

# David (Dowon) Baek

✉ dbaek@mit.edu | 🌐 david-baek | in dbaek-ai

## EDUCATION

### Massachusetts Institute of Technology (MIT)

Cambridge, MA, USA

*Ph.D. in Electrical Engineering & Computer Science (EECS), GPA: 5.0/5.0*

*Sep 2023 – Current*

- 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\*, E. Michaud\*, J. Engels, X. Sun, M. Tegmark, “The Geometry of Concepts: Sparse Autoencoder Feature Structure,” 2024, [arXiv](#).
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.

## HONORS & AWARDS (SELECTED)

- Silver Medal, University Physics Competition, 2018
- Finalist, Samsung Collegiate Programming Cup (SCPC), 2018
- Silver Medal, Korean Mathematical Olympiad (High School Division), 2016
- Silver Medal, International Junior Science Olympiad (IJSO), 2014

## TECHNICAL SKILLS

**Programming:** Python, C/C++, Java, Matlab, Mathematica, L<sup>A</sup>T<sub>E</sub>X, HTML, Javascript

**Libraries:** PyTorch, Tensorflow<sup>†</sup>, Numpy, Scipy, QuTiP, Vue.js/Vuetify, etc.

## COMMUNITY SERVICE

- Chair of Publicity & Communications Committee @ Ashdown House (MIT Graduate Housing) Nov 2023 - Present
- Vice President of Publicity @ MIT EECS Graduate Student Association Jan 2024 - Present