

Yanan Long

ylong@uchicago.edu | Website | GitHub | LinkedIn

Machine learning researcher and consulting CTO working on AI evaluation, Bayesian statistics and biomedical ML. Research spans benchmark validity, evaluation reporting, uncertainty quantification and participatory AI governance, alongside experience building training, tuning and experiment-tracking pipelines.

Research and engineering experience

StickFlux Labs | 2024-present

Consulting CTO

- Built a training, hyperparameter-tuning and versioning pipeline using Weights & Biases for multivariate time-series classification, comparing engineered-feature, deep-learning and tabular foundation-model approaches and mentoring students in reproducible ML workflows.
- Contributed Bayesian statistical analyses to research on first- and third-party social-impact evaluation reporting and benchmark saturation, resulting in two ICML 2026 papers.
- Developed an archive-and-adjudication protocol for public frontier-AI evaluations, testing candidate inference against recovery, prediction, transfer and calibration criteria and documenting the limits of the candidate model.
- Co-authored Evaluation Cards and Every Eval Ever on evaluation reporting and shared result schemas, and developed Bayesian nonparametric methods for ambiguity-aware safety evaluation.
- Designed and led multilingual mechanistic-interpretability research, developing tooling to compare model behaviour and internal mechanisms across languages.

University of Chicago | 2018-2024

Researcher Data Scientist

- Worked with a physician-led team on frailty prediction from accelerometry and clinical covariates, translating domain requirements into features and predictive models.
- Investigated molecular fingerprints and neural-network baselines for drug-drug interaction prediction, comparing performance and computational requirements with graph-based methods.
- Applied Bayesian data analysis to large biomedical datasets, including research on human sex ratios at birth published in PLOS Computational Biology.

Education

University of Chicago | 2023

PhD, Chemistry (Biomedical Data Science)

Colby College | 2016

BA, Chemistry (Distinction) and Mathematics; Music minor; cum laude

St Edmund Hall, University of Oxford | 2015

Visiting student; tutorials in Biochemistry and Mathematics

Technical skills

Programming: Python, R, C/C++, Julia, Scala, Java, Bash, Perl

Statistics: Stan, PyMC, NumPyro; Bayesian modelling, uncertainty quantification, calibration

Machine learning: PyTorch, JAX/Flax, TensorFlow/Keras, scikit-learn, Transformers, PyTorch Geometric, gradient boosting

Data and MLOps: SQL, Spark, Weights & Biases, Neptune, Lightning, Airflow, Docker, GCP, AWS

Languages: English; Chinese (Mandarin, Cantonese); German; French

Selected publications

More publications: yananlong.github.io/publications/. Author lists marked as abbreviated do not indicate full author order.

Conference papers

Who Evaluates AI's Social Impacts? Mapping Coverage and Gaps in First and Third Party Evaluations

Anka Reuel, Avijit Ghosh, Jenny Chim, Yanan Long and collaborators. **ICML 2026**. Study of 186 first-party release reports and 248 third-party evaluation sources. Author list abbreviated.

When AI Benchmarks Plateau: A Systematic Study of Benchmark Saturation

Mubashara Akhtar, Anka Reuel, Yanan Long and collaborators. **ICML 2026**. Study of saturation across 60 language-model benchmarks. Author list abbreviated.

"Taking Stock at FAccT": Using Participatory Design to Co-Create a Vision for the Fairness, Accountability and Transparency Community

Shiran Dudy, Jan Simson and Yanan Long. **ACM FAccT 2026**. Participatory design for research-community governance.

Challenges to Grassroots Organization Engagement with AI Policy

Carter Buckner, Jennifer Mickel, Yanan Long and collaborators. **ACM FAccT 2026**. Author list abbreviated.

Workshop papers

Bayesian Inference and Decision Audits for Public Archives of Frontier AI Evaluations

Yanan Long. **ICML 2026 workshops CTB and HT**. Archive reconstruction and validation gates, including negative results for the candidate inference model.

Embracing Ambiguity: Bayesian Nonparametrics and Stakeholder Participation for Ambiguity-Aware Safety Evaluation

Yanan Long. **AAAI 2026 workshop MURE**. Revised version of the CHI 2025 STAIG position paper.

Preprints

Evaluation Cards: An Interpretive Layer for AI Evaluation Reporting

Avijit Ghosh, Anka Reuel, Jenny Chim, Yanan Long and collaborators. **arXiv, 2026**. Author list abbreviated. Preprint.

Every Eval Ever: A Unifying Schema and Community Repository for AI Evaluation Results

Jan Batzner, Sree Harsha Nelaturu, Yanan Long and collaborators. **arXiv, 2026**. Author list abbreviated. Preprint.

Queer NLP: A Critical Survey on Literature Gaps, Biases and Trends

Sabine Weber, Yanan Long and collaborators. **arXiv, 2026**. Author list abbreviated. Preprint.

Service and recognition

- Core organizer, Queer in AI, 2023-present, including workshop organization and participatory AI-governance research.
- Co-author through Organizers of Queer in AI of Queer In AI: A Case Study in Community-Led Participatory AI, ACM FAccT 2023 Best Paper Award.
- Reviewing includes ICLR, NeurIPS, ACM FAccT, AIES and PLOS Computational Biology.