

Sarubi Thillainathan

sarubi@lst.uni-saarland.de | sarubi.com | github.com/sarubi | [Google Scholar](https://scholar.google.com/citations?user=sarubi) | [LinkedIn](https://www.linkedin.com/in/sarubi)

PhD candidate in Computational Linguistics at Saarland University, specialising in controllable and personalised text generation and the adaptation of large language models. Hands-on experience developing, fine-tuning, and evaluating transformer-based language models, with a research focus on reinforcement learning, reward design and alignment, model merging, and rigorous model evaluation, including building evaluation frameworks and analysing model failures, and on shaping model behaviour to individual users. Strong engineering profile in PyTorch and Hugging Face, comfortable in large ML codebases and training models at scale, with experience contributing to production systems.

EDUCATION

Saarland University, Saarland Informatics Campus

Saarbrücken, Germany

PhD in Computational Linguistics

2023 – Present

Research on controllable and personalised text generation: model merging, reinforcement learning, and parameter-efficient adaptation of large language models, with reader-adaptive text complexity.

Advisors: Prof. Dr. Alexander Koller, Prof. Dr. Vera Demberg.

Dissertation: *Adapting Text Generation to Individual Users*.

University of Moratuwa

Katubedda, Sri Lanka

MSc by Research in Computer Science & Engineering

2020 – 2022

Advisors: Dr. Surangika Ranathunga, Prof. Dr. Sanath Jayasena.

Thesis: *Pre-training and fine-tuning multilingual seq-to-seq models for domain-specific low-resource NMT*.

University of Ruhuna

Matara, Sri Lanka

Bachelor of Computer Science (Hons)

2014 – 2018

Thesis: *Text Editing with Dynamic Prediction and Spell Checking Using Machine Learning for Sri Lankan Languages*.

RESEARCH EXPERIENCE

Saarland University, Language Science & Technology

Advisors: A. Koller, V. Demberg

Doctoral Researcher

2023 – Present

- Proposed **AuthorMix**, a modular authorship style-transfer framework that builds a target-style model by mixing style-specific LoRA adapters. Used reinforcement learning (GRPO) to learn fine-grained layer-wise (rather than adapter-wise) mixing weights against a designed joint style-and-meaning reward, the first model-merging approach to learn continuous, signed mixing weights with RL; ranked first on the combined style–meaning score among all baselines including GPT-5.1, and was the only method best-or-tied on every dimension in human evaluation (**EMNLP 2026**, Main Conference).
- Developed an in-context-learning method for controllable text adaptation that precisely controls fine-grained linguistic features (dependency depth, dependency length, word count, difficult words) to rewrite text to target reading levels, without fine-tuning or large parallel data (**AAAI 2025** Workshop on AI for Education).
- Designed a controlled reading study that independently varies syntactic and lexical complexity and tests reading comprehension across readers with different cognitive profiles (working-memory capacity and vocabulary size), connecting reader cognitive profiles with NLP-driven text-complexity control (**AMLaP 2026**; ongoing).
- Steering LLM generation with reinforcement learning, using behavioural signals to adapt text to individual readers (ongoing).

NLPC, University of Moratuwa

Advisors: S. Ranathunga, E.-S. A. Lee

Research Assistant

2020 – 2022

- Demonstrated the viability of fine-tuning self-supervised multilingual seq-to-seq models for extremely low-resource NMT in Sinhala, Tamil and English, including non-English-centric fine-tuning (Best Paper, MERCon 2021).
- Designed and ran all mBART-based experiments for the multilingual seq-to-seq transfer study on low-resource translation (Findings of ACL 2022).
- Conceived and led a low-resource MT project on parameter-efficient fine-tuning, building the initial experiments and codebase and mentoring two students (Lee Lab, U. of Toronto) through to publication at Findings of NAACL 2024.
- Proposed continual pre-training for domain adaptation and multi-stage fine-tuning; contributed to a production MT system for the government sector of Sri Lanka.

- Implemented API-driven Identity & Access Management features (SCIM2 provisioning).

Research Intern

2016 – 2017

- Optimised a distributed event-based streaming solution for the DEBS 2016 Grand Challenge, including JVM auto-tuning and queueing-theory performance modelling (co-authored two DZone technical articles).

