Syllabus 2025 Fall Version 2.2 (2025.10.16)
Kyung Hee Univ, Graudate School Big Data Analytics

Data & Algorithm Governance

교수 : 최은창 techlaw@khu.ac.kr
시간 : 월요일 2pm-5pm (오비스홀 254 호)

I. 강의 개요

- **강의 목표:** AI(Machine Learning, Deep Learning, Computer Vision, Large Language Model, AI Agents) 기술의 적용사례는 각 산업 분야에서 급증하고 있으나 반면 AI 시스템의 불완전성으로 인한 공정성, 강건성, 안전성, 책무성 문제도 발생하고 있다. 이 강의는 Predictive AI, Generative AI 의 사용과 관련한 AI 안전성 가이드라인, 윤리적 법적 이슈를 실무적 차원에서 이해하고 신뢰 가능한 AI 의 요건을 학습한다.
- **강의 방식:** ACM FaccT (Conference on Fairness, Accountability, and Transparency)에서 발표된 논문들과 실제로 문제된 AI risk, AI system failure 사례에 대한 분석에 중점을 둔다. 학생들은 각자 특정 산업 분야에 적용된 AI 사례를 선택하여 발표하고 이를 바탕으로 기말 페이퍼를 소논문 스타일로 작성함으로써 전문적 역량을 기르도록 한다.

II. 수업 목표

- AI 시스템을 실무에 적용될 때 발생하는 리스크 (편향, 공정성, 데이터 프라이버시, 예측의 부정확성) 관련한 이슈를 학습하고 공정성, 강건성, 안전성, 투명성, 설명가능성 등 Trustworthy AI 의 요소를 이해한다
- 산업 분야별 사례 분석: 자율주행, 상담, 금융, 의료, 공공영역 등에 AI 가 적용된 사례와 문제점을 분석함으로써 AI 비즈니스 가치창출 전략에 수반되는 문제점을 대비한다.
- 자동화된 판단의 편향·예측값의 오류, data security, privacy leak, reliability and accuracy, ethical and compliance issues, 학습 데이터의 무단 이용 사례를 학습한다.
- 학습 데이터의 취합 및 알고리즘 설계 단계에서 AI 의 법적/윤리적 리스크를 최소화하기 위한 방법을 이해한다.

III. 성적평가 및 유의사항

- **수업 참여(10%), 발표(10%), 과제 I,II(20%) 중간고사(20%), 기말고사(Paper Writing 40%)**
- 공결 출석인정 절차는 전산화되었으므로 온라인 신청. 공결 사유는 학칙에 열거되어 있음
- 중간고사(Mid-term) : 문제는 e-Campus 게시판에 공시되며 손으로(handwriting) 작성 후 스캔하여 pdf 파일로 제출함. Pdf 파일 명칭에 구체적 제목, 학생 이름, 제출일을 기재.
- 과제 및 중간고사의 분석은 학생이 독립적으로 수행해야 함. 집단협업 및 외부 조력 금지.
- 출처표기에 대한 강조: Presentation, 중간고사, Final Paper 작성할 때 인용한 자료에 대한 출처인용 (reference)을 정확하게 각 페이지의 하단에 구체적으로 기재해야 함.
- LLM 을 자료 조사 목적으로 사용할 수 있으나 학생은 상세한 문헌 자료를 직접 찾아 읽어보고 분석해야 함. 또한 인용한 자료의 출처 reference 를 상세히 각주에 기재해야 함. LLM 출력을 그대로 베끼거나 출처 표기가 부정확한 경우는 표절로 간주하여 D- 부여
- 강의 계획서의 Reading Materials 를 반드시 미리 읽고 강의에 참석해야 합니다. 자료를 읽어오고, 수업에 적극적으로 참여하면 학점에 긍정적으로 반영됩니다
- 강의 Slides 에는 저작권이 있으므로 공유 금지(전송, 전달, 배포, 클라우드 업로드 금지)

- https://khcanvas.khu.ac.kr/courses/79641 e-Campus "데이터알고리즘거버넌스" 게시판에서 중간고사 공지/ 과제물 공지/ 기말 페이퍼 작성방법/ 휴강 일정을 확인해야 함

IV. 커뮤니케이션 규칙 준수

- 수강생은 수강 이유와 '자기소개' 이메일을 9 월 10 일까지 techlaw@khu.ac.kr 로 보내세요 (포함내용: 전공, 이름, 수강 이유, 관심분야 키워드, 희망하는 발표 테마)
- 발표할 테마와 cases 를 9 월 20 일까지 미 이메일을 보내서 교수에게 컨펌받아야 함.
- Email Header 제목 작성시 아래의 커뮤니케이션 규칙을 준수해야 함

