NICK SHAPIRO

Berkeley, CA | nick@nickshapiro.com | nickshapiro.com | github.com/sfgeekgit | linkedin.com/nickshapiro

Independent AI safety researcher and full-stack engineer with 20+ years shipping production systems. Foresight Institute AI Safety grantee (June 2026); AFFINE Superintelligence Alignment Fellow (May 2026). Grand Prize winner, Future of Life Institute AI Safety Creative Contest. Former founder with successful exit (Game Craftsmen LLC, 1M+ users, $3M+ revenue). Empirical approach to alignment: building the testbed infrastructure theorists need to run reproducible experiments on real agent behavior.

AI SAFETY RESEARCH & FELLOWSHIPS

Foresight Institute Grantee Summer 2026

- Full grant awarded: Stipend, expenses, and compute budget. Working from Foresight's San Francisco node, embedded in the AI safety research community.
- Building an open-source experimental framework for reproducible multi-agent AI safety research. Agent-agnostic design (prototype used OpenClaw; framework supports any agent backend).
- Fork-and-test methodology: snapshot running world environments, spin up copies under different experimental conditions, compare behavior from identical starting states.
- Designed to surface surprising emergent behavior for alignment theorists.
- Menu-driven control interface: view running worlds and their budgets, add funds, and spin up new worlds with configurable parameters (agent LLM backends, scenario types, constraints, starting budgets). Designed to make launching new experimental configurations fast and frictionless.
- Configurable scenarios: prisoner's dilemmas, shared resource commons, deception games, prediction markets. Custom budget/credit layer for fine-grained resource control and full observability. (currently under development 2026)

AFFINE Superintelligence Alignment Seminar Fellow May 2026

- Month-long intensive AI safety fellowship at Chateau Hostačov, Czech Republic.
- Peer-teaching curriculum included foundational alignment theory: decision theory, instrumental convergence, Goodhart's Law, orthogonality, CEV, etc. Focus on noticing one's own confusion.
- Engaged substantively with visiting alignment theorists; significantly redesigned the agentspace experimental framework based on feedback from researchers at the seminar.

EXPERIENCE

Game Craftsmen LLC Founder & CEO 2011–2025

- Founded and operated a profitable browser-game studio; successful exit with full investor return.
- 1M+ users, $3M+ revenue, 24/7 availability. Community thriving to this day.
- Built entire initial product stack: custom full-stack framework, backend systems, infrastructure, and security. Scaled from prototype to production. Infrastructure still in use under new owner.
- Raised Kickstarter, angel, and institutional funding; reached cash-flow positive. Hired and managed 20+ employees and 100+ freelancers; established engineering processes across the org.
- This experience has translated very well into managing modern agentic coding workflows.

KlickNation (Electronic Arts) Senior Software Engineer 2009–2011

- Rebuilt backend infrastructure for scale: server migration, caching improvements, query optimizations.
- Designed and shipped monetization events that increased revenue 20–50×. Developed A/B tests and engagement features to boost player retention.

Metaweb (acquired by Google) Software Engineer 2006–2008

- Built application-layer features for Freebase.com — the precursor to the Google Knowledge Graph.
- Developed interactive web UI for exploring complex structured data at scale; implemented typeahead autocomplete with progressive streaming for high-latency workloads.

COG1 Engineering Lead 2004–2006

- Led a team of four engineers; owned all technology operations and strategy. Automated mission-critical systems, redesigned Oracle database architecture, executed full infrastructure migration to Linux servers.

Earlier Roles

Contract engineer across multiple startups and agencies, including Direct Hit Technologies (Ask.com), Online Partners (100M+ pageviews/month), TryMedia Systems (10M+ downloads/year) and Publicis & Hal Riney. Built production systems in Perl, PHP, and Python across diverse environments.

SELECTED PROJECTS

- **The Choice Before Us —** Grand Prize ($10K) winner of the Future of Life Institute AI Safety Creative Contest (300+ entries). Solo-built TypeScript incremental game teaching AGI risk concepts. https://thechoicebeforeus.com
- **Auditron —** Multi-agent RL environment for scalable oversight (OpenEnv Hackathon, March 2026). Supplier agents compete with real incentives to cheat; an auditor agent (GRPO/Unsloth, Qwen2.5-7B) detects fraud from indirect signals only. All agents adapt simultaneously. True positive rate from 24% to 91% over training. https://sfgeekgit.github.io/PyTorchOpenEnv_Auditron_trainer/
- **"Agentspace" Multi-Agent Framework —** Open-source experimental testbed for reproducible AI safety research. Fork-and-test methodology, configurable scenarios, observability layer, agent-agnostic design. https://github.com/sfgeekgit/agentspace (work in progress)
- **Document Brain —** RAG system (Python, FAISS, LangChain) ingesting legal documents in multiple formats; answers always include citations to original source.
- **Hero Stories —** LLM-powered real-time story generation in a live game with 10,000+ daily PvP battles. Grounded in real player histories. "Players absolutely love it." https://www.heroesrisinggame.com/game/hero_stories.game
- **Dandelions —** PyTorch Deep Q-Learning agent built from scratch for a web game. https://nickshapiro.com/dandelion

TECHNICAL SKILLS

Languages: Python, TypeScript/JavaScript, SQL, PHP, Perl
ML / RL: LLM application development, LangChain, multi-agent RL, PyTorch, Deep Q-Learning
Databases: Postgres, MySQL, Oracle
Infrastructure: Linux/Debian administration, Docker, Git, monitoring, server migration, API design
Architecture: Scaling web services, caching strategies, real-time systems, customer-facing platforms

OTHER

- Top 1% finish (Blind Mode), Astral Codex Ten "Who Predicted 2023?" Forecasting Contest
- OEIS A336748 and A336759 Published in the Online Encyclopedia of Integer Sequences (2020).
- 800/800 Analytical GRE.

EDUCATION & PROGRAMS

Forecasting Modeling and Shaping AI Futures	Lens Academy	2026
Technical AI Safety Intensive	BlueDot Impact	2026
Cooperative AI		2026
Recurse Center		2023
MS Computer Science	Michigan State University	2000
BS Astrophysics	Michigan State University	1997