AI 안 전 · 신 뢰 성 학 과
Department of AI Safety and Trustworthiness

교 육 목 표

① AI안전성·신뢰성 전공은 기술과 전략, 윤리와 안전을 통합적으로 이해하고, AI 기술의 윤리적·법제도적·안전성 확보 역량을 강화한다. [지]
② 산업, 공공, 교육 등 다양한 영역에서 인공지능 전략을 수립하고 조직의 AI 전환을 주도할 수 있는 리더십을 함양한다. [덕]
③ 데이터 거버넌스, AI 표준화, 산학협력 프로젝트 관리 역량을 종합적으로 배양함으로써, 실제 산업 현장에서 발생하는 복합적 문제를 해결할 수 있는 실전적 전문 인재와 인공지능책임자(CAIO)를 양성하는 것을 목표로 한다. [술]

AI안전·신뢰성 전공(Major of AI Safety and Trustworthiness)

〈기초 공통 과목 및 종합 시험 과목〉

기초 공통 과목			종합 시험 과목		
교 과 목	학점	비 고	교 과 목	학점	비 고
인공지능개론	3		인공지능개론	3	택 3
AI 윤리	3		AI 윤리	3	
AI 거버넌스와 법제	3		AI 거버넌스와 법제	3	
AI 안전성 및 신뢰성 기술	3		AI 안전성 및 신뢰성 기술	3	

교 과 과 정

〈석사과정〉

0000000 인공지능개론 (AI Fundamentals) 3학점
인공지능의 기본 개념, 역사, 발전 과정을 이해하고 주요 응용 분야를 폭넓게 탐구한다. 기계학습, 딥러닝, 자연어처리 등 AI의 핵심 기술과 알고리즘을 학습하여 지능형 시스템의 작동 원리를 이해한다. Python 기반의 AI 라이브러리(TensorFlow, PyTorch, Scikit-learn 등)를 활용한 실습을 통해 기본적인 예측·분류 모델을 구현하고 성능을 평가한다. 더불어 AI의 신뢰성, 안전성, 설명가능성 등 사회적·윤리적 쟁점을 함께 다루며, 기술적 이해와 사회적 책임의 균형을 강조한다.

0000000 AX 전략 및 전환관리 (AX Strategy & Transformation Management) 3학점
조직과 사회 전반의 AI 전환(AI Transformation, AX)을 전략적 관점에서 이해하고, 기술·조직·정책이 융합된 전환관리 역량을 함양하는 것을 목표로 한다. 인공지능 기술이 산업, 공공, 교육 등 다양한 영역에서 혁신을 촉진하는 과정을 분석하고, 이를 효과적으로 추진하기 위한 전략 수립 방법론을 익힌다. 또한 디지털 전환과 AI 전환의 차이점을 이해하며, AI 생태계 조성, 인적 역량 개발, 데이터 거버넌스 구축 등 AX의 핵심 요소를 체계적으로 다룬다. 실제 사례를 통해 기업과 공공조직의

AI 도입 성공·실패 요인을 비교 분석하고, 리더십, 변화관리, 윤리적 의사결정 등 전환 과정에서 요구되는 거버넌스 역량을 실습 중심으로 학습한다.

0000000 AI 거버넌스와 법제 (AI Governance and Law) 3학점

인공지능 기술의 발전이 사회·경제·정치 전반에 미치는 영향을 다각도로 탐구하고, 책임 있고 신뢰할 수 있는 AI 생태계를 구축하기 위한 거버넌스와 법제의 원리를 심층적으로 다룬다. AI 기술의 투명성, 공정성, 안전성, 프라이버시 보호 등 핵심 가치와 이를 실현하기 위한 정책적·제도적 접근 방식을 학습한다. 또한 AI 알고리즘 편향, 자동화 의사결정, 데이터 주권, 지식재산권, 책임소재 등 실제 사회에서 발생하는 윤리·법적 쟁점을 사례 중심으로 분석한다. 더불어 OECD, EU, UNESCO 등 국제기구의 AI 윤리 가이드라인과 주요 국가의 법·제도 프레임워크를 비교하며, 글로벌 표준화 동향과 국내 대응 전략을 함께 고찰한다.

