

# Hayley Ross

Computational Linguist | Researcher | Organizer & Project Manager

PhD Student, Harvard University

## Education

### PhD in Linguistics

**Harvard University** 2020 – 2025

- › Probed large language models' ability to generalize the meaning of adjective-noun pairs to pairs unseen during pretraining
- › Evaluated GPT-2 and GPT-3 on "between the lines" meaning
- › Designed & conducted 10 experiments on 600+ total participants to study pronoun ambiguity, in collaboration with theoreticians
- › Created & taught introductory course on large language models
- › Organized weekly talk series incl. a panel with 50+ attendees

### MSc in Computational Linguistics

**Brandeis University** 2018 – 2020

### Master's in Mathematics

**University of Oxford** 2011 – 2015

## Experience (selected)

### Deep Learning Scientist Intern

**NVIDIA, Remote** May–Aug 2024

- › Designed and synthetically generated a new benchmark & dataset for LM tool use, improving performance of our 4B model by 70%
- › Fine-tuned model at <1 week's notice from CEO for public blog post

### Research Intern

**Raytheon BBN Technologies, Cambridge, MA** Jun–Aug 2022

- › Fine-tuned XLM-R for cross-lingual transfer learning over events
- › Implemented, improved & published a temporal relation parser based on BERT (achieved +8 points F1 over SotA) Jun–Aug 2019
- › Co-authored a survey on pre-trained LMs with 700+ citations

### Software Developer

**Ghyston, Bristol, UK** 2015 – 2018

- › Developed 14 full-stack projects across industries & languages, including a 10x faster CSV-to-SQL processor and a live bus tracker
- › Managed a two-person project, including estimation, budget, staffing, software development and final delivery
- › Mentored junior & client developers, including code reviews

## Publications (selected)

Min, Bonan, **Hayley Ross**, [6 others], and Dan Roth (2023). 'Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey.' ACM Computing Surveys. <https://doi.org/10.1145/3605943>

**Ross, Hayley**, Jonathon Cai, and Bonan Min (2020). 'Exploring Contextualized Neural Language Models for Temporal Dependency Parsing.' In *Proceedings of EMNLP 2020*. <https://aclanthology.org/2020.emnlp-main.689/>

[hayley.ross@zoho.com](mailto:hayley.ross@zoho.com)

<https://rossh2.github.io>

## Skills

### NLP & Machine learning

Large language models  
Few-shot prompting & probing  
Data exploration & analysis  
Annotation design

### Linguistics

Behavioral experiment design  
Language-to-meaning mapping

### Software development

Full-stack development  
Software architecture  
Clean code & version control

### Project management

Personnel & budget management  
Client communication

### Technical communication

Written & oral presentation  
Non-technical audiences  
Workshop/course design

## Technology

Python (incl. PyTorch, TensorFlow)  
R, SQL, Regex  
Java, C#, JavaScript/TypeScript  
Git, Unix/Linux, LaTeX, Markdown

## Languages

English	native
German	native
French	fluent (B2)
Latin	