

# Thennal D K

📍 Kottayam, IN    ✉ thennal10@gmail.com    🌐 thennal10.github.io    📄 thennal10

## Education

---

**Indian Institute of Information Technology Kottayam** Dec 2021 – May 2025  
*Bachelor in Computer Science (B.Tech with Honours)* CGPA: 9.11

## Experience

---

**Language Technology Group, University of Hamburg** May 2024 – Aug 2024  
*Research Intern*

- Developed an optimal few-shot fine-tuning regime for topic modeling.
- Devised a novel pruning procedure for LLM-based embedding models, reducing model size by 21% with negligible performance drop.
- Wrote a research paper based on findings, currently in review.
- Internship funded by the DAAD WISE scholarship.

**International Centre for Free and Open Source Software (ICFOSS)** Mar 2018 – Nov 2022  
*Research Intern*

- Created IMaSC - The ICFOSS Malayalam Speech Corpus, a 50-hour text-to-speech dataset.
- Supervised data collection, speaker recording, and quality control.
- Trained and evaluated multiple models, achieving an average MOS score of 4.50.

**Kerala Police Intelligence Department** Apr 2023 – Apr 2024  
*Machine Learning Intern*

- Trained and deployed a state-of-the-art automatic speech recognition (ASR) model for Malayalam.
- Designed and implemented a Malayalam news extraction system using web scraping and OCR.
- Developed a scalable face recognition system optimized for fast inference.
- Led a 15-person team for the task, ensuring seamless collaboration.

**Production and Maintenance Division, Institute of Human Resource Development** Dec 2022 – Mar 2023  
*Data Scientist Intern*

- Rebuilt web stack to streamline employee workflows and replace outdated systems.
- Created visualizations and reports to communicate insights to internal stakeholders and government agencies.
- Collaborated with the procurement team to optimize purchasing decisions.

## Publications

---

**Advocating Character Error Rate for Multilingual ASR Evaluation** Nov 2024  
 Thennal D K, Jesin James, Deepa P Gopinath, Muhammed Ashraf K  
 Accepted to NAACL 2025

**Large Language Models Are Overparameterized Text Encoders** Oct 2024  
 Thennal D K, Tim Fischer, Chris Biemann  
 arXiv Preprint

**Fisher Mask Nodes for Language Model Merging** May 2024  
 Thennal D K, Ganesh Nathan, Suchithra M S  
 LREC-COLING 2024

**IMaSC – ICFOSS Malayalam Speech Corpus** Nov 2022  
 Deepa P Gopinath, Thennal D K, Vrinda V Nair, Swaraj K S, Sachin G  
 arXiv Preprint

**Performance Enhancement of Deep Neural Network Based Automatic Voice Disorder Detection System with Data Augmentation – A Case Study**

Sept 2022

Thennal D K, Vrinda V Nair, R Indudharan, Deepa P Gopinath

Biomedical Engineering: Applications, Basis and Communications, Vol. 35

**Memory Based Speech Duration Model using Exemplar Theoretic Approach**

Nov 2019

Thennal D K, Deepa P Gopinath, Vrinda V Nair

International Conference on Artificial Intelligence & Speech Technologies (AIST 2019)

## Projects

---

**Model Merging for Automatic Speech Recognition**

Aug 2024 – present

- Investigating model merging techniques for ASR models fine-tuned in different languages with Dr. Manu Madhavan.
- Establishing a rigorous benchmark for model merging with clear metrics.
- Developing novel model merging methods applicable to inter-lingual merging.

**Advocating for Character Error Rate in Automatic Speech Recognition**

Aug 2024 – Oct 2024

- Documented shortcomings of the commonly used word error rate metric for multilingual evaluation with Dr. Jesin James.
- Conducted multilingual surveys collecting human preferences among different ASR models.
- Calculated metric correlations, providing experimental evidence in favor of Character Error Rate.
- Paper accepted at NAACL 2025.

**Fisher Mask Nodes for Model Merging**

July 2023 – Oct 2023

- Developed a novel and compute-efficient model merging algorithm with Dr. Suchithra M S.
- Evaluated performance on various BERT family models, achieving a performance improvement of +6.5%.
- Achieved speedups between 57.4x and 321.7x.
- Published in LREC-COLING 2024.

**Data Augmentation for Automatic Voice Disorder Detection**

May 2018 – Feb 2019

- Evaluated data augmentation techniques for automatic voice disorder detection, focusing on leukoplakia with Dr Vrinda V Nair.
- Developed a custom data augmentation strategy that increased dataset size by 8x while preserving data diversity.
- Achieved a 46.9% increase in accuracy.
- Published in a peer-reviewed journal.

## Volunteer

---

**Anannia**

Feb 2018 – present

*Communications Lead*

- Organized events, campaigns, and workshops for the queer community in Kerala.
- Developed and launched a public website, expanding international outreach.
- Served as the main liaison with donors, securing over \$3M in grants.
- Taught math and programming to marginalized children.

## Technologies

---

**Languages:** Python, HTML/CSS, JavaScript, SQL, C/C++, C#, Java

**Software:** PyTorch, TensorFlow, Scikit-learn, Vue.js, React, Unity

**Miscellaneous:** Linux, Shell (Bash/Zsh), LaTeX, Git, Docker, PostgreSQL

## Awards

---

**DAAD WISE Scholarship:** German Academic Exchange Service