

Thennal D K

📍 Hamburg, DE ✉ thennal10@gmail.com 🌐 thennal10.github.io 📄 thennal10

Education

University of Hamburg Oct 2026 – present
MSc. in Intelligent Adaptive Systems

Indian Institute of Information Technology Kottayam Dec 2021 – May 2025
B.Tech (Hons.) in Computer Science CGPA: 9.16

Experience

Language Technology Group, University of Hamburg June 2025 – present
Independent Contractor

- Implemented pipelines to automatically apply the actantial narrative model using LLMs, with paper submitted to **CHR 2026**.
- Designed LabelPigeon, a novel label projection technique utilized for low resource cross-lingual transfer, with paper accepted at **ACL 2026 Findings**.
- Quantified the creativity of LLM-generated stories via a novel narrative similarity evaluation scheme, with paper submitted to **EMNLP 2026**.

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

- Conducted a general investigation on embedding models and LLM representations, developing an optimal few-shot fine-tuning regime for topic modeling.
- Devised L3Prune, a novel pruning procedure for LLM-based embedding models reducing model size by 21% with a negligible performance drop, with paper published in **RepL4NLP 2025**.

Institute of Human Resource Development Apr 2023 – Apr 2024
Machine Learning Intern

- Worked with the Kerala Police Intelligence Department, leading a team of 15 to develop AI systems for policing.
- Trained and deployed a state-of-the-art automatic speech recognition (ASR) model for Malayalam, along with a speaker extraction and identification system.
- Designed and implemented a Malayalam news extraction system utilizing optical character recognition.
- Developed a scalable face recognition system optimized for fast inference.

Institute of Human Resource Development Dec 2022 – Mar 2023
Data Science Intern

- Rebuilt web stack to streamline employee workflows, significantly improving productivity.
- Created visualizations and reports to communicate insights to stakeholders and government agencies.
- Optimized purchasing decisions based on statistical modeling.

Publications

Do Large Language Models Always Tell The Same Stories? June 2026
Thennal D K, Hans Ole Hatzel
arXiv Preprint, Submitted to EMNLP 2026

Just Use XML: Revisiting Joint Translation and Label Projection. July 2026
Thennal D K, Chris Biemann, Hans Ole Hatzel
Findings of the Association for Computational Linguistics: ACL 2026 (pp. 34461–34478)

Large Language Models Are Overparameterized Text Encoders May 2025

Thennal D K, Tim Fischer, Chris Biemann

Proceedings of the 10th Workshop on Representation Learning for NLP (RepL4NLP-2025) (pp. 170-184)

Advocating Character Error Rate for Multilingual ASR Evaluation

Apr 2025

Thennal D K, Jesin James, Deepa P Gopinath, Muhammed Ashraf K

Findings of the Association for Computational Linguistics: NAACL 2025 (pp. 4926-4935)

Fisher Mask Nodes for Language Model Merging

May 2024

Thennal D K, Ganesh Nathan, Suchithra M S

Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024) (pp. 7349–7355)

IMaSC: A Malayalam Speech Corpus for High-Quality Text-to-Speech Synthesis

Nov 2022

Deepa P Gopinath, **Thennal D K**, Vrinda V Nair, Swaraj K S, Sachin G

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026) (pp. 5864-5872)

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

Proceedings of the International Conference on Artificial Intelligence & Speech Technologies (AIST 2019) (pp. 108–112)

Projects

Advocating for Character Error Rate in Automatic Speech Recognition

Aug 2024 – Oct 2024

- Documented the shortcomings of the commonly used word error rate metric for multilingual evaluation in collaboration with Dr. Jesin James.
- Collected human preference data in 3 languages and calculated metric correlations, providing experimental evidence in favour of Character Error Rate.
- Wrote a paper arguing our position, accepted at **NAACL 2025 Findings**.

Fisher Mask Nodes for Model Merging

July 2023 – Oct 2023

- Devised a novel and compute-efficient model merging algorithm under Dr. Suchithra M S.
- Evaluated performance of our method across various models in the BERT family, with performance improvement of **+6.5%** and a speedup between 57.4x and 321.7x.
- Documented research outcomes in an academic paper, published in **LREC-COLING 2024**.

ICFOSS Malayalam Speech Corpus

Mar 2018 – Nov 2022

- Collaborated on 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.
- Wrote a research paper compiling our results, accepted at **LREC 2026**.

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 (Overleaf/R Markdown), Git, Docker, PostgreSQL

Awards

DAAD WISE Scholarship: German Academic Exchange Service