Sumukh Vasisht Shankar

New Haven, Connecticut, USA • (857) 206-5194 • sumukhmys1999@gmail.com • www.svs.codes

EDUCATION

Northeastern University   Boston, MA
Master of Science in Data Science – **GPA: 3.92/4.00**   May 2023
Relevant Courses: Data Management and Processing, Supervised and Unsupervised ML, Geometric Deep Learning, Natural Language Processing

National Institute of Engineering   Mysuru, India
Bachelor of Engineering in Information Science – **GPA: 9.05/10.00**   Aug 2021
Relevant Courses: Big Data Analytics, Recommender Systems, Artificial Intelligence, Graph Theory, Neural Networks

RESEARCH EXPERIENCE

Yale University   New Haven, CT
Data Scientist, Cardiovascular Data Science Lab   May 2023 – Present

- Leverage machine learning and natural language processing to advance cardiovascular data science research.
- My work encompasses diverse research projects, including Generative Adversarial Networks (GANs) for synthetic data generation, Retrieval-Augmented Generation (RAG) methods, and machine learning models for ECG image analysis.
- Led efforts to design and implement data architectures and infrastructures for several large-scale clinical studies, including DETECT-AS (a multi-center NIH-funded clinical study), the ID-SHD, PRECARDIA and DETECT-AS clinical studies.
- Developing secure, scalable platforms (web and mobile applications) for data collection, management, and analysis, which streamlined participant recruitment, data integration, and real-time monitoring across multiple study sites.
- PI: Dr. Rohan Khera

Northeastern University   Boston, MA
Research Apprentice, Geometric Deep Learning Lab   Aug 2022 – Sep 2023

- Applied Geometric Deep Learning and Equivariant Neural Networks to fluid mechanics and material sciences.
- Analyzed the symmetries inherent in fluid dynamics and optics.
- Developed novel scale, translation, and rotation symmetric equivariant neural networks that integrates physical priors, enhancing physical realism and data efficiency.
- PI: Dr. Robin Walters

Northeastern University   Boston, MA
Research Assistant, Cybersecurity and Privacy Institute   Jan 2022 – Dec 2022

- Designed, collected, and analyzed Internet measurement campaigns to illuminate data localization practices in the European Union.
- Implemented proof-of-concept models for various research papers.
- Developed machine learning models to predict the physical locations of IP addresses.
- PI: Dr. David Choffness, Advisor: Dr. Alexander Gamero Garrido

Northeastern University   Boston, MA
Research Data Scientist, Lazer Lab   Nov 2021 – May 2023

- Implemented and monitored scalable data pipelines to identify and characterize social behavior from social media data, primarily from Twitter.
- Developed and refined machine learning models to classify tweets as Misinformation or Disinformation.
- Created comprehensive Misinformation Dashboards for various events, providing real-time insights and visualizations. This work contributed to understanding the spread of mis-information on social media platforms.
- Collaborated with cross-functional teams to ensure the robustness and accuracy of the pipelines and models, enhancing the overall efficacy of the project.
- PI: Dr. David Lazer

PUBLICATIONS

1. Oikonomou, E.K., Sangha, V., **Vasisht Shankar, S.**, Coppi, A., Krumholz, H.M., Nasir, K., Miller, E.J., Gallegos Kattan, C., Al-Mallah, M.H., Al-Kindi, S. and Khera, R., 2025. Artificial intelligence-enabled electrocardiography and echocardiography to track preclinical progression of transthyretin amyloid cardiomyopathy. *European Heart Journal*, p.ehaf450.
2. **Vasisht Shankar, S.**, Dhingra, L.S., Aminorroaya, A., Adejumo, P., Nadkarni, G.N., Xu, H., Brandt, C., Oikonomou, E.K., Pedroso, A.F. and Khera, R., 2025. Automated transformation of unstructured cardiovascular diagnostic reports into structured datasets using sequentially deployed large language models. *European Heart Journal-Digital Health*, p.ztaf030.
3. Aminorroaya, A., Dhingra, L.S., Pedroso, A.F., **Vasisht Shankar, S.**, Coppi, A., Khunte, A., Foppa, M., Brant, L.C., Barreto, S.M., Ribeiro, A.L.P. and Krumholz, H.M., 2025. Development and multinational validation of an ensemble deep learning algorithm for detecting and predicting structural heart disease using noisy single-lead electrocardiograms. *European Heart Journal-Digital Health*, p.ztaf034.
4. Dhingra, L.S., Aminorroaya, A., Sangha, V., Pedroso, A.F., **Vasisht Shankar, S.**, Coppi, A., Foppa, M., Brant, L.C., Barreto, S.M., Ribeiro, A.L.P. and Krumholz, H.M., 2025. Ensemble deep learning algorithm for structural heart disease screening using electrocardiographic images: PRESENT SHD. *Journal of the American College of Cardiology*, *85*(12), pp.1302-1313.
5. **Vasisht Shankar, S.***, Wang, R.*, D'Souza, D., Singer, J.P. and Walters, R. 2024. "Equivariant Neural Networks for Controlling Dynamic Spatial Light Modulators." Integrating Materials and Manufacturing Innovation, November. https://doi.org/10.1007/s40192-024-00383-1.
6. Aminorroaya, A., Dhingra, L.S., Oikonomou, E.K., Saadatagah, S., Thangaraj, P., **Vasisht Shankar, S**., Spatz, E.S. and Khera, R., 2024. Development and multinational validation of an algorithmic strategy for high Lp (a) screening. Nature Cardiovascular Research, *3*(5), pp.558-566.
7. **Vasisht Shankar, S**., D'Souza, D., Singer, J.P. and Walters, R., Controlling Dynamic Spatial Light Modulators using Equivariant Neural Networks. In Workshop on "Machine Learning for Materials" ICLR 2023.