[이메일 헤더작성시 포함할 정보] 강좌명, 제출자료 명칭 (축약)

- 적절: [AI Governance], (중간고사 답안 제출) 홍길동, 제출일자
- 적절: [AI Governance], (과제물 1 번) 홍길동, 제출일자
- 부적절: 교수님 접니다. 기다리시던 보고서예요 ㅋㅋㅋ (민수 올림)
- 적절: 첨부파일 파일명 → (과제물 1 번)홍길동, 제출일 20251018.pdf
- 부적절: 첨부파일 파일명 → "과제 1.pdf"로만 기재한 경우. 이름과 날짜가 없음

V. 주차별 강의주제 및 필독자료

1 주차 (Sept 1). Overview on AI Risks, LLM Failures, Machine Learning Pipeline

- The AI Risk Repository (MIT Future Tech, 2025) https://arxiv.org/pdf/2408.12622
- AI Risks Catergories https://ai.wharton.upenn.edu/white-paper/artificial-intelligence-risk-governance
- OECD definition of AI system https://oecd.ai/en/wonk/ai-system-definition-update
- Machine Learning Pipeline https://broutonlab.com/blog/how-machine-learning-pipelines-work/
- Deep Learning Pipeline https://d2iq.com/blog/kubecon2018-data-pipeline
- Model Drift https://ahha.ai/2024/12/09/model-drift/
- Data Drift, Concept Drift https://www.ibm.com/think/topics/model-drift https://www.evidentlyai.com/ml-in-production/data-drift

2 주차 (Sept 8). Computer Vision Failure, Data Privacy in AI System Development

- Computational Limits of Deep Learning https://arxiv.org/pdf/2007.05558
- Faulty or Ready? Handling Failures in Deep-Learning Computer Vision Models https://dl.acm.org/doi/pdf/10.1145/3544548.3581555
- Why Do Multi-Agent LLM Systems Fail? https://arxiv.org/pdf/2503.13657?
- Detecting hallucinations in large language models using semantic entropy https://www.nature.com/articles/s41586-024-07421-0
- What could go wrong? Discovering and describing failure modes in computer vision https://arxiv.org/pdf/2408.04471
- Computer Vision Models Fail https://encord.com/blog/why-do-computer-vision-models-fail-in-production/
- Investigation of Tesla's Autopilot Crash Risks https://www.wsj.com/business/autos/tesla-autopilot-crash-investigation-997b0129 https://www.wsj.com/business/autos/regulators-probing-tesla-recall-tied-to-autopilot-a1af6d67?mod=article_inline

- 데이터 프라이버시의 중요성 https://www.ibm.com/kr-ko/topics/data-privacy
- 우리 기업을 위한 유럽연합 EU 일반 개인정보보호법(GDPR) 가이드북(2022.12. 개정) https://www.privacy.go.kr/front/bbs/bbsView.do?bbsNo=BBSMSTR_000000000049&bbscttNo=14518
- 약학정보원(KPIC)이 암호화된 환자 개인정보를 환자들의 동의를 받지 않고 한국 IMS 헬스데이터에게 공급한 사례 : 약국관리 프로그램 'Pharm Manager 2000'
  (민사판례 1) https://www.boannews.com/media/view.asp?idx=80239
  (민사판례 2) https://www.doctorsnews.co.kr/news/articleView.html?idxno=129093
  (형사판례) http://www.dailypharm.com/Users/News/NewsView.html?ID=313507
- OpenAI's ChatGPT is violating Europe's privacy laws, Italian Data Protection Authority https://techcrunch.com/2024/01/29/chatgpt-italy-gdpr-notification/ https://techcrunch.com/2023/04/01/chatgpt-blocked-in-italy/
- Survey on Data Collection for Machine Learning https://arxiv.org/pdf/1811.03402
- Preserving Privacy While Sharing Data 2022 MIT Sloan Management Review https://sloanreview.mit.edu/article/preserving-privacy-while-sharing-data/
- 차분 프라이버시(Differential Privacy)의 가능성과 한계 https://sapi.co.kr/wp-content/uploads/2020/05/차분-프라이버시의-가능성과-한계-구본효.pdf
- FTC Releases 2023 Privacy and Data Security Update https://www.ftc.gov/news-events/news/press-releases/2024/03/ftc-releases-2023-privacy-data-security-update
- Preserving Privacy in AI Applications through Anonymization of Sensitive Data https://www2.deloitte.com/content/dam/Deloitte/de/Documents/Innovation/Deloitte_Trustworthy%20AI%20_Data%20Anonymization_Feb2022.pdf
- 개인정보보호위, 생성형 인공지능(AI) 개발·활용 위한 개인정보 처리 안내서 (2025.8) https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=11410#LINK
- 개인정보보호위, 인공지능(AI) 개발·서비스를 위한 공개된 개인정보 처리 안내서 (2024.7) https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=10362
- 개인정보보호위, 가명정보 처리 가이드라인 https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS217&mCode=D010030000&nttId=9900#LINK ; https://www.shinkim.com/kor/media/newsletter/2342
- 개인정보보호위원회, AI 개발·서비스를 위한 공개된 개인정보 처리 안내서 (2024.7) https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=10362#LINK
- 개인정보보호위원회,'이루다'개발사 ㈜스캐터랩에 과징금·과태료 등 제재 처분 (2021) https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=7298

