David (Dowon) Baek

(+1) 617-583-2851 | dbaek@mit.edu | david-baek | dbaek-ai

Education

Massachusetts Institute of Technology (MIT) — Cambridge, MA, USA
M.S. in Electrical Engineering & Computer Science (EECS), GPA: 5.0/5.0 — *Sep 2023 – May 2026 (Expected)*
- Advisor: Max Tegmark
- Research Area: LLM Interpretability, Representation Learning, AI Safety

Seoul National University (SNU) — Seoul, Korea
B.S. in Physics and Computer Science, Summa Cum Laude, GPA: 4.23/4.3 — *Mar 2017 – Aug 2023*
- Presidential Award (Ranked **1st** among graduating cohort in College of Natural Sciences)
- Includes two years on leave for compulsory military service (2020–21, Job: Cyber Security Specialist)

Publications

1. D. Baek, Y. Li, M. Tegmark, "Generalization from Starvation: Representations in LLM Knowledge Graph Learning," manuscript in preparation for submission to ICLR 2025.
2. D. Baek, Z. Liu, M. Tegmark, "GenEFT: Understanding Statics and Dynamics of Model Generalization via Effective Theory," *ICLR 2024 Workshop on Bridging the Gap Between Practice and Theory in Deep Learning*, [arXiv].
3. S. H. Park, D. Baek, I. Park, S. Hahn, "Design of Scalable Superconducting Quantum Circuits using Flip-chip Assembly," *IEEE Transactions on Applied Superconductivity*, 33(5), pp.1-6, 2023 [Link].

Experience

Tegmark AI Safety Group — Dec 2023 - Present
Graduate Research Assistant (Advisor: Prof. Max Tegmark) — *Cambridge, MA, USA*
- Studied geometrical structure of knowledge representations in Large Language Models (LLMs), with experience in fine-tuning LLMs and Sparse Autoencoders (SAEs) using PyTorch and Transformers package
- Proposed and empirically verified physics-inspired effective theory of neural network generalization

Applied Superconductivity Laboratory — Feb 2022 – Feb 2023
Undergraduate Research Assistant (Advisor: Prof. Seungyong Hahn) — *Seoul, Korea*
- Studied neural network-based control pulse optimization and geometry optimization strategies for superconducting qubits, utilizing FEM simulations and Python.

Projects

Filtering Data to Improve NeRF for Driving Scenes — Mar 2024 - May 2024
Class Project, Advances in Computer Vision — *Cambridge, MA, USA*
- Proposed and evaluated a filter-based data-preprocessing strategy for NeRF that enables autonomous driving scene reconstruction with reduced compute time and memory, with the minimal performance loss possible

Automated System for Effectively Managing Leaves of Soldiers — Sep 2020 - Nov 2020
2020 Open Source Online Hackathon, hosted by Korean Open Source Academy for Military. — *Seoul, Korea*
- Used Node.js, Express, MongoDB, Passport.js for backend and Vue/Vuetify, Chart.js for frontend

Wanderlust: Community for Hikers — Sep 2019 - Dec 2019
Class Project, Mobile Computing and its Applications — *Seoul, Korea*
- Used Flutter, Google Maps, and Google Firebase to build the application

Technical Skills

Programming: Python, C/C++, Java, Matlab, Mathematica, LaTeX, HTML, Javascript
Libraries: PyTorch, Tensorflow†, Numpy, Scipy, QuTiP, etc.

Honors & Awards (Selected)

- Silver Medal, University Physics Competition, 2018
- Finalist, Samsung Collegiate Programming Cup (SCPC), 2018
- Silver Medal, International Junior Science Olympiad (IJSO), 2014