0000000 AI 안전성 및 신뢰성 기술 (AI Safety and Trustworthiness) 3학점

인공지능 기술이 사회 전반에 확산됨에 따라 요구되는 AI의 안전성(Safety)과 신뢰성(Trustworthiness) 확보 방안을 기술적·정책적 관점에서 심층적으로 다룬다. AI 시스템의 위험요소를 식별하고, 설계·개발·운영 단계에서 발생할 수 있는 오류, 편향, 보안 취약성을 최소화하기 위한 접근법을 학습한다. 또한 신뢰 가능한 AI를 위한 핵심 요소인 설명가능성, 투명성, 공정성, 책임성의 개념을 이해하고, 이를 기술적으로 구현하는 다양한 프레임워크와 평가 지표를 분석한다. 실제 사례를 통해 의료, 금융, 자율주행 등 고위험 영역에서의 AI 안전성 확보 전략을 검토하며, 국제 표준(ISO/IEC, NIST)과 법제 동향을 함께 비교한다.

0000000 AI 윤리 (AI Ethics) 3학점

AI 기술이 인간 사회와 제도, 문화에 미치는 영향을 폭넓게 탐구한다. 학습자는 알고리즘 편향, 자동화로 인한 일자리 변화, 개인정보 침해와 같은 사회적 쟁점을 실제 사례를 통해 분석한다. 또한 자율주행, 의료 AI, 범죄 예측 시스템 등에서 발생하는 윤리적 딜레마를 시뮬레이션하며, 다양한 이해관계자 관점에서 문제해결 방안을 모색한다. 최종적으로 학생들은 기업이나 정부의 AI 윤리 가이드라인을 비교·비판하고, 새로운 윤리 기준을 제안하는 정책 보고서를 작성한다.

0000000 AI 보안 설계 및 평가 (Design and Evaluation of AI Security) 3학점

인공지능 시스템의 보안취약점을 이해하고, 안전하고 신뢰할 수 있는 AI 모델을 설계·평가하는 능력을 기르는 것을 목표로 한다. AI 시스템의 구조적 취약성, 데이터 및 모델 공격 유형(예: 적대적 공격, 데이터 포이즈닝, 모델 도용 등)을 분석하고, 이를 예방·완화하기 위한 보안 설계 원리를 학습한다.

0000000 AI 보안 및 리스크 관리 (AI Security and Risk Management) 3학점

AI 시스템이 초래할 수 있는 잠재적 보안 위험 요인을 식별하고, 이를 체계적으로 관리하는 방법을 다룬다. 학습자는 AI의 불안정성, 적대적 공격, 데이터 피드백 루프 등 다양한 보안 리스크를 분석하며, 보안 설계와 모니터링 체계를 학습한다. AI 사고 사례를 검토하고, 위험 완화 계획을 직접 수립하여 실무적 보안관리 능력을 기른다.

0000000 설명가능한 AI(XAI)와 해석 가능성 (Explainable and Interpretable AI(XAI)) 3학점

AI의 의사결정 과정을 사람에게 이해 가능한 방식으로 설명하는 기술적 접근을 다룬다. LIME, SHAP, Grad-CAM 등 주요 설명가능성 기법을 실습하며, 블랙박스 모델의 불투명성이 사회적 신뢰에 미치는 영향을 비판적으로 논의한다. 학생들은 특정 모델을 선정해 해석가능성을 개선하는 실험을 수행하고, 그 결과를 연구 보고서로 제출한다.

0000000 **AI 품질보증 및 성능평가 (AI Quality Assurance and Performance Evaluation) 3학점**
AI 시스템의 품질을 과학적으로 검증하는 방법을 학습한다. 정확성, 일관성, 강건성, 지속가능성과 같은 품질 속성을 중심으로, 테스트·검증 절차를 실습한다. ISO/IEC 25010 및 ISO/IEC 42001 표준을 참조해 품질지표를 설계하고, 실제 모델에 적용하여 품질평가 보고서를 작성한다. 이 과정을 통해 학생들은 AI 시스템의 신뢰성과 성능을 객관적으로 평가하는 능력을 갖추게 된다.