3주차 (Sept 15) Emotional Recognition, The Limits of Automated Job Interviews

- Where Automated Job Interviews Fall Short, Harvard Business Review www.hbr.org/2022/01/where-automated-job-interviews-fall-short
- 86% of Organizations Are Conducting Automated Video Interviews (AVIs). www.gartner.com/en/newsroom/press-releases/2020-04-30-gartner-hr-survey-shows-86--of-organizations-are-cond

- Public Debate on Facial Recognition Technologies in China https://mit-serc.pubpub.org/pub/public-debate-on-facial-recognition-technologies-in-china/release/1
- AI models fed AI-generated data quickly spew nonsense https://www.nature.com/articles/d41586-024-02420-7
- Best Practices and Lessons Learned on Synthetic Data for Language Models https://arxiv.org/pdf/2404.07503v1
- 안면인식기술의 법적 쟁점 : 데이터 수집 단계의 법적 리스크 https://sapi.co.kr/wp-content/uploads/2021/09/%EC%95%88%EB%A9%B4%EC%9D%B8%EC%8B%9D%EA%B8%B0%EC%88%A0%EC%9D%98-%EB%B2%95%EC%A0%81-%EC%9F%81%EC%A090_0923_2.pdf
- Artificial intelligence, algorithms, and social inequality https://compass.onlinelibrary.wiley.com/doi/full/10.1111/soc4.12962
- Racial Discrimination in Face Recognition Technology https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/
- Facebook to Pay $550 Million to Settle Facial Recognition Suit https://www.nytimes.com/2020/01/29/technology/facebook-privacy-lawsuit-earnings.html
- ACLU v. Clearview AI https://www.aclu.org/cases/aclu-v-clearview-ai
- Sam Altman's Eye-Scanning Worldcoin Venture Blocked in Spain https://www.wsj.com/tech/cybersecurity/sam-altmans-eye-scanning-worldcoin-venture-blocked-in-spain-057b7b14
- Managing Data Privacy Risk in Advanced Analytics, MIT Sloan Management Review https://sloanreview.mit.edu/article/managing-data-privacy-risk-in-advanced-analytics/
- Rite Aid Corporation, FTC v. www.ftc.gov/legal-library/browse/cases-proceedings/2023190-rite-aid-corporation-ftc-v
- FTC Signals Tough Line in First AI Discrimination Case Under Section 5 www.perkinscoie.com/en/news-insights/ftc-signals-tough-line-in-first-ai-discrimination-case-under-section-5.html
- New York City Local Law 114 for Automated Employment Decision Tools(AEDT) www.nyc.gov/site/dca/about/automated-employment-decision-tools.page
- NYC Tackles AI and Automated Decision-making in Employment and Recruiting https://ktslaw.com/en/Insights/Alert/2023/8/NYC-Tackles-AI-and-Automated-Decision-making-in-Employment-and-Recruiting
- 개인정보보호위원회고시 '자동화된 결정에 대한 개인정보처리자의 조치 기준' (2024.9) https://www.law.go.kr/LSW/admRulLsInfoP.do?admRulSeq=2100000247380
- AI 채용 문제점 ① 편향된 데이터로는 공정한 채용이 불가능 https://service.prism.work/insight/?q=YToyOntzOjEyOiJrZXl3b3JkX3R5cGUiO3M6MzoiYWxsIjtzOjQ6InBhZ2UiO2k6MTt9&bmode=view&idx=14131229&t=board
- AI 채용 문제점 ② AI 면접 결과에 대한 설명 불가능 https://service.prism.work/insight/?q=YToxOntzOjEyOiJrZXl3b3JkX3R5cGUiO3M6MzoiYWxsIjt9&bmode=view&idx=14403189&t=board
- 고용노동부, 개인정보호호 가이드라인 [인사노무편] https://moel.go.kr/policy/policydata/view.do?bbs_seq=20230101495

