Sam Adam-Day *(he/him)*

Email: **me@samdamday.com** **samadamday.com** GitHub: **SamAdamDay**

Innovative AI safety and mathematics researcher with over 20 years of programming experience

RESEARCH EXPERIENCE

Neural Interactive Proofs, Department of Computer Science, University of Oxford 2023–

- Game-theoretic approach to multi-agent AI safety and scalable oversight.
- Built large, well-documented and tested codebase for multi-agent reinforcement learning.
- Ran 100s of experiments with large language models, and conducted novel analysis.
- Produced joint first-author paper, and presented at ICML and NeurIPS 2024 workshops.
- Funded by Long Term Future Fund and OpenAI Superalignment grants.

Postdoctoral Research Assistant, Department of Computer Science, University of Oxford 2023–2024

- Collaborative investigation of the expressive power of graph neural networks.
- Advised PhD student on mathematical aspects of their research project.
- Helped supervise undergrad students with project on learning with constraints.
- First-author conference papers published in NeurIPS 2023 and 2024.

Causal Alignment in Transformer Models, ML Alignment & Theory Scholars programme 2023

- Investigated procedure for testing hypotheses in mechanistic interpretability.
- Produced codebase for automated experimentation using TransformerLens.

Team Lead, OxAI Safety Hub Labs internship 2022

- Research internship performing active learning using large language models.
- Led multidisciplinary team of investigators, managing upskilling process.
- Contributed over 10,000 lines of Python code, and ran over 500 GPU experiments.

PhD Research, Institute of Mathematics, University of Oxford 2019–2023

- Resolved 2004 open problem in geometric group theory using set-theoretic techniques.
- Spearheaded project investigating asymptotic behaviour of graph neural networks.

MSc Research, University of Amsterdam 2017–2019

- Devised novel techniques combining logic, geometry and combinatorics.

EDUCATION

DPhil (PhD) in Mathematics, **University of Oxford** 2019–2023
Branchwise-real trees and bisimulations of potentialist systems

MSc Master of Logic, **University of Amsterdam** 2017–2019
Cum Laude; GPA: 9.1/10

MMath Master of Mathematics, **University of Oxford** 2012–2016
Master's part: First Class 92%; top in year.

SELECTED PUBLICATIONS

Neural Interactive Proofs, Hammond and Adam-Day (*equal contribution*), *Preprint under review*, arXiv:2412.08897 2024

Almost Surely Asymptotically Constant Graph Neural Networks, Adam-Day, Benedikt, Ceylan and Finkelshtein, *NeurIPS 2024*, arXiv:2403.03880 2024

Zero-One Laws of Graph Neural Networks, Adam-Day, Iliant and Ceylan, *NeurIPS 2023*, arXiv:2301.13060 2023

The Intermediate Logic of Convex Polyhedra, Adam-Day, Bezhanishvili, Gabelaia, and Marra, *Prepreint submitted to the Annals of Pure and Applied Logic*, arXiv:2307.16600 2023

On the continuous gradability of the cut-point orders of R-trees, Adam-Day, *Topology and its Applications*, doi:10.1016/j.topol.2021.107937 2022

Uniform, rigid branchwise-real trees, Adam-Day, *To appear in the Israel Journal of Mathematics*, arXiv:2206.15344 2022

Polyhedral completeness of intermediate logics: the Nerve Criterion, Adam-Day, Bezhanishvili, Gabelaia and Marra, *The Journal of Symbolic Logic*, doi:10.1017/jsl.2022.76 2022

Bisimulations of potentialist systems, Adam-Day, *Preprint submitted to The Journal of Symbolic Logic*, arXiv:2206.10359 2022

OTHER EXPERIENCE

Academic talks presented

- "Almost Surely Asymptotically Constant Graph Neural Networks", *NeurIPS poster session*, December 2024.
- "Neural Interactive Proofs", *SATA Workshop at NeurIPS*, December 2024.
- "Almost Surely Asymptotically Constant Graph Neural Networks", *Logic Advanced Class*, November 2024.
- "Asymptotically useless? The outputs of these GNNs converge to a constant function", *Oxford Learning on Graphs seminar*, October 2024.
- "Zero-One Laws of Graph Neural Networks", *NeurIPS poster session*, December 2023.
- "Prover-Verifier Games", *Oxford AI Safety WIP Sessions*, June 2023.
- "Polyhedral Completeness of Intermediate and Modal Logics", *Logic Algebra and Truth Degrees*, September 2022.
- "Uniform, rigid branchwise-real tree orders", *European Set Theory Conference*, August 2022.

Academic events co-organised

- Oxford AI Safety Work-In-Progress Sessions, a bi-weekly research seminar. 2023.
- British Postgraduate Model Theory Conference, University of Oxford, April 2021.
- Oxford Set Theory Seminar series. 2020–2021.
- Set Theory in the UK 4, University of Oxford, December 2019.

Web developer and server administrator, self-employed 2010–

- Developed websites in Python and PHP, working directly with clients.
- Maintained and secured websites receiving millions of monthly visitors.