

AADITYA K. SINGH

1070 Bridgeview Way #1808, SF, CA | +1 703-731-2036 (US) | awesomeaaditya@gmail.com | <https://aadityasingh.github.io/>

EDUCATION

University College London

Gatsby Computational Neuroscience Unit, Ph.D.

London, UK

Sep. 2021 - Nov. 2024

Massachusetts Institute of Technology

GPA: 5.0/5.0


Cambridge, MA

Sep. 2017 - Jun. 2021


- M.Eng. and B.Sc. in Computer Science and Engineering, B.Sc. in Brain and Cognitive Sciences

PUBLICATIONS

Llama team, AI@Meta (Contributors: [A. K. Singh](#), ...). **The Llama 3 herd of models**. <https://arxiv.org/abs/2407.21783>. 


[A. K. Singh](#), T. Moskovitz, S. Dragutinovic, F. Hill, S. C. Y. Chan[†], A. M. Saxe[†]. **Strategy cooptation explains the emergence and transience of in-context learning** ICML 2025 (Oral). <https://arxiv.org/abs/2503.05631>. 

[A. K. Singh](#)^{*}, M. Y. Kocyigit^{*}, A. Poulton, D. Esiobu, M. Lomeli, G. Szilvasy, D. Hupkes. **Evaluation data contamination in LLMs: how do we measure it and (when) does it matter?** *In submission*. <https://arxiv.org/abs/2411.03923>

[A. K. Singh](#), T. Moskovitz, F. Hill, S. C. Y. Chan[†], A. M. Saxe[†]. **What needs to go right for an induction head? A mechanistic study of in-context learning circuits and their formation**. ICML 2024 (Spotlight). <https://arxiv.org/abs/2404.07129>. 

[A. K. Singh](#), DJ Strouse. **Tokenization counts: the impact of tokenization on arithmetic in frontier LLMs**. <https://arxiv.org/abs/2402.14903>. 

[A. K. Singh](#), Y. Yang, K. Tirumala, M. Elhoushi, A. S. Morcos. **Brevity is the soul of wit: Pruning long files for code generation**. ICML DMLR workshop (2024). <https://arxiv.org/abs/2407.00434>

[A. K. Singh](#)^{*}, S. C. Y. Chan^{*}, T. Moskovitz, E. Grant, A. M. Saxe[†], F. Hill[†]. **The transient nature of emergent in-context learning in transformers**. NeurIPS 2023. <https://arxiv.org/abs/2311.08360>. 

[A. K. Singh](#), D. Ding, A.M. Saxe, F. Hill, A. K. Lampinen. **Know your audience: specializing grounded language models with the game of Dixit**. EACL 2023. <https://arxiv.org/abs/2206.08349>

A. S. Yue, L. Madaan, T. Moskovitz, DJ Strouse, [A. K. Singh](#). **HARP: A challenging human-annotated math reasoning benchmark** *In submission*. <https://arxiv.org/abs/2412.08819>. 

S. Dragutinovic, A. M. Saxe, [A. K. Singh](#). **Softmax \geq linear: Transformers may learn to classify in-context by kernel gradient descent** *In submission*. <https://arxiv.org/abs/2510.10425>. 

J. H. Lee, A. K. Lampinen, [A. K. Singh](#)[†], A. M. Saxe[†]. **Distinct computations emerge from compositional curricula in in-context learning** *In submission*. <https://arxiv.org/abs/2506.13253>

Y. Zhang, [A. K. Singh](#), P. E. Latham[†], A. M. Saxe[†]. **Training dynamics of in-context learning in linear attention** ICML 2025 (Spotlight). <https://arxiv.org/abs/2501.16265>


C. Wang^{*}, A. U. Yaari^{*}, [A. K. Singh](#), V. Subramaniam, D. Rosenfarb, J. DeWitt, P. Misra, J. R. Madsen, S. Stone, G. Kreiman, B. Katz, I. Cases, A. Barbu. **Brain treebank: large-scale intracranial recordings from naturalistic language stimuli** NeurIPS 2024 (Oral). <https://arxiv.org/abs/2411.08343>

L. Madaan, [A. K. Singh](#), R. Schaeffer, A. Poulton, S. Koyejo, P. Stenetorp, S. Narang, D. Hupkes. **Quantifying variance in evaluation benchmarks**. *In submission*. <https://arxiv.org/abs/2406.10229v1>

