Messi H.J. Lee

messihjlee@gmail.com | 개인 웹사이트 | 구글 스칼라 | LinkedIn

연구 관심 분야

AI의 사회심리학(Social Psychology of AI), 머신 사이콜로지(Machine Psychology), AI 편향(예: 대규모 언어 모델 및 시각-언어 모델의 고정관념), LLM 행동 평가, 계산사회과학, 텍스트 데이터 분석(Text-as-data)

경력

MTS Company | 전문연구요원 | 대한민국 2025년 6월–현재

- 다섯 개 암종에서 유사분열(mitosis) 계수를 수행하는 병리 AI 소프트웨어 두 종(2등급 및 3등급 의료기기)의 식품의약품안전처(MFDS) 인허가에서 기술 부문 총괄
- 전체 슬라이드 이미지(WSI) 80,000장 이상과 그 주석(영역 및 유사분열 좌표)으로 구성된 사내 데이터베이스 관리
- 기술특례상장을 위한 지식재산(IP) 전략 관리: 특허 출원 24건, 진행 중 38건, 등록 1건
- 전체 슬라이드 이미지로부터 유방암의 림프절 전이를 예측하는 다중 인스턴스 학습(MIL) 모델 개발, 두 병원의 외부 검증 코호트에서 AUC 0.78 달성

학력

박사 | Washington University in St. Louis | 미국 2025년 5월

- 계산 및 데이터 과학 박사(Computational and Data Sciences, 심리 및 뇌과학 트랙)
- 학위논문: "Stereotyping in Language (Technologies): An Examination of Racial and Gender Stereotypes in Natural Language and Language Models" [발표 영상]
- 맥도넬 국제 장학생(McDonnell International Scholar)

학사 | 홍콩과학기술대학교(HKUST) | 홍콩 2017–2020

- 정량사회분석 학사(BSc in Quantitative Social Analysis)
- 졸업 프로젝트: "Part-time Electoral Campaigns on Social Media"
- 학업성취 메달(Academic Achievement Medal) 및 최우등 졸업(First Class Honors)

학회 및 학술지 논문 (게재 확정)

Lee, M. H. J. & Lai, C. K. (2026). Implicit-bias-like patterns in reasoning models. *Nature Machine Intelligence*. [링크] [프리프린트]

Lee, M. H. J., Jang, J., Seo, J., Cho, N. H., Cha, Y. J., Park, S., Kim, J. M., Choe, M., Jung, H. R., Nahm, J. H., Yoo, S. H., Park, H., Park, J. H., Jang, M., Kim, E., Kim, T., Hong, M., Shin, S., Noh, J., Ha, J., & Shin, S.-J. (2026). Deep learning-based mitotic quantification for objective adjunct for grading of urothelial carcinoma: a multi-institutional validation study. *Scientific Reports*. [링크]

Chun, J., Kang, H., Chung, H., Jang, J.-M., Seo, J., Kim, T., Lee, W., Park, C., Hong, M., Kim, H.-M. B., **Lee, M. H. J.**, Jang, K., Jung, C. K., Kim, S. W., & Lee, A. (2026). Deep Learning-Based Diagnosis of Epithelial Ovarian Cancer from Whole-Slide Histopathology Images. *Diagnostics*, 16(10), 1470. [링크]

Jeon, Soyeon, **Lee, M. H. J.**, Montgomery, J. M., & Lai, C. K. (Political Analysis). From Faces to Politics: Vision-Language Models (Sometimes) Link Visual Demographic Characteristics to Ideological Labels. *Political Analysis*. Published online 2026:1-9. [링크] [재현 자료]

Lee, M. H. J., Montgomery, J. M., & Lai, C. K. (June 2024). Large Language Models Portray Socially Subordinate Groups as More Homogeneous, Consistent with a Bias Observed in Humans. *FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. [링크] [재현 자료]

Lee, M. H. J., Montgomery, J. M., & Lai, C. K. (2024). America's racial framework of superiority and Americanness embedded in natural language. *PNAS Nexus*, pgad485. [링크] [재현 자료]

심사 중

Lee, M. H. J. (심사 중). Agent Personas Shift Allocation Decisions About People: What Matters Is Whether the Two Strings Match. [프리프린트]

Lee, M. H. J. (심사 중). The Minimal Persona: Rarer and Shared Labels Both Skew Allocation Among AI Agents. [프리프린트]

Lee, M. H. J. (심사 중). Token-Level Entropy Reveals Demographic Disparities in Language Models. [프리프린트]

