

# Dmitry Titkov

Israeli citizen · available to work from Israel · titkov.mitia@gmail.com · @insearchoffame · github.com/yolojewjitsu

Software engineer building AI evaluation systems, tool-using agents, and the backend infrastructure around them. My recent work spans RL-style environments, reward functions and graders, adversarial QA, rollout systems, provider reliability, and operated agent tooling.

## EXPERIENCE

---

### Founding Engineer, Parsewave

Oct 2025 - Jun 2026 · Remote (San Francisco)

Early engineer at an AI-evaluation / RL-data company supplying eval tasks and training data to frontier AI labs.

- Designed and shipped agentic evaluation tasks and RL-style environments across many domains: multi-turn tool use, coding, long-horizon, multimodal, finance (GDPval), and hardware sensors. Built them both through automated generation pipelines and by hand.
- Engineered reward functions and graders resistant to specification-gaming (Goodhart), using semantic verifiers and an anti-leakage discipline so reward reflects real capability rather than shortcuts.
- Worked extensively on the CI validation and reward-scoring service that ran agent rollouts against benchmark tasks and aggregated their reward scores.
- Operated fleets of autonomous task-crafting agents and ran adversarial QA over their output.
- Built backend and fleet infrastructure for the eval platform: provisioning automation, an LLM-provider gateway, and monitoring.

## INDEPENDENT ENGINEERING & PUBLIC ARTIFACTS

---

- Built and operate a self-hosted multi-bot system in Rust: JSON-schema-constrained host tools, bounded WebSearch-only subagents, persistent memory, and a human-reviewed browser bridge. Telegram-facing model processes do not receive shell or unrestricted file tools; the trust boundaries and residual risks are documented in a public case study.
- Built a live full-stack jewelry design product with Next.js and FastAPI, generated-mesh and constrained-CAD paths, geometry preparation, artifact storage, automated tests, and explicit castability limitations.
- Published four focused, zero-dependency Python utilities: LoopGuard, EvalGuard, FailGuard, and TokenGuard.

## SKILLS

---

**AI / agents / eval:** RL-environment and eval-harness design · reward and grader design · anti-reward-hacking (Goodhart / specification-gaming) · adversarial QA · agent-trajectory and failure analysis · multi-agent orchestration · typed tool contracts · human-approval boundaries

**Backend / systems:** Rust · Python · TypeScript · FastAPI · async / concurrent systems · Postgres · Redis · distributed systems / SRE · Docker · Prometheus

**Infra / DevOps:** Hetzner fleet automation · provisioning CLIs · CI/CD · Linux

## EARLIER

---

**Web and smart-contract developer** (2022-2025): React, Node.js, Nest.js, Solidity / Vyper / Tact, blockchain integrations, Docker, CI/CD, Linux.

## EDUCATION

---

**Bauman Moscow State Technical University (BMSTU)**, Specialist Degree, Computer Science · 2017-2023