

Shaomu Tan

✉ shaomu.tan.edu@gmail.com 🏠 HomePage 🌐 shaomutan 🎓 G-Scholar

Professional Summary

Researcher specializing in machine translation, multilingual LLMs, reasoning systems, and reward modeling. Extensive experience in training 7B–72B parameter LLMs using distributed systems, reward modeling from human preferences, agentic evaluate-and-improve pipelines. Published 7 papers at EMNLP, ACL, and NeurIPS, and earned Kaggle medals.

- Expected to complete my Ph.D. in 2026 and currently seeking Research Scientist roles in Large Language Models training, alignment, and evaluation.
- Research Interests: Multilingual and Multimodal LLMs, Vision-Language World Models, LLM Reasoning, Evaluation and reward modeling, Reinforcement Learning from Human Feedback.

Education

University of Amsterdam, Netherlands 2022 – 2026
Ph.D. in Computer Science. Specializing in Machine Translation, mentored by [Christof Monz](#).

Utrecht University, Netherlands 2020 – 2022
Master of Science in Artificial Intelligence. Dutch GPA – 8/10 (US scale: 4.0/4.0)

Shandong University, China 2016 – 2020
Bachelor of Science in Information System.

Professional Experience

Applied Scientist Research Intern, Amazon AGI, Berlin, Germany Oct 2025 – Jan 2026

- Conducted research on test-time scaling for long-context (book-level) machine translation, bridging research and production LLM systems ([Amazon Nova](#)).
- Designed and led a large-scale human evaluation pipeline (~1M words, ~\$90K USD) to assess long-context generation quality. Research output currently under review at ACL 2026.

Research Intern, Sony (Speech & Language AI Lab), Tokyo, Japan Jun 2025 – Sep 2025

- Conducted research on Reinforcement Learning (RLVR) for Machine Translation evaluation.
- Designed and trained reasoning-based LLM translation evaluation models, a multi-role Large Reasoning Model for quality estimation and refinement. Research output currently under review at ACL 2026, [preprint](#).

Competitions

Kaggle Gold Medal (Ranking: 11/1,315 teams) — in Google’s 2023 American Sign Language Fingerspelling Recognition. I design and implement Conformer-Transformer architecture, use sign language landmarks for training, and fuse CTC with cross-entropy for learning objectives, [\[link\]](#).

Kaggle Silver Medal (Ranking: 19/950 teams) — in 2024 WSDM Cup: Multilingual Chatbot Arena Human Preference Reward Modeling. I train 7-14B reward models for multilingual human preferences, optimized with DeepSpeed, VLLM, quantization, pseudo labeling, and model merging, [\[link\]](#).

WMT23 1st place — in General Machine Translation Competition: English-Hebrew. We train small Encoder-Decoder system (0.5B) that performs on-par with GPT-4-5shot, [\[link\]](#).

Publications

Shaomu Tan, Dawei Zhu, Ke Tran, Michael Denkowski, Sony Trenous, Bill Byrne, Leonardo Ribeiro, Felix Hieber. "What Does LLM Refinement Actually Improve? A Systematic Study on Document-Level Literary Translation", **ACL 2026 Under Review**.

Shaomu Tan, Ryosuke Mitani, Ritvik Choudhary, Qiyu Wu, Toshiyuki Sekiya, Christof Monz. "Remedy-R: Generative Reasoning for Machine Translation Evaluation without Error Annotations", **ACL 2026 Under Review**, [\[link\]](#).

Shaomu Tan and Christof Monz. "Remedy: Learning Machine Translation Evaluation from Human Preferences with Reward Modeling", **EMNLP 2025 Main Conference**, [\[link\]](#).

Shaomu Tan, Di Wu, and Christof Monz. "Neuron Specialization: Leveraging intrinsic task modularity for multilingual machine translation.", **EMNLP 2024 Main Conference**, [\[link\]](#).

Di Wu, **Shaomu Tan**, and Christof Monz. "How Far Can 100 Samples Go? Unlocking Overall Zero-Shot Multilingual Translation via Tiny Multi-Parallel Data.", **ACL Findings 2024 Conference**, [\[link\]](#).

Shaomu Tan, and Christof Monz. "Towards a Better Understanding of Variations in Zero-Shot Neural Machine Translation Performance.", **EMNLP 2023 Main Conference**, [\[link\]](#).

Baohao Liao, **Shaomu Tan**, and Christof Monz. "Make Pre-trained Model Reversible: From Parameter to Memory Efficient Fine-Tuning.", **NeurIPS 2023 Main Conference**, [\[link\]](#).

Service & Teaching

Reviewing

- **2023-2025**: ACL Rolling Review (ACL, EMNLP, EACL, NAACL), Transactions on Audio, Speech, and Language Processing (TASLP), WMT, Multilingual Representation Learning (MRL).

Talks

- "Remedy-R: Large Reasoning Models for Machine Translation Evaluation" at **University of Tokyo**, invited by Prof. [Yoshimasa Tsuruoka](#), August 2025.
- "The Second Half of Machine Translation" at Nara Institute of Science and Technology (**NAIST**), invited by Prof. [Taro Watanabe](#), July 2025.

Guest Lectures

- "[Interference and Knowledge Transfer in Multilingual Translation Models](#)" at **University of Amsterdam** for AI Master course "Deep Learning for NLP", co-presented with Prof. Christof Monz, October 2024.
- "[A Journey on Multilingual Neural Machine Translation](#)" at **Utrecht University** for AI Bachelor course "Models for Language Processing", co-presented with Prof. Denis Paperno, June 2024.

Technical Skills

- **Programming related**: Python, Bash and Linux, Git, Docker, Slurm, AWS.
- **Framework**: PyTorch, HF (Transformers, TRL, PEFT), SentencePiece, Fairseq.
- **LLMs**: Verl, OpenRLHF, vLLMs, Megatron-LM, TRL, Sentencepiece, Llama-Recipes, NeMo.
- **Scaling**: DeepSpeed, FSDP, Flash-Attention, Megatron-LM.
- **Large-Scale Experience**: Distributed training and inference for up to 72B LLMs.
- **Languages**: English (full professional proficiency). Chinese (native speaker).