- 개인정보보호위원회, '자동화된 결정에 대한 개인정보처리자의 조치 기준 https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=10174

4 주차 (Sept 22). Copyright Infringement of Foundation Models, Fair Learning

- AI Copyright Wars, Copyright, Map of AI copyright lawsuits https://chatgptiseatingtheworld.com/2024/09/06/updated-maps-of-us-copyright-lawsuits-related-cases-sept-6-2024/
- Will we run out of data? Limits of LLM scaling based on human-generated data https://arxiv.org/pdf/2211.04325
- AI produces gibberish when trained on too much AI-generated data (Nature) https://www.nature.com/articles/d41586-024-02355-z
- AI firms will soon exhaust most of the internet's data (The Economist) www.economist.com/schools-brief/2024/07/23/ai-firms-will-soon-exhaust-most-of-the-internets-data
- 콘텐츠 기업들, 자사 웹사이트에서 크롤링 금지 데이터 수집 자체를 원천 차단 www.chosun.com/economy/tech_it/2024/07/23/A2PQH6G3ZZEUJOUG3FQGQZHO7U/
- AI 학습과 뉴스 저작권 간의 '공정이용' 갈등 https://www.lawtimes.co.kr/LawFirm-NewsLetter/196993
- AI Companies and Copyright Lawsuits in the U.S. https://chatgptiseatingtheworld.com/2024/04/01/map-of-20-copyright-lawsuits-v-ai-companies/
- 생성형 AI 와 저작권 쟁점 https://www.datocms-assets.com/45669/1701306602-231128-12.pdf
- Eight newspaper publishers sue Microsoft and OpenAI over copyright infringement https://www.cnbc.com/2024/04/30/eight-newspaper-publishers-sue-openai-over-copyright-infringement.html
- The New York Times v.s OpenAI, Harvard Law Review https://harvardlawreview.org/blog/2024/04/nyt-v-openai-the-timess-about-face/
- Reexamining "Fair Use" in the Age of AI https://hai.stanford.edu/news/reexamining-fair-use-age-ai
- Mark A. Lemley, Bryan Casey (2021) Fair Learning, Texas Law Review https://texaslawreview.org/fair-learning/
- Jenny Quang (2021) Does Training AI Violate Copyright Law https://btlj.org/wp-content/uploads/2023/02/0003-36-4Quang.pdf

5 주차 (Sept 29). AI in Financial Services: Use Cases and Applications

- AI: These are the biggest risks to businesses and how to manage them https://www.weforum.org/agenda/2023/07/ai-biggest-risks-how-to-manage-them/
- AI's Trust Problem, Harvard Business Review 2024 https://hbr.org/2024/05/ais-trust-problem
- 금융산업에서의 AI 활용 방안에 따른 리스크 요인 분석 (자본시장연구원) https://www.kcmi.re.kr/common/downloadw.php?fid=26494&fgu=002002&fty=004003

- 금융분야 AI 활용 활성화 및 신뢰확보 방안 (금융위원회 FSC) https://www.fsc.go.kr/no010101/78235?srchCtgry=&curPage=&srchKey=&srchText=&srchBeginDt=&srchEndDt=
- Eliminating Algorithmic Bias Is Just the Beginning of Equitable AI, Harvard Business Review https://hbr.org/2023/09/eliminating-algorithmic-bias-is-just-the-beginning-of-equitable-ai
- Fair Lending Compliance(미국) https://www.fdic.gov/banker-resource-center/fair-lending

6 주차 (Oct 6). (추석 휴강)

7 주차 (Oct 13). AI Agents: Structure, Applications, Controlling its Misbehaviors

- AI 에이전트란 무엇일까요? https://aws.amazon.com/ko/what-is/ai-agents/
- What are AI agents? https://www.ibm.com/think/topics/ai-agents
- We Need to Control AI Agents Now (Jonathan Zittrain) www.theatlantic.com/technology/archive/2024/07/ai-agents-safety-risks/678864/
- 2010 Flash Crash in The Stock Market in the U.S. https://corporatefinanceinstitute.com/resources/equities/2010-flash-crash/

8 주차 (Oct 20). Mid-Tersm Exam

e-Campus 게시판에 공지된 문제를 분석하여 수기(hand-writing)로 작성한 이후에 스캔 또는 사진으로 촬영 후 pdf 파일로 이메일 제출. 과제의 해결과 분석은 학생이 독립적으로 수행해야 함.

