Ryan Teehan

rsteehan@gmail.com • 973-460-1672 • rteehas.github.io • GitHub

Education

2014–2018 | **Joint B.A. in Mathematics & M.S. in Computer Science**
University of Chicago, Chicago, IL, USA

Research Experience

Prove AI

Jacqueline Perry AI Fellow ***11/5/21 - Present***

- Collaborating with other fellows on a large-scale human kinematics dataset along with a survey of existing datasets
- Culminating in a **dataset aggregation paper**

HuggingFace BigScience

Working Group Co-Chair ***7/12/21 - Present***

- Leading a group of international researchers on a broad set of research projects investigating emergent properties of large language models
- Coordinating with the engineering team on checkpointing and model recovery during large language model training
- Led the writing of a survey paper on the current state of the art on research about emergent properties of large neural networks **in submission to ACL Rolling Review**

Contributor ***4/28/21 - Present***

- Evaluation Working Group
  - Few-Shot Evaluation Subgroup:
    - Contributed wishlist evaluation tasks for analogical reasoning, symbolic rules-based generalization, linguistic rules-based generalization, and compositionality
    - Helped compile dataset information for initial evaluation suite for the large language model
  - Bias and Fairness Subgroup:
    - Contributing to a survey paper on bias and fairness evaluations for large language models
- Modeling Working Group
  - Contributed to early talks about architecture for the To large language model
  - Contributed prompts to the Hackaprompt effort to prompt the train and test datasets
  - Culminated in a **paper accepted to ICLR 2022**

Toyota Technological Institute at Chicago

Visiting Student ***7/15/21 - Present***

- Advised by Professor Bradly Stadie on research within Reinforcement Learning
- Working on NLP in the context of robotics. Specifically, we are working on goal-conditioned pretraining used in conjunction with affordance models for RL agents

Eleuther AI

Contributor and Researcher ***3/1/21 - Present***

- Collaborated with other independent researchers on text-text contrastive models for fiction criticism, culminating in a **preprint paper**
- Implementing a solution to the fractal inverse problem for equivariant deep learning algorithms (Code)
- **Project Lead** focusing on contrastive learning for code and code review. Currently, we are developing a dataset of granular code modifications and corresponding natural language rationales

Independent Research

Collaborated with Dr. Vinay Prabhu ***1/20/21 - Present***

- Worked on a paper studying contronymy and gender bias in NMT algorithms, awarded a **Best Paper Award** during the *Africa NLP Workshop at EACL 2021*
- Worked on a paper on a novel representation for survey papers, accepted to and presented at the *Rethinking ML Papers Workshop at ICLR 2021*
- Collaborated on multiple submissions to Google BIG-Bench, focusing on humor, riddles in the Kannada language, and reference resolution
- Collaborated on multiple submissions to the NL-Augmenter project involving natural language transformers to improve robustness, culminating in a **preprint paper**

Deep Skies Lab

Research Assistant for Professor Brian Nord ***6/15/18 - 9/11/18***

- Generated simulations of strongly lensed galaxies to train new neural networks in a high performance computing environment using the Lenstronomy Python package along with Keras

Industry Experience

Charles River Analytics

Software Engineer II ***5/2020 - Present***

- Developed probabilistic supply chain models in **Pyro** to infer the existence of missing nodes
- Implemented a multi-resolution Bayesian time series model for real-time maintenance of the condition of complex machinery at coarse and fine time scales
- Developed probabilistic models of satellite movements using the **Scruff** probabilistic programming language
- Contributed to NLP models to be used for the Center for Open Science TOP Factor Score
- Utilized copulas to model dependences between univariate distributions for a Monte-Carlo model that computed the relative cost of uncertainties
- Wrote the probabilistic modeling section of a **SBIR Phase I application** which was awarded by the Defense Logistics Agency
- Lead writing a white paper for the **Emergent III Broad Agency Announcement (BAA)**
- Implementing NLP algorithms for aspect-sentiment analysis

Software Engineer I ***1/2019 - 5/2020***

- Implemented a Multi-Objective Monte Carlo Tree Search algorithm in Scala for repair schedule optimization
- Used the **Figaro** probabilistic programming language to model satellite movements

Infinite Analytics

Data Scientist ***9/2018 - 1/2019***

- Reduced latency of Spark computations, implemented in Scala, by 25%
- Evaluated and implemented algorithms for dimensionality reduction of large, sparse, binary matrices
- Optimized word embeddings to improve search results for customer search

Publications

Ryan Teehan*, Vinay Prabhu*, and Eniko Srivastava*. "Did they direct the violence or admonish it? A cautionary tale on contronomy, androcentrism and back-translation foibles", *EACL 2021 Workshop on African NLP Poster Session*, 2021. https://openreview.net/pdf?id=hUzjN3Sjrc. **Best Paper Award winner**.