T. Moskovitz, [A. K. Singh](#), DJ Strouse, T. Sandholm, R. Salakhutdinov, A. D. Dragan, S. McAleer. **Confronting reward model overoptimization with constrained RLHF**. ICLR 2024 (Spotlight). <https://arxiv.org/abs/2310.04373>

Y. Yang, [A. K. Singh](#), M. Elhoushi, A. Mahmoud, K. Tirumala, F. Gloeckle, B. Roziere, C. Wu, A. S. Morcos, N. Ardalani. **Decoding data quality via synthetic corruptions: embedding-guided pruning of code data**. NeurIPS ENSLP workshop 2023 (Oral spotlight). <https://arxiv.org/abs/2312.02418>

[A. K. Lampinen](#), S. C. Y. Chan, [A. K. Singh](#), M. Shanahan **The broader spectrum of in-context learning** *In submission*. <https://arxiv.org/abs/2412.03782>.

S. C. Y. Chan, A. Santoro, [A. K. Lampinen](#), J. X. Wang, [A. K. Singh](#), P. H. Richemond, J. McClelland, F. Hill. **Data distributional properties drive emergent in-context learning in transformers**. NeurIPS 2022 (Oral). <https://arxiv.org/abs/2205.05055>. 

RESEARCH EXPERIENCE

OpenAI

Member of Technical Staff

San Francisco, CA

Nov. 2024 - Present

- Researcher on the science and foundations of reinforcement learning team. Figuring out how to spend the next 100x in compute.
- Key contributor on frontier model releases such as o3 and GPT-5 Thinking.

Gatsby Computational Neuroscience Unit

PhD Student

London, UK

Sep. 2021 - Nov. 2024

- Co-supervised by Prof. Andrew Saxe and Dr. Felix Hill.
- Researching the transience of emergent few-shot learning in transformers from an empirical, mechanistic, and theoretical lens.
- Exploring the effects of number tokenization on math reasoning.
- Mentored many younger students: Sara Dragutinovic on her Master's thesis (received the Gibbs Project Prize at Oxford), Yedi Zhang (spotlight paper at ICML), and Jin Hwa Lee.
- Lead student-faculty representative: Aggregate and voice student concerns to faculty, and work to solve them.

Meta AI Research

Research Scientist Intern

Menlo Park, CA

Jun. 2023 - Dec. 2023

- Part of the Data Curation team in FAIR Labs, led by Dr. Ari S. Morcos, and then the Llama 3 team.
- Developed embedding-based and heuristic methods for pruning code data. NeurIPS and ICML workshop papers.
- Contributed to LLaMa 3 efforts, at all parts of the pipeline: Data preprocessing (for math reasoning), Scaling Laws, Evaluations

DeepMind

Research Engineering Intern

London, UK

Jun. 2021 - Apr. 2022

- Part of the Grounded Language team led by Dr. Felix Hill and Dr. Jane X. Wang.
- Led a 5-person project on finetuning grounded language models without direct supervision. EACL 2023 paper.
- Contributed to a larger project characterizing properties of data crucial for emergent few-shot learning. NeurIPS 2022 paper.
- Presented final results directly at organization-wide Research Lead meeting (RPM).

MIT InfoLab

Research Assistant

Cambridge, MA

Jun. 2020 - Jun. 2021

- Part of MIT CSAIL and MIT CBMM. Supervised by Prof. Boris Katz and collaborated with Prof. Ila Fiete.
- Led a project on bio-inspired deep attentional modulation for few-shot learning in object recognition. Master's Thesis (Grade: A)
- Core contributor to a project relating human intracranial recordings to language features (e.g., part of speech). In submission.
- Investigated Expander Hopfield Networks and their applicability for few-shot learning. Presented at lab meeting.
- Co-mentored three undergrad students and one high school student.

Citadel Securities

Quantitative Research Intern

Chicago, IL

Jun. 2019 - Aug. 2019

- Developed novel model selection techniques that led to 10% improvement and a 5x speed-up in equities alpha generation.
- Pioneered genetic algorithm methods for continuous blackbox optimization that led to 4% improvement over baseline alphas.
- Implemented and optimized Hidden Markov Model variants for prediction.
- Summarized work in 5 internal reports. Helped port algorithms to production.

Orbital Insight

Computer Vision Research Intern





Boston, MA

Jan. 2019

- Implemented a Discriminative Autoencoder for unsupervised change detection of satellite images using Tensorflow.
- Developed a weakly supervised model and experimented with different architectures, loss metrics and hyperparameters.
- Presented initial results to the Computer Vision team and VP of Tech Research.