9 주차 (Oct 27). AI in Healthcare, How to Test Medical AI Systems

- Dissecting racial bias in an algorithm used to manage the health of populations https://www.science.org/doi/10.1126/science.aax2342
- “Millions of black people affected by racial bias in health-care algorithms” Nature https://www.nature.com/articles/d41586-019-03228-6
- A Health Care Algorithm Offered Less Care to Black Patients www.wired.com/story/how-algorithm-favored-whites-over-blacks-health-care
- How Health Care Algorithms and AI Can Help and Harm https://publichealth.jhu.edu/2023/how-health-care-algorithms-and-ai-can-help-and-harm
- The testing of AI in medicine is a mess. Here’s how it should be done https://www.nature.com/articles/d41586-024-02675-0
- Does AI Help or Hurt Human Radiologists’ Performance? It Depends on the Doctor https://hms.harvard.edu/news/does-ai-help-or-hurt-human-radiologists-performance-depends-doctor
- 미국 건강보험 이동성 및 책임법(HIPAA)에 의한 보건의료데이터 개인정보 보호 https://www.kiri.or.kr/pdf/연구자료/연구조사자료/rs-2014-07_09.pdf
- 미국 HIPAA 규정을 준수하며 Google 애널리틱스를 사용하기 https://support.google.com/analytics/answer/13297105?hl=ko

- 보건복지부,「보건의료데이터 활용 가이드라인」: 가명처리 방법 2024 www.mohw.go.kr/board.es?mid=a10501010000&bid=0003&list_no=1480106&act=view
- Patient data for commercial companies- ethical framework for sharing patients' data https://jme.bmj.com/content/early/2023/05/24/jme-2022-108781
- DeepMind's Streams app faces legal action over NHS data use https://www.cnbc.com/2021/10/01/google-deepmind-face-lawsuit-over-data-deal-with-britains-nhs.html
- Prediction of respiratory failure risk in patients with pneumonia in the ICU https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0291711

10 주차 (Nov. 3). Student Presentations 각자 주제와 Cases 를 선택하여 심층분석 (30 분)

11 주차 (Nov 10). Algorithms in the Public Sector. Automated decisionmaking Systems

- Algorithms in the public sector: four case studies of ADMS in Spain https://digitalfuturesociety.com/report/algorithms-in-the-public-sector-four-case-studies-of-adms-in-spain/
- This Algorithm Could Ruin Your Life (Dutch city of Rotterdam's welfare algorithm) https://www.wired.com/story/welfare-algorithms-discrimination
- When—and—Why You Should Explain How Your AI Works, Harvard Business Review. https://hbr.org/2022/08/when-and-why-you-should-explain-how-your-ai-works
- Investigation into France CNAF's Algorithm for Mass Profiling in Welfare System https://www.lighthousereports.com/methodology/how-we-investigated-frances-mass-profiling-machine
- The flawed algorithm at the heart of Robodebt (Australia's automated system) https://pursuit.unimelb.edu.au/articles/the-flawed-algorithm-at-the-heart-of-robodebt
- 3.3 *Transparency and Explainability* in Artificial Intelligence Index Report 2024 - Stanford University (pp. 180-185) https://aiindex.stanford.edu/wp-content/uploads/2024/04/HAI_2024_AI-Index-Report.pdf
- Requirements for Trustworthy Artificial Intelligence https://scholarworks.indianapolis.iu.edu/server/api/core/bitstreams/c748d7f0-5d6b-4f09-9467-f69db4eb6f07/content
- AI-Related Risks Test the Limits of Organizational Risk Management (Pros/ Cons) https://sloanreview.mit.edu/article/ai-related-risks-test-the-limits-of-organizational-risk-management/
- 설명가능한 AI 란 무엇인가? https://www.ibm.com/kr-ko/topics/explainable-ai
- 신뢰할 수 있는 AI 개발 안내서- AI 모델 명세 및 추론 결과에 대한 설명 https://tta-trustworthy-ai.gitbook.io/general/part2/11#id-11-2
- The Foundation Model Transparency Index 2024 https://crfm.stanford.edu/fmti/May-2024/index.html
- "알고리즘 통제: 투명성의 구체적 기준" https://sapi.co.kr/wp-content/uploads/2021/09/알고리즘-통제-투명성의-구체적-기준_0923_2.pdf
- Algorithm Accountability : France, the Digital Republic Law https://www.ogpstories.org/algorithm-accountability-what-government-can-do-right-now/