PRE-PRINTS

1. Thangaraj, P.M.*, **Vasisht Shankar, S**.*, Huang, S., Nadkarni, G., Mortazavi, B., Oikonomou, E.K. and Khera, R., 2024. A Novel Digital Twin Strategy to Examine the Implications of Randomized Control Trials for Real-World Populations. medRxiv. (*Under Review – Nature Cardiovascular Research*)
2. Khunte, A., Sangha, V., Oikonomou, E.K., Dhingra, L.S., Aminorroaya, A., Coppi, A., **Vasisht Shankar, S**., Mortazavi, B.J., Bhatt, D.L., Krumholz, H.M. and Nadkarni, G., 2024. Automated diagnostic reports from images of electrocardiograms at the point-of-care. medRxiv, pp.2024-02. (*Under Review - Nature*)
3. Oikonomou, E.K., Craig, N.J., Holste, G.I., **Vasisht Shankar, S.**, White, A., Mahendran, M., Newby, D.E., Dweck, M.R. and Khera, R., 2025. Artificial intelligence-enabled echocardiography as a surrogate for multi-modality aortic stenosis imaging: post-hoc analysis of a clinical trial. *medRxiv*, pp.2025-03.

4. Adejumo, P., Thangaraj, P.M., **Vasisht Shankar, S**., Dhingra, L.S., Aminorroaya, A. and Khera, R., 2024. Retrieval-Augmented Generation for Extracting CHA2DS2VASc Features from Unstructured Clinical Notes in Patients with Atrial Fibrillation. medRxiv, pp.2024-09.
5. Gamero-Garrido, A., Yu, K., **Vasisht Shankar, S.**, Singh, S.K., Balasubramanian, S., Wilcox, A. and Choffnes, D., 2025. Empirically Measuring Data Localization in the EU. *arXiv preprint arXiv:2504.09019*.
6. Aminorroaya, A., Dhingra, L.S., Sangha, V., Oikonomou, E.K., Khunte, A., **Vasisht Shankar, S.**, Camargos, A.P., Haynes, N.A., Hofer, I., Ouyang, D. and Nadkarni, G.N., 2023. Deep learning-enabled detection of aortic stenosis from noisy single lead electrocardiograms. medRxiv, pp.2023-09.
7. **Vasisht Shankar, S.**, Oikonomou, E.K. and Khera, R., 2023. CarDS-Plus ECG Platform: Development and Feasibility Evaluation of a Multiplatform Artificial Intelligence Toolkit for Portable and Wearable Device Electrocardiograms. medRxiv.

*Equal Contribution

PRESENTATIONS

1. A Novel Application of Open-Source Large Language Models to Detech Visible and Hidden Clinical Features in 12-Lead Electrocardiograms, ACC Scientific Sessions, Chicago, Illinois, March 2025; **Poster Presentation**
2. Using LLMs to extract tabular EHR data of heart failure cohorts across multicenter, non-interoperable health systems, AHA Scientific Sessions, Chicago, Illinois, November 2024; **Moderated Poster Presentation**
3. Digital Twins for Cardiovascular Care, IEEE BHI 2024, Houston, Texas, November 2024; **Invited Speaker**
4. Generative AI in Healthcare, Yale Healthcare Data Science Workshop, New Haven, Connecticut, September 2024; **Invited Speaker**
5. Equivariant Neural Networks for Controlling Dynamic Spatial Light Modulators. TMS AI in Materials International Conference, Cleveland, Ohio, June 2024; **Oral Presentation**
6. CarDSPlus ECG Platform. AI in Medicine Symposium, Yale University, February 2024; **Poster Presentation**
7. A Novel Digital Twin Strategy to Examine the Implications of Randomized Control Trials for Real-World Populations. ML for Health Conference, New Orleans, Louisiana, December 2023; **Poster Presentation**
8. Controlling Dynamic Spatial Light Modulators using Equivariant Neural Networks. ML4Materials Workshop, The International Conference on Learning Representations (ICLR) (Remote); **Poster Presentation**
9. Controlling Dynamic Spatial Light Modulators using Equivariant Neural Networks. Khoury Research Apprenticeship Showcase, Boston, MA; **Oral Presentation**