0000000 **데이터 거버넌스와 책임 데이터 관리 (Data Governance and Responsible Data Management) 3학점**
AI의 윤리적 책임이 데이터 수준에서 어떻게 구현되어야 하는지를 다룬다. 데이터의 수집, 저장, 활용, 폐기에 이르는 전 과정에서의 책임 있는 관리 원칙을 배우며, 데이터 편향, 프라이버시 침해, 메타데이터 관리의 중요성을 분석한다. 실제 사례를 통해 데이터 거버넌스 정책을 설계하고, AI 학습용 데이터의 품질 확보 방안을 제시한다.

0000000 **휴먼-AI 인터랙션 디자인 (Human-AI Interaction Design) 3학점**
인간의 인지적 특성과 감정 반응을 고려해 AI 인터페이스를 설계하고, 대화형 에이전트·챗봇·추천 시스템 등 다양한 응용 사례를 실습한다. 특히 설명가능성과 사용성의 균형, 투명한 피드백 구조, 인간의 의사결정 지원이라는 관점에서 AI 인터페이스의 윤리적 설계를 연구한다.

0000000 **AI 사회적 실습(캡스톤 프로젝트) (AI Ethics and Trust Capstone Project) 4학점**
학생들은 실제 기업, 공공기관, 혹은 비영리단체와 협력하여 AI 윤리 진단 프로젝트를 수행한다. 윤리·신뢰·안전성의 관점에서 AI 시스템을 분석하고 개선방안을 제시하며, 프로젝트 결과를 연구보고서와 프레젠테이션 형태로 제출한다. 이를 통해 학생들은 학문적 이해를 실제 사회문제 해결에 적용하는 실무 능력을 함양한다.

0000000 **석사학위논문연구 (Research for the Master's Degree) 4학점**
석사과정생 논문 작성 지도이며, 매 학기 개설된다.

0000000 **AI 필드 리포트 세미나 (AI Field Report Seminar) 4학점**
재직자나 현장 실무 경험이 있는 학생이 근무 기관, 산학협력 기업, 혹은 공공기관 등에서 AI 관련 프로젝트를 수행하고, 그 결과를 학문적으로 분석·보고서 형태로 제출하는 현장연계형 교과목이며, 매 학기 개설된다. 학생은 실제 산업 현장에서의 AI 활용, 신뢰성 확보, 윤리적 의사결정, 데이터 관리 등의 문제를 탐구하고, 이를 학문적 연구와 연계한다.

0000000 **AI 윤리·책임성과 법적 책임 연구 세미나**
(AI Ethics, Accountability, and Legal Responsibility Seminar) 3학점
인공지능 사고의 법적 책임 문제를 심층적으로 다룬다. 자율주행차, 의료진단 AI, 챗봇 등에서 발생한 실제 사례를 법적 관점에서 분석하고, 책임 귀속의 원칙과 공동책임 구조를 논의한다. 학생들은 사례 분석을 토대로 학술 논문 초안을 작성한다.

0000000 **스마트 제조 AI 혁신과 디지털 트윈 실습**
(AI-Driven Smart Maunfacturing & Digital Twin Practice) 3학점
제조 산업의 디지털 전환과 AI 기반 스마트 제조 기술을 통합적으로 다룬다. 센서·공정·설비·품질 데이터와 IoT/IIoT 기반 실시간 데이터 스트림을 수집·분석하고, 머신러닝·딥러닝·시뮬레이션·최적화 알고리즘을 활용하여 공정 제어, 품질 예측, 생산 스케줄링 자동화 등 실제 제조 문제를 해결한다. 디지털 트윈·스마트 팩토리 플랫폼·로봇 자동화·컴퓨팅 기술을 학습하고, 실제 제조 프로세스를 가상환경에서 설계–시뮬레이션–검증–실증하는 프로젝트를 수행한다. 또한 인간–AI 협업 기반의 생산 시스템, 지속 가능한 제조, 산업 안전, 데이터 윤리 및 산업 AI 확산의 사회·경제적 영향을 논의함으로써 기술·산업·윤리가 결합된 지능형 제조 혁신 역량을 함양한다.

