

법정책 연구실

Law and Policy
Laboratory



박도현
교수

gray@gist.ac.kr
062-715-6393

Education

- 2021. 2 서울대학교 일반대학원 법학과 졸업 (Ph.D.)
- 2016. 2 서울대학교 법학전문대학원 졸업 (J.D.)
- 2011. 8 서울대학교 경제학부 졸업 (B.A.)

Experience

- 2023. 5 ~ GIST AI대학원 조교수
- 2023. 1 ~ 2023. 2 스위스 취리히 연방공과대학교(ETH Zurich) 방문연구원
- 2021. 9 ~ 2023. 2 서울대학교 컴퓨터공학부 강사 ('AI와 법' 강의 담당)
- 2021. 3 ~ 2023. 2 서울대학교 박사후연구원

Awards and Honors

- 2021. 11 제6회 흥진기법률연구상 대상 수상
- 2019. 8 서울대학교 기초교육원 우수조교상 수상
- 2016. 4 제5회 변호사시험 합격

연구실 소개

오늘날 인류가 활용하는 모든 제품과 서비스는 이를 뒷받침하는 법정책과 밀접한 관계를 형성하고 있다. 현재 세계 각국은 인적, 물적 인프라를 형성하는 진흥법과 윤리적, 사회적 문제를 예방하는 규제법을 바탕으로 인공지능 기술의 경쟁력과 수용성을 확보하기 위한 다양한 정책을 펼쳐나가고 있다. 우리나라의 인공지능 산업은 이와 같은 국제적 추세와 국내의 고유한 특성을 골고루 반영한 법정정책 연구가 바탕이 되었을 때 비로소 결실을 맺을 수 있을 것이다. (Law for AI)

한편 인공지능은 여타의 신기술과 달리 인간의 의사결정을 대체하는 특성을 가지고 있다. 인공지능과 다른 인지적 특성을 가진 인간에게는 이것이 불확실성의 원천이 되는 동시에 인간이 기존에 해결하지 못한 난제를 풀어낼 수 있는 새로운 기회가 되기도 한다. 따라서 인간과 인공지능의 인지적 특성을 비교하여 인공지능이 법정정책 연구사항을 더 원활히 구현할 수 있는 영역과 그 구체적 방법론을 탐구해보고자 한다. (AI for Law)

(1) 신뢰할 수 있는 인공지능 법정정책 연구

- '신뢰할 수 있는 인공지능(Trustworthy AI)' 개념의 역사적 배경과 이론적 원리 연구
- 해외의 신뢰할 수 있는 인공지능 법정정책 연구: EU, US, OECD, Google, Microsoft 등
- 국내의 신뢰할 수 있는 인공지능 법정정책 연구: 정부, 기업, 학계, 시민사회 등
- 영역별 신뢰할 수 있는 인공지능 법정정책 연구: 의료, 금융, 교통, 고용 등

(2) 신뢰할 수 있는 인공지능 기술 연구

- 데이터 프라이버시 연구: 비식별 조치, 차분 프라이버시, 연합학습, 합성 데이터 등
- 설명가능 인공지능 연구: DAPRA 프로젝트, 화이트박스 접근, 블랙박스 접근 등
- 공정한 인공지능 연구: 벤치마크 데이터셋, 공정성 지표(metrics) 등
- 법률 인공지능 연구: 인공지능 기반 법률 플랫폼, 인공지능 법률가 등

(3) 신뢰할 수 있는 인공지능 융합 연구

- 공학적으로 현존하는 기술의 구현가능성을 전제로 한 최적의 법정정책에 관한 연구
- 인지과학적으로 인간이 신뢰성을 인지하도록 하는 데 필요한 조건에 관한 연구
- 인공지능 기술의 부정적 파급력을 최소화하는 재분배와 사회보장 법정정책에 관한 연구
- 인공지능의 미래지향적 이슈 연구 (파운데이션 모델, 강인공지능, 초인공지능 등)

연구 성과

수행중인 주요 연구과제 (주요과제경력)

- 박사급연구원 & 공동연구원, 인공지능 신뢰성 (2022. 11 ~ 현재), 한국연구재단 & 과학기술정보통신부
- 연구책임자, 스위스의 인공지능 신뢰성 거버넌스에 관한 연구 (2022. 11 ~ 2023. 10), 한국연구재단
- 연구보조원 & 박사급연구원, 기술적, 의학적, 윤리적, 법적 관점에서 바라본 인공지능의 책임성과 거버넌스 (2019. 7 ~ 2022. 6), 한국연구재단 & 교육부
- 연구보조원, 보건의료 빅데이터 보호 및 활용을 위한 법적 기반 강화 연구 (2019. 1 ~ 2019. 3), 보건복지부
- 연구보조원, 인공지능을 이용한 의사결정과 설명가능성의 문제 (2018. 8 ~ 2019. 7), 서울대학교

주요논문 (대표실적)

- 인공지능 공정성의 이론과 실제, 법철학연구 제26권 제2호 (2023. 8, 단독저자)
- 신뢰할 수 있는 인공지능의 이론적 고찰, 법철학연구 제25권 제2호 (2022. 8, 단독저자)
- 인간 편향성과 인공지능의 교차, 서울대학교 법학 제63권 제1호 (2022. 3, 단독저자)
- 인공지능 윤리의 두 가지 가치, 법학논문집 제45권 제1호 (2021. 4, 단독저자)
- 인공지능 윤리규범과 규제 거버넌스의 현황과 과제, 경제규제와 법 제13권 제1호 (2020. 5, 제1저자)
- 인공지능과 자율성의 역학관계, 홍익법학 제20권 제3호 (2019. 9, 단독저자)
- 인공지능과 고용차별의 법경제학, 법경제학연구 제16권 제1호 (2019. 4, 제1저자)
- 인공지능과 차별, 저스티스 통권 제171호 (2019. 4, 공동저자)

저서

- 포스트 챗GPT, 한빛비즈 (2023. 5, 복채터 공저)
- 개인정보 판례백선, 박영사 (2022. 7, 복채터 공저)
- 인공지능 시대의 개인정보 보호법, 박영사 (2022. 5, 공저)
- 인공지능 원론: 설명가능성을 중심으로, 박영사 (2021. 7, 공저)

G.I.S.T. AI for X 융합연구

