

Robert Sneiderman

Mathematical Statistician | Production ML Systems | LLM Evaluation and Formal-Methods Research

robbynsneiderman@gmail.com | github.com/Robby955 | linkedin.com/in/robbynsneiderman | robbynsneiderman.com | Canada

Authorized to work in the United States (U.S. citizen by descent) and Canada.

EDUCATION

- M.Sc. Statistics, University of Victoria (2018-2020). Thesis stream with Co-op. Final GPA 8.81 / 9.00. M.Sc. thesis completed 31 Dec 2020. JJEM Graduate Award in Mathematics and Statistics.
- B.Sc. Mathematics, McGill University (Sep 2013-Feb 2018). Cumulative GPA 3.61 / 4.00; Major GPA 3.82 / 4.00. Tomlinson Engagement Award for Mentoring.

RESEARCH AND PUBLICATIONS

- **Channel-Stability Auditing for AI-Mediated Integrity Measurement: A Pre-Registered Equivalence Study on ToxicChat** - KDD 2026 · 6th Workshop on Integrity in Social Networks and Media, Jeju, Korea - 2026. Workshop paper (accepted, ORAL; presentation August 9, 2026). <https://openreview.net/forum?id=BSP5yVgTDQ>. Pre-registered audit of channel stability for LLM-as-judge integrity evaluation. N=150 balanced ToxicChat sample; two judges (GPT-4o-mini, Claude Haiku 4.5) across three to four measurement channels. Channel-to-judge variance ratio 1.4x; 3 of 6 within-judge contrasts cross a 0.03-Brier minimum-effect bar after Holm-Bonferroni. Within-model channel swaps moved Brier as much as replacing the judge with a deployed moderation API. Audit cost USD 0.16; pre-registration commit 96e6fc4.
- **Consensus Is Not Enough: Disagreement-Preserving Evaluation for Cultural AI** - ICML 2026 Culture × AI Workshop (DoingAIDifferently), Seoul, South Korea - 2026. Workshop paper (accepted, poster; forthcoming July 2026). Submission #25. Introduces Disagreement-Preserving Evaluation (DP-EVAL): a framework for auditing whether cultural-AI systems preserve meaningful disagreement in user-facing responses. Five-dimensional rubric covering Visibility, Tradeoffs, Calibration, Boundaries, and Actionability. Single-rater pilot study: DP-style responses raised the composite by +1.73 (Wilcoxon $p = 0.004$ over 9 paired prompt-by-model cells).
- **When Top-k Truncation Is Not the Bottleneck** - ICML 2026 Hypothesis Testing Workshop (testing.ml) - 2026. Workshop paper (accepted, poster). Equivalence-testing analysis of where top-k truncation does and does not change LLM token-probability calibration across the API stack. Applies two one-sided tests (TOST) and BCa bootstrap confidence intervals over provider, model, and quantization conditions, built on the logprobe measurement toolkit.
- **Establishing the Signal above the Noise: Accounting for an Environmental Background in the Detection and Quantification of Salmonid Environmental DNA** - Fishes (MDPI) - 2022. Peer-reviewed journal article. <https://doi.org/10.3390/fishes7050266>. Binomial-Poisson methods for detecting salmonid eDNA at low copy number against environmental background noise. Controlled experiments at Goldstream Hatchery.

MANUSCRIPTS UNDER REVIEW

- **From Agents to Axioms: A Verified Multi-Agent Lean Workflow for Statistical Learning Theory** - ICML 2026 AI4Math Workshop - 2026. Submitted 2026-05-24, under review. A verified multi-agent Lean 4 workflow that formalizes statistical learning theory results end to end, building on the FormalSLT library, with a failure-mode taxonomy and a reproducibility recipe.
- **Silent Inconsistencies: A Kernel-Verified Audit of LLM Logical Consistency** - IJCAI-ECAI 2026 LogiSymb Workshop - 2026. Submitted 2026-05-31, under review (notification 2026-06-14). A kernel-verified audit of large language model logical consistency that measures a silent-error rate: cases where a model asserts a claim and its negation without surfacing the contradiction.
- **From Agents to Axioms: Verifier-Gated Acceptance as a Neurosymbolic Pattern for Lean Formalization** - NeSy 2026 (PMLR proceedings) - 2026. Submitted 2026-06-03, under review. Frames verifier-gated acceptance, where only kernel-checked Lean proofs enter the library, as a general neurosymbolic pattern for LLM-driven formalization, evaluated on the FormalSLT build.
- **AIMS** - COLM 2026 - 2026. In preparation (submission-candidate-v2, tag 2f6549c); held for the 2026-06-20 to 2026-06-23 submission window. Conference submission in preparation for COLM 2026, with an EBNF grammar and four firing predicates evaluated across three domains.

