강의계획서

<table>
<tr><td rowspan="2">과목명</td><td>국문</td><td colspan="5">공정한 인공지능</td></tr>
<tr><td>영문</td><td colspan="5">Fair Artificial Intelligence</td></tr>
<tr><td>운영대학</td><td colspan="2">충남대학교</td><td>교과구분(교과목코드)</td><td>일반(313011)</td><td rowspan="4">담당교수</td><td>성명</td><td>김효은</td></tr>
<tr><td>운영학과</td><td colspan="2">일반선택</td><td></td><td></td><td>소속</td><td>한밭대학교</td></tr>
<tr><td>학점시수</td><td colspan="2">3/3/0</td><td>개설년도 / 학기</td><td>2023년 2학기</td><td>연락처</td><td></td></tr>
<tr><td></td><td colspan="2"></td><td></td><td></td><td>이메일</td><td>hyoekim26@hanbat.ac.kr</td></tr>
<tr><td>교과목표 및 개요</td><td colspan="7">이 교과는 인공지능 윤리의 주제들 중 가장 핵심적인 기반이 되는 주제인 인공지능 편향과 공정성이라는 주제를 다룬다. 편향이나 공정성이라는 주제는 그동안 인문사회 영역에서 주로 다루어져 왔던 주제이지만 최근에는 인공지능 기술의 한 영역이 되었다. 인공지능 구성 단계들에서 윤리적 요소가 이미 내장돼 있는 만큼 인공지능 편향을 잘 인식하고 필요에 따라 잘 조정하는 것이 필요하
다. 이는 사회적 영향, 이해관계자들 간의 조정 및 인공지능 시스템의 강건성을 확보하는 문제이기도 하다. 이 교과에서는 인지능 윤리의 주제들 뿐만 아니라 인공지능의 구성 절차에서 개입되는 편향이 어떤 양상을 가지는지, 편향을 판단하는 여러 유형의 공정성 기준들이 어떤 것들이 있는지, 편향 완화를 하는 기술적, 비기술적 방법들을 학습한다. 이러한 내용들을 통해서 인공지능이 내린
의사결정이나 데이터의 편향성 여부를 역 추론할 수 있는 역량을 가지는 것이 목표이다. 인공지능과 관련한 공정성이란 단순한 나누기를 넘어선 복잡한 사회적 과학적 맥락과 사회적 구성원들의 논의가 중요하다는 점을 자연스럽게 알게 될 것이다.</td></tr>
<tr><td>주 핵심역량과 교과목간 연계성</td><td colspan="7">이 교과목의 주 핵심역량은 융합적 해결 역량이다. 인공지능이 구성되는 과정에서 편향, 투명성 등의 윤리적 요소가 개입되는 메커니즘과 그 완화 방법을 이론적 뿐만 아니라 기술적으로도 이해하는 능력이 필요하다. 이러한 융합적 해결역량을 추상적 개념의 이해 뿐만 아니라 인공지능의 구성 단계들을 기초적으로 이해하는 것이 필요하며, 이를 통해 인공지능이 적용되는 모든 영역, 모든 위치에서 업무경쟁력을 향상시킬 수 있을 뿐만 아니라 한 시민으로서 권리와 의무를 다할 수 있게 된다.</td></tr>
<tr><td rowspan="2">핵심</td><td colspan="2">모듈화</td><td colspan="2">통합</td><td colspan="2">확장</td></tr>
<tr><td colspan="6"></td></tr>
</table>

역량(%)	ICT 기술 활용	시스템 사고	프로젝트 실행	융합적 해결	창의적 혁신	테크니컬 커뮤니케이션	진로 학습	지역사회 공헌	심미적감성
	25	0	0	50	0	25	0	0	0

역량 기반 학습 목표	핵심역량	학습목표
	ICT 기술활용	인공지능의 편향 인식 및 완화 방법을 기술적 차원, 이론적 차원에서 배운다.
	융합적 해결	인공지능이 구성되는 과정에서 편향, 투명성 등의 윤리적 요소가 개입되는 메커니즘과 그 완화 방법을 이론적 뿐만 아니라 기술적으로 이해하며, 사회의 이해관계자들 간의 거버넌스가 개입되는 것을 안다.
	테크니컬커뮤니케이션	인공지능의 윤리적 요소들을 기술적 차원에서 소통하여 문제해결을 할 줄 안다.

수업방법(%)	강의	토의/토론	실험/실습	현장학습	발표	기타
	100	0	0	0	0	0

교수법(선택)	문제중심학습	프로젝트기반학습	플립러닝
	O		

성적평가(%)	출석	중간고사	기말고사	과제	토론	기타
	60	20	20	0	0	0

기타 안내 사항	교재: [인공지능과 윤리], 김효은 저, 커뮤케이션북스 인공지능총서. - 수업 내용에 더 많은 내용이 있으므로 교재는 참고로써 활용함. 참고자료: 1. "인공지능 편향식별의 공정성 기준과 완화(Fairness Criteria and Mitigation of AI Bias)", 한국심리학회지: 일반, 40: 4, 459-485. (이 자료를 포함한 여타의 참고자료들 개강 후 소개) 2. 대량살상수학무기 ? 캐시 오닐 저, 김정혜 옮김, 흐름 3. 모두 거짓말을 한다 ? 세스 스티븐스 다비도위츠 저, 더퀘스 강좌 홍보 영상 https://youtu.be/zKUv_nIhNr4

주차	수업내용	교재범위 및 과제물	비고
1	인공지능에 내재한 윤리 -왜 인공지능윤리인가 -빅데이터 활용의 비윤리적 알고리즘 -인공지능윤리의 이슈들		
2	AI의 인지모형, 의사결정, 도덕적 딜레마 -인공지능의 인지모형 -인공지능의 의사결정		
3	인공지능편향과 설명가능성 -데이터 편향 -알고리즘 편향 -편향을 어떻게 감소시킬까		
4	AI윤리 질의응답 -AI윤리 가능한 질문과 설명1 -AI윤리 가능한 질문과 설명2 -AI윤리 가능한 질문과 설명3		
5	인공지능 윤리정책과 방향 -한국의 AI윤리정책과 기업들의 AI윤리 정책 -세계 대학들의 인공지능 윤리교육 -자율주행차와 자율살상무기 관련 윤리		

6	AI윤리 심화설명 심화설명 1 심화설명 2 심화설명 3		
7	중간고사		
8	도덕적 로봇 만들기 -인공도덕성의 의미 -하향식 인공도덕성 구현 -상향식 인공도덕성 구현		
9	인간 편향과 인공지능 -기업들의 AI윤리 정책 -인간 편향과 AI편향 -인공지능 구성단계에서의 편향		
10	공정성의 다양한 기준과 AI편향의 식별 -편향 판단기준으로서 공정성 -편향식별 기술들		
11	인공지능 편향 완화방법 -기계학습 전처리 단계에서의 편향완화 -처리과정 단계에서의 편향완화 -결과 산출 후 단계에서의 편향완화		
12	AGI(범용인공지능)의 가능성과 편향 -AGI의 의미 -AGI 시도의 사례 -인간수준의 지능이란 가능한가		
13	인공지능과 자유의지 -인간 수준의 AI와 자연/인공의 구분 -자율성의 속성과 인공지능 -인공지능은 어떤 기능/속성까지 가질 수 있는가		
14	AGI, GPT-3에 대한 심화설명 -GPT 인과추론에 대한 심화설명 -AGI, GPT-3윤리 에 대한 심화설명		
15	기말고사		