Alexa R. Tartaglini

alexart@stanford.edu
https://www.alexatartaglini.com/

Education

Stanford University, Stanford, CA — Sept. 2024 – present
Ph.D. in Computer Science
Advisors: Christopher Potts, Judith E. Fan

New York University, New York, NY — Sept. 2018 – May 2023
B.A. with Honors in Computer Science
B.A. in Mathematics
GPA: 3.938 / 4.0; *summa cum laude*

Research Experience

Brown University, LUNAR Lab — 2023 – 2024
Research Scientist
Mentor: Ellie Pavlick
Project: Finding abstract visual relation circuits in Vision Transformers

New York University, Human & Machine Learning Lab — 2019 – 2024
Research Scientist (2023 – 2024), Undergraduate Researcher (2019 – 2023)
Mentors: Brenden M. Lake, Wai Keen Vong
Honors Thesis: "Human-Machine Perceptual Divergence: Two Investigations on How Neural Networks See the World."
Projects:
- Probing shape versus texture bias in deep neural networks
- Modeling human visual category learning with CNNs

NIH, Training Program in Computational Neuroscience — 2020 – 2021
Mentor: Wei Ji Ma

Work Experience

New York University, Courant Institute — 2020 – 2023
Course Assistant
Class: Introduction to Computer Programming
Supervisor: Joshua Clayton

Publications

[1] Qinan Yu, **Alexa R. Tartaglini**, Peter Hase, Christopher Potts & Carlos Guestrin. "Outcome Rewards Do Not Guarantee Verifiable or Causally Important Reasoning." *Under review*, 2026.

[2] Position: Humans are Missing from AI Coding Agent Research. Zora Wang, John Yang, Kilian Lieret, **Alexa R. Tartaglini**, Valerie Chen, Yuxiang Wei, Zijian Wang, Lingming Zhang, Karthik Narasimhan, Ludwig Schmidt, Graham Neubig, Daniel Fried, Diyi Yang. *Under review*, 2026.

[3] Addressing Divergent Representations from Causal Interventions on Neural Networks.
Satchel Grant, Simon Jerome Han, **Alexa R. Tartaglini**, Christopher Potts. *ICLR*, 2026. **Oral**.

[4] Diagnosing Bottlenecks in Data Visualization Understanding by Vision-Language Models.

Alexa R. Tartaglini, Satchel Grant, Daniel Wurgaft, Christopher Potts, Judith E. Fan. *Under review*, 2025.

[5] Control and Predictivity in Neural Interpretability.
Satchel Grant & **Alexa R. Tartaglini**. *NeurIPS MechInterp Workshop*, 2025.

[6] Deep neural networks can learn generalizable same-different relations.
Alexa R. Tartaglini*, Sheridan Feucht*, Michael A. Lepori, Wai Keen Vong, Charles Lovering, Brenden M. Lake, and Ellie Pavlick. *Proceedings of the 8th Annual Conference on Computational Cognitive Neuroscience*, 2025. Poster.

[7] Beyond the Doors of Perception: Vision Transformers Represent Relations Between Objects.
Michael A. Lepori*, **Alexa R. Tartaglini***, Wai Keen Vong, Thomas Serre, Brenden M. Lake, Ellie Pavlick. *NeurIPS*, 2024. Poster presentation.

[8] A Mechanistic Analysis of Same-Different Judgements in Vision Transformers.
Alexa R. Tartaglini* & Michael A. Lepori*. Awarded **Best Talk in AI** at the *5th International Convention on Mathematics of Neuroscience and AI*, 2024.

[9] A developmentally-inspired examination of shape versus texture bias in machines.
Alexa R. Tartaglini, Wai Keen Vong, and Brenden M. Lake. *Proceedings of the Annual Meeting of the Cognitive Science Society 44*, 2022. **Oral presentation.**

[10] Modeling artificial category learning from pixels: Revisiting Shepard, Hovland, and Jenkins (1961) with deep neural networks.
Alexa R. Tartaglini, Wai Keen Vong, and Brenden M. Lake. *Proceedings of the Annual Meeting of the Cognitive Science Society 43*, 2021. **Poster presentation.**

Selected Honors

2024 **MIT Presidential Fellowship**, awarded by MIT to the most promising prospective EECS Ph.D. admits.

2024 **Gordon Wu Fellowship**, awarded by Princeton University to the most promising prospective Ph.D. admits in Computer Science.

2023 **Robert J. Glushko Prize for Outstanding Undergraduate Honors Thesis in Minds, Brains, and Machines**, awarded by the NYU Minds, Brains, and Machines initiative.

2023 **Mathematics Award for Academic Achievement**, awarded to the top-performing graduating senior in Mathematics by the NYU Courant Institute.

2023 **Computer Science Prize for Academic Excellence in the Honors Program**, awarded to the top-performing graduating senior in Computer Science by the NYU Courant Institute.

2023 **Phi Beta Kappa**

2022 **Barry M. Goldwater Scholarship**

2021 **Computer Science Prize for the Most Promising Student in the Junior Year**, awarded to the top-performing junior by the NYU Courant Institute.

2020 **NIH grant**, "Blueprint Training in Computational Neuroscience: From Biology to Model and Back Again." (R90DA043849)

2019 **NYU Presidential Honors Scholars**

2016 **Rensselaer Medal Scholarship**, awarded for excellence in STEM.

Service & Activities

NYU Women in Computing (WinC), member. 2018 – 2023
New York Cares, volunteer. 2018 – 2022
Girl Scouts of America, member. Silver Award recipient in 2014. 2012 – 2015