Gayathri Ethiraj

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

SUMMARY

AI/ML engineer with **4+** years of R&D work experience in **Large Language Models (LLM) and Automatic Speech Recognition (ASR)** at Zoho. Pursuing MS in Data Science at UW–Madison; seeking **Summer 2026 internships** in machine learning, deep learning, or related fields.

EDUCATION

University of Wisconsin–Madison, WI, USA – MS in Data Science, Expected Dec 2026

Amrita Vishwa Vidyapeetham University, Coimbatore, IN – Bachelor's 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** : PyTorch, Git, TensorBoard, Neptune, NumPy, pandas, Matlab, Scikit-learn, Gradio, Flask, Postman, Label Studio, DVC, WebDataset, Poetry, FFmpeg, OpenFst, Conda, NFS, Slurm, pytest, Selenium, soundfile, Google Cloud Storage (GCS)
- **Soft skills** : Team collaboration, Interpersonal communication, Research, Mentoring

WORK EXPERIENCE

Zoho Corporation, Chennai, India

AI/Data Engineer, LLM, R&D Jan 2024–July 2025

- Modified Lilac's open-source **clustering** pipeline, replacing OpenAI with open-weights LLMs (Llama, Mistral 7B); iterated on prompt engineering for custom cluster labeling; benchmarked clustering quality, scalability, identifying high time costs & false-positives.
- Developed a **rate-limited Google Classification API pipeline** to produce reliable data–category pairs for downstream model training.
- Built a Streamlit app using **Hugging Face APIs** to load data, tokenize, filter, ingest meta, generate recipes & 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.

ML/Deep Learning Engineer, ASR, R&D July 2021–Dec 2023

- Integrated a **FST based Inverse Text Normalization (ITN)** system into ASR post-processing using **NVIDIA NeMo** and added 4 custom grammar rules using **Pynini**. Compiled into FAR, reducing processing time by **140×** and memory use by **80%**.
- Generated **5M+** synthetic spoken-written pairs for capitalization using **OpenAI APIs** and led **4**-member team for ITN data annotation.
- Benchmarked F1 scores for internal punct-cap models; Refactored & retrained punct-cap **transformer-encoder** in **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 with a **torchaudio** based 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, ASR, R&D 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 – *Firebase, Pyrebase, Bulma* | Mar–May 2020

- Developed a Flask-based 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 & troubleshot network topologies on Packet Tracer and real equipment with hands-on Cisco routing & switching experience.