David A. Dalrymple

☎ +1 22 4444 9614   🐦 @davidad   🌐 davidad   in davidad   ✉ davidad@alum.mit.edu

Formal Education

2011–2012	**Harvard University**, Partial Ph.D., Biophysics	Took most of the neuroscience classes.
2008–2010	**Massachusetts Institute of Technology**, Partial Ph.D.	AI under Minsky and Sussman.
2006–2008	**Massachusetts Institute of Technology**, **S.M.**, Media Technology	Massively parallel computing.
	Youngest-ever recipient of an MIT graduate degree (age 16).	
2000–2005	University of Maryland, Baltimore County, **B.S., Mathematics**	Analysis, matrix theory, neural nets…
2000–2005	University of Maryland, Baltimore County, **B.S., Computer Science**	Algorithms, data structures…

Skills and Interests

Primary **areas of working knowledge**: Applied category theory, AI safety, mathematical modelling, interactive theorem proving, formal verification, modal and substructural logic, measure-theoretic probability, stochastic processes, machine learning, nervous system imaging and simulation, information theory, decision theory, game theory, mechanism design, numerical methods, control theory, philosophy of mind, metaethics

Computer science **areas of expertise:** Abstraction design (interfaces, visualizations, relational schemas, languages, or APIs), type theory, categorical semantics, data structures (distributed, in-core, or out-of-core), databases, numerical optimization algorithms, efficiency optimization *of* algorithms, data analysis, network and consensus protocols.

Selected **programming skills:**   Rust, Rocq, C++17, C, `git`, ARMv8, Metal, Intel 64, (La)TeX/TikZ, PostgreSQL

Experience & Informal Education

2023–Present — Programme Director, **UK Advanced Research + Invention Agency**. Directing a £59M R+D programme on Safeguarded AI, and the broader Mathematics for Safe AI 'opportunity space'.

2021–2023 — Research Fellow, **Oxford University**, Faculty of Philosophy, Future of Humanity Institute. Metaethics with an eye toward aligning AIs with moral truth (in case that exists); then shifted to a sociotechnical safety-engineering approach which became "the Open Agency Architecture" and thence "Safeguarded AI."

2017–2023 — Research Scientist, **Protocol Labs** (half-time). Co-inventor of the Filecoin protocol suite. Filecoin's market cap is presently over $1B, and at its peak was over $10B. More recently invented Hypercerts mechanism for public goods funding. Contributed to research on potential future Protocol Labs projects.

2015–2017 — Intermittent software engineering & machine learning consulting; research at MIRI workshops.

2015 — MIRI Summer Fellow (San Leandro, CA). Caught up on AI safety.

2014 — Sr. Software Engineer, **Twitter**. Worked on the cache system, which serves 90% of Twitter's traffic. Added low-overhead performance monitoring to production cache services using reservoir sampling and Hoare's selection algorithm, and implemented a buffer type with tunable resize policies and a "zero-copy" specialization.

2014 — Recurse Center participant. Produced many one-off hacks, mostly in C and assembly (visible on github); wrote technical articles on `davidad.github.io`, including two that reached the front page of Hacker News; taught several classes to fellow Recursers, ranging from analog circuit design to cloud app deployment techniques.

2012–2013 — MIT Research Affiliate (San Francisco), funded by a personal grant from Peter Thiel. During this time I coordinated efforts in labs at MIT, UCSF, and University of Vienna aimed at creating a complete data-driven simulation of the nervous system of the nematode worm *C. elegans* using a mix of genetic engineering, custom electro-optical systems, and high-performance computing.

2010–2011 — Independent scientist (Cambridge, MA), funded by a personal grant from Larry Page. Self-studied probability theory and machine learning, and implemented a PCA algorithm based on singular value decomposition for reliably extracting cellular signals from video of fluorescent *E. coli*. I also created a moderately sophisticated HTML5 interface for exploring the video data as well as the PCA results.

2010 — Attended **Singularity University**'s summer program for graduate students, and got a whirlwind introduction to the state of the art in fields ranging from synthetic biology to robotics; began studying neuroscience.

2005–2006 — Consulting programmer working with companies in Maryland and New York on B2B Web products.

2001, 2002, 2003 — Summer intern at **Kurzweil Technologies**, contributing to chatbot and image processing projects.

Academic Publications

2025 — "*Flexible Hardware-Enabled Guarantees for AI Compute*," J. Petrie, O. Aarne, N. Ammann, D. Dalrymple, **arXiv**, doi:10.48550/arXiv.2506.15093

2024 — "*Towards guaranteed safe AI: A framework for ensuring robust and reliable AI systems*," D. Dalrymple, J. Skalse, Y. Bengio, S. Russell, M. Tegmark, S. Seshio, S. Omohundro, C. Szegedy, B. Goldhaber, N. Ammann, A. Abate, J. Halpern, C. Barrett, D. Zhao, T. Zhi-Xuan, J. Wing, J. Tenenbaum, **arXiv**, doi:10.48550/arXiv.2405.06624

2019 — "*Dioptics: A common generalization of open games and gradient-based learners*," D. Dalrymple, **SYCO** 5: 5th Symposium on Compositionality, U. Birmingham, UK, 4–5 Sep 2019.

