

Bumjin Park

✉ bumjin.kaist.ac.kr

Seoul, Republic of Korea

🏠 leo-bpark.github.io | [in leo-bpark](https://www.linkedin.com/in/leo-bpark) | [@ bumjini](https://www.instagram.com/bumjini)

RESEARCH



I study the mind of AI through rigorous analysis of **neural representations** — the computational brain of artificial intelligence. My research focuses on **Interpretability, Cognitive Architectures, and Neural Representation**, exploring how internal structures of AI models reveal the underlying principles of cognition shared between humans and machines.




Ultimately, I aim to develop neural reasoning systems that **integrate** human and computational forms of reasoning, advancing toward a unified form of **general intelligence**.

EDUCATION

- **KAIST (Korea Advanced Institute of Science and Technology)** Sep. 2023 – Present
Ph.D. Student in Artificial Intelligence
 - Advisor: Prof. [Jaesik Choi](#)
 - Topic: Integrating Cognitive Architectures into Large Language Models
- **KAIST (Korea Advanced Institute of Science and Technology)** Aug. 2023
M.S. in Artificial Intelligence
 - GPA: 4.17 / 4.3
 - Advisor: Prof. [Jaesik Choi](#)
 - Thesis: Partitioned Channel Gradient for Reliable Saliency Map in Image Classification
- **Chung-Ang University** Aug. 2020
B.S. in Mathematics (Double Major in Software Engineering)
 - GPA: 4.39 / 4.5

PUBLICATIONS

- **Deontological Keyword Bias: The Impact of Modal Expressions on Normative Judgments of Language Models** ACL Main, 2025
Bumjin Park, Jinsil Lee, Jaesik Choi, Annual Meeting of the Association for Computational Linguistics, 2025 [Paper](#) | [Project](#)
- **Memorizing Documents with Guidance in Large Language Models** IJCAI, 2024
Bumjin Park, Jaesik Choi, International Joint Conference on Artificial Intelligence, 2024 [Paper](#) | [Project](#)
- **Identifying the Source of Generation for Large Language Models** ICPRAI 2024
Bumjin Park, Jaesik Choi, Pattern Recognition and Artificial Intelligence. ICPRAI, 2024 [Paper](#) | [Project](#)
- **Message Action Adapter Framework in Multi-Agent Reinforcement Learning** Applied Sciences, 2025
Bumjin Park, Jaesik Choi, Applied Sciences, 2024 [Paper](#)

- **Cooperative Multi-Robot Task Allocation with Reinforcement Learning** *Applied Sciences, 2022*
Bumjin Park, Cheongwoong Kang, Jaesik Choi, Sensors, 2024  Paper
- **Scheduling PID Attitude and Position Control Frequencies for Time-Optimal Quadrotor Waypoint Tracking under Unknown External Disturbances** *Sensors, 2021*
Cheongwoong Kang, Bumjin Park, Jaesik Choi  Paper
- **Generating Multi-Agent Patrol Areas by Reinforcement Learning** *ICCAS / IEEE, 2021*
Bumjin Park, Cheongwoong Kang, Jaesik Choi, International Conference on Control, Automation and Systems (ICCAS)  Paper

PROJECTS


- **NYU Global AI Frontier Lab — Mechanistic Interpretability Research** *Aug 2024 – Oct 2024*
Tools: Python, PyTorch, Transformer Models (Llama, Gemma), Activation Patching, LoRA
 - Mechanistic interpretability of Large Language Models to investigate bias representations.
 - NYU Global AI Frontier Lab is co-directed by Prof. Yann LeCun and Prof. Kyunghyun Cho.
- **ADD (Agency for Defense Development) — Unmanned Swarm CPS Research Lab** *Oct 2021 – Mar 2025*
Tools: ROS, Gazebo, Webots, PyTorch, Python, UAV/UGV Simulation
 - Developed multi-agent reinforcement learning algorithms for patrol and communication tasks in unmanned swarms.
 - Implemented Sim-to-Real transfer using domain adaptation to bridge physical and simulated environments.
 - Built ROS-based communication pipelines between Gazebo and Webots for UAVs (DJI) and UGVs (Husarion Rosbot) .
 - Achieved 2 journal, 1 conference, 1 domestic journal, and 4 domestic conference publications; filed 2 patents (1 registered).
- **Kolmar — AlchemyNet & Domain Knowledge (Phase 2)** *Nov 2024 – May 2025*
Tools: Python, PyTorch, Transformer, Node.js
 - Advanced Phase 2 of AlchemyNet by incorporating domain knowledge into cosmetic property prediction.
 - Built a web-based AI service for cosmetic composition design and property inference.
 - Submitted 1 journal and 1 conference paper (under review).
- **Kolmar — AI for Cosmetic Composition (Phase 1)** *Aug 2023 – Mar 2024*
Tools: Transformer, Python, Scikit-learn, Pandas, Visualization Libraries
 - Developed AlchemyNet, a neural model inspired by the concept of an alchemist, to encode cosmetic formulations.
 - Predicted multiple physicochemical properties such as viscosity, pH, density, and hardness.
 - Designed explainable embedding spaces linking chemical composition with sensory attributes.
- **X-Ray Object Detection and Saliency** *Aug 2024*
Tools: Python, PyTorch, Grad-CAM, Explainable AI (XAI)
 - Applied attribution-based XAI methods to improve interpretability of X-ray object detection models.
 - Analyzed decision-making in overlapping object cases to enhance model transparency.

EXPERIENCES

- **Visiting Research Scholar — NYU Global AI Frontier Lab**
Mechanistic Interpretability, Bias in LLMs


 - Conducted research at New York University's Global AI Frontier Lab (co-directed by Prof. Yann LeCun and Prof. Kyunghyun Cho).
 - Investigated bias representation in large language models through neuron- and circuit-level analysis.
 - Collaborated with researchers on interpretability frameworks for scalable model analysis.

Aug 2024 – Oct 2024


- **Teaching Assistant — Deep Learning Course (KAIST AI)**
Tools: PyTorch, Colab, Git, Python


 - Assisted in teaching and grading for the graduate-level Deep Learning course at KAIST AI.
 - Provided guidance to students on neural network architectures, optimization, and experimental design.
 - Led tutorial sessions and supported lab assignment implementation.

Sep 2024 – Jun 2025


- **Lab Representative — SAILAB, KAIST AI**
Focus: Research Organization and Student Coordination

 - Served as Ph.D. student representative for the SAILAB research group.
 - Organized internal seminars and managed communication between students and faculty.
 - Coordinated collaboration initiatives and lab activities.

Sep 2023 – Aug 2024


- **GPU Server Management — SAILAB, KAIST AI**
Tools: Linux, Docker, Git

 - Managed GPU servers and computing resources for SAILAB's research infrastructure.
 - Optimized resource allocation and maintained high-performance computing environments.
 - Automated server monitoring and scheduling with Slurm and Docker containers.

Jan 2022 – Dec 2023

