Ph.D. Candidate

Research Blog
bumjin@kaist.ac.kr
bumjini42@gmail.com

Bumjin Park

My research focuses on understanding AI’ s **mind** through rigorous and precise analysis based on scientific assumptions about the ”brain” of AI, neural representations. Just as **epistemology** has long sought to interpret human cognition, I explore the fundamental principles that shape both human and artificial intelligence.

I work on Cognitive Science, Mechanistic Interpretability, Explainable AI, Large Language Models, Multi-Agent Systems, and Communication, leveraging mathematics and programming to drive my research.

My long-term goal is to uncover the **General Principles of Mind** that underlie both human and AI intelligence and to advance philosophical and scientific research that enables their efficient utilization.

Education

Ph.D. in Artificial Intelligence **09/2023 – Present**
Korea Advanced Institute of Science and Technology (*KAIST*), *AI Graduate School*
- **Proposal**: Integrating Cognitive Architectures into Large Language Models [Drive]

M.S. in Artificial Intelligence (**4.17/4.3**) **09/2021 – 08/2023**
Korea Advanced Institute of Science and Technology (*KAIST*), *AI Graduate School*
- **Thesis**: Partitioned Channel Gradient for Reliable Saliency Map in Image Classification [Drive]

B.S. in Mathematics (**Double Major in Software Engineering**) (**4.39/4.5**) **03/2018 – 08/2020**
Chung-Ang University, Korea

Publication

ACL-Main **Deontological Keyword Bias: The Impact of Modal Expressions on Normative Judgments of Language Models** [Arxiv]
Bumjin Park, Jinsil Lee, Jaesik Choi
Annual Meeting of the Association for Computational Linguistics (*Main Conference*), *2025*

IJCAI **Memorizing Documents with Guidance in Large Language Models** [IJCAI]
Bumjin Park, Jaesik Choi
International Joint Conference on Artificial Intelligence, 2024

ICPRAI **Identifying the Source of Generation for Large Language Models** [Springer Nature]
Bumjin Park, Jaesik Choi
Pattern Recognition and Artificial Intelligence, 2024

Applied Sci. **Message Action Adapter Framework in Multi-Agent Reinforcement Learning** [Appl. Sci.]
Bumjin Park, Jaesik Choi
Applied Sciences, 2025

Applied Sci. **Cooperative Multi-Robot Task Allocation with Reinforcement Learning** [Appl. Sci.]
Bumjin Park, Cheongwoong Kang, Jaesik Choi
Applied Sciences, 2022

Sensors **Scheduling PID Attitude and Position Control Frequencies for Time-Optimal Quadrotor Waypoint Tracking under Unknown External Disturbances** [Sensors]
Cheongwoong Kang, Bumjin Park, Jaesik Choi
Sensors, 2021

ICCAS **Generating Multi-Agent Patrol Areas by Reinforcement Learning** [IEEE]
Bumjin Park, Cheongwoong Kang, and Jaesik Choi
2021 21st International Conference on Control, Automation and Systems (ICCAS), IEEE

UnderReview **AlchemyNet: Learning Material Composition Distributions for Property Prediction** [Drive]
Bumjin Park, Jihyun Jun, Jungeun Lee, Hyungjin Bae, Jinmo Kim, Sangkeun Han, Jaesik Choi
Pattern Recognition Letters

Project

ADD (Agency for Defense Development) - Unmanned Swarm CPS Research Lab 10/2021 - Present

Multi-agent swarm intelligence interacting with
physical and simulated environments (Cyber-Physical System).

Tasks:

- Multi-agent reinforcement learning for patrol and communication.
- Simulation-to-reality (Sim-to-Real) transfer using domain adaptation.
- ROS-based communication between physical and simulated environments (Gazebo and Webots) for both UAVs (DJI) and UGVs (Husarion Rosbot). [Youtube Demo]

Achievements:

- Publications: Journal (2) / Conference (1) / Domestic Journal (1) / Domestic Conference (4)
- Patent Applications (2), Registered Patent (1)
- Software Registrations (2)
- KTL Certified Evaluations (2)

Kolmar - AlchemyNet & Domain Knowledge (Phase 2) 11/2024 – Present

Phase 2 of cosmetic AI leveraging **domain knowledge** for AlchemyNet.

- Publications: Journal (Under Review: 1) / Conference (Under Review: 1)
- Web Application (1)

Kolmar - AI for Cosmetic Composition (Phase 1) 08/2023 – 03/2024

Development of **AlchemyNet** (inspired by the concept of an alchemist) to encode
formulation compositions and predict multiple properties, including viscosity, pH, density, and hardness.

Global AI Frontier LAB 08/2024 – Present

Interpretation of bias representation in Large Language Models based on mechanistic Interpretability.

Minor Projects - Short Term Projects

X-Ray Object Detection and Saliency 08/2024
This study applies XAI techniques, such as attribution maps, to enhance the interpretability of X-ray object detection models. By analyzing decision-making in overlapping object scenarios, it improves model transparency and aids multi-object detection.

Experiences

Lab Representative **09/2023 - 08/2024**

Organized SAILAB as a Ph.D. student.

GPU Server Management **01/2022 - 12/2023**

Organized GPU resources in SAILAB

Skills

Programming **Python / Torch / C++**

- Pytorch
- Pytorch Hook:

Paper (Domestic)

정보과학회	**대형언어모델 생성 텍스트의 원천 문서 추적** 박범진, 최재식 정보과학회지 특집원고, *2024*	[Drive]
KSC	**대형 언어 모델의 문장 표현의 설명가능적 해석** 박범진, 최재식 한국정보과학회 (*KSC*), *2024*	[Drive]
KIMST	**자연어 명령어 기반 군집 로봇 제어 프레임워크** 박범진, 최재식 한국군사과학기술학회 (*KIMST*), *2024*	[Drive]
KIMST	**배터리 효율성을 위한 강화학습 로봇 제어** 장원준, 박범진, 최재식 한국군사과학기술학회 (*KIMST*), *2024*	[Drive]
ICROS	**멀티 에이전트 강화학습에서의 게이트 기반 메시지 교환** 박범진, 강청웅, 최재식 *Journal of Institute of Control, Robotics and Systems* (*ICROS*), *2023*	[Drive]
KROS	**샘플 필터링을 통한 효율적인 강화학습 모델 지식 증류** 박범진, 강청웅, 최재식 한국로봇학회 (*KROS*), *2023*	[Drive]
KIMST	**계층별 관련도 전파 기반 무인이동로봇 강화학습 의사결정 과정 자동 분석** 강청웅, 박범진, 최재식 한국군사과학기술학회 (*KIMST*), *2022*	[Drive]
KIMST	**심층강화학습 기반 목표물 추적 시스템에서 다중 목적 함수 분석** 강청웅, 박범진, 최재식 한국군사과학기술학회 (*KIMST*), *2021*	[Drive]