 이를 통해 인공지능의 원리를 정보보호 분야의 다양한 문제에 유연하게 적용할 수 있는 융합적 사고력을 배양하며, 최종적으로 현장에서 즉시 활용 가능한 지능형 보안 솔루션 개발 역량을 강화하고자 한다.

TA00006 전선 휴먼-AI인터랙션디자인

이 교과목은 인공지능 시스템에서의 불확실성, 자동화 수준, 설명가능성과 관련하여 인간-AI 상호작용의 특성을 탐구한다. 학생들은 위험을 줄이고 인간과 인공지능 간의 안전하고 효과적인 협력을 지원하는 상호작용 설계의 핵심 원리를 학습한다. 또한 사용자 과신, 자동화 편향, 책임의 분산 등 인간-AI 환경에서 자주 발생하는 문제를 분석하고, 이러한 문제를 완화할 수 있는 인터페이스와 피드백 구조를 설계하는 능력을 기른다. 특히 인공지능을 인간의 판단을 대체하는 존재가 아니라, 이를 지원하고 설명하는 보조적 도구로 위치시키는 데 중점을 둔다. 상호작용 설계, 위험 인식, 인간 중심 인공지능의 개념을 통합적으로 다룸으로써, 이 교과목은 인간의 의사결정과 통제에 부합하는 책임 있는 인공지능 시스템을 설계하기 위한 기초를 마련하는 데 목적이 있다.

TA00007 전선 AI거버넌스와위험관리

이 교과목은 정책 프레임워크, 책임 구조, 의사결정 절차, 감사, 통제 메커니즘 등을 포함한 인공지능 거버넌스의 개념과 핵심 구성요소를 소개한다. 학생들은 기관 차원에서 인공지능을 관리하기 위한 조직 모델을 설명하고 평가하는 방법을 학습한다. 또한 인공지능 시스템의 도입과 운영 전 과정에서 발생할 수 있는 기술적, 윤리적, 법적, 사회적 위험을 식별하고, 이에 대한 위험 평가 및 완화 전략을 설계하는 능력을 기른다. 특히 판단을 자동화 시스템에 전적으로 위임하기보다, 인간의 책임과 통제를 중심으로 인공지능 의사결정 권한을 구조화하는 데 중점을 둔다. 거버넌스, 위험 분석, 조직 리더십의 관점을 통합적으로 다룸으로써, 이 교과목은 책임 있는 인간-AI 공존을 위한 지속가능한 관리 체계를 제안할 수 있는 역량을 함양하는 데 목적이 있다.

3학년 교과내용

DT00022 전선 AI모델최적화

데이터 모델을 최적화하고, 최적의 솔루션을 찾기 위한 수학적 기법을 배우는 과정이다. 최적화는 주어진 문제에서 가장 적합한 해를 찾는 과정으로, 머신러닝 및 데이터 분석에서 모델을 개선하는 중요한 기법이다. 구체적으로, 테일러 전개, 컨벡스 함수, 그래디언트 디센트 등을 포함하여 학습하며, 이를 통해 최적화 이론의 기초 개념을 이해하고, 데이터 모델링 및 머신러닝 알고리즘에 적용할 수 있는 능력을 배양한다.

TA00008 전선 AI와소셜이노베이션실습

이 교과목은 실제 사회문제를 대상으로 문제 정의, 데이터 활용, 인공지능 적용, 결과 검증에 이르기까지 인공지능 프로젝트의 전 과정을 직접 수행하는 실습 중심의 경험을 제공한다. 학생들은 이해관계자의 관점, 윤리적·법적 제약, 기술적 한계를 통합적으로 고려하여 AI 기반 해결책을 설계하는 방법을 학습한다. 또한 인공지능을 단순한 기술 도구로 보는 데 그치지 않고, 복잡한 사회문제를 책임 있게 해결하기 위한 수단으로 접근하도록 강조한다. 아울러 공공적 가치, 책임성, 장기적 파급효과를 함께 고려함으로써 사회적으로 의미 있고 지속가능한 혁신을 기획하는 역량을 기른다. 프로젝트 기반 학습을 통해 이 교과목은 사회혁신을 위한 책임 있는 인공지능 솔루션을 설계하고 구현할 수 있는 실천적 역량을 함양하는 데 목적이 있다.

