

Lucas Torroba Hennigen

lucastor@mit.edu | ltorroba.github.io

EDUCATION

PhD Computer Science

Sep. 2021 – Present

Massachusetts Institute of Technology

Cambridge, MA

- Investigating efficient training and inference for LLMs, supervised by Yoon Kim.
- Selected courses:** Algorithms for Inference, Intro to Numerical Methods, Accelerated Computing (CUDA)

MPhil Advanced Computer Science (Distinction)

Oct. 2019 – Jul. 2020

University of Cambridge

Cambridge, UK

- Thesis topic on “Intrinsic Probing through Dimension Selection,” supervised by Ryan Cotterell.
- Selected courses:** Natural Language Syntax and Parsing, Machine Learning for Language Processing, Advanced Topics in Machine Learning and NLP, Machine Learning for Programming.

BSc (Hons) Artificial Intelligence and Computer Science (First-class)

Sep. 2016 – Jul. 2019

University of Edinburgh

Edinburgh, UK

- NLP honors project on “Historical Event Ordering with Contextual Information,” supervised by Shay Cohen.
- Selected courses:** Machine Learning and Pattern Recognition; Natural Language Understanding, Generation, and Machine Translation; Reinforcement Learning.

EXPERIENCE

Research Intern

May 2023 – Aug. 2023

MIT-IBM Watson AI Lab

Cambridge, MA

- Explored the scaling of instruction mining datasets via data mining, supervised by Akash Srivastava.
- Contributed to the development of alignment methods for IBM models.

Research Intern

Oct. 2020 – Jul. 2021

Québec Artificial Intelligence Institute (Mila)

Remote

- Modeling contextual representations using Dirichlet process mixture models, supervised by Siva Reddy.
- Presented EMNLP work at McGill/Mila Computational Linguistics reading group.

PUBLICATIONS (* \triangleq EQUAL CONTRIBUTION)

Peer-reviewed:

Lucas Torroba Hennigen*, Shannon Shen*, Aniruddha Nrusimha, Bernhard Gapp, David Sontag, Yoon Kim. 2024. Towards Verifiable Text Generation with Symbolic References. In *The First Conference on Language Modeling*, Philadelphia, PA. COLM.

Li Du, Afra Amini, **Lucas Torroba Hennigen**, Xinyan Velocity Yu, Holden Lee, Jason Eisner, Ryan Cotterell. 2024. Principled Gradient-Based MCMC for Conditional Sampling of Text. In *The First Conference on Language Modeling*, Vienna, Austria. ICML.

Lucas Torroba Hennigen, Yoon Kim. 2023. Deriving Language Models from Masked Language Models. In *Proceedings of the 61st Annual Meeting of the Association of Computational Linguistics*, Toronto, Canada. ACL.

Kevin Du, **Lucas Torroba Hennigen**, Niklas Stoehr, Alexander Warstadt, Ryan Cotterell. 2023. Generalizing Backpropagation for Gradient-based Interpretability. In *Proceedings of the 61st Annual Meeting of the Association of Computational Linguistics*, Toronto, Canada. ACL.

Li Du, **Lucas Torroba Hennigen**, Tiago Pimentel, Clara Meister, Jason Eisner, Ryan Cotterell. 2023. A Measure-Theoretic Characterization of Tight Language Models. In *Proceedings of the 61st Annual Meeting of the Association of Computational Linguistics*, Toronto, Canada. ACL.

Niklas Stoehr, **Lucas Torroba Hennigen**, Josef Valvoda, Robert West, Aaron Schein, Ryan Cotterell. 2023. In *Proceedings of the 61st Annual Meeting of the Association of Computational Linguistics*, Toronto, Canada. ACL.

Peihao Wang, Rameswar Panda, **Lucas Torroba Hennigen**, Philip Greengard, Leonid Karlinsky, Rogerio Feris, David Daniel Cox, Zhangyang Wang, Yoon Kim. 2023. Learning to Grow Pretrained Models for Efficient Transformer Training. In *The Eleventh International Conference on Learning Representations*, Kigali, Rwanda. ICLR.

Karolina Stańczak*, **Lucas Torroba Hennigen***, Adina Williams, Ryan Cotterell, Isabelle Augenstein. 2023. A Latent-Variable Model for Intrinsic Probing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Washington, DC. AAAI.

Karolina Stańczak, Edoardo Ponti, **Lucas Torroba Hennigen**, Ryan Cotterell, Isabelle Augenstein. 2022. Same Neurons, Different Languages: Probing Morphosyntax in Multilingual Pre-trained Models. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics*, Seattle, WA. NAACL.

Alexander Immer*, **Lucas Torroba Hennigen***, Vincent Fortuin, Ryan Cotterell. 2022. Probing as Quantifying the Inductive Bias of Pre-trained Representations. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, Dublin, Ireland. ACL.

Niklas Stoehr, **Lucas Torroba Hennigen**, Samin Ahabab, Robert West, Ryan Cotterell. 2021. Classifying Dyads for Militarized Conflict Analysis. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Punta Cana, Dominican Republic. EMNLP.

Lucas Torroba Hennigen, Adina Williams, Ryan Cotterell. 2020. Intrinsic Probing through Dimension Selection. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, Online. EMNLP.

Or Honovich*, **Lucas Torroba Hennigen***, Omri Abend, Shay B. Cohen. 2020. Machine Reading of Historical Events. In *Proceedings of the 58th Conference of the Association for Computational Linguistics*, Online. ACL.

Under Review & In Preparation:

Lucas Torroba Hennigen, Hunter Lang, Han Guo, Yoon Kim. Memory-Efficient Language Model Training with Gradient Projections. In preparation.

AWARDS

Michael Athans Fellowship
Funding for 1st year of PhD.

Mitacs Globalink Research Award
Funding to do an in-person internship at Mila (Unable to claim due to COVID-19 pandemic).

Graduate Tutors Prize for Distinction in Masters degree
Achieved a Distinction in MPhil.

Class Prize for Artificial Intelligence and Computer Science
Graduated in 1st place in Edinburgh AI & CS cohort. 1 recipient.

Howe Prize for Top Performance in UG4 Artificial Intelligence
Highest aggregate grade in 4th year undergraduate AI cohort. 1 recipient.

British Computing Society Prize for Top Performing Student in the Professional Issues Course
Highest grade in 3rd year undergraduate Professional Issues course. 2 recipients.