Paolo **Perrone**

Research Scientist · Probability, Information Theory & AI Safety

University of Oxford · Department of Computer Science · Oxford, U.K.

paolo.perrone@cs.ox.ac.uk | www.paoloperrone.org

Profile

Mathematician working at the interface of **probability, information theory, and statistics**, as well as **empirical AI alignment**.

My research develops rigorous foundations for probabilistic and statistical reasoning, such as entropy and information flow, Bayesian inference, stochastic processes, optimal transport, and the semantics of probabilistic programming, with publications in journals such as the IEEE Transactions on Information Theory, LICS, and Advances in Mathematics.

Based in Computer Science at the University of Oxford, I pair theory with building: I co-founded an **ARIA-funded lab** and contribute to a **UK AI Security Institute–funded** alignment project, designing formal languages, developer tools, and methods for keeping AI agents monitorable and aligned.

Experience

University of Oxford, Department of Computer Science — *Oxford, U.K.*

Research Scientist (Research Associate) — *2021 – present*

- Foundations and semantics of probabilistic programming, in the group of Sam Staton.
- Categorical probability and information theory; entropy, Bayesian inference, and stochastic processes.

Adjoint Labs — *Oxford, U.K.*

Co-founder & Research Engineer — *2025 – present*

- Oxford research company funded by ARIA under the *Safeguarded AI* programme, building mathematically grounded approaches to AI safety.
- Design and build formal languages and developer tools for diagrams, structure, and code, such as *DOTS*, a DSL for compositional recursive diagrams, and *Sidecar*, round-trip code ↔ architecture visualization.

Recursive Safeguarding — *Oxford, U.K.*

Technical contractor — *2025 – present*

- Alignment of long-horizon AI agents: inferring agent plans with uncertainty-aware representations to keep humans in the loop.
- Funded by the UK AI Security Institute (*Alignment Project*); previously ARIA *Safeguarded AI*. Team spanning empirical ML (ex-Mila), probabilistic programming, and formal methods.

Massachusetts Institute of Technology (MIT) — *Cambridge, MA, U.S.A.*

Postdoctoral Associate, Mathematics — *2020*

- Group of David Spivak.

Max Planck Institute for Mathematics in the Sciences — *Leipzig, Germany*

Postdoctoral Researcher, Mathematics — *2019*

- Group of Jürgen Jost.

Selected Publications

Full list at www.paoloperrone.org.

2026	**Empirical Measures and Strong Laws of Large Numbers in Categorical Probability**, Logical Methods in Computer Science	*probability*
2024	**Starting Category Theory**, World Scientific	*category theory*
2024	**Markov Categories and Entropy**, IEEE Transactions on Information Theory	*information theory*
2023	**Dilations and information flow axioms in categorical probability**, Mathematical Structures in Computer Science	*information theory*
2023	**Representable Markov Categories and Comparison of Statistical Experiments in Categorical Probability**, Theoretical Computer Science	*bayesian*
2022	**Probability monads with submonads of deterministic states**, Proceedings of LICS	*prob programming*
2021	**De Finetti's Theorem in Categorical Probability**, Journal of Stochastic Analysis	*exchangeability*
2020	**Stochastic order on metric spaces and the ordered Kantorovich monad**, Advances in Mathematics	*optimal transport*
2019	**A Probability Monad as the Colimit of Spaces of Finite Samples**, Theory and Applications of Categories	*optimal transport*
2016	**Hierarchical Quantification of Synergy in Channels**, Frontiers in Robotics and AI	*information theory*

Education

2018	**PhD in Mathematics** (*magna cum laude*), MPI for Mathematics in the Sciences & University of Leipzig	*Leipzig, Germany*
2013	**MSc in Physics** (110/110 cum laude), Università degli Studi di Milano	*Milan, Italy*
2011	**BSc in Physics** (110/110 cum laude), Università degli Studi di Milano-Bicocca	*Milan, Italy*

Skills

Probability & Statistics	Probabilistic & Bayesian modeling, information theory & entropy, stochastic processes, optimal transport; measure-theoretic and categorical foundations
Machine Learning	Probabilistic ML, generative models, probabilistic programming, statistical foundations of learning
AI Safety	Formal methods and tooling for monitorable, aligned AI (ARIA *Safeguarded AI*, UK AISI *Alignment Project*)
Engineering	Design of formal languages & DSLs, developer tooling, research software
Programming	Python, TypeScript / JavaScript, C, C++ (Rust, Haskell: working knowledge), LaTeX
Languages	Italian (native), English (fluent), German (fluent), Spanish (fluent)

Selected Invited Talks

2026	**Expectations in Markov Categories**, Compositional Foundations of Statistics & Machine Learning, Univ. of Warwick	*U.K.*
2024	**Markov categories, symmetries, and generative AI**, ACT for Engineering Design (MIT)	*Cambridge, MA*
2024	**Markov categories: towards user-friendly probability theory**, Finance and Stochastics Seminar, Imperial College	*London, U.K.*
2024	**Markov categories, Bayesian inversions, and statistical experiments**, Tweag	*Paris, France*
2023	**Probability, Symmetry, and Entropy with Markov Categories**, Computational Statistics & ML Seminar (OxCSML)	*Oxford, U.K.*
2023	**Probabilistic programming: challenges and opportunities**, Alan Turing Institute	*London, U.K.*
2023	**Universal properties in probability theory**, keynote, Category Theory 2023	*Louvain-la-Neuve, Belgium*