- EU Artificial Intelligence Act 1) https://artificialintelligenceact.eu/high-level-summary/ 2) 한국어 번역문 https://www.klri.re.kr/kor/data/S/1026/view.do

12 주차 (Nov 17). Algorithmic Ranking Manipulation, Dynamic Pricing, Fair Trade

- 비즈니스 기회 창출을 위한 AI 알고리즘의 활용 https://assets.kpmg.com/content/dam/kpmg/kr/pdf/kr-im-ai-algorithm-201806.pdf
- '알고리즘 조작' 논란 확산…플랫폼 보험 중개, 공정경쟁 저해 우려 www.fins.co.kr/news/articleView.html?idxno=95997
- 플랫폼의 보험상품 취급과 알고리즘 사전 검증 https://shorturl.at/HFBVQ
- Shoppers Are Caught Off Guard as Prices on Everyday Items Change More Often https://www.wsj.com/articles/shoppers-are-caught-off-guard-as-prices-on-everyday-items-change-more-often-11643970606
- Retailers, Are You Getting the Full Value of Your Dynamic Pricing Strategy? https://www.bain.com/how-we-help/retailers-are-you-getting-the-full-value-of-your-dynamic-pricing-strategy/
- 알고리즘으로 가격 차별…'다이나믹 프라이싱' https://www.lawtimes.co.kr/news/181985
- Amazon used an algorithm to essentially raise prices on other sites, the FTC says https://www.nbcnews.com/business/business-news/amazon-used-algorithm-essentially-raise-prices-rcna123410
- 알고리즘을 이용한 Amazon.com 의 PB(Private Brand) 랭킹 조작: FTC, State of Washington vs. Amazon.com https://www.ftc.gov/news-events/news/press-releases/2023/09/ftc-sues-amazon-illegally-maintaining-monopoly-power
- 알고리즘을 이용한 쿠팡과 씨피엘비(CPLB)의 PB(Private Brand) 랭킹 조작 (공정거래위 과징금 부과 사례) https://t.ly/a_1CE
- 전자상거래에서의 소비자 오인 가능성과 알고리즘 공개의무 https://sapi.co.kr/wp-content/uploads/2021/09/전자상거래에서의-소비자-오인-가능성과-알고리즘-공개의무_0923_2.pdf
- Beyond Dark Patterns: A Concept-Based Framework for Ethical Software Design https://arxiv.org/pdf/2310.02432
- Dark Patterns will now be heavily regulated https://www.williamfry.com/knowledge/dark-patterns-not-a-new-concept-but-will-now-be-heavily-regulated
- '다크패턴'에 대한 공정거래위원회 자율관리가이드라인 https://www.ftc.go.kr/www/selectReportUserView.do?key=10&rpttype=1&report_data_no=10140
- Dark Patterns, the tricks websites use to make you say yes, explained https://www.vox.com/recode/22351108/dark-patterns-ui-web-design-privacy
- Dark Patterns : Legal Cases https://www.deceptive.design/cases
- Dark Patterns used by online travel agency websites (OTA) https://doi.org/10.1016/j.annals.2020.103055

13 주차 (Nov 24). Predictive Policing, Predicting the Risk of Recidivism

- Analysis of the *Correctional Offender Management Profiling for Alternative Sanctions (COMPAS)* Recidivism Algorithm https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

- The accuracy, fairness, and limits of predicting recidivism https://www.science.org/doi/10.1126/sciadv.aao5580
- Can you make AI fairer than a judge? Play our courtroom algorithm game https://www.technologyreview.com/2019/10/17/75285/ai-fairer-than-judge-criminal-risk-assessment-algorithm/
- Societal Impacts of Predictive Policing Algorithms (PPAs) in Crime Prevention https://libraetd.lib.virginia.edu/downloads/x059c832h?filename=Smith_Joshua_STS_Research_Paper.pdf
- Artificial Intelligence in Predictive Policing Issue Brief https://naacp.org/resources/artificial-intelligence-predictive-policing-issue-brief
- Predictive policing algorithms are racist. They need to be dismantled. https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/
- Law Enforcement and Data-Driven Predictions https://t.ly/Y3Tan
- Robert Williams Arrest Case : Facial Recognition Led to Wrongful Arrests www.nytimes.com/2024/06/29/technology/detroit-facial-recognition-false-arrests.html

