Isaac Thompson

Provo, UT | isaac@imthomp.com | github.com/imthomp | scholar.google.com |
www.linkedin.com/in/imthomp

EDUCATION

Doctor of Philosophy in Computer Science September 2025 – expected April 2029
Brigham Young University, Provo, UT

- 3.95 GPA
- Advisor: Dr. Eric Ringger
- Focus: Low-Resource Machine Translation and Multilingual NLP

M.S. in Computer Science (en route to Ph.D.) September 2025 – expected April 2027
Brigham Young University, Provo, UT

- 3.95 GPA
- Advisor: Dr. Eric Ringger
- Thesis focus: Machine Translation Evaluation and Metric Reliability

Bachelor of Science in Computer Science and Spanish August 2022 – April 2025
Brigham Young University, Provo, UT

- 3.89 GPA
- Minors in Mathematics, Statistics, and Translation & Localization

PUBLICATIONS

01 **AmanaMT: Translation Direction Predicts Automatic Metric Reliability** WMT 2026 @ EMNLP
FIRST AUTHOR

Published a 730,748-segment study of 10 automatic MT metrics across 33 languages and three annotation tiers. A correction in progress recomputes 725,566 segments with eight metrics and withdraws the headline direction claim: expert-MQM controls show that annotation protocol and label quality explain much of the apparent effect.

02 **Bringing Mapudungun into the Modern MT Ecosystem** AMERICASNLP 2026 @ ACL
FIRST AUTHOR

Trained and released 48 NLLB models for Mapudungun–Spanish across eight tokenization conditions, three model sizes, and both translation directions. A correction in progress restores dropped target-language tags and finds that morphology-aware tokenization does not reliably outperform Standard BPE.

RESEARCH SOFTWARE & SYSTEMS

01 **vLLM Batch Invariance**

SYSTEMS RESEARCH

Studies whether vLLM returns the same output regardless of batch neighbors and prefix-cache state, and what that guarantee costs. Merged tuned H100 batch-invariant kernels upstream and proposed a generic Hopper fallback that produced 0 differing rows in 292,020 checks while recovering 94.4% of Qwen2.5-7B's batch-32 tuning gap; measured 40.5–93.4% concurrency-32 decode gains across tested untuned models.

02 **GLM-5.2/5.3 Research Review and Model Serving**

ENGINEERING PROJECT

Serves roughly 700 GB GLM-5.2/5.3 checkpoints through an OpenAI-compatible vLLM API with eight-GPU tensor parallelism on SLURM, and orchestrates ten reviewer roles for paper review and pre-experiment red-teaming, with response validation, conditional retries, resumable execution, unit tests, and CI.

03 **Agentic MT**

RESEARCH SOFTWARE

Controlled study of when retrieval, multi-step reasoning, quality estimation, and iterative refinement help multilingual MT. Combines LaBSE/FAISS translation-memory retrieval, COMET-KIWI threshold routing, a translate, estimate, and refine loop, prompt traces, and cached outputs. Retrieval alone was the strongest condition for en→de and en→ha.

WORKING PAPERS

01 **Drawing Legs on a Snake: Meaning, Not Structure, for Chinese Characters a Pretrained MT Model Cannot Read** — IN PREPARATION

FIRST AUTHOR

Found that pretrained MT models can't represent many Chinese characters, and showed that a new embedding per missing character fixes names and ordinary words while radical-based decomposition never helps; proposed a dictionary-plus-pronunciation initialization for characters with no training data.

02 **Right Answer, Wrong Language: How Span F1 Changes Conclusions in Cross-Lingual QA Evaluation** — IN PREPARATION

FIRST AUTHOR

Multilingual needle-in-a-haystack tests score correct English answers as wrong. This artifact explains most of the non-Latin-script gap. It also manufactures cross-script transfer from fine-tuning and reshuffles MLQA rankings across 33 models in lm-evaluation-harness. We propose script-aware scoring as a drop-in fix.

03 **Reward Reliability under Machine-Translation Optimization** — IN PREPARATION

FIRST AUTHOR

Built frozen policy-relative audits for MT rewards across controlled best-of-N and contextual-bandit studies plus generative GRPO. In 7B language-generalization runs, hack calls held in 4/4 tasks (12/12 cells) and safe calls in 27/34 tasks after correcting and rerunning a rollout-padding artifact.

