

Nathaniel R. Robinson

[linkedin.com/in/nathaniel-robinson](https://www.linkedin.com/in/nathaniel-robinson)
x.com/robinson_n8
n8rrobinson.wixsite.com/mysite

[Google Scholar](https://scholar.google.com/citations?user=n8rrobinson)
n8rrobinson@gmail.com
+1(385)477-9025

Education

PhD, Computer Science **Summer 2027** (expected)
Johns Hopkins University, Center for Language and Speech Processing *Baltimore, MD*

MS, Language Technologies (MLT) **Aug 2023**
Carnegie Mellon University, School of Computer Science *Pittsburgh, PA*

BS, Applied Computational Mathematics **Aug 2021**
Brigham Young University *Provo, UT*

Concentration: **Machine Learning**; accompanying minors in **Linguistics** and **Arabic**

- GPA: 3.99/4.00
- Academic scholarships

Relevant coursework:

Advanced NLP ¹	Speech Recognition ³	ML for Structured Data	Multiling. NLP
Machine Translation ²	Algorithms & Optim.	Data & Uncertainty	Parallel Processing

¹With Graham Neubig, Robert Frederking, and Deryle Lonsdale ²With Steve Richardson ³With Shinji Watanabe

Skills

- Python⁴, Unix, Regex, Algorithms, Data structures, C, C++, Web Dev, Agile, Docker, SQL, Pro Translation
- **Soft skills:** Especially talented at communication, presentation, technical writing, and collaboration
- **Languages:** *Superior:* Haitian; *Advanced:* Arabic⁵, French, Spanish; *Novice:* Mandarin, Quechua

⁴PyTorch, CUDA, Scikit-learn, Keras, NumPy, SciPy, Spark, HuggingFace ⁵Modern Standard, Levantine, and Egyptian Arabic

Experience

Language Technologies Researcher, JHU (and prev. CMU and BYU) **May 2019–present**

- Develop technologies for low-resource languages, 19 peer-reviewed publications to date
- **Current advisors:** [Sanjeev Khudanpur](#), [Kenton Murray](#); **Past advisors:** [David Mortensen](#), [Nancy Fulda](#)

AI Research Engineer Intern, Duolingo **May–Aug 2025**

- Designed interpretable metrics, led needed meta-prompting/design experiments for low-resource MT

Multilingual Research Intern, Cohere **June–Aug 2024**

- Led creation of new LLM dialectal Arabic benchmark, under Kelly Marchisio & Sebastian Ruder

Applied Research Mathematician, National Security Agency **May–Aug 2019**

- Implemented C library for arithmetic capabilities in post-quantum public-key digital signing algorithm

Projects and Publications (See [Google Scholar](#))

Machine Translation (MT) and Language Modeling (LM) for Low-resource Languages

- Build first LLM benchmarks for MT of 203 languages ([WMT](#)), and dialectal Arabic modeling ([ACL](#))
- Led 18-researcher, 9-institution team to build largest MT datasets yet for 42 Creole languages ([NAACL](#))
- Chiefly organized first shared tasks for Creole language MT ([WMT](#)) and dialectal Arabic LM ([VarDial](#))
- Explored cross-lingual transfer in MT for closely related languages ([ACL](#), [ACL](#), [AfricaNLP](#), [LoResMT](#))

Low-resource Speech Recognition (SR), Speech Translation (ST), and speech representations

- Led the winningest submission in the low-resource ST task at IWSLT 2024 ([IWSLT](#))
- Developed synthetic data augmentation techniques for low-resource ([Interspeech](#)) & accented ([TSD](#)) SR
- Phonetics-inspired methods for language documentation and Comp. Ling. ([ACL](#), [ACL](#), [COLING](#))

Controllable Text Generation with Large Language Models (LLMs) and Embeddings for NLP

- Developed novel algorithms for LLM control and embeddings ([NeurIPS](#), [CtrlGen](#), [NEJLT](#), [SAMII](#))

More detailed Curriculum Vitae available on request