Thomas Andrew Lamb

PhD Student in Machine Learning, University of Oxford

thomas.lamb@eng.ox.ac.uk | https://tomalamb.github.io/

LinkedIn | GitHub | Google Scholar

Oxford, UK

Research Overview

AI safety and reliable machine learning, focusing on uncertainty quantification, calibration, and generalisation in LLMs. My work develops methods for LLM reliability that combine empirical performance with formal guarantees, drawing on Bayesian inference and statistical learning theory. I study semantic calibration, uncertainty-aware inference, and adaptive steering of LLMs, alongside theoretical questions on generalisation and approximation, aiming to enable robust and trustworthy deployment of foundation and scalable oversight models in high-risk settings.

Experience

- **Apple Machine Learning Research [🌐]** *May 2026 – Present*
  Research Scientist Intern (Incoming) Paris, France
  - Working with Sinead Williamson and Michael Kirchhof on uncertainty quantification for LLMs.
- **DeepMind / Torr Vision Group, University of Oxford [🌐]** *Jul 2022 – Jan 2023*
  Machine Learning Research Intern Oxford, UK
  - Designed a framework for faithful knowledge distillation.
  - Derived LP-based bounds on teacher–student confidence disagreement enabling formal analysis.
  - Developed methods yielding more calibrated student models; resulted in our paper *Faithful Knowledge Distillation*.
- **University of Durham [🌐]** *Jul 2019 – Sept 2019*
  Summer Research Student Durham, UK
  - Proved new identities relating polylogarithmic integrals to multiple zeta values with Prof. Herbert Gangl.

Education

- **University of Oxford** *Oct 2023 – Present*
  PhD (DPhil) in Machine Learning, Department of Engineering Science Oxford, UK
  - Supervisors: Tim G. J. Rudner (University of Toronto & Vector Institute) and Philip H. S. Torr (University of Oxford)
  - Research on AI safety and reliable ML, developing methods for LLM uncertainty, calibration, and generalisation
- **University of Edinburgh** *Sept 2022 – Aug 2023*
  MSc Artificial Intelligence (Distinction) Edinburgh, UK
  - MSc Artificial Intelligence Class Prize (ranked 1st overall)
  - Dissertation: *Self-Supervised Learning of Tractable Generative Models*
- **University of Durham** *Oct 2016 – Jul 2020*
  MMath Mathematics (First-Class Honours) Durham, UK
  - Percy Heywood Prize; John Crowther Prize
  - Dissertation: *An Introduction to Modular Forms and the Eichler–Shimura Isomorphism*

Publications

Conference Papers

- **Focus On This, Not That! Steering LLMs with Adaptive Feature Specification** *2025*
  T. A. Lamb, A. Davies, A. Paren, P. Torr, F. Pinto ICML
- **Universal In-Context Approximation by Prompting Fully Recurrent Models** *2024*
  A. Petrov, T. A. Lamb, A. Paren, P. Torr, A. Bibi NeurIPS
- **Hidden in Plain Sight: Evaluating Abstract Shape Recognition in Vision-Language Models** *2024*
  A. Hemmat, A. Davies, T. A. Lamb, J. Yuan, P. Torr, A. Khakzar, F. Pinto NeurIPS
- **Improving Semantic Uncertainty Quantification in Language Model Question-Answering via Token-Level Temperature Scaling** *2025*
  T. A. Lamb, D. Ivanova, P. Torr, T. Rudner AISTATS
- **Detecting LLM Hallucination through Layer-wise Information Deficiency: Analysis of Unanswerable Questions and Ambiguous Prompts** *2025*
  H. Kim, T. A. Lamb, A. Bibi, P. Torr, Y. Gal EMNLP

Workshops and Preprints

- **Towards Label-Free Biological Reasoning Synthetic Dataset Creation via Uncertainty Filtering** *2025*
  J. L. Stoisser, L. Phillips, A. Misra, T. A. Lamb, P. Torr, M. B. Martell, J. Fauqueur, K. Martens NeurIPS ER
- **Can Large Language Models Express Uncertainty Like Humans?** *2025*
  L. Tao, Y. F. Yeh, B. Kai, M. Dong, T. Huang, T. A. Lamb, J. Yu, P. Torr, C. Xu arXiv
- **Faithful Knowledge Distillation** *2023*
  T. A. Lamb, R. Brunel, K. Dvijotham, M. P. Kumar, P. H. S. Torr, F. Eiras arXiv

Talks

- **Improving Semantic Uncertainty Quantification in LLMs via Token-Level Temperature Scaling** *Sept 2025*
  Oxford Novo Nordisk
- **Semantic-Level Confidence Calibration of Language Models via Temperature Scaling** *Mar 2025*
  Oxford University of Oxford
- **Towards Safety and Transparency: Addressing Bias and Uncertainty Quantification in LLMs** *Oct 2024*
  London Huawei Noah's Ark Lab

Teaching

- **University of Oxford** *Sept 2023 – Jun 2024*
  OXAi Education Team Member Oxford, UK
  - Created blog posts and teaching materials on machine learning, ranging from introductory concepts to advanced topics for undergraduates and researchers.
- **University of Oxford and Online** *Sept 2020 – Present*
  Mathematics and AI Tutor Oxford, UK & Remote
  - Teach MSc-level Reinforcement Learning, Statistical Learning Theory (OPUS program) and tutor GCSE/A-Level and undergraduate mathematics.
  - Supported students across proof-based mathematics and foundational ML topics.
- **University of Durham** *Oct 2019 – Jun 2020*
  Mathematics Teaching Assistant Durham, UK
  - Marked and provided detailed feedback on second-year Algebra coursework, ensuring clarity and rigor.

Awards and Prizes

- **MSc Artificial Intelligence Class Prize** *2023*
  University of Edinburgh
  - Awarded to the student who attained the highest overall mark in the MSc Artificial Intelligence course.
- **Percy Heywood Prize** *2020*
  Department of Mathematical Sciences, University of Durham
  - Awarded to a student graduating with an MMath whose performance is outstanding in the final year.
- **John Crowther Prize** *2019*
  University College, University of Durham
  - Awarded for performance in third-year Mathematics examinations.

Technical Skills

- **Programming Languages:** Python, Bash, Java, SQL, Matlab
- **Machine Learning:** PyTorch, GPyTorch, Transformers, OpenAI, WandB, scikit-learn, NumPy, Pandas, SciPy
- **Research Areas:** AI safety, LLM generalisation, statistical learning theory, LLMs, NLP, uncertainty quantification, calibration, Bayesian methods, variational inference
- **Other Tools:** Git, GurobiPy, NLTK, Datasets

References

- **Dr. Tim G. J. Rudner**
  Assistant Professor University of Toronto / Vector Institute
  - Contact details: available upon request
- **Prof. Philip H. S. Torr**
  Professor of Engineering Science University of Oxford
  - Contact details: available upon request