0000000 **금융데이터 분석과 투자 (Financial data analytics & investment) 3학점**

금융 투자 분야에서의 인공지능 활용에 요구되는 경영 이론과 분석 방법론을 배우고, 데이터 기반 투자 전략인 퀀트를 수행하기 위한 알고리즘 활용 유형을 탐구한다. 또한, 실제 금융 투자 분야에서 활용이 확대되고 있는 Boosting, DQN 등의 예측 및 의사결정 알고리즘 활용 기법과 기업의 투자 의사결정 프로세스를 학습한다. 본 교과목을 통해 투자 분야의 AX전략에 대한 이해를 높이고 금융 기업이 요구하는 AI 리터러시 역량을 갖출 수 있다.

0000000 **개인 맞춤형 디지털 헬스케어와 AI 전략 (Personalized Digital Healthcare & AI Strategy) 3학점**

미래 헬스케어의 진화 방향과 AI 기반 건강·의료·웰니스 기술의 융합적 적용을 탐구한다. 의료 데이터, 디지털 바이오마커, 퍼스널 헬스 데이터, 디지털 치료제, 헬스케어 IoT, 감정·행동 인식 AI 등 기술–데이터–인간 경험이 결합된 의료·웰니스 혁신을 학습한다. 이를 통해 학생들은 개인 맞춤형 케어, 예방 중심 의료, 지속가능한 보건 생태계 설계 등 미래 헬스케어 시스템의 기획–설계–평가–책임 있는 구현 능력을 기른다.

0000000 **바이오AI 혁신과 데이터 거버넌스 (BioAI Innovation and Data Governance) 3학점**

바이오 분야에서 인공지능(AI)이 활용되는 최신 기술 동향과 응용 사례를 탐구하고, 유전체 분석, 신약 개발, 질병 예측, 맞춤의료 등 다양한 생명과학 분야에서의 혁신적 변화를 살펴본다. 또한 바이오 데이터의 수집·분석·활용 과정에서 발생하는 개인정보보호 및 데이터 윤리 이슈를 다루며, 의료·유전체 데이터 등 민감정보의 AI 활용 시 요구되는 법제도적 규제, 국제 표준, 정책 동향을 비교·분석한다. 본 교과목을 통해 바이오 분야 AI에 대한 기술적 이해와 함께 정책적 시각에서 책임 있는 데이터 활용 방안을 모색한다.

0000000 **스마트팜 지능화 기술과 실증 연구 (Smart Farm Intelligence & Field Practice) 3학점**

스마트팜의 지능화 기술과 실제 농업 환경에서의 실증을 통합적으로 다룬다. 센서 기반 환경 데이터, 생육·수분·병해충 데이터, 영상 데이터 등 다양한 농업 데이터를 수집·해석하고, 머신러닝·딥러닝·환경제어 알고리즘을 활용하여 작물 생육 최적화·병해충 탐지·생산성 향상 등의 문제를 탐구한다. 학생들은 데이터 기반 농업 시스템의 구조를 이해하고, AI 모델 설계–평가–실증 과정을 수행한다. 또한 인간–AI 협력 관점에서 농업의 미래, 지속가능한 식량 시스템, 생태·윤리적 관점을 함께 논의함으로써 기술과 사회적 가치가 결합된 지능형 농업 혁신 역량을 함양한다.

0000000 **AI 기반 글로벌 K-컬처 콘텐츠 디자인 (AI-Driven Global K-Culture Content Design) 3학점**

K-컬처(K-Culture)의 창의성과 정체성을 기반으로, AI 기술을 활용하여 글로벌 문화 콘텐츠를 설계·제작·전략화하는 방법을 탐구한다. 음악·영상·웹툰·캐릭터·패션·관광·공연·로컬 문화 등 다양한 K-컬처 영역을 다루며, 생성형 AI, 멀티모달 AI, 디지털 휴먼, 가상 세계, AI 에이전트 기반의 지능형 콘텐츠 제작을 학습한다. 학생들은 콘텐츠의 기획, 스토리텔링, 문화데이터 분석, 마케팅 및 유통 전략 설계, 글로벌 문화 수용성 분석 등 기획–제작–확산의 전 과정을 AI와 함께 실습한다. 뿐만 아니라 인간 창작자의 감성과 가치, 문화 다양성, 윤리,창작자 권리 및 정체성 보존 등 AI 시대의 문화적 책임과 인간–AI 협력의 철학적 문제를 함께 성찰한다.