04 **Donor or Seed? Separating Donor Effects from Training Noise in Low-Resource Translation** IN PREPARATION

EQUAL-CONTRIBUTION CO-LEAD

Shows that comparing donor languages for low-resource NLLB fine-tuning on single runs mostly measures seed noise. A donor × seed variance decomposition across six targets finds donor effects are real but small, and similarity proxies fail to predict them.

---

05 **Speech Registers for Multimodal MT** IN PREPARATION

CONTRIBUTOR

Collaborative project applying source-token registration (Qu et al. 2025) to speech and text translation. My contribution: Whisper ASR and NLLB translation baselines on FLEURS and Europarl, COMET, BLEU, and WER scoring, and SLURM workflows for the experiments.

EXPERIENCE

Research Assistant
February 2025 – Present

BYU Machine Translation Lab (MATRIX), Provo, UT

- Conducting research across multiple active projects; see Research for details.
- Running large-scale multilingual experiments on BYU's supercomputer (SLURM): fine-tuning models up to 3.3B parameters and serving approximately 700 GB GLM checkpoints across eight H200/B200 GPUs.
- Building automated evaluation pipelines (Python, Hugging Face Transformers) computing BLEU, chrF, COMET, and BERTScore across multilingual outputs, enabling comparison across language pairs.
- Led eight researchers producing MQM human quality judgments across 12 languages, and built batched LLM-judge workflows with few-shot prompts, structured-output cleanup, resumable JSONL caching, and severity aggregation.

MT Research Engineer Intern
August 2023 – April 2024, May 2026 – Present

B. H. Roberts Foundation, Holladay, UT

- Designing a scalable multilingual MT pipeline for Latter-day Saint scholarly content in 10 languages, integrating domain-adapted models and custom evaluation to ensure translation quality.
- Developing terminology-aware MT workflows and domain-adapted evaluation methods for religious text, building glossaries and style guides to enforce consistency across languages at scale.
- Building publication-ready translation infrastructure and terminology databases for independent operational use, enabling the foundation to evaluate and deploy MT systems without technical support.

Senior Student Developer
June 2024 – August 2025

BYU Translation Research Group, Provo, UT

- Led development of ataexamcorpus.net, a 10,000-entry ATA certification exam corpus (Python, SQL, AWS), enabling the first large-scale computational study of professional translation exam patterns.
- Redesigned database schema and migrated 10,000+ items into AWS RDS (EC2, Elastic IP); built a researcher-facing access interface enabling programmatic and web-based corpus queries.
- Coordinated a 5-member team of linguists, translators, and engineers to align corpus design across technical and non-technical stakeholders, producing a schema adopted for all subsequent releases.

Teaching Assistant, CS 111 & CS 270

January 2023 – August 2023, September 2025 – April 2026
BYU Computer Science Department, Provo, UT

- Supported 100+ students as TA for Introduction to Machine Learning (CS 270): perceptrons, k-fold CV, decision trees, backpropagation, regression.
- Supported 100+ students in Introduction to Computer Science (CS 111) covering Python fundamentals, debugging, and algorithmic thinking through office hours, lab sessions, and grading.
- Collaborated with 5+ TAs to improve courses and grading workflows for future cohorts.

SKILLS

ML
Python, PyTorch, Hugging Face Transformers, PEFT, LoRA, QLoRA, bitsandbytes, scikit-learn, pandas, NumPy, R

SYSTEMS & DATA
vLLM, OpenAI-compatible REST APIs, FastAPI/Flask, SLURM/HPC, multi-GPU inference, LaBSE/FAISS, SQL, AWS EC2/RDS

ENGINEERING
pytest, unit and integration testing, coverage, Git/GitHub Actions, CI/CD, code review, packaging and dependency management, linters, formatters, and type checking, Bash/Linux, LaTeX

RESEARCH
experimental design, negative-result analysis, multilingual evaluation, reward-model auditing, bootstrap and paired tests, reproducible datasets, code, and papers

LANGUAGES
Native English, fluent Spanish and Portuguese, basic Mandarin Chinese and ASL

RELEVANT COURSEWORK

CURRENT
Reinforcement Learning (CS 401R), Readings in Machine Translation (CS 704R)

GRADUATE
Evaluation Metrics for Language Models (CS 501R), Transformers in NLP (CS 574), Advanced Machine Translation (CS 601R), AI Model Interpretability (CS 601R), Advanced Deep Learning (CS 674), Readings in Prosocial Language Models (CS 704R)

UNDERGRADUATE
Intro to Machine Learning (CS 270), Deep Learning (CS 474), Intro to Machine Translation (CS 479), full BS CS curriculum