OPEN-SOURCE CONTRIBUTIONS

- OpenAI parameter-golf - merged PR #1412 at 1.08354 BPB on the public leaderboard. The 2026-04-06 SP8192_HessianSDClip_ProgressiveRecurrence entry (Hessian SDClip plus progressive recurrence) held the 10min/16MB track record for a period.
- EleutherAI lm-evaluation-harness - merged PR #3691, a GPQA preprocessing regex fix removing bracket-stripping

that corrupted chemical nomenclature, physics notation, and math expressions across 12 files; verified 12 corrupted answers in Diamond, 18 in Main, 22 in Extended, closing issue #2907.

- PrimeNumberTheoremAnd (Tao and Kontorovich) - merged Lean 4 PR #1442 contributing a 2bc-normalized B the lemma into the BKLNW table-bound theorem; CI passed.
- Stanford Centaur PyPantograph - merged PR #174 fixing MCTS search-state advancement, where tactics were applied to the root goal instead of the selected search node, plus a dead test-assertion finding; merged by maintainer lenianiva on 2026-06-01.
- FormalSLT v0.1 - public, machine-checked Lean 4 and Mathlib statistical learning theory library at github.com/Robby955/FormalSLT (MIT). 0 sorry, 0 admit, axioms restricted to propext, Classical.choice, and Quot.sound. The theorem spine runs from finite-sample ERM through Rademacher, VC, sub-Gaussian chaining, stability, and finite PAC-Bayes, and now carries a sharp two-sided McDiarmid bounded-differences inequality on a reusable Bennett and Bernstein MGF module. 60 FormalSLT/ Lean modules (85 checked Lean files including examples/), 33,427 lines. Targeting CPP 2027.
- PAC-Bayes certificate compiler - a generator that turns a trained model's concrete finite statistics (data law, hypothesis prior, bounded loss, sample size, target confidence) into a self-contained Lean module that machine-checks a PAC-Bayes generalization guarantee for those numbers, by instantiating the verified finite Catoni-style bound from FormalSLT. Scope: finite hypothesis class, bounded loss in $[0,1]$, rational inputs (proof-of-concept verified).
- formal-martingales - public Lean 4 library at github.com/Robby955/formal-martingales (Apache 2.0). An axiom-clean formalization of Ville's inequality for nonnegative supermartingales.
- memory-bench - public controlled benchmark (v0.1.0) of four memory mechanisms (Persistent Memory, RMT, TTT-Linear, Gated DeltaNet) inside a 12-layer 286M-parameter nanochat GPT under identical data and optimizer. Persistent Memory is the only mechanism with a statistically significant gain at 2048 context (-0.0017 BPB, $p = 0.012$, Cohen's $d = -2.72$).
- newton-muon-ddp - public study testing the Newton-Muon optimizer on a parameter-golf budget (8xH100 DDP, 27M-parameter GPT). An honest negative result: full Newton-Muon is 0.007 BPB worse than baseline Muon, traced to torch.compile integration overhead rather than the algorithm, with per-layer condition-number diagnostics and a multi-seed protocol.
- logprobe - public Rust CLI and library (cargo install logprobe). Strict bits-per-byte computation and diagnostics for normalization errors, raw-logit mistakes, and entropy bias across OpenAI, vLLM, Together, and JSONL logprob formats. Companion crate cargo-build-rx is published on crates.io.
- partition-function-inflation - public repository for the partition-function normalization line of work connected to the parameter-golf and BayesGPT compression research.

PRODUCTION PRODUCTS

- Prophet Hacks 2026 - calibrated forecasting agent submitted to the Anthropic / Prophet Arena hackathon. Live at agent.forecastingpath.com/predict with the multi_outcome_retrieval variant (Opus 4.7 plus Brave retrieval plus Kalshi market floor). Brier 0.0379 on the 26-event sample-resolved backtest. Logit-pool blending, disagreement-driven abstention, Fay-Herriot borrowed-strength shrinkage, and a Kalshi-longshot guard. Walkthrough at youtu.be/1ON-WAurV_0.
- TheoremPath.com and TheoremPath iOS - mathematical-statistics and ML reference platform. Next.js, KaTeX, Lean 4 integration through FormalSLT, real-time SSE queue updates, BKT mastery tracking, FSRS spaced repetition, transparent recommender reason codes, and an iOS companion app.
- ConfigPulse - iOS Configuration Profile compliance platform. Live on the App Store. DNS-over-HTTPS filtering, signed .mobileconfig profiles, and gambling-site blocking on iOS non-MDM.
- MoodSpan.org - clinical psychiatry and psychology reference platform. Approximately 793 articles, hybrid retrieval, a groundedness gate, and semantic-entropy hallucination detection.
- Smart-Trends.io - news intelligence pipeline. Approximately 500,000 articles, 36 Airflow DAGs, embedding search, and automated daily reports.
- SmartBreeds and SmartBookshelf.io - computer-vision and multimodal identification apps using YOLO, Swin classifiers, OCR, FastAPI serving, and GPU inference.
- ComputationPath.com and related Path properties - interactive reference pages and labs across statistics, probability, computation, linguistics, and actuarial topics.

