Thomas Andrew Lamb

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

Education

- **University of Oxford** — Oxford, UK
  DPhil (PhD) in Machine Learning, Department of Engineering Science. — *Oct. 2023 - Present*
  - Supervised by Prof. Philip H.S. Torr (University of Oxford) and Dr. Tim G.J. Rudner (NYU).
  - My general research interests concern AI safety. In particular, how can we elicit, produce and give guarantees for the safe and reliable deployment of ML models, specifically LLMs and variants such as reasoning models.
  - Have several publications at top-tier ML conference including ICML, NeurIPS and ICLR. These include papers on: the semantic calibration of and uncertainty quantification for LLMs; steering LLMs adaptively at inference time; biases in VLMs; and giving formal statements on universal in-context approximation of fully recurrent models including SSMs.
- **University of Edinburgh** — Edinburgh, UK
  MSc Artificial Intelligence: Distinction. — *Sept. 2022 - Aug. 2023*
  - Courses: Machine Learning and Pattern Recognition, Reinforcement Learning (91%), Bayesian Theory (97%), Probabilistic Modelling and Reasoning (85%), Accelerated Natural Language Processing, and Natural Language Understanding, Generation and Machine Translation (84%). Audited Targeted Causal Learning, Methods of Causal Inference, and Algorithmic Game Theory.
  - My MSc dissertation was titled 'Self-Supervised Learning of Tractable Generative Models' and was supervised by Dr. Antonio Vergari. This project looked at using conditional composite log-likelihood estimation (CCLE) with various novel patching schemes as an alternative method of training EiNet models, a class of probabilistic models that allows for exact forms of inference. Here, we aimed to investigate if a CCLE objective function can act as a form of implicit regularisation, as well as if CCLE objective can aid in improving inpainting performance on image datasets.
  - Received the MSc Artificial Intelligence Class Prize.
- **University of Durham** — Durham, UK
  MMath Mathematics: First-Class Honours — *Oct 2016 - July 2020*
  - Specialised in Pure Mathematics. Selection of courses taken: Algebraic Topology, Riemannian Geometry, Representation Theory, Algebraic Curves, Partial Differential Equations, Algebraic Number Theory, Elementary Geometry, Topology, Complex Analysis, Galois Theory, Statistics, Analysis in Many Variables, Algebra, Introduction to Programming, Computer Systems and Numerical Analysis.
  - My MMath thesis was titled 'An Introduction to Modular Forms and the Eichler-Shimura Isomorphism' and was supervised by Prof. Herbert Gangl. I aimed to provide a comprehensive introduction to the theory and results of classical Modular Forms. This included detailed discussions on Hecke operators, Maeda's conjecture, L-functions and was concluded by proving the Eichler-Shimura Isomorphism.
  - Averaged 90% overall and received the John Crowther and the Percey Heywood prizes.

Publications and Preprints

- Tom A. Lamb, Desi Ivanova, Philip H. S. Torr, and Tim G. J. Rudner. Semantic Calibration of LLMs Through the Lens of Temperature Scaling." ICLR Workshop: Quantify Uncertainty and Hallucination in Foundation Models (2025).
- Kim, Hazel, Lamb, Tom A., Adel Bibi, Philip Torr, and Yarin Gal. "Detecting llm hallucination through layer-wise information deficiency: Analysis of unanswerable questions and ambiguous prompts." arXiv preprint arXiv:2412.10246 (2025).
- Tom A. Lamb, Adam Davies, Alasdair Paren, Philip H. S. Torr, and Francesco Pinto. (2024). "Focus On This, Not That! Steering LLMs With Adaptive Feature Specification." ICML 2025,.
- Arshia Hemmat, Adam Davies*, Tom A. Lamb*, Jianhao Yuan*, Philip Torr, Ashkan Khakzar, Francesco Pinto."Hidden in Plain Sight: Evaluating Abstract Shape Recognition in Vision-Language Models". NeurIPS 2024 Track Datasets and Benchmarks, Poster (2024)
- Aleksandar Petrov, Tom A. Lamb, Alasdair Paren, Philip H. S. Torr, and Adel Bibi. "Universal in- context approximation by prompting fully recurrent models." NeurIPS Poster (2024)
- Petrov, A., Torr, P., and Bibi, A. "Prompting a Pretrained Transformer Can Be a Universal Approximator". In Forty-first International Conference on Machine Learning (2024). *Acknowledged for correcting and rectifying key proofs within the paper.*
- Lamb, T. A., Brunel, R., Dvijotham, K. D., Kumar, M. P., Torr, P. H., & Eiras, F. (2023)." Faithful Knowledge Distillation". arXiv preprint arXiv:2306.04431.

