Gayathri Ethiraj

Madison, WI | +1 (608)421-8827 | gethiraj@wisc.edu | https://gayu-thri.github.io/ | linkedin.com/in/gethira

EDUCATION

University of Wisconsin–Madison – *WI, USA*
Masters in Data Science | Expected Dec 2026
Relevant Coursework: Introduction to Algorithms, Statistical Models for Data Science, Statistical Methods for Data Science

Amrita Vishwa Vidyapeetham University – *Coimbatore, IN*
Bachelors in Computer Science & Engineering | June 2021
GPA: 3.75/4.0 (*Distinction*)

SKILLS

- **Languages** : Proficient: Python, SQL, MySQL | Familiar: Java, C, C++, HTML/CSS
- **Technical** : Data Engineering, NLP, LLM, ASR, ITN; RESTful APIs, HTTP(s), Linux, Shell Scripting (Bash), CI/CD, CCNA(R&S)
- **Tools** : HuggingFace APIs, PyTorch, Tensorboard, Neptune, NumPy, Pandas, Git, MATLAB, Scikit-Learn, Streamlit, Gradio, Flask, Postman, OAuth, Label Studio, Youtube-dl, DVC, WebDataset, Poetry, FFmpeg, Soundfile, Pynini, OpenFst, Conda, NFS, Slurm, Pytest, Selenium, Google Cloud Storage (GCS), Google ASR
- **Soft Skills** : Team collaboration, Interpersonal communication, Research, Mentoring

WORK EXPERIENCE

Zoho Corporation, Chennai, India

AI/Data Engineer, LLM, R&D – *HuggingFace APIs, Data Engineering, LLM* Jan 2024–July 2025

- Performed large-scale data categorization using Lilac clustering; Benchmarked scalability, identifying high time costs and false-positive rates. Developed a rate-limited Google Classification API pipeline to produce reliable data–category pairs for downstream model training.
- Built a Streamlit app using HuggingFace APIs to load data, tokenize, filter, ingest meta, generate recipes and perform regex-based search.
- Researched various open-source data tools for LLMs. Generated 100s of billions of tokens of Indic data by translating open-source English datasets using an internal CPU-based translation model, optimized with multi-threading for faster processing.

AI/Data Engineer, ASR, R&D – *ITN, Pynini, PyTorch, RESTful APIs, DVC, Flask* July 2021–Dec 2023

- Integrated a rule-based ITN system into ASR post-processing using NVIDIA NeMo and added 4 custom grammar rules. Compiled into FAR, reducing processing time by **140×** and memory use by **80%**. Led a **4**-member annotation team for ITN data collection.
- Evaluated F1 scores of existing punct-cap models and co-developed a transformer-encoder replacement using PyTorch Lightning.
- Generated transcriptions using Whisper to fine-tune an internal ASR model. Built & deployed a Flask inference demo as a Linux service.
- Preprocessed ASR datasets, boosting training data by **16%**. Enhanced ASR functionality by implementing an audio I/O module.
- Built internal tools (*pystratus*, *dvc-stratus*) for dataset and model management, enabling secure cross-team adoption with OAuth and resource policies. Maintained ZWAF for OneAuth token validation in cross-team Zoho API communication.
- Integrated ZWAF into FastAPI-based ASR web server middleware, securing cross-team usage.

Project Trainee (Internship), ASR, R&D – *Python, PyTorch Dataset, Pytest, Streamlit* Jan-June 2021

- Collected and preprocessed datasets for ASR model, assisting peer ML Engineers with data requirements, increased benchmark data suite by **83%** (400 hours of audio). Added PyTorch Iterable Dataset classes for each dataset and unit tests using pytest.
- Processed open-source datasets using youtube-dl for YouTube data, Google ASR for synthetic transcriptions and developed a Streamlit tool for audio recording. Organized team sessions to create a limited benchmark with real-time recordings.

ONGC, Chennai, India

Intern – *CSS, HTML, XAMPP, MySQL* | May 2019

- Built a web-based Intranet Issue Tracker on ONGC Intranet for RCC users to create/retrieve issues, approve/deny solutions, and track resolution status. Collaborated with software engineers to deliver a real-world product.

PROJECTS

Gesture-based game control using a low-resolution camera – *Deep Learning, TensorFlow, Keras, ConvLSTM* July-Dec 2020

- Developed a gesture-based control system using a 3D ConvLSTM model with efficient pre-processing techniques to replace keyboard inputs in games, achieving 80% accuracy with minimal latency.

Web app for College Admission Management System – *Python, Flask, Firebase, Pyrebase, Bulma* | Mar-May 2020

- Developed an admin-user application where admins can add colleges, set cut-offs, and register students. Users can log in, submit admission applications, and receive status notifications. Deployed using Google's Firebase Hosting.

Network configuration and troubleshooting – *Packet Tracer, CCNA Lab Pods* | Jun 2018-Jun 2019

- Configured and troubleshot complex network topologies on Packet Tracer and real equipment during CCNA lab training, gaining strong understanding of configuration, operation, and maintenance of Cisco routing and switching devices.