2014 and 2012 — Served on programme committee of **Artificial General Intelligence (AGI)** conference.

2013 — "*Physical principles for scalable neural recording*," A. Marblestone, B. Zamft, Y. Maguire, M. Shapiro, T. Cybulski, J. Glaser, D. Amodei, P. Stranges, R. Kalhor, D. Dalrymple, *et al.*, **Frontiers in Computational Neuroscience** 7(137), doi:10/shp. Preprint featured in **Nature Physics** 9(525), doi:10.1038/nphys2736.

2013 — "*Conneconomics: the economics of large-scale neural connectomics*," A. Marblestone, E. Daugharthy, R. Kalhor, I. Peikon, J. Kebschull, S. Shipman, Y. Mishchenko, D. Dalrymple, B. Zamft, K. Kording, E. Boyden, A. Zador, G. Church, **bioRxiv**, doi:10.1101/001214

2012 — "*Turing: brain model still incomplete*," D. Dalrymple, **Nature** (March 15), doi:10/shq

2010 — "*Reconfigurable Asynchronous Logic Automata*," N. Gershenfeld, D. Dalrymple, K. Chen, A. Knaian, F. Green, E. Demaine, S. Greenwald, P. Schmidt-Nielsen, ACM SIGACT-SIGPLAN Symposium on **Principles of Programming Languages (POPL)**, pp. 1–6; doi:10/dv4xbp.

2008 — "*Asynchronous Logic Automata*," D. Dalrymple, N. Gershenfeld, and K. Chen, **AUTOMATA**-2008: Theory and Applications of Cellular Automata. pp. 313–322.

Patents

U.S. Patent #10,708,071: Consensus Protocols in Distributed Computing Systems. Issued 2020 Jul 7.
U.S. Patent #10,615,979: Replication-Related Protocols for Decentralized Networks. Issued 2020 Apr 7.
U.S. Patent #8,013,629: Reconfigurable Logic Automata. Filed 2009 Sep 16; issued 2011 Sep 6.
U.S. Patent #8,035,414: Asynchronous Logic Automata. Filed 2009 Apr 13; issued 2011 Oct 11.

Teaching and Speaking Engagements

2025 — **Simons Institute for the Theory of Computing**, UC Berkeley: Invited talk entitled *Safeguarded AI Workflows*.

2025 — **High-Confidence Software & Systems**, Annapolis, Md.: Invited talk on *AI for formal verification*.

2025 — **AI Security Forum**, Paris: Keynote entitled *flexHEG: Flexible Hardware-Enabled Guarantees*.

2024 — **CODE BLUE**, Tokyo: Keynote entitled *AI for formal verification; formal verification for AI*.

2023 — **Institut Henri Poincaré**, Paris: Invited talk entitled *An Open Agency Architecture for Safe Transformative AI*.

2023 — **Estonian Academy of Sciences**, Tallinn, Estonia: Invited talk entitled *An Open Agency Architecture for Safe Transformative AI*.

2023 — **DeepMind UK Alignment Meetup**, London: *An Open Agency Architecture for Safe Transformative AI*.

2022 — **Funding the Commons**, New York: Invited talk entitled *Hypercerts, on-chain primitives for impact markets*.

2019 — **Eminent Thinkers in AI Governance**, Perth, Australia: *Cross-cultural alignment on AI governance concepts*.

2018 — Protocol Labs' **Lab Day**, San Francisco: Participated in panel discussion *Open Problems in Science* with Michael Nielsen and Juan Benet; presented on *IPLD Research & Future Directions*.

2018 — **Midlands Graduate School**, U. Nottingham: *Type theory in decentralized computing*.

2012 — **TEDxMontreal**: TED talk entitled *A new type of mathematics*.

2010–2011 — **Teaching assistant to Marvin Minsky** for three iterations of his famous class "The Society of Mind." Delivered guest lecture entitled *Mind vs. Brain: Confessions of a Defector*, which has over 35,000 views.

2007–2009 — Teaching assistant for MIT's **How to Make Almost Anything**, and its less catchily-titled sequel, **How to Make Something That Makes Almost Anything** (on, *e.g.*, 3D printer design).

2002 — **Microsoft Research**, Redmond, Wash.: Introduced future directions for embedded and distributed computing; had the largest attendance of any guest speaker at MSR.

2001 — **TED**, Monterey, Calif.

2000 — **Smithsonian Institute**, Washington, DC: *Multigenerational panel on the future of technology* and *Technology showcase*, both nationally televised.

Trade Publications and Honors

Invited Delegate, *Meeting of Eminent Thinkers on Artificial Intelligence Governance*, Tsinghua University, Beijing (2019)

Member, Industrial Advisory Board, University of Birmingham School of Computer Science (2018–)

Essay *The Principle of Least Action* published in "This Idea Is Brilliant: Underappreciated Scientific Concepts Everyone Should Know" (2018)

Essay *Differentiable Programming* published in "Know This: Today's Most Interesting & Important Scientific Ideas" (2017)

Interviewed in 2010 movie *The Singularity is Near*, giving Erdős–Bacon number of 5