Lee, M. H. J., Jeon, S., Montgomery, J. M., & Lai, C. K. (준비 중). Visual Cues of Gender and Race are Associated with Stereotyping in Vision-Language Models. [프리프린트]

PSA005 Team (*Nature Human Behavior* 심사 중). Stereotype Threat in Black College Students Across Many Operationalizations. [Stage 1 보고서]

진행 중인 연구

Lee, M. H. J. & Jeon S. (준비 중). Vision-Language Models Generate More Homogeneous Stories for Phenotypically Black Individuals. [프리프린트]

Lee, M. H. J. & Lai, C. K. (준비 중). Probability of Differentiation Reveals Brittleness of Homogeneity Bias in GPT-4. [프리프린트]

워크숍 논문 (약식 심사)

Lee, M. H. J., Montgomery, J. M., & Lai, C. K. (2023). The Effect of Group Status on the Variability of Group Representations in LLM-generated Text. *NeurIPS 2023 Socially Responsible Language Modelling Research (SoLaR) Workshop*. [OpenReview]

발표

Lee, M. H. J. & Lai, C. K. (2025년 2월). Implicit Bias-like Patterns in Reasoning Models. 구두 발표, Political Data Science Lab Meeting, Washington University in St. Louis.

Lee, M. H. J., Montgomery, J. M., & Lai, C. K. (2024년 10월). *GPT-4 Associates Distinctively Black Faces with Political Liberalism: Uncovering Political Bias at the Intersection of Race and Gender in AI*. 구두 발표, Political Data Science Lab Meeting, Washington University in St. Louis.

Lee, M. H. J., Montgomery, J. M., & Lai, C. K. (2024년 6월). *Large Language Models Portray Socially Subordinate Groups as More Homogeneous, Consistent with a Bias Observed in Humans*. 구두 발표, ACM Conference on Fairness, Accountability, and Transparency 2024, 브라질 리우데자네이루.

Lee, M. H. J., Montgomery, J. M., & Lai, C. K. (2024년 5월). *Large Language Models Portray Socially Subordinate Groups as More Homogeneous, Consistent with a Bias Observed in Humans*. 포스터 발표, STL DataFest 2024, 미국 미주리주 세인트루이스.

Lee, M. H. J. (2024년 3월). *Large Language Models Portray Socially Subordinate Groups as More*

Homogeneous, Consistent with a Bias Observed in Humans. 구두 발표, TRIADS "Messy" Data Meeting, Washington University in St. Louis.

Lee, M. H. J., Lai, C. K., & Montgomery, J. M. (2024년 2월). *Outgroup Homogeneity-like Effect in Large Language Models*. 포스터 발표, Society for Personality and Social Psychology Conference, 미국 캘리포니아주 샌디에이고. [포스터]

Lee, M. H. J., Lai, C. K., & Montgomery, J. M. (2023년 12월). *The Effect of Group Status on the Variability of Group Representations in LLM-generated Text*. 포스터 발표, NeurIPS 2023 Socially Responsible Language Modelling Research (SoLaR) Workshop, 미국 루이지애나주 뉴올리언스. [포스터]

Lee, M. H. J., Montgomery, J. M., & Lai, C. K. (2023년 2월). *A Test of the racial position model in natural language*. 포스터 발표, Society for Personality and Social Psychology Conference, 미국 조지아주 애틀랜타. [포스터]

Lee, M. H. J. (2022년 4월). *Racial/ethnic stereotypes in natural language*. 구두 발표, Social and Personality Psychology Brown Bag, Washington University in St. Louis.

Lee, M. H. J., & Lai, C. K. (2022년 2월). *The spread of stereotypes on social media*. 온라인 포스터 발표, Society for Personality and Social Psychology Conference.

Lee, M. H. J., Keller, F., & Yang, J. (2020년 8월). *Central coordination patterns distinguish AstroTurfing from grassroots campaigns*. 온라인 포스터 발표, Politics and Computational Social Science & Political Networks.