Prizes

- **MSc Artificial Intelligence Class Prize (School of Informatics, University of Edinburgh):** Awarded to the student who gained the highest overall mark in the MSc Artificial Intelligence course.
- **The Percy Heywood Prize (Department of Mathematical Sciences, University of Durham):** Awarded to a student graduating with an MMath Master of Mathematics whose performance is outstanding in the final year.
- **The John Crowther Prize (University College, University of Durham):** Awarded by University College for performance in my third year Mathematics exams.

Research Experience and Internships

- **G-Research** Saint-Jean-Cap-Ferrat, France
  Spring into Quant Finance 2025, G-Research *April 2025*
  - Competitive entrance to a week consisting of workshops around quantitative finance.
  - Attended sessions on mathematical finance, machine learning in finance, and participated in coding workshops and hackathons.
- **Oxford University** Oxford, UK
  UNIQ+ DeepMind Machine Learning Research Intern *July 2022 - July. 2024*
  - Research intern in machine learning at Oxford University based in the Torr Vision Group (TVG). Here, I was supervised by Francisco Girbal Eiras and Prof. Philip H.S. Torr.
  - Our research introduced a faithful imitation framework in which to discuss the relative calibration of teacher-student neural network pairs.
  - We computed upper and lower bounds on the maximum difference in confidences of teacher-student pairs in perturbation neighbourhoods surrounding images using LP methods. Finally, we introduced a new form of distillation that produced empirically and verifiably more faithful teacher-student pairs than other forms of knowledge distillation.
  - This work can be found via the following link: https://arxiv.org/pdf/2306.04431
- **University of Durham** Durham, UK
  Summer Research Student *July 2019 - Sept. 2019*
  - Summer research student working with Prof. Herbert Gangl.
  - Inspired by Prof. Michael Hoffman's paper 'Polylogarithmic Integrals and MZV (Multiple Zeta Values)', we conjectured and subsequently proved novel results that relate integrals over the unit square of polylogarithms multiplied with different two variable polynomial fractions to linear combinations of MZVs.

Talks

- **Semantic-Level Confidence Calibration of Language Model via Temperature Scaling.** LLM Seminar, Department of Statistics, University of Oxford.
- **Towards Safety and Transparency: Addressing Bias and Uncertainty Quantification in Neural Models.** Huawei Noah's Arc Lab, London, UK.

Reviewing Duties

- **ICML 2024 workshop:** Mechanistic Interpretability.
- **ICLR 2025 workshop:** Quantify Uncertainty and Hallucination in Foundation Models.
- **ICML 2025 workshop:** Long Context Foundation Models.
- **NeurIPS 2025:** Main conference, reviewing papers on the general theme of UQ for LLMs.

Teaching and Outreach

- **OXAi Education Team Member**
  University of Oxford, UK *Sep. 2023 - Present*
  - Responsible for creating resources including blog posts on ML covering technical through to introductory topics. Moreover, I am responsible for developing talks for later year groups of secondary school students in the UK as part of OxAI outreach.
- **Mathematics and AI Tutor**
  University of Oxford and Online *Sep. 2020 - Present*
  - I currently give AI tutorials at a MSc level on Reinforcement Learning as part of the OPUS program within the University of Oxford.
  - I have tutored Mathematics for several years. This involved teaching GCSE and A Level Mathematics and Further Mathematics, as well as university level Mathematics.
- **Mathematics Homework Assessor**
  University of Durham, UK *Oct. 2019 - June 2020*
  - Marked and assessed the homework of second year Mathematics students at the University of Durham who were taking the second year Algebra course.

Technical Skills

- **Programming/Frameworks:** Python, Bash, git, WandB, Java, Matlab, SQL
- **Specific Python Packages Used:** PyTorch, GPyTorch, Transformers, OpenAI, Pandas, GurobiPy, SciPy, Scikit-learn, NLTK, Datasets.