EMPLOYMENT

- Mathematical Statistician, Statistics Canada (Aug 2021-Present). Methodology research on small-area estimation (Fay-Herriot with spatial augmentation for AAFC crop yields), equivalence-testing for revision significance,

imputation and variance estimation; R and Python tooling for production survey pipelines.

- Freelance Data Consultant (Dec 2020-Aug 2021). Vaccination analytics for Valid Research UK; population-ecology statistical analyses for the Wildlife Conservation Society of Canada.
- Data Consultant, WHO Regional Office for Europe, Copenhagen / Remote (Oct 2020-Jul 2021). Behavioral Insights team. Vaccination hesitancy and antibiotic uptake analyses across multiple member states.
- Data Scientist (Co-op), BC Public Service, Ministry of Transportation and Infrastructure, Victoria (May 2020-Jan 2021). Trip-level analysis of tens of millions of records; PowerBI dashboards; recurring analytics in R.
- Data Analyst, University of Victoria (Sep 2018-Aug 2020). Environmental DNA / qPCR collaboration with EcoFish Research Ltd.; statistical analysis in R.

SERVICE & VOLUNTEER

- **Peer review (volunteer service):** Workshop Reviewer, ICML 2026 workshops: AI4GOOD and Pluralistic Alignment.
- **Open-source and public-library service:** Four merged pull requests at research-grade organizations (OpenAI parameter-golf, EleutherAI lm-evaluation-harness, the Tao and Kontorovich PrimeNumberTheoremAnd project, and Stanford Centaur PyPantograph), plus the public Lean libraries FormalSLT (MIT) and formal-martingales (Apache-2.0), contributed as unpaid work to public research infrastructure (detailed under Open-Source Contributions above).
- **Community:** Research Assistant, University of Manitoba, Department of Biomedical Engineering, with Dr. Zahra Moussavi (Jun 2012-Aug 2015). Across multiple high-school and undergraduate summers, independently administered several dozen Montreal Cognitive Assessment (MoCA) examinations to Alzheimer's-study participants and supported virtual-reality protocol research.

PROFESSIONAL CREDENTIALS

- Society of Actuaries exams P, FM, IFM, FAM, PA, and SRM - all first-attempt passes.
- Selected certificates: Google TensorFlow Developer Certificate (2021); Deep Learning Specialization (2020); Machine Learning Specialization (Stanford, 2020); How Google Does Machine Learning (Google Cloud, 2020); Natural Language Processing in TensorFlow (2020); Build Basic Generative Adversarial Networks (GANs) (2020); Developing Data Products (Johns Hopkins, 2019).

SKILLS

- Languages: Python, R, TypeScript, Swift, Lean 4, SQL.
- ML stack: PyTorch, FastAPI, Docker, GCP, OpenAI API, Anthropic API, vLLM, Hugging Face transformers, YOLO, Swin.
- Statistical inference: equivalence testing, Bayesian methods, MLE / MAP, bootstrap, MCMC, GLM, PAC-Bayes, survey methodology, small-area estimation, imputation, variance estimation.
- Formal methods and infrastructure: Lean 4, Mathlib4, axiom audits, Vercel, Cloudflare, Postgres / Neon / pgvector, Airflow, GitHub Actions, Clerk, Sentry.
- Current research interests: equivalence-testing methodology for LLM evaluation, disagreement-preserving evaluation frameworks, and formal verification of statistical learning theory.

LANGUAGES, AWARDS, AND INTERESTS

- Languages: English native; French limited working proficiency.
- Awards: JJEM Graduate Award in Mathematics and Statistics, University of Victoria; Tomlinson Engagement Award for Mentoring, McGill University.
- Interests: Saxophone and a music background that informs exploratory work in AI-generated music (Suno); ongoing French language study.