HONORS & AWARDS

1. Graduate Student Award for Significant Research Impact, Khoury College of CS, 2023
2. Global Nomination and Recognition, NASA SpaceApps Challenge, 2020
3. Mysuru Physics Lecturers Association Award for Excellence in Physics, 2017
4. All India Rank 1 in RIE Physics and Mathematics Stream Exam, 2017

GRANTS & FELLOWSHIPS

1. ALHI Early Career Travel Grant, ML4H Conference, 2023
2. Khoury Research Apprenticeship Fellowship ($14,000), 2022

TEACHIING EXPERIENCE

National Institute of Engineering   Mysuru, India
Teaching Assistant, Dept. of Information Science   Sep 2019 – Dec 2019

- Assisted Prof. CK Vanamala in lectures for Data Analytics and conducted hands on workshops for building recommender systems to a cohort of 75 students.

NIE IEEE Summer of Code   Mysuru, India
Lead Resource Person   Oct 2019
- Resource person for a 1-day workshop for a technical fest - Rubix 2019. Taught the ways of game development using Python and helped the attendees develop a total of six games.

NIE IEEE Student Branch   Mysuru, India
Resource Person, Game Development in Python   May 2019
- Taught Progressive Web Application development and Data Mining to a cohort of over 50 students over a course 7 sessions.

LEADERSHIP & OUTREACH

1. Co-organizer at Healthcare Data Science Workshop 2024, Yale University   Sep 2024
2. Career Peer Advisor, Northeastern University   Sep 2023 - May 2024
3. Co-Organizer, Biblus 2020, National Institute of Engineering   Feb 2020
4. Co-Organizer, NISB IEEE Women in Engineering 2019   Mar 2019
5. Student Activity Committee Coordinator, NISB   May 2019 – May 2020
6. Student Branch Networking Coordinator, NISB   May 2018 – May 2019

PROFESSIONAL EXPERIENCE

Walmart Inc.   Dallas, TX
Data Engineer III Intern   May 2022 – Aug 2022
- Worked on an initiative to generate real-time reports on Walmart+ customer data, focusing on store signups.
- Built near real-time data streaming pipelines using Apache Spark and Apache Kafka to stream petabytes of click data into Google Cloud Storage.
- Developed comprehensive reports and interactive dashboards using Looker, providing actionable insights, and enhancing decision-making processes for the marketing and operations teams.
- Improved data accessibility and reporting efficiency, allowing for timely analysis and response to customer behavior trends.

FICO   Bangalore, India
Software Engineering Intern   Jan 2021 – Jun 2021
- Implemented and integrated a Regular Expression Validation feature for a Mortgage Decision Management Platform tool, enhancing data accuracy and validation efficiency.
- Developed and deployed multiple APIs to facilitate the conversion of data into various formats for optimal storage across different database systems.
- Successfully increased the project's code coverage by over 11\% through comprehensive regression testing, improving the reliability and maintainability of the codebase. This work significantly contributed to the robustness and functionality of the platform.

Antriksh Labs Pvt. Ltd.   Mysuru, India
AI Researcher   Jan 2020 – Dec 2020
- Initiated active methodological research practices to identify optimal visualization techniques for both structured and unstructured data.
- Conducted qualitative research and developed intelligent data pipelines for an AutoML SaaS product.
- Leveraged distributed systems such as Spark and Hadoop to accelerate data processing and statistical analysis, significantly enhancing the efficiency and effectiveness of data workflows.

- This work contributed to advancing the capabilities of the AutoML platform and improving its user experience.

TECHNICAL SKILLS

Languages: Python, R, SQL, C++, HTML, CSS, JavaScript, Swift
Database: MySQL, PostgreSQL, Google Firebase, Mongo DB
Cloud: Google, AWS, Azure
Data Science: Supervised and Unsupervised ML, Deep Learning, Natural Language Processing, Retrieval Augmented Generation (Tensorflow and PyTorch)
Frameworks: Angular, React, Node, Vue
Data Engineering Tools: Spark, Hadoop, Kafka, Zookeeper, Looker