수상 및 연구비

연구비(Small Grant) 2024
The Center for the Study of Race, Ethnicity & Equity, WashU ($1,000)

학회 참가 여비 지원(Travel Grant) 2024
ACM Conference on Fairness, Accountability, and Transparency ($1,000)

OpenAI API Researcher Access Program 2024
OpenAI ($5,000 상당 크레딧)

우수 연구상(Outstanding Research Award) 2024
The Society for Personality and Social Psychology

대학원생 학회 참가 여비 지원(Graduate Student Conference Travel Award) 2023
The Center for the Study of Race, Ethnicity & Equity, WashU ($750)

디지털 전환 트랙 우승 및 청중상(Winner and People's Choice) 2023
McDonnell International Scholars Academy ($3,000)

연구비(Small Grant) 2022
The Center for the Study of Race, Ethnicity & Equity, WashU ($900)

다양성 대학원생 여비 및 등록비 지원(Diversity Graduate Travel/Registration Award) 2022
The Society for Personality and Social Psychology ($500)

학업성취 메달(Academic Achievement Medal) 2020
The Hong Kong University of Science and Technology

학부 연구상 3위(Undergraduate Research Award Second Runner Up) 2020
The Hong Kong University of Science and Technology

재학생 장학금(Scholarship for Continuing Undergraduate Students) 2019
The Hong Kong University of Science and Technology

재학생 장학금(Scholarship for Continuing Undergraduate Students) 2017, 2018
Ng Hong Mun Educational Foundation

강의 경력

Washington University in St. Louis, 강의 조교(Assistant in Instruction) 2023
Text as Data (담당 교수: Dr. Brandon Stewart)

Washington University in St. Louis, 강의 조교(Assistant in Instruction) 2022–2023
Quantitative Methods I & II (담당 교수: Dr. Shelly Cooper)

학술 활동

학술지 심사

- Nature Communications (2026년 9월)
- Journal of Language and Social Psychology (2025년 5월)
- Ethics and Information Technology (2025년 4월)
- Computers in Human Behavior: Artificial Humans (2024년 11월, 2025년 1월)
- PLOS ONE (2024년 12월)
- Personality and Social Psychology Bulletin (2024년 3월)

학회 심사

- ACM Conference on Fairness, Accountability, and Transparency (FAccT) 2027
- ACL Rolling Review (ARR) 2026년 10월
- Conference on AI, Ethics, and Society (AIES) 2026 (프로그램 위원)
- Conference on Neural Information Processing Systems (NeurIPS) 2026
- Conference on AI, Ethics, and Society (AIES) 2025 (프로그램 위원)
- Society for Personality and Social Psychology (SPSP) 2024 학생상 심사
- Socially Responsible Language Modelling Research (SoLaR) 2.0 Workshop at NeurIPS 2024
- Association for Computational Linguistics (ACL) Student Research Workshop 2024
- Trustworthy Multi-modal Foundation Models and AI Agents (TiFA) Workshop at ICML 2024
- Socially Responsible Language Modelling Research (SoLaR) Workshop at NeurIPS 2023

기술 역량

대규모 언어 모델, 자연어 처리 및 텍스트 데이터 분석

- 검색 증강 생성(RAG), 벡터 데이터베이스(ChromaDB)
- 대규모 언어 모델의 행동 평가
- 텍스트 데이터 분석 기법(예: 구조적 토픽 모델, 텍스트 분류, 감성 분석)

컴퓨터 비전 및 계산 병리학

- 전체 슬라이드 이미지 분류를 위한 다중 인스턴스 학습(MIL): Gated-Attention MIL (ABMIL), DTFD-MIL, 계층적 MIL (HIPT 방식)
- 병리 파운데이션 모델을 특징 추출기로 활용(H-optimus-0, UNI2, Kaiko Midnight), timm 및 HuggingFace transformers 사용
- OpenSlide를 이용한 전체 슬라이드 이미지 처리, scikit-image를 이용한 조직 분할 및 염색 공간 분석
- RF-DETR (DINOv2 백본과 DETR 디코더) 기반 객체 탐지, Roboflow를 통한 COCO/YOLO 형식 주석 파이프라인, ONNX 및 TensorRT로 모델 배포

계산사회과학 및 심리학

- Qualtrics, CloudResearch, MTurk, Prolific을 활용한 설문 실험

주로 사용하는 프로그래밍 언어 및 패키지

- python: PyTorch, torchvision, timm, HuggingFace, bitsandbytes, peft, sentence-transformers, transformers, openai, OpenCV, albumentations, scikit-learn, gensim, numpy, pandas
- R: text, tidyverse, dplyr, ggplot2, psych, lme4, lmerTest, afex, simr

연구 보도

Rutgers Today, Rutgers University (링크)	2026
Tech Xplore (링크)	2026
ALLATRA Media (링크)	2026
JoongAng Ilbo (링크)	2026
Bioengineer.org (링크)	2026
Department of Political Science, Washington University in St.Louis (링크)	2024
Center for the Study of Race, Ethnicity & Equity (링크)	2024