Undergraduate Research

Cambridge, MA

- Compositional ensemble learning. HackMIT Best Use of Data prize (2019). 
- Cross-movement art generation with a variational autoencoder. MIT College of Computing Launch Poster Session. 
- Physical-based audiovisual simulation in automatic online perception. Class project and runnable MTurk experiment. 
- Comparisons to Bayesian neural networks for weight pruning. Class project. 

Naval Research Labs

SEAP Researcher, Lab for Computational Physics and Fluid Dynamics

Washington, DC

Jun. 2017 - Aug. 2017

Metron, Inc.

Analyst Intern, Category Theory

Reston, VA

Jun. 2016 - Aug. 2016

ENGINEERING EXPERIENCE

FeatureX

Software Engineering Intern

Boston, MA

Jun. 2018 - Aug. 2018

Bublup

Software Engineering Intern

McLean, VA

Jun. 2015 - Aug. 2015

TEACHING EXPERIENCE

UCL Gatsby PhD courses

Sep. 2022 - Present

- New student "last black box" bootcamp, Systems and Theoretical Neuroscience, Probabilistic and Unsupervised Learning
- Created new problem sets on: expectation maximization, variational inference, expectation propagation, deep linear networks

MIT 6.867 Graduate Machine Learning

Sep. 2019 - Dec. 2019

- Ideated and wrote exam questions, comprising a third of the midterm and half of the final.
- Led recitation sessions for over 40 students and mentored eight project teams. Rated 7/7 by students in the course evaluation.

INVITED TALKS

- Gatsby Foundation Tricentre Meeting (Jun. 2024) - Learning dynamics of in-context learning in transformers
- Sainsbury Wellcome Center Symposium (Mar. 2024) - Neuroscience on neural networks
- Gatsby Foundation Scientific Advisory Board (Feb. 2024) - Learning dynamics of in-context learning in transformers
- MIT, Fiete Lab (Jan. 2024) - Learning dynamics of in-context learning circuits
- University of Witwatersrand, NLP Advanced Topics lecture (Oct. 2023) - the industrial LLM pipeline + mechanistic interpretability
- DeepMind Analysis Group (May 2023) - Emergence and transience of in-context learning in transformers
- Summerfield Lab (Jan. 2023) - Concept formation in puzzles

AWARDS

- Third place poster, Citadel PhD Summit (2024)
- Hertz Fellowship Finalist (2021)
- MIT Nominee for Marshall and Rhodes Scholarships (2021)
- MIT Brain and Cognitive Sciences Academic Achievement Award (2020)
- MIT CS+HASS Undergraduate Research and Innovation Scholar (2019-2020)
- MIT SHASS Burchard Scholar (2020)
- Top 16 in the US, MIT Battlecode (2018)
- 2nd place team in the US, National Science Bowl (2017)
- US National Olympiads: Silver Medalist (Physics, 2017), 14th in the Nation (Computational Linguistics, 2017), Top 50 in the Nation (Chemistry, 2017), USA(J)MO Qualifier (Math, 2014-2017), Platinum Division (Computing, 2017)

PROGRAMMING SKILLS

- Proficient: Python, Numpy, JAX, PyTorch, Tensorflow, Java, MATLAB
- Competent: Bash, pyspark, \LaTeX , ffmpeg, JavaScript, C, Ruby, Rails, HTML5, JQuery, CSS

RELEVANT COURSEWORK

Artificial Intelligence

Probabilistic & Unsupervised Learning, Kernels, Reinforcement Learning, Bayesian Inference, Statistical Learning Theory, Machine Learning, Natural Language Processing, Artificial Intelligence*

Math

Theory of Computation, Stochastic Processes, Matrix Methods for Machine Learning, Differential Equations*, Complex Analysis*, Probability Theory*, Linear Algebra*

Neuroscience

Systems & Theoretical Neuroscience, Neural Circuits for Cognition, Computational Cognitive Science, Systems Neuroscience Lab, Molecular & Cellular Neurobiology, Organic Chemistry

Software

Computer Security, Computer Systems Engineering, Elements of Software Construction, Design & Analysis of Algorithms, Parallel Computing*

** indicates UG-level classes taken at Thomas Jefferson High School for Sci/Tech, 2015-2017*