Michael Hu

+1 (678) 908 2031
michael.hu@nyu.edu
michahu.github.io
michahu

Education

Fall 2022 – Present
PhD in Data Science, *NYU*
- Research interests: NLP, representation learning, and interpretability.

Fall 2017 – Spring 2021
BSE in Computer Science, *Princeton University*
- Minors: Statistics and Machine Learning, Robotics and Intelligent Systems
- GPA: 3.91 / 4
- Coursework: *Theoretical Machine Learning, *Theory of Deep Learning, *Advanced NLP, Robotics, Bayesian Modeling, Advanced Algorithms, Differential Equations, Real Analysis, Probability Theory

* indicates graduate-level class

Selected Research

Dec 2021 – May 2022
Using Natural Language and Program Abstractions to Instill Human Inductive Biases in Machines
- Observed a mismatch between human and machine inductive biases on a simple search task.
- Injected human priors into machines using natural language supervision.

Fall 2020 – Spring 2021
Case Studies on the Interaction Between Machine Learning and Language
Senior Thesis. Advisors: Tom Griffiths, Karthik Narasimhan
- Learned commentaries of reinforcement learning (RL) environments to induce fast adaptation.
- Used meta-learning to model human cognition over word analogies.

Spring 2020 – Summer 2021
Safe Reinforcement Learning with Natural Language Constraints
Junior Independent Work. Advisor: Karthik Narasimhan
- Designed RL agents that understand natural language constraints, such as “Don’t step in puddles.”
- Created HazardWorld, a new open-source RL environment, to test agent performance.

Industry

Sept 2022 – Present
Research Scientist, *Yobi (part-time)* New York, NY
- Studying the limits of what machines can predict.

Sept 2021 – Sept 2022
Software Engineer, *Yobi* New York, NY
- Acted as the primary architect and engineer for Yobi’s data pipelines and machine learning models.
- Implemented data pipelines and models at scale, processing terabytes of data within hours.

Summer 2021
Software Engineering Intern, *Roblox* Remote
- Designed, implemented, and deployed a data pipeline that surfaces trending text in Roblox chat. Applied the pipeline towards catching spammers and large-scale misinformation campaigns.

Summer 2019
Software Engineering Intern, *Roblox* San Mateo, CA
- Engineered a high-throughput (40k queries per second) probabilistic filter that censors inappropriate text across the entire Roblox platform. Led 5 engineers to test and deploy the text filter.
- Trained BERT, a deep language model, to identify mean language and bullying. Automated the labeling of training data for BERT using Snorkel, a data programming package.

Publications

* indicates equal contribution

[1] Sreejan Kumar, Carlos G. Correa, Ishita Dasgupta, Raja Marjieh, Michael Y. Hu, Robert D. Hawkins, Nathaniel D. Daw, Jonathan D. Cohen, Karthik Narasimhan, and Thomas L. Griffiths. Using natural language and program abstractions to instill human inductive biases in machines. *NeurIPS*, 2022.

[2] Marcin Biesiada, Michael Y. Hu, Loren Dean Williams, Katarzyna J Purzycka, and Anton S.

Petrov. rRNA expansion segment 7 in eukaryotes: from Signature Fold to tentacles. *Nucleic Acids Research*, 50(18):10717–10732, 10 2022.

[3] Sreejan Kumar, Ishita Dasgupta, Michael Hu, Raja Marjieh, Robert D. Hawkins, Nathaniel Daw, Jonathan Cohen, Karthik R Narasimhan, and Thomas L. Griffiths. Using natural language to guide meta-learning agents towards human-like inductive biases. In *ACL Workshop on Learning with Natural Language Supervision*, 2022.

[4] Tsung-Yen Yang*, Michael Hu*, Yinlam Chow, Peter J. Ramadge, and Karthik Narasimhan. Safe reinforcement learning with natural language constraints. *NeurIPS (spotlight)*, abs/2010.05150, 2021.

[5] Santi Mestre-Fos, Petar I. Penev, Suttipong Suttapitugsakul, Michael Y. Hu, Chieri Ito, Anton S. Petrov, Roger M. Wartell, Ronghu Wu, and Loren Dean Williams. G-quadruplexes in human ribosomal rna. *Journal of Molecular Biology*, 431(10):1940–1955, 2019.

[6] Lizzette M. Gómez Ramos, Natalya N. Degtyareva, Nicholas A. Kovacs, Stefany Y. Holguin, Liuwei Jiang, Anton S. Petrov, Marcin Biesiada, Michael Hu, Katarzyna J. Purzycka, Dev P. Arya, and Loren Dean Williams. Eukaryotic ribosomal expansion segments as antimicrobial targets. *Biochemistry*, 56(40):5288–5299, 2017. PMID: 28895721.

[7] Kai Wang, Anthony K. Guzman, Zi Yan, Shouping Zhang, Michael Y. Hu, Mehdi B. Hamaneh, Yi-Kuo Yu, Seda Tolu, Jinghang Zhang, Holly E. Kanavy, Kenny Ye, Boris Bartholdy, and Eric E. Bouhassira. Ultra-high-frequency reprogramming of individual long-term hematopoietic stem cells yields low somatic variant induced pluripotent stem cells. *Cell Reports*, 26(10):2580–2592.e7, 2019.

Honors and Awards

2022–2027	NSF Graduate Research Fellowship ($45,000 per year for 3 years)
2021	NSF Graduate Research Fellowship Program (GRFP) Honorable Mention
2021	Outstanding Computer Science Senior Thesis Prize ($600)
2021	Summa Cum Laude, Princeton Computer Science
2020	Princeton Center for Statistics and Machine Learning Summer Research Award ($4,000)

Service

Fall 2019 – Spring 2021 **Peer Academic Advisor**

- Advise freshmen and sophomores on academics, extracurriculars, and career.
- Offer emotional support, especially during stressful times in the academic year.

Spring 2019 – Spring 2021 **Undergraduate Teaching Assistant**

Spring 2021	Grader	Fundamentals of Machine Learning (COS 424)
Fall 2020	Teaching Assistant	Computer Networks (COS 461)
Spring 2020	Grader	Introduction to Machine Learning (COS 324)
Fall 2019	Head Grader	Reasoning about Computation (Math for CS, COS 340)
Spring 2019	Grader	Reasoning about Computation (Math for CS, COS 340)

Skills

Software Development

- Proficient with Python, Go, Scala.
- Familiar with R, Java, C#, C, SQL, JavaScript, HTML, CSS.
- PyTorch, TensorFlow, Cython, Mechanical Turk, Spark, Airflow, Docker.

Hobbies Yoga, cooking, basketball

Last updated October 20, 2022