TA00009 전필 AI보안설계및신뢰성평가

이 교과목은 인공지능 시스템의 설계 단계에서 보안 요구사항과 신뢰성 기준을 어떻게 도출하고, 이를 초기부터 시스템 아키텍처에 반영할 수 있는지를 다룬다. 학생들은 성능, 안정성, 강건성, 잠재적 장애와 위협에 대한 보호를 함께 고려하면서, 보안 중심의 관점에서 인공지능 시스템 설계에 접근하는 방법을 학습한다. 또한 인공지능 시스템의 신뢰성을 종합적으로 평가하고, 그 결과를 구조적이고 전문적인 형식으로 문서화하여 보고하는 능력을 기른다. 특히 기술적 분석 결과를 인간 의사결정자와 조직이 실제로 활용할 수 있는 이해 가능한 기준과 지표로 연결하는 데 중점을 둔다. 설계, 평가, 문서화, 책임성의 관점을 통합적으로 다룸으로써, 이 교과목은 실제 조직 환경에서 인공지능 시스템의 책임 있는 운영을 지원할 수 있는 역량을 함양하는 데 목적이 있다.

DT00031 전선 빅데이터플랫폼

빅데이터 관리와 처리의 핵심 개념을 배우고, 거대 기업(Google, Facebook, Amazon 등)들이 어떻게 대규모 데이터를 다루는지에 대한 사례를 분석하는 과정이다. 빅데이터 시스템이 기업 운영과 의사결정에 얼마나 중요한 역할을 하는지 이해하고, 이를 구현하는 플랫폼과 기술들을 학습한다. 이를 통해 실제 빅데이터 시스템을 구축하고 운영하는 데 필요한 지식과 기술을 습득한다.

TA00010 전선 설명가능한AI(XAI)

이 교과목은 설명가능한 인공지능의 원리를 소개하며, 설계 단계에서부터 해석 가능성을 확보하는 모델 기반 접근과 LIME, SHAP와 같은 사후적 설명 기법을 함께 다룬다. 학생들은 다양한 설명 기법이 어떤 원리로 작동하는지, 어떤 조건에서 적절하게 적용될 수 있는지, 그리고 실제 활용에서 어떠한 한계를 지니는지를 학습한다. 또한 인공지능 예측의 근거를 분석하고, 영향력이 큰 변수를 식별하며, 모델 출력에 수반되는 불확실성을 평가하는 능력을 기른다. 특히 설명 결과의 타당성을 비판적으로 검토하고, 설명에 대한 오해나 과도한 확신이 발생할 가능성을 함께 성찰하는 데 중점을 둔다. 아울러 비전문가, 의사결정자, 기타 이해관계자도 이해할 수 있는 형태로 기술적 설명을 전달하는 방법을 익힘으로써, 책임 있고 정보에 근거한 인공지능 판단을 지원하는 역량을 함양한다. 기술적 이해, 비판적 분석, 의사소통 실습을 통합적으로 다룸으로써, 이 교과목은 설명가능하고 책임 있는 인공지능을 위한 기초 역량을 기르는 데 목적이 있다.

TA00011 전선 글로벌AI인증및표준

이 교과목은 인공지능 관련 주요 국제표준과 인증 프레임워크의 목적, 구조, 실무적 의미를 소개한다. 학생들은 안전성, 신뢰성, 윤리성, 보안성과 관련된 요구사항을 체계적으로 해석하는 방법을 학습하고, 이러한 표준이 글로벌 규제 및 제도 환경 속에서 어떠한 기능을 수행하는지 이해하게 된다. 또한 국제 규범의 관점에서 인공지능 시스템을 평가하고, 인증, 적합성 평가, 감사 요구사항이 인공지능 시스템의 기술적·운영적 요소와 어떻게 연결되는지를 설명하는 능력을 기른다. 특히 인공지능이 국제적 기대 수준에 부합하도록 개발·배포·소통될 수 있도록, 기술 설계와 조직 프로세스, 거버넌스 구조를 유기적으로 연계하는 데 중점을 둔다. 기술, 제도 이해, 거버넌스 관점을 통합적으로 다룸으로써, 이 교과목은 국제 환경에서 책임 있는 인공지능 관리를 수행할 수 있는 실천적 역량을 함양하는 데 목적이 있다.