TECHNICAL SKILLS

Machine Learning, Deep Learning & NLP: Neural Networks; Transformers; Large Language Models (LLMs); Reinforcement Learning (RLHF, PPO, GRPO); Generative Modelling; Representation Learning; Model Evaluation (automatic & human); Controllable text generation; user modeling & personalization; authorship style transfer; machine translation; language modeling; parameter-efficient fine-tuning (LoRA, adapters); model merging; in-context learning; pre-training & continual pre-training

Frameworks & Tools: PyTorch, Hugging Face Transformers / PEFT / TRL, scikit-learn, Weights & Biases, Git

Infrastructure: Training & fine-tuning on HPC cluster (SLURM, HTCondor); Docker; Linux

Programming: Python, Java, C, Bash

Languages: Tamil (native), English (fluent), Sinhala (fluent)

SELECTED PUBLICATIONS (full list at [Google Scholar](#))

- [1] **Sarubi Thillainathan**, Ji-Ung Lee, Michael Sullivan, Alexander Koller. “[AuthorMix: Modular Authorship Style Transfer via Layer-wise Adapter Mixing](#).” In Proceedings of the 2026 Conference on Empirical Methods in Natural Language Processing (EMNLP 2026), Main Conference.
- [2] **Sarubi Thillainathan**, Margarita Ryzhova, Alexander Koller, Vera Demberg. “Disentangling Syntactic and Lexical Complexity in Text Adaptation: Effects on Readers with Different Working Memory Capacity and Vocabulary Size.” **AMLAP 2026** (Architectures and Mechanisms for Language Processing), Saarbrücken, September 2–4, 2026.
- [3] **Sarubi Thillainathan**, Alexander Koller. “[Controllable Text Adaptation Using In-Context Learning with Linguistic Features](#).” **AAAI 2025** Workshop on AI for Education, 2025.
- [4] Tong Su, Xin Peng, **Sarubi Thillainathan**, David Guzmán, Surangika Ranathunga, En-Shiun Lee. “[Unlocking Parameter-Efficient Fine-Tuning for Low-Resource Language Translation](#).” In Findings of the Association for Computational Linguistics: **NAACL 2024**, pages 4217–4225, Mexico City, Mexico.
- [5] En-Shiun Annie Lee, **Sarubi Thillainathan**, Shravan Nayak, Surangika Ranathunga, David Ifeoluwa Adelani, Ruisi Su, Arya D. McCarthy. “[Pre-Trained Multilingual Sequence-to-Sequence Models: A Hope for Low-Resource Language Translation?](#)” In Findings of the Association for Computational Linguistics: **ACL 2022**, pages 58–67, Dublin, Ireland.
- [6] **S. Thillainathan**, S. Ranathunga, S. Jayasena. “[Fine-Tuning Self-Supervised Multilingual Sequence-To-Sequence Models for Extremely Low-Resource NMT](#).” In 2021 Moratuwa Engineering Research Conference (**MERCon**), pages 432–437, Moratuwa, Sri Lanka. **[Best Paper Award]**

AWARDS & HONORS

Best Paper Award, MERCon 2021 (Big Data / ML / Cloud Computing track)

Outstanding Excellence Award, WSO2 (2019 performance)

2nd Place, Undergraduate Research Symposium, University of Ruhuna (2018)

Merit Award, University of Ruhuna Invention & Innovation Exhibition (2017)

TALKS & TUTORIALS

- **Tutorial & Workshop:** “Pre-Trained Multilingual Sequence-to-Sequence Models for NMT: Tips, Tricks and Challenges” and “Neural Machine Translation for Low-Resource Languages: Fine-tuning Using Multilingual Sequence-to-Sequence Pre-trained Models.” Toronto Machine Learning Micro-Summit on NLP (TMLs), 2022.
- **Presentation:** “Pre-Trained Multilingual Sequence-to-Sequence Models: A Hope for Low-Resource Language Translation?” CoSubmitting Summer (CSS) Workshop at ICLR 2022.
- **Tutorial:** “Neural Machine Translation for Low-Resource Languages.” ICON 2021: 18th International Conference on Natural Language Processing.
- **Presentation:** “Multilingual Neural Machine Translation for Sinhala–Tamil–English.” 6th Annual Symposium on Natural Language Processing, University of Moratuwa, 2020.

SERVICE & MENTORING

Peer review: Information Processing & Management (journal); ACL Rolling Review (ARR); WMT 2026; LREC

Thesis supervision: W. LaCroix (MSc, 2026, *Grade-Specific Text Simplification via Weighted Adapter Merging*; with A. Koller, J.-U. Lee); N. Herath & T. Tennakoon (BSc, 2022, *Pivot-based Transfer Learning for Low-Resource NMT*; with S. Ranathunga)

Teaching: Instructor / TA, University of Moratuwa (2020–2022): OOP & Software Development, Graph Theory for Computing, Research & Report Writing, Bioinformatics; Visiting Evaluator, BIT final-year projects (2020–2021)