Zachary S. Siegel 4554 Frist Center, Princeton, NJ 08544

☑ zss@princeton.edu✔ www.zacharysiegel.org GithubGoogle Scholar

EDUCATION	Princeton University , Princeton, New Jersey, USA B.S.E. in Computer Science, Minor in Philosophy (3.9/4.0)	Aug 2021 – May 2025
PUBLICATIONS	[1] CORE-Bench: Fostering the Credibility of Published Research Th Reproducibility Agent Benchmark Zachary S. Siegel, Sayash Kapoor, Nitya Nagdir, Benedikt Stroebl, Arvir	rough a Computational Id Narayanan
	 Transactions on Machine Learning Research (2024) [2] Language Guided Operator Learning for Goal Inference Zachary S. Siegel, Jiayuan Mao, Nishanth Kumar, Tianmin Shu, Jacob A 	ndreas
	 Workshop on Learning Effective Abstractions for Planning @ CORL (202 [3] Learning Grounded Action Abstractions from Language Lionel Wong, Jiayuan Mao, Pratyusha Sharma, Zachary S. Siegel, Jiahai F B Tenenbaum, Jacob Andreas 	24) eng, Noa Korneev, Joshua
	 The Twelfth International Conference on Learning Representations (ICLR [4] Characterizing the Implicit Bias of Regularized SGD in Rank Minimi Tomer Galanti, Zachary S. Siegel, Aparna Gupte, Tomaso Poggio Workshop on Mathematics of Machine Learning @ NeurIPS (2024) 	2024) zation
	 [5] Superimposing height-controllable and animated flood surfaces into for risk communication Zachary S. Siegel, Scott A Kulp 	street-level photographs
	 [6] BRIGHT: A Realistic and Challenging Benchmark for Reasoning-Int Hongjin Su, Howard Yen, Mengzhou Xia, Weijia Shi, Niklas Muennigh Liu, Quan Shi, Zachary S. Siegel, Michael Tang, Ruoxi Sun, Jinsung Yo Chen, Tao Yu 	ensive Retrieval off, Han-yu Wang, Haisu on, Sercan O Arik, Danqi
	 Preprint, Under Review at ICLR (2024) [7] AI Agents That Matter Sayash Kapoor, Benedikt Stroebl, Zachary S. Siegel, Nitya Nadgir, Arvir Preprint, Under Review at TMLR (2024) 	nd Narayanan
RESEARCH	Senior Thesis	May 2024 – Present
EXPERIENCE	 Department of Computer Science, Princeton University <i>Modeling Open-Ended Goal Inference.</i> I am working with Professors Jacob Andreas to model how people infer the goals of others in open-e We hypothesize that people use a learned transition model of the env with predicting goals, which we will validate by creating a new domai experiments, and building a computational model. I am also the first author how to learn operators both symbolically and with language models for goa was published at the Learning Effective Abstractions for Planning Worksh Undergraduate Researcher Center for Information Technology Policy, Princeton University <i>Building and Evaluating Agent Benchmarks</i>. I am working with Professor to evaluate and build agent benchmarks for automating aspects of scied am the first author of CORE-Bench, published at TMLR, which eval can computationally reproduce existing scientific papers, a co-author Matter, under review at TMLR, which argues many existing agent evalu are methodologically flawed, and I am building a new benchmark to evalu implement research ideas in code given a high-level description. 	Tom Griffiths and ended goal spaces. ironment to assist n, running human of a paper showing l prediction, which op @ CORL 2024. Nov 2023 – Present Arvind Narayanan entific research. I luates how agents of AI Agents that iation benchmarks ate how agents can May 2023 – Sep 2023
	Language & Intelligence Group, MIT	

Learning Grounded Abstractions from Language. I worked with Professors Jacob Andreas and Josh Tenenbaum to use large language models (LLMs) to learn operators for task and motion planning systems (TAMP). I implemented policy learning approaches for low-level motion planning in the Alfred domain and integrated the pipeline of using LLMs for TAMP. I am a co-author on the paper, which was published at ICLR 2024.

Undergraduate Researcher

Department of Computer Science, Princeton University

Building Retrieval Benchmarks. I worked with Professor Danqi Chen to build a retrieval benchmark for tasks where semantic similarity is not sufficient for matching queries to documents. I helped select a domain and developed a pipeline, from the Art of Problem Solving wiki, to scrape questions for use in the retrieval benchmark. I am a co-author of the paper currently under review at ICLR.

Junior Independent Work

Department of Computer Science, Princeton University

Training LLMs on Podcasts. I worked with Professor Danqi Chen to use podcasts for improving the conversational abilities of LLMs. I found podcast data sources, developed methods to transcribe hundreds of thousands of hours of audio in parallel, fine-tuned LLaMA-2-7B, and compared performance to base models. Additionally, I investigated the quality of podcast audio as a training source.

Visiting Undergraduate Researcher

Center for Brains, Minds, and Machines, MIT Investigating the Low-Rank Bias of Neural Networks. I worked with Professors Tomer Galanti and Tomaso Poggio to investigate why the ranks of weight matrices of neural networks are minimized during stochastic gradient descent. I helped develop a theoretical bound on the rank and ran extensive experiments with different network architectures to investigate the rank and singular values of weight matrices. I am a co-author on the paper published at the

M3L Workshop @ NeurIPS 2024.

Summer Intern

Climate Central

Visualizing Flood Surges for Risk Communication. I worked with Dr. Scott Kulp to build a system that superimposes water surfaces on street view images to communicate the risk of flood surges to vulnerable communities. From LIDAR depth maps and RGB images, the method uses depth completion techniques to generate a 3D model of the environment, and then superimposes a flood surface using Blender. The system is now being deployed across the country to communicate risk. I am the first author on the paper published in the Weather and Climate Extremes journal.

SELECTED	 MAT 203: Advanced Vector Calculus 	Fall 2021
COURSES	 MAT 204: Advanced Linear Algebra 	Spring 2022
	 ORF 309: Probability and Stochastic Systems 	Fall 2022
	PSY 255: Cognitive Psychology	Fall 2022
	COS 484: Natural Language Processing	Spring 2023
	 ORF 307: Optimization 	Spring 2023
	 COS 597Q: AI Safety and Alignment 	Fall 2023
	 COS 345: Robotics 	Fall 2024
TEACHING	 COS 226 (Algorithms) Precept Assistant 	Spring 2022
EXPERIENCE	 MAT 203 (Multivariable Calculus) Course Assistant 	Fall 2022
	 MAT 204 (Linear Algebra) Course Assistant 	Spring 2023
	 COS 484 (Natural Language Processing) Course Assistant and Grader 	Spring 2024
	 COS 426 (Computer Graphics) Course Assistant 	Fall 2024
CAMPUS	• Center for Jewish Life – VP of Shabbat, Education, and Holiday Planning	Jan 2022 – Jan 2024
ACTIVITIES	• The American Whig-Cliosophic Society – <i>Deputy President of the Senate</i>	Sep 2022 – May 2023

Apr 2024 - May 2024

Sep 2023 – May 2024

May 2022 – Aug 2022

May 2019 - Jun 2021

[CV compiled on 2024-11-29]