Ryan Teehan*, Vinay Prabhu*, and Matthew McAteer*. "SPICES: Survey Papers as Interactive Cheatsheet

Embeddings", *Rethinking ML Papers Workshop at ICLR 2021*, 2021. https://openreview.net/forum?id=1sysg9hi3KS.

Vinay Uday Prabhu, **Ryan Teehan**, Sanghyun Han, Nicholas Roberts, Eniko Srivastava, Aremu Anuoluwapo, Ọpẹyemi Adémólá, Rohan Bisariya, and Ana Carolina da Hora, "Directing the violence or admonishing it? A survey of contronymy and androcentrism in Google Translate and some recommendations" *Preprint*, 2021 https://openreview.net/forum?id=13LjoyYWcaw. *Note: This is a significant extension of the workshop paper*

Shahbuland Matiana*, JR Smith*, **Ryan Teehan***, Louis Castricato*, Stella Biderman, Leo Gao, and Spencer Frazier, "Cut the CARP: Fishing for zero-shot story evaluation". *Preprint*, 2021. https://arxiv.org/abs/2110.03111.

Victor Sanh*, Albert Webson*, Colin Raffel*, Stephen Bach*, and 38 others (including **Ryan Teehan**), "Multitask Prompted Training Enables Zero-Shot Task Generalization", *Preprint*, 2021. https://arxiv.org/abs/2110.08207. *Accepted to ICLR 2022.*

Ryan Teehan*, Natasha Seelam*, Oleg Serikov*, Miruna Clinciu*, Shachar Mirkin*, Eliza Szczechla*, and Aaron Gokaslan, "Emergent Structures and Training Dynamics in Large Language Models", 2021, *Currently under review at ACL Rolling Review.*

Kaustubh D. Dhole and 124 others (including **Ryan Teehan**), "NL-Augmenter: A Framework for Task-Sensitive Natural Language Augmentation," *Preprint to be submitted to NAACL*, 2021. https://arxiv.org/abs/2112.02721.

Open Source Projects

Google BIG-Bench: A benchmark aimed at testing the abilities of large language models. I collaborated on the following accepted benchmark task submissions:

- Ruin A Name With One Edit: A task designed to measure whether large language models are able to identify and utilize humor
- Kannada Riddles: A task designed to test whether large language models can answer riddles in the Kannada language
- Hyperbaton Identification: a task using the inversion of normal word order (hyperbaton) and reference resolution to test the robustness of large language models

NL-Augmenter: A project aimed at developing natural language transformations to augment text datasets. I collaborated on the following accepted transformation submissions:

- Yoda Transformation: A natural language transformation that modifies sentences to flip the clauses such that it reads like "Yoda Speak"
- Gender Neopronoun Substitution: A natural language transformation specific that substitutes gendered pronouns with their neopronoun counterparts (ex. ze/hir/hirs)
- Pig Latin: A natural language transformation that translates the input text into pig latin

Research Proposals

Defense Logistics Agency Emergent III Broad Agency Announcement White Paper, **Lead Author**, *In Preparation*

Defense Logistics Agency SBIR Proposal, **Co-First Author**, *Awarded*

Computing Skills

Programming:	Python, Scala, Julia, SQL, Bash scripting
Machine Learning:	Pandas, Scikit-learn, Numpy, Scipy, TensorFlow, Pytorch, OpenAI Gym, Mujoco
Probabilistic Programming:	Figaro, Scruff, Pyro

Awards and Honors

2014 *University of Chicago Presidents Scholar*
2015 *Eagle Scout*
2014-2018 *Dean's List*
2021 *Best Paper Award Winner – AfricaNLP Workshop at EACL*
2021 *Accepted into the Advanced Language Processing School (ALPS) 2022*
2021 *Jacquelin Perry AI Fellow*

Teaching Experience

2018-2020 **Python Tutor** for a number of students in basic to intermediate programming and intermediate machine learning
2019 **Statistics Tutor** for multiple students, focusing on introductory statistics

Service and Leadership

2021 *Co-organized a social on Open Collaboration in Machine Learning Research with the Machine Learning Collective (MLC) at ICLR 2021*
2021 *Working Group Chair – HuggingFace BigScience*
2021 *Project Lead at Eleuther AI*

Languages

English: Native
Farsi: Intermediate