Brendan Long

Seattle, WA
self@brendanlong.com
brendanlong.com
lesswrong.com/users/brendan-long
github.com/brendanlong

Generalist software engineer (14+ years, media codecs to datacenter infrastructure, startups to Big Tech) moving into AI safety. On LessWrong since 2009 (5,000 karma), with independent interpretability research and writing on tokenization, sampling, and positional encoding.

Research and Writing

How Far Apart Does a Model Think Its Tokens Are? Jun 2026

Trained models to learn RoPE position increments in order to inspect the learned structure. Models learned larger gaps at word and punctuation boundaries; middle layers recovered Chinese word boundaries without spaces.

Shorter Tokens Are More Likely Aug 2025

Showed temperature and top-k sampling are biased toward tokens with more completions. Forcing a separate ‘c’ token moved c-initial word frequency from 4% to 10%.

Filler Tokens Don’t Allow Sequential Reasoning Dec 2025

Argued that filler tokens let a model scale parallel but not sequential computation, making “Let’s Think Dot by Dot” evidence *against* hidden long-term reasoning.

Training a Transformer to Compose One Step Per Layer (and Proving It) Apr 2026

Transformers trained normally on a composition task learned a lookup table rather than a multi-step algorithm. An answer-only → full-sequence curriculum recovered the sequential algorithm, verified with logit-lens and ablation.

Work Experience

Production Engineer (IC6) — Meta — Nov 2023 - Aug 2026

- Noticed that the org’s largest and highest-risk stateful services (>10k hosts) had no working load tests. Proposed and ran a cross-org workstream with 10+ teams, raising test coverage from 20% to 60% of OpEx. Promoted to IC6 on this work.
- Built dashboards to identify efficiency opportunities, then worked with service teams to implement and verify them. Reduced server costs by ~$35M/yr while improving reliability.
- Optimized an internal capacity-analysis web app used by 10+ teams; cut p100 load time from 100s to 7s and payload from 35MB to 1MB.
- Mentored four engineers (one advanced to team lead) and a summer intern (received a full-time offer).

Staff Software Engineer — Atlas Exploration — Oct 2021 - Oct 2023

- Designed and built a high-uptime, scalable, low-latency integration with a card processor.
- Improved iOS app signup funnel completion, reducing errors during signup by 95% and improving overall user onboarding rate by ~10% while reducing system complexity.
- Migrated database clients from Prisma to Sequelize. Improved performance and reliability. Reduced server costs by ~75%.

Senior Software Engineer — Arena Analytics — Dec 2016 - Oct 2021

- Refactored an ETL job processing millions of events per run, reducing runtime by 2 orders of magnitude while also improving code readability, memory usage, and debuggability.
- Designed and implemented Dockerized services on ECS with automated deployments to fix biweekly issues with unreliable manual/stateful deployments.
- Maintained and/or wrote OCaml open source libraries (PostgreSQL database driver, Excel file reader, FreeTDS-based SQL Server library).
- Mentored junior engineers on career development and technical topics (one advanced to team lead).

Senior Software Engineer — CableLabs — Jan 2012 - Dec 2016

- Added HTML5 audio, video and text tracks to WebKit (C++) and GStreamer (C). Collaborated with Apple, Google, and others and the HTML5 mailing list.
- One of the editors for MPEG's "MPEG-DASH Conformance and Reference Software" spec.

Software Engineering Intern — RealGo — May 2010 - Dec 2011

Projects

Lion Reader — Dec 2025 - Current

- RSS reader and save later app with AI features (summarization, MCP, narration) and custom handling for Markdown, Google Docs, arXiv and LessWrong posts.
- Built fully AI-assisted (>3,000 commits) with high performance and reliability.

Education

B.Sc., Computer Science — **Colorado State University** — 2009 - 2011

Technologies and Languages

Languages	Python, TypeScript, OCaml, Java, SQL, JavaScript, C++, C, Hack (PHP)
Technologies	PostgreSQL, MySQL, AWS, Git, Docker, GraphQL, REST API's, Node.js, Spring, PyTorch, RunPod