Ben Shapiro

Applied AI · Product Builder · Mindful Technologist

San Diego, CA
hello@shapyro.com · shapyro.com
linkedin.com/in/benshapyro

AI strategy and applied AI lead. Four years of AI work inside real organizations, starting before ChatGPT. At Cadre, an AI consultancy and Anthropic partner, I worked with more than 50 organizations across 20+ industries. I take a messy business problem, find the few workflows where AI actually pays, and get people using it. I also build: Claude skills, an AI assistant suite, and model evaluations.

EXPERIENCE

Cadre AI — Co-Founder & Head of Product & AI Strategy

Dec 2024 – Sep 2026

AI consultancy for mid-market organizations

- **Turned an investment office's AI platform evaluation into two follow-on contracts.** Broke 18 loosely written use cases into single, specified workflows and a 20-source data matrix the team adopted, held six AI vendors to the same seven pass/fail gates, presented two executive reviews, and scoped both follow-on contracts.
- **Closed a 12-month AI program for a federal supply contractor with two operating companies.** Ran department deep dives at both companies and a leadership session, built the use-case library, led three executive readouts, then priced a 29-workstream program with seat budgets and a training plan. One finding, that a quarter of order lines never got a vendor response, needed no AI at all.
- **Led AI discovery and roadmaps for a community bank, a university, a private-credit firm and a marketing agency.** Interviewed four to seven departments at each and gave leadership a ranked opportunity map and roadmap. The bank went on to a ten-month engagement.
- **Trained a private-equity firm to build its own AI tools.** Designed and taught an eight-session Claude program for a 22-person firm, built nine starter skills, and coached staff who built 12 more. The firm ended with 23 skills and a rule that a second person tests each skill first.
- **Directed an AI executive-assistant suite through 46 releases in six weeks.** Daily briefing, meeting prep and a writing assistant in the user's own voice, with Claude writing the code, tested in about a dozen recorded sessions with users.
- **Built about 70 Claude skills and led Cadre's internal skills marketplace.** About 35 were in active use by mid-2026, including a Claude coaching skill that a colleague set up for an organization with no help from me.
- **Led or co-led 47 AI training sessions for teams at 16 organizations** (Jun 2025 – Jul 2026), plus workshops for EO and YPO business owners and SDSU students. Created a data-safety traffic light for AI use and a six-question readiness test.

Tide Rock — Director of AI · private-equity firm

May 2024 – Jan 2025

Brought AI into the firm's own work: deal sourcing and diligence first, then operations, corporate development, HR and accounting.

SHPR (shpr.ai) — AI Product and Tech Consultant · my own studio

Nov 2022 – Present

- **Automated operations for a vehicle-repair company.** Built a system that catches missed service opportunities and automated customer intake, dispatching and reporting: about $30,000 more revenue and $20,000 less cost a month.
- **Shaped products for two AI startups.** For an AI marketplace, validated a new product with a written test plan, tested a prototype with users and reset the roadmap with the CEO. For a sports-prediction startup, led a site redesign that lifted conversions 250%.

Endera Motors — Founding Team, Product Manager · electric school and shuttle buses

Jan 2020 – Jan 2023

TekMate — Founder & CEO · technology tutoring for older adults

Sep 2019 – Apr 2021

SELECTED BUILDS AND EVALUATIONS

- **Tycho, an open-source voice co-pilot for Star Citizen** (github.com/benshapyro/tycho-ai). A Claude brain, chosen on an 82-case eval (97% of decisions right), turns speech into keypresses. A safety gate outside the model holds irreversible actions for a spoken "confirm".
- **Personal context layer with a trust gate**, and Canon, its company-scale design (spec only). 5,760 sources and 52,982 claims, each tied to a quoted source span and checked by a mechanical floor plus an AI judge before it can be served. Calibrated the judge against my own labels: precision 1.000 with zero false passes; one tuning step doubled recall from 0.343 to 0.686, still with zero false passes.
- **Blind-label study that found the judge's blind spot.** Planted 70 fake claims across six kinds of corruption among 20 real ones. The production judge passed all 20 real claims but also 27 of the 70 fakes; I traced the cause to a context window that let it ground claims in text it hadn't seen. A span-only judge agreed with my rulings at kappa 0.878. Of 282 claims cited in real briefs, 0 were wrong.
- **Pre-registered comparison against plain connectors.** 31 frozen tasks, three setups, predictions written first: 29 of 31 held.
- **Gauntlet, an eval harness for coding agents.** Real TypeScript tasks graded against hidden tests in sealed, network-off containers, reporting pass^k and cost. After a test agent escaped its work folder and applied the answer key, I jailed every agent shell in Docker.
- **Financial-aid estimator rebuild.** Rebuilt the tool from a spec with Claude Code in about an hour and a half. The model produces no dollar figures; fixed, tested code does all the math, and it matched a real advisor call to the dollar.

SKILLS AND TOOLS

AI strategy: discovery interviews, use-case scoring, vendor evaluations, roadmaps, pricing, executive readouts.

Building with AI: Claude Skills, Claude Code, the Claude Agent SDK, MCP, n8n, prompt and context design.

Evaluation: judge calibration against human labels, planted-negative tests, Cohen's kappa, pre-registered comparisons, pass^k reliability.

Software: TypeScript, Next.js, Python, Postgres, Docker, Vercel. **Teaching:** workshop design, train-the-trainer programs, coaching.

Education: B.S. in Administration and Entrepreneurship, San Diego State University.