

CHENYU GAO

+1 (667) 910-5300 | chenyu.gao@vanderbilt.edu | chenyugoal.github.io | [LinkedIn](#) | [Google Scholar](#)

EDUCATION

Vanderbilt University

Ph.D. in Electrical & Computer Engineering

Nashville, TN

July 2022 – Nov 2026 (expected)

- **Research Focus:** Applying multimodal representation learning and generative models to solve challenges in medical image analysis and computer vision.

Johns Hopkins University

M.S. in Biomedical Engineering

Baltimore, MD

Aug 2020 – May 2022

Sun Yat-sen University

B.S. in Biomedical Engineering

Guangzhou, China

Aug 2016 – June 2020

EXPERIENCE

Data Science and Machine Learning Intern

insitro

South San Francisco, CA

June 2025 – Aug 2025

- Designed and implemented a multimodal framework applying natural language processing (NLP) to encode DNA sequencing data, guiding self-supervised learning (SSL) on images to extract robust features for gene discovery.
- Collaborated with research scientists to design model architecture and partnered with software engineers to use and help debug infrastructure pipelines for large-scale computing.
- First-authored a technical paper, currently pending internal review for its potential to accelerate drug discovery by identifying new gene targets.

Graduate Research Assistant

Vanderbilt University

Nashville, TN

July 2022 – Present

- Engineered a cascaded diffusion model for a privacy red-team attack, reconstructing high-fidelity 3D facial geometry from defaced MRI data to quantify and expose critical re-identification risks. [Blog]
- Designed a first-of-its-kind system for brain age estimation from multi-modal MRI to enable early detection of neurodegenerative diseases, resulting in a provisional patent. [GitHub]
- Architected a scalable data processing pipeline using Singularity, local and HPC to ingest, harmonize, and quality-assure 100,000 MRI scans from 40+ datasets for the world's largest diffusion and structural MRI database.
- Implemented a conditional GAN in PyTorch to synthetically extend the MRI field-of-view, rescuing previously incomplete scans and increasing effective dataset size by an estimated 10-15%.

Graduate Research Assistant

Johns Hopkins University

Baltimore, MD

Dec 2020 – May 2022

- Adapted a lifelong learning algorithm from vision to speech and validated its omnidirectional knowledge transfer on a spoken digit benchmark, contributing to a publication in the top-tier journal *IEEE TPAMI*. [GitHub]
- Developed and benchmarked a suite of deep learning and classical computer vision algorithms for MR image analysis, establishing performance baselines that guided subsequent research on MRI defacing.

RECENT PUBLICATIONS & INTELLECTUAL PROPERTY

Note: For a complete list, please see my [Google Scholar](#) profile. Visit my [website](#) for showcases.

- C Gao, et al. "Pitfalls of defacing whole-head MRI: re-identification risk with diffusion models and compromised research potential." *Computers in Biology and Medicine*. 2025.
- C Gao, et al. "Brain age identification from diffusion MRI synergistically predicts neurodegenerative disease." *Imaging Neuroscience*. 2025.
- **U.S. Patent Pending:** System and Method of Brain Age Identification... (Non-provisional application filed Sept. 2025)

TECHNICAL SKILLS

Languages

Python, Bash

Frameworks

PyTorch, TensorFlow, Scikit-learn, Hugging Face, Pandas, NumPy, Polars

DevOps & Cloud

Git, Docker, Singularity, HPC/Slurm, AWS (S3, EC2), Weights & Biases