

Alexa R. Tartaglino

alexart@stanford.edu
<https://www.alexatartaglino.com/>

EDUCATION	Stanford University , Stanford, CA Ph.D. in Computer Science <i>Advisors:</i> Christopher Potts, Judith E. Fan Sept. 2024 – present
	New York University , New York, NY B.A. with Honors in Computer Science; B.A. in Mathematics <i>Summa cum laude</i> ; GPA: 3.938 / 4.0 Sept. 2018 – May 2023
RESEARCH EXPERIENCE	Google DeepMind ♦ Student Researcher ♦ <i>Mentors:</i> Yuxuan Li, Been Kim, Martin Wattenberg ♦ <i>Project:</i> Visual meaning and communication in vision-language models June 2026 – Nov. 2026
	Brown University, LUNAR Lab ♦ Research Scientist ♦ <i>Mentor:</i> Ellie Pavlick ♦ <i>Project:</i> Finding abstract visual relation circuits in Vision Transformers 2023 – 2024
	New York University, Human & Machine Learning Lab ♦ Research Scientist (2023 – 2024), Undergraduate Researcher (2019 – 2023) ♦ <i>Mentors:</i> Brenden M. Lake, Wai Keen Vong ♦ <i>Projects:</i> Shape vs. texture bias in DNNs; modeling human category learning with CNNs 2019 – 2024
	NIH Blueprint Training Program in Computational Neuroscience ♦ <i>Mentor:</i> Wei Ji Ma 2020 – 2021
	New York University, Courant Institute ♦ Course Assistant, Introduction to Computer Programming 2020 – 2023
PUBLICATIONS	<ul style="list-style-type: none">[1] A Heart Means Recursion: Exploring Visual Meaning and Communication in VLMs. Alexa R. Tartaglino, Zhengxuan Wu, Thaddäus Wiedemer, Martin Wattenberg, Been Kim, Yuxuan Li. <i>Under review</i>, 2026.[2] Outcome Rewards Do Not Guarantee Verifiable or Causally Important Reasoning. Qinan Yu, Alexa R. Tartaglino, Peter Hase, Christopher Potts & Carlos Guestrin. <i>ICML</i>, 2026. Poster.[3] I Walk the Line: Examining the Role of Gestalt Continuity in Object Binding for Vision Transformers. Alexa R. Tartaglino & Michael A. Lepori. <i>CVPR Workshop on How Do Vision Models Work? (HOW)</i>, 2026.[4] Position: Humans are Missing from AI Coding Agent Research. Zora Wang, John Yang, Kilian Lieret, Alexa R. Tartaglino, Valerie Chen, Yuxiang Wei, Zijian Wang, Lingming Zhang, Karthik Narasimhan, Ludwig Schmidt, Graham Neubig, Daniel Fried, Diyi Yang. <i>Under review</i>, 2026.[5] Addressing Divergent Representations from Causal Interventions on Neural Networks.

- Satchel Grant, Simon Jerome Han, **Alexa R. Tartaglini**, Christopher Potts. *ICLR*, 2026. **Oral**.
- [6] [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.
- [7] [Control and Predictivity in Neural Interpretability](#).
Satchel Grant & **Alexa R. Tartaglini**. *NeurIPS MechInterp Workshop*, 2025.
- [8] [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. *Conference on Computational Cognitive Neuroscience (CCN)*, 2025. Poster.
- [9] [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.
- [10] [A developmentally-inspired examination of shape versus texture bias in machines](#).
Alexa R. Tartaglini, Wai Keen Vong, and Brenden M. Lake. *CogSci*, 2022. **Oral**.
- [11] [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. *CogSci*, 2021. Poster.

TALKS

- ❖ [Do World Models Emerge in Neural Networks Trained for Prediction?](#) Generative Adversarial Collaboration (GAC), *Conference on Computational Cognitive Neuroscience (CCN)*, 2026.
- ❖ Diagnosing Bottlenecks in Data Visualization Understanding by Vision-Language Models. Keynote, VISxVISION Workshop, *Vision Sciences Society (VSS)*, 2026.
- ❖ [A Mechanistic Analysis of Same-Different Judgements in Vision Transformers](#) (with Michael A. Lepori). *5th International Convention on the Mathematics of Neuroscience and AI*, 2024. Awarded **Best Talk in AI**.

SELECTED HONORS

- 2024 **MIT Presidential Fellowship**
- 2024 **Gordon Wu Fellowship**, Princeton University
- 2023 **Robert J. Glushko Prize** for Outstanding Undergraduate Honors Thesis
- 2023 **Mathematics Award for Academic Achievement**, NYU (top graduating senior)
- 2023 **Computer Science Prize for Academic Excellence**, NYU (top graduating senior)
- 2023 **Phi Beta Kappa**
- 2022 **Barry M. Goldwater Scholarship**
- 2021 **Computer Science Prize for the Most Promising Junior**, NYU
- 2020 **NIH Blueprint Training Grant** in Computational Neuroscience (R90DA043849)
- 2019 **NYU Presidential Honors Scholar**

SERVICE & ACTIVITIES

- NYU Women in Computing (WinC)**, member. 2018 – 2023
- New York Cares**, volunteer. 2018 – 2022