14 주차 (Dec 1). Data Bias, Algorithmic Fairness

- MIT FutureTech, AI Risk Repository: https://cdn.prod.website-files.com/669550d38372f33552d2516e/66bc918b580467717e194940_The%20AI%20Risk%20Repository_13_8_2024.pdf (6-7 pages, 31-32 pages)
- Machine Bias, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
- Shedding light on AI bias with real world examples https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples
- Why AI bias can hurt your business, The Wired https://www.wired.com/sponsored/story/why-ai-bias-can-hurt-your-business-ey
- What Do We Do About the Biases in AI? Harvard Business Review https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai
- Auditing large language models for race & gender disparities https://5harad.com/papers/auditing-llms.pdf
- Data, Power and Bias in Artificial Intelligence https://crcs.seas.harvard.edu/sites/projects.iq.harvard.edu/files/crcs/files/ai4sg_2020_paper_81.pdf
- AI bias : Why fair artificial intelligence is so hard to make www.vox.com/future-perfect/22916602/ai-bias-fairness-tradeoffs-artificial-intelligence
- Shedding light on AI bias with real world examples,
- https://www.ibm.com/blog/shedding-light-on-ai-bias-with-real-world-examples/
- Why AI bias can hurt your business, The Wired https://www.wired.com/sponsored/story/why-ai-bias-can-hurt-your-business-ey/
- What Do We Do About the Biases in AI? Harvard Business Review https://hbr.org/2019/10/what-do-we-do-about-the-biases-in-ai
- Data, Power and Bias in Artificial Intelligence https://crcs.seas.harvard.edu/sites/projects.iq.harvard.edu/files/crcs/files/ai4sg_2020_paper_81.pdf

- Manage AI Bias Instead of Trying to Eliminate It, MIT Sloan Management Review https://sloanreview.mit.edu/article/manage-ai-bias-instead-of-trying-to-eliminate-it/
- US Federal Trade Commission (FTC) Guideline Using Artificial Intelligence and Algorithms https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-algorithms

15 주차 (Dec 8). Management of AI Risks, Algorithmic Auditing

- Algorithmic Auditing: The Key to Making Machine Learning in the Public Interest https://www.businessofgovernment.org/sites/default/files/Algorithmic%20Auditing.pdf
- Auditing Algorithmic Risk, MIT Sloan Management Review (2024) https://sloanreview.mit.edu/article/auditing-algorithmic-risk/
- Avoid ML Failures by Asking the Right Questions, MIT Sloan Management Review https://sloanreview.mit.edu/article/avoid-ml-failures-by-asking-the-right-questions
- AI & Algorithmic Risks Report Netherlands https://www.autoriteitpersoonsgegevens.nl/uploads/2024-01/AI%20%26%20Algorithmic%20Risks%20Report%20Netherlands%20-%20winter%202023%202024.pdf
- Algorithmic governance https://policyreview.info/concepts/algorithmic-governance
- Managing algorithmic risks Safeguarding the use of complex algorithms and machine learning https://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-risk-algorithmic-machine-learning-risk-management.pdf
- Assessing Risks and Impacts of AI (ARIA), US NIST https://ai-challenges.nist.gov/uassets/6; https://ai-challenges.nist.gov/aria
- ISO/IEC 23894 표준 Artificial intelligence — Guidance on risk management https://tta-trustworthy-ai.gitbook.io/cat/cat-2.0/aisystem
- 신뢰할 수 있는 AI 개발 안내서(일반, 분야별: 의료, 채용, 생성 AI, 공공 사회) https://ai.kisdi.re.kr/aieth/main/contents.do?menuNo=400041
- AI 윤리기준 자율점검표 / AI 윤리영향평가 https://ai.kisdi.re.kr/aieth/main/contents.do?menuNo=400008
- The 4 Asian Countries: AI Governance Comparison Table (Korea Singapore Japan China) https://shorturl.at/EFQYu
- SHAP(Shapley Additive exPlanations) https://shap.readthedocs.io/en/latest/
- Summary of EU AI Act https://artificialintelligenceact.eu/high-level-summary/
- 유럽연합 인공지능법(국문 번역자료) ; EU Artificial Intelligence Act https://www.klri.re.kr/kor/data/S/1026/view.do
- The EU AI Act Explained: Tracking Developments for Responsible AI https://www.responsible.ai/the-eu-ai-act-explained-tracking-developments-for-responsible-ai/
- California Legislature Approves Bill Proposing Sweeping A.I. Restrictions https://www.nytimes.com/2024/08/28/technology/california-ai-safety-bill.html Machine Bias, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing
- Organizations Face Challenges in Timely Compliance With the EU AI Act https://sloanreview.mit.edu/article/organizations-face-challenges-in-timely-compliance-with-the-eu-ai-act/

