

Woochang Sim

 whatchang |  woo-chang-sim |  woochang@gm.gist.ac.kr

RESEARCH INTERESTS

I study **abstract reasoning in language models**, guided by the principle that AI capabilities are functions of architecture, scale, data, and optimization. My research has progressed from (i) **diagnosing reasoning gaps** through cognitive science-based evaluation frameworks (published in TIST), to (ii) **data-driven interventions** through synthetic datasets encoding human-intuitive analogies (GIFARC), and now to (iii) **adaptive reasoning**, leveraging prior knowledge acquired during pretraining to generalize to genuinely unseen tasks. I currently approach this through Bayes-Adaptive MDPs and meta-reinforcement learning, testing these ideas on the ARC-AGI-3 benchmark. Building on this, I ultimately aim to extend rapid adaptation toward continual learning, where models accumulate and refine knowledge over time rather than learning statically.

EDUCATION

2025 - Present **Ph.D. in AI Convergence**, GIST (Advisor: Sundong Kim)
2023 - 2025 **M.S. in AI Graduate School**, GIST (Advisor: Sundong Kim)
2016 - 2023 **B.S. in Computer Engineering**, Chosun University

RESEARCH EXPERIENCE

GIFARC: Synthetic Dataset for Human-Intuitive Analogies

Apr 2025 - Present

first Author, Arrive

- **Created GIFARC dataset:** 10K ARC-style tasks across 20 analogy categories (e.g., Rotational Symmetry, Wave & Diagonal Flow, Kaleidoscope & Symmetry Expansion) extracted from real-world dynamic GIF scenes
- **Developed three-stage pipeline:** Visual abstraction extraction using o4-mini → ARC-style task concept and description design → Python code generation and validation
- **Conducted systematic experiments:** (i) In-context learning with GPT-4.1-mini evaluating three context conditions (full description, description without analogy, no solution) and (ii) Supervised fine-tuning with Mistral-Nemo-Minitron-8B-Base using LoRA (4 epochs) on 1K sampled tasks
- **Key finding:** Models trained with analogical context showed clear preference for high-level reasoning over pixel-level operations, achieving lower negative log probability on ARC-AGI-2 benchmark (more difficult than ARC-AGI-1)
- **Team leadership:** Managed task distribution, coordinated weekly meetings with agenda preparation, and facilitated consensus-building among collaborators

Reasoning Abilities of LLMs on Abstraction and Reasoning Corpus [Code]

Nov 2023 - Jan 2025

Co-first Author, Published in ACM TIST

- **Operationalized LoTH-based evaluation framework** into Python-based experimental settings for three reasoning properties: Logical Coherence, Compositionality, and Productivity
- **Led implementation** of Logical Coherence experiments (CoT, Least to Most, ToT prompting) and Compositionality environment (action-based setting inspired by RAP paper; custom GUI for human baseline) on GPT-4-turbo and GPT-4o
- **Key finding:** Revealed significant performance gaps between state-of-the-art LLMs and human-level reasoning—Logical Coherence: 10.2% model vs. 4.0% with valid reasoning; Compositionality: 14% model vs. 86% human accuracy; Productivity: 17.12% valid instances generated

TECHNICAL SKILLS

Languages	Python, Korean (Native), English (Intermediate)
Frameworks	PyTorch, Hugging Face Transformers, Git, Weights & Biases, Notion
Specializations	LLM Fine-tuning, Prompt Engineering, Benchmark Design, Inter-annotator Agreement Analysis, Statistical Evaluation
Research Tools	Experimental Design, Ablation Studies, Human Evaluation Protocols, Data Pipeline Development

LEADERSHIP & SERVICE

Student Representative, AI Convergence Department, GIST

Jan - Dec 2025

- Organized sports day and welcome reception for incoming graduate students
- Facilitate communication between students and faculty members on department initiatives

