Chenyu Gao

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

Education

Vanderbilt University — Nashville, TN
Ph.D. in Electrical & Computer Engineering — *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 — Baltimore, MD
M.S. in Biomedical Engineering — *Aug 2020 – May 2022*

Sun Yat-sen University — Guangzhou, China
B.S. in Biomedical Engineering — *Aug 2016 – June 2020*

Experience

Data Science and Machine Learning Intern — South San Francisco, CA
insitro — *June 2025 – Aug 2025*
- Designed and implemented a framework for multimodal self-supervised representation learning to enable the discovery of novel gene targets. First-authored a technical paper, currently pending internal review for publication.

Graduate Research Assistant — Nashville, TN
Vanderbilt University — *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]
- Architected a scalable data processing pipeline using Pandas, Singularity, and HPC to ingest, harmonize, and quality-assure over 100,000 MRI scans from 40+ datasets for the world's largest diffusion MRI database.
- Developed a state-of-the-art deep learning system for brain age estimation from multi-modal MRI to enable early detection of neurodegenerative diseases, resulting in a provisional patent. [GitHub]
- 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 — Baltimore, MD
Johns Hopkins University — *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.
- **Provisional Patent:** System and Method of Brain Age Identification... (U.S. Patent 63/701,861)

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