

불임 1**강의계획서**

○ 강의계획서 입력정보

교과목번호	M3225.001100	강좌번호	001	교과목명 (부제명)	거대언어모델과 대화형 인공지능	학점	3-3-0	전자출결사용여부	Y	
대표교수	성명	조요한		직급	조교수					
	E-mail	yohan.jo@snu.ac.kr		공개여부	O					
*강좌 해시태그	국문/영문	언어모델, 자연어처리, 인공지능, 에이전트, 대화시스템								
시간표 편성내역	국문/영문	43-2동 B102호, 화목 15:30-16:45								
*1. 수업목표	1. 거대언어모델 및 대화형 인공지능과 관련된 중요한 개념들을 정의하고 이해한다. 2. 거대언어모델 및 대화형 인공지능 분야의 다양한 연구 주제들을 파악한다. 3. 각 연구 주제에 대해 주요 논문들을 분석하고, 핵심 아이디어와 방법론을 이해하고 장단점을 평가한다. 4. 간단한 대화형 시스템을 구현해본다. 5. 대화형 인공지능과 관련된 문제를 선택해, 이를 해결하기 위한 전략을 개발하고 적용해본다.									
*2. 교재 및 참고문헌	국문/영문									
*3. 평가방법	성적부여 방식	절대평가		등급제여부	A~F					
	성적영역구분	출석	과제	중간	기말	수시평가	태도	기타	합계	
	성적비율		90%	5%	5%				100%	
	비고		학기 중 3회	퀴즈	퀴즈					
	출석 규정	수업일수의 1/3을 초과하여 결석하면 성적은 "F" 또는 "U"가 됨(학칙 85조). 결석에 대하여 교원에게 별도로 출석인정을 받은 경우 예외로 할 수 있음. (학업성적처리규정. 조기취업자 출석 및 성적처리 지침)								
기타 사항	부정행위 규정, 대체시험 여부 및 방법, 과제 또는 시험 피드백 제공 여부 등 평가방법 관련 기타 사항									
*4. 생성형 AI 도구 활용 방침		본 강의에서는 생성형 AI의 활용이 가능하다. 수강생은 수업의 목적과 학습 성과를 저해하지 않는 범위 내에서 생성형 AI를 사용할 수 있다. 다만, 과제의 최종 코드와 프레젠테이션의 핵심 내용은 수강생 본인의 힘으로 도출해야 하며, 퀴즈에서는 AI를 사용할 수 없다.								

5. 정원 외 신청	수용가능인원					
6. 수강생 참고사항	국문/영문					
	면담시간/장소					
*7. 강의계획	주요 수업방식	플립러닝	이론 위주	토론위주	프로젝트	기타
			O		O	
<p>본 교과목은 기존의 “대화형 인공지능”과 동일한 과목임.</p> <p>본 교과목은 오늘날 인공지능의 중심을 이루는 거대언어모델(Large Language Models)과 이를 기반으로 한 대화형 인공지능을 다룬다. 특히 거대언어모델의 학습부터 응용까지 이어지는 전체 파이프라인을 아래와 같이 체계적으로 다룬다.</p> <p>주차별 내용 (변동 가능)</p> <ol style="list-style-type: none"> 1. Introduction - Dialogue Systems and Societal Impact 2. Language Models - Word2vec and Seq2seq 3. Language Models - Transformer 4. Language Models - BERT 5. Language Models - GPT-1, GPT-2, and GPT-3 6. Language Models - T5 and Decoding Methods 7. Supervised Fine-tuning - Dialogue Fine-tuning 8. Supervised Fine-tuning - Instruction Tuning and Reasoning 9. Alignment - Reinforcement Learning from Human Feedback 10. Alignment - Direct Preference Optimization 11. Alignment - Advanced Learning Algorithms 12. Alignment - Learning from AI Feedback 13. Alignment - Pluralistic Values 14. Alignment - ChatGPT and API Programming 15. Interpretability - Mechanistic Interpretability 16. Interpretability - Knowledge Editing 17. Interpretability - AI Psychometrics 18. Grounding - Personas 19. Grounding - Documents 						

		<p>20. Grounding - Images 21. Grounding - Knowledge Graphs 22. Grounding - Tool Agents (1)-(6) 23. Grounding - User Simulators 24. Applications - Persuasion 25. Applications - Mental Health</p> <p>과제 및 프로젝트</p> <p>1. (개인 과제) 언어모델을 학습하고 디코딩하는 실습 2. (팀 과제) ChatGPT API를 활용해 특색있는 대화형 시스템을 구현하는 실습 3. (팀 프로젝트) 언어모델 및 대화형 인공지능 연구 주제 중 하나를 골라 개선하고 최종 결과를 발표</p> <p>수업에서는 각 연구 주제별로 핵심 논문들의 아이디어, 데이터, 방법론 등을 강의한다. 실습으로는 언어모델 학습, 언어모델 API를 이용한 대화시스템 구현, 대화형 인공지능과 관련된 작은 연구 프로젝트를 수행한다. 이 과목을 통해 수강생들은 거대언어모델 연구 동향을 큰 틀에서 파악하고, 핵심적인 연구 주제들과 방법론을 이해하며, 이를 실제 연구에 적용하는 능력을 기르게 된다.</p> <p>수강신청 시 참고사항</p> <ul style="list-style-type: none"> 본 수업에서는 기계학습과 딥러닝에 관한 기본적인 개념들은 다루지 않으며, 수강생들이 이미 관련된 지식을 갖고 있다고 가정하고 수업을 진행할 예정임. 기준에 "데이터사이언스 특강 - 대화형 자연어처리" 또는 "대화형 인공지능" 과목을 수강한 학생은 이 과목을 수강할 수 없음. 정원외 수강신청을 원하는 학생들은 정원외 수강신청 기간에 시스템에서 신청을 하고, 수강을 원하는 이유를 상세히 적을 것.
8. 장애학생 지원사항 ※ 필요에 따라 내용 수정 가능	강의수강 관련	<ul style="list-style-type: none"> 시각장애: 교재 제작(디지털교재, 점자교재, 확대교재 등), 대필도우미 허용 지체장애: 교재 제작(디지털교재), 대필도우미 및 수업보조 도우미 허용 청각장애: 대필 및 문자통역 도우미 활동 허용, 강의 녹취 허용 건강장애: 질병 등으로 인한 결석에 대한 출석 인정, 대필도우미 허용 학습장애: 대필도우미 허용 지적장애/자폐성장애: 대필도우미 및 수업 멘토 허용
	과제 및 평가 관련	<ul style="list-style-type: none"> 시각장애/지체장애/청각장애/건강장애/학습장애: 과제 제출기한 연장, 과제 제출 및 응답 방식의 조정, 평가 시간 연장, 평가 문항 제시 및 응답 방식의 조정, 별도 고사실 제공 지적장애/자폐성장애: 개별화 과제 제출 및 대체 평가 실시

	비고	본 강의를 수강하는 장애학생들에게는 이상의 지원 서비스 이외에도 장애학생 개개인의 특성과 요구에 따라, 지도교수 및 장애학생지원센터와의 상담을 통하여 적절한 수준의 지원 서비스를 제공합니다. 장애학생에 대한 지원서비스와 관련하여 문의사항이 있는 학생들은 담당 교수 ***(02-880-****) 혹은 장애학생지원센터(02-880-8787)로 문의바랍니다.
--	----	--

- ◇ * 은 필수 입력 항목으로 반드시 입력
- ◇ 강의계획서(논문제외) 미입력시 출석부 출력 불가