DT00023 전선 데이터공학

데이터공학은 데이터 분석에 필요한 데이터를 수집, 설계, 관리하는 방법, 데이터 분석에 필요한 다양한 도구의 개발 및 평가방법, 그리고 이를 기반으로 데이터 파이프라인을 구축하는 방법에 대해서 학습한다. 또한, 데이터 분석 프로젝트의 수행, 운영, 관리에 필요한 조직 프로세스, 데이터 거버넌스, 데이터 프로덕트 개발 방법론 및 개발 플랫폼 활용 지식을 현장 업무 세례 중심으로 학습한다.

TA00012 전선 MLOps및AI운영관리

데이터공학은 데이터 분석에 필요한 데이터를 수집, 설계, 관리하는 방법, 데이터 분석에 필요한 다양한 도구의 개발 및 평가방법, 그리고 이를 기반으로 데이터 파이프라인을 구축하는 방법에 대해서 학습한다. 또한, 데이터 분석 프로젝트의 수행, 운영, 관리에 필요한 조직 프로세스, 데이터 거버넌스, 데이터 프로덕트 개발 방법론 및 개발 플랫폼 활용 지식을 현장 업무 세례 중심으로 학습한다.

TA00013 전선 시리스크평가및손실관리

이 교과목은 인공지능 시스템의 전 생애주기 전반에서 발생할 수 있는 기술적, 윤리적, 법적, 운영적 위험을 다루고, 이러한 위험을 발생 가능성과 영향도를 기준으로 체계적으로 식별하고 평가하는 능력을 기르는 데 중점을 둔다. 학생들은 인공지능 사고와 실패로 인해 발생할 수 있는 재무적 손실, 평판 훼손, 법적 책임, 운영 중단 등 다양한 손실 유형을 분석하고, 이에 대한 예방, 완화, 대응, 복구 전략을 설계하는 방법을 학습한다. 또한 인공지능 위험 평가 결과를 의사결정자와 조직이 실제로 이해하고 활용할 수 있는 지표, 보고 체계, 의사소통 구조로 전환하는 역량도 함께 기른다. 위험 분석, 손실 관리, 조직 커뮤니케이션의 관점을 통합적으로 다룸으로써, 이 교과목은 책임 있는 인공지능 운영과 장기적 지속가능성을 지원할 수 있는 실천적 역량을 함양하는 데 목적이 있다.

TA00014 전필 AI안전성리눅스캡스톤디자인

이 교과목은 조직의 요구와 비즈니스 목적에 부합하는 안전한 인공지능 솔루션을 설계하고 구현하기 위해 필요한 전략적 과정을 종합적으로 이해하도록 한다. 학생들은 조직의 요구사항, 위험 인식, 실제 적용 가능한 인공지능 도입 전략을 논리적으로 연결하는 설계 역량을 기르게 된다. 또한 학생들이 스스로 발굴하고 탐색한 사례를 바탕으로, 실제 문제 맥락이 반영된 구체적 프로젝트 성과물을 도출하도록 지원한다. 아울러 최종 프로젝트가 현장 수요와 실제 적용 환경에 더욱 부합할 수 있도록 멘토링을 통해 지속적으로 보완·고도화하는 과정을 포함한다. 더불어 전문가 초청 특강을 통해 실무적 통찰, 산업 현장의 맥락, 프로젝트 완성도를 높이기 위한 전문적 조언을 제공한다. 전략적 기획, 프로젝트 기반 학습, 멘토링, 실무자 연계를 통합적으로 운영함으로써, 이 교과목은 실제 조직 환경에서 적용 가능한 안전한 인공지능 솔루션을 설계할 수 있는 실천적 역량을 함양하는 데 목적이 있다.