16 주차 (Dec 15). Research Tersm Paper Submission

- 소논문 형태(참고문헌 및 표지 제외하고 본문만 25~ 36 pages)로 작성하여 반드시 Print 하여 제출. 중요한 문장에 밑줄을 쳐서 가독성을 높여야 함. 각자 20 분씩 발표
- 최종 수정한 기말 페이퍼 PDF 파일은 12 월 18 일 자정까지 이메일로 제출.
- 소논문 형태의 기말 페이퍼 작성시 학생들은 인용한 문헌자료는 출처표기를 철저하게 각 페이지의 하단에 기재해야 합니다. 출처표기는 미주가 아니라 각주에 상세하게 달아야 하며 엄밀해야 합니다. 출처표기가 정확하게 하지 않거나 미흡한 경우 D- 학점을 부여함.

VI. 발표 주제 선정 / 기말 페이퍼 작성

- 학생은 발표할 테마(AI 가 적용된 산업분야를 선택)와 분석할 사례들(cases)을 10 줄로 요약하여 미리 9 월 20 일 자정까지 교수에게 이메일로 송부. (중복 방지를 위한 절차)
- 발표한 테마와 사례를 심층 분석하여 소논문 형태의 기말 페이퍼로 발전시켜야함
- 소논문 주제를 이메일로 송부 9 월 28 일 자정까지: 포함시킬 내용 (1) 구체적 제목 (2) Abstract (연구의 필요성/기존연구의 한계/ 분석의 초점을 요약한 200 자) (3) 주요 목차 (4) 분석하려는 cases 또는 dataset (5) 참고문헌 목록
- 10 월 6 일부터 소논문 draft 를 매주 업데이트 후 Print 로 제출하면 피드백을 받을 수 있음. Print 원고에는 제출일/Version 명칭/이름/학번을 기록. 페이지 기재 필수
- 발표 주제 선정에 참고할 Journals : Harvard Business Review, MIT Sloan Management Review, Big Data & Society https://journals.sagepub.com/home/bds ; Internet Policy Review https://policyreview.info/ ; AI & SOCIETY https://link.springer.com/journal/146
- 발표 주제 선정에 참고할 학회 Proceedings : AAAI/ACM Conference on AI, Ethics, and Society (AIES) Proceedings https://dl.acm.org/conference/aies/proceedings ; Fairness, Accountability, and Transparency (ACM FAccT) ACM Conference, Accpeted Papers & Proceeedings https://facctconference.org 2024-2020

VII. 참고자료

- 개인정보보호위, 생성형 인공지능(AI) 개발·활용 위한 개인정보 처리 안내서 (2025.8) https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS074&mCode=C020010000&nttId=11410#LINK
- 개인정보보호위원회, 가명정보 처리 가이드라인(2024.2.4 개정) https://www.pipc.go.kr/np/cop/bbs/selectBoardArticle.do?bbsId=BS217&mCode=D010030000&nttId=9900
- 개인정보위, AI 개발·서비스를 위한 공개 개인정보 처리안내서(2024.7)https://t.ly/w_Y_u
- 보건복지부, 「보건의료데이터 활용 가이드라인」 (2024.1) www.mohw.go.kr/board.es?mid=a10501010000&bid=0003&list_no=1480106&act=view
- Artificial Intelligence Index Report 2025, 2024 – Stanford HAI https://aiindex.stanford.edu/wp-content/uploads/2024/05/HAI_AI-Index-Report-2024.pdf https://hai.stanford.edu/ai-index/2025-ai-index-report
- 유럽연합 인공지능법(국문 번역자료) ; EU Artificial Intelligence Act https://www.klri.re.kr/kor/data/S/1026/view.do
- 유럽연합 일반개인정보보호법(EU GDPR) 가이드북 (국문 번역자료) https://www.privacy.go.kr/front/bbs/bbsView.do?bbsNo=BBSMSTR_000000000049&bbscttNo=14518

- 우리 기업을 위한 EU 일반 개인정보보호법(GDPR) 가이드북(2022.12. 개정)
  https://www.privacy.go.kr/front/bbs/bbsView.do?bbsNo=BBSMSTR_000000000049&bbscttNo=14518