

RYAN TEEHAN

rst306@nyu.edu • 973-460-1672 • rteeahas.github.io • [GitHub](#)

EDUCATION

2022–PRESENT | **PhD, Center for Data Science**
| New York University, New York, NY, USA
2014–2018 | **Joint B.A. in Mathematics & M.S. in Computer Science**
| University of Chicago, Chicago, IL, USA

EXPERIENCE

New York University | **PhD Student 09/22 – Present**

- Developed a method for few-shot learning of new concepts in language models, **accepted to COLM**.
- Developing optimal guidance methods for diffusion models (collaboration with students at NYU and Columbia), **preprint forthcoming**.
- Researching hierarchical representations in language models, **in progress**.
- Researching neural program synthesis and latent algorithm execution for the ARC challenge, **in progress**.

CarperAI | **Cofounder and Researcher 07/22 – 09/23**

- Conducted research related to Socratic questions for debugging code with LLMs, **accepted to SIGCSE 2024**.
- Conducted research on measuring the diversity of generative model outputs

HuggingFace BigScience | **Working Group Co-Chair 7/21 - 05/22**

- Led a group of international researchers on a broad set of research projects investigating emergent properties of large language models (culminated in **accepted as a workshop paper**)

Charles River Analytics | **Software Engineer 1/19 - 05/22**

- Used probabilistic programming languages to model supply chains (**Pyro**) and satellite movements (**Scruff** and **Figaro**).
- Implemented a Multi-Objective MCTS algorithm in Scala for repair schedule optimization

SELECTED PUBLICATIONS

Raghav Singhal*, Zachary Horvitz*, and **Ryan Teehan***. Realizing their Potential: Improving Diffusion Models with Particle-Based Inference Scaling. *In preparation*.

Ryan Teehan*, Mengye Ren, Brenden Lake. "CoLLEGe: Concept Embedding Generation for Large Language Models". *Accepted to COLM 2024*.

Erfan Al-Hossami, Razvan Bunescu, Justin Smith, **Ryan Teehan**. "Can Language Models Employ the Socratic Method? Experiments with Code Debugging". *Accepted to SIGCSE 2024*.

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 as a Spotlight 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, *Accepted to the Challenges & Perspectives in Creating Large Language Models Workshop at ACL 2022*.

Kaustubh D. Dhole et. al. (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>.

Aarohi Srivastava et. al. (including **Ryan Teehan**), "Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models," *Accepted to TMLR*, 2022. <https://arxiv.org/abs/2206.04615>