Teaching Assistant, Machine Learning & Deep Learning Course, GIST

Sep - Dec 2025

- Grade assignments on bias-variance tradeoffs, statistical learning theory, and core ML concepts with detailed feedback
- Taught fundamental deep learning concepts covering NN, CNN, and RNN architectures

PROJECTS

Open-Domain Question Answering System [\[Code\]](#)

Oct - Nov 2021

- Implemented hybrid retrieval system combining sparse methods (BM25, ElasticSearch) and dense retrieval (DPR) over 60K Wikipedia documents
- Fine-tuned klue/roberta-large reader model with 5-fold cross-validation, applying data preprocessing, context concatenation strategies, and question masking techniques
- Achieved 68.89 Exact Match (EM) and 78.69 F1 score on Korean QA benchmark (6th place out of 19 teams)

Relation Extraction with Typed Entity Markers [\[Code\]](#)

Sep - Oct 2021

- Implemented state-of-the-art typed entity marker (punct) method for sentence-level relation extraction, marking entities with special tokens (@, #, *, ^)
- Applied masked language model (MLM) pre-training on task data with optimal configuration (40% no-relation data, 2 epochs) and custom additional layers on RoBERTa-large
- Achieved 72.6 F1 score through systematic experimentation with different pre-training strategies and model architectures

K-pop Relation Extraction Dataset Construction

Nov 2021

- Built annotated corpus from Wikipedia K-pop raw text with 2,468 instances across 10 relation classes (e.g., Singer:Work, Singer:Member, Work:Track, Singer:Career)
- Designed comprehensive annotation guidelines and relation mapping templates, conducting pilot tagging and iterative refinement
- Achieved high inter-annotator agreement (Fleiss' Kappa of 0.842) through careful guideline development and multiple discussion rounds

AWARDS

Mar 2024 The Chungshim Fellowship

Jul 2023 Best Presentation Award, Korea Computer Congress 2023

PUBLICATIONS

Lee, Seungpil*, **Sim, Woochang***, Donghyeon* Shin, Wongyu Seo, Jiwon Park, Seokki Lee, Sanha Hwang, Sejin Kim, and Sundong Kim (Jan. 2025). "Reasoning Abilities of Large Language Models: In-Depth Analysis on the Abstraction and Reasoning Corpus". In: *ACM Trans. Intell. Syst. Technol.* Just Accepted. ISSN: 2157-6904. DOI: [10.1145/3712701](https://doi.org/10.1145/3712701). URL: <https://doi.org/10.1145/3712701>.

Sim, Woochang*, Hyunseok* Ryu, Kyungmin* Choi, Sungwon Han, and Sundong Kim (May 2025). "GIFARC: Synthetic Dataset for Leveraging Human-Intuitive Analogies to Elevate AI Reasoning". In: *arXiv preprint arXiv:2505.20672*. Under review.

Lee, Jihwan*, **Sim, Woochang***, Sejin Kim, and Sundong Kim (2024). "Enhancing Analogical Reasoning in the Abstraction and Reasoning Corpus via Model-Based RL". In: *IJCAI Workshop on the Interactions between Analogical Reasoning and Machine Learning*.

Sim, Woochang, Hyebin Jin, Sejin Kim, and Sundong Kim (Feb. 2024). "The Possibility of Prompt Engineering for ARC Problem Solving". In: *The Korean Institute of Information Scientists and Engineers* 30.2, pp. 63–69. URL: <https://www.kci.go.kr/kciportal/ci/sereArticleSearch/ciSereArtiView.kci?sereArticleSearchBean.artiId=ART003050661>.

Gu, Gyojoon, **Sim, Woochang**, Jaegyun Im, Sejin Kim, and Sundong Kim (Dec. 2023). "Using Contrastive Learning for Abstraction and Reasoning Task". In: *Korea Software Congress*, pp. 828–830. URL: <https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE11705255>.

Seo, Wongyu, **Sim, Woochang**, and Sundong Kim (Dec. 2023). "Augmenting Few-Shot Demonstrations with Large Language Model". In: *Korea Software Congress*, pp. 341–343. URL: <https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE11705093>.