

# Atul Saxena

Research & Product Engineer | RL / Agents | ML Systems | Technical Product  
atulhoonmain988@gmail.com | +91-9889852511 |  LinkedIn |  GitHub | Portfolio

## Profile

High-agency engineer and researcher who works across the full MTS problem surface: **research, systems, product judgment, customer understanding, and execution**. Owns ambiguous problems from user/research discovery and technical framing through implementation, evaluation, productionization, and feedback loops; comfortable making tradeoffs with researchers, engineers, and stakeholders instead of waiting for a fully specified ticket. Currently an **RL Production Environment Engineer at Snorkel AI**; previously led RL, agent-evaluation, and data-curation work at **Bespoke Labs**, with founder-level experience taking ML/software products from zero to production. Published mechanistic-interpretability research, built RL/simulation and reasoning systems, contributes across major open-source codebases, and communicates technical work through public maintainership, documentation, stakeholder conversations, and conference speaking.

## Core MTS Strengths

|                    |                                                                                                                                                                                                                                       |
|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| RL / Agents        | RL environments, agent evaluation, reward and verifier design, task generation, long-horizon trajectories, Gymnasium, RL/RLVR workflows, failure-mode analysis                                                                        |
| ML Research        | PyTorch, mechanistic interpretability, activation patching, causal interventions, probing, ablations, MCTS-guided search, multimodal/VLA experimentation                                                                              |
| ML Systems         | Training/inference pipelines, evaluation harnesses, dataset/versioning, model packaging, synthetic data generation, Hugging Face, Parquet, LeRobot/LIBERO adapters                                                                    |
| Systems            | Python, Rust, Go, Zig, TypeScript/Node.js, FastAPI/Flask, PostgreSQL, Redis, Docker, Linux, CI/CD, unit/integration testing, production debugging                                                                                     |
| Product / Customer | User/stakeholder discovery, problem framing, product judgment, requirements-to-technical-scope translation, engineering handoff, tradeoff communication, feedback loops, async collaboration, external-facing technical communication |
| Security           | OAuth2/OIDC, JWT, RBAC, secrets handling, input validation, OWASP Top 10, least privilege, STRIDE threat modeling                                                                                                                     |

## Experience

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                            |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------|
| <b>RL Production Environment Engineer</b>   <i>Snorkel AI</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                          | <b>2026 – Present</b>      |
| Remote                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                            |
| <ul style="list-style-type: none"><li>Build and debug production-grade RL environments for model training and evaluation: task design, reward/verifier logic, failure-case construction, reproducible execution, and agent-behavior analysis.</li><li>Turn complex software and ML tasks into reliable environments that expose genuine capability gaps rather than superficial pass/fail behavior.</li></ul>                                                                                                          |                            |
| <b>Product &amp; AI Systems (Contract)</b>   <i>Chalkbyte / Pyndra</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                 | <b>2026 – Present</b>      |
| Remote                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                            |
| <ul style="list-style-type: none"><li>Research business and user needs, define product requirements, and translate them into actionable technical scope for engineering.</li><li>Iterate with engineering from stakeholder feedback and contribute to product and AI-capability development.</li></ul>                                                                                                                                                                                                                 |                            |
| <b>Pod Lead - RL, Agent Evaluation &amp; Data Curation</b>   <i>Bespoke Labs</i>                                                                                                                                                                                                                                                                                                                                                                                                                                       | <b>Mar 2026 – Jun 2026</b> |
| Remote                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |                            |
| <ul style="list-style-type: none"><li>Led a pod focused on RL workflows, agent-evaluation frameworks, task quality, and data curation for frontier-model development.</li><li>Designed and reviewed evaluation harnesses and failure cases, with emphasis on whether tasks and graders measured the intended capability rather than exploitable shortcuts.</li><li>Contributed upstream to Bespoke Curator (PR #717), adding normalized Mistral <code>finish_reason</code> handling and regression coverage.</li></ul> |                            |
| <b>Co-Founder &amp; ML Systems Engineer</b>   <i>PKBoost AI Labs</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                   | <b>2025 – Present</b>      |
| Remote / Hybrid                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                            |
| <ul style="list-style-type: none"><li>Lead ML systems work in Rust across training/inference pipelines, reproducible evaluation harnesses, dataset/versioning conventions, and production-oriented model packaging.</li><li>Build boosting/GBDT-style components and drift-aware workflows emphasizing latency, memory efficiency, deterministic runs, monitoring, and safe rollback.</li></ul>                                                                                                                        |                            |
| <b>Co-Founder &amp; CTO</b>   <i>Boostr AI by GradeBoostHub</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                        | <b>Feb 2020 – May 2026</b> |
| Remote / Hybrid                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |                            |
| <ul style="list-style-type: none"><li>Owned product and engineering end-to-end across database design, APIs, authentication, dashboards, production deployments, observability, and incident response.</li><li>Shipped and operated Docker-based infrastructure with RBAC, JWT/OAuth, backups, monitoring, structured logging, and hardened endpoints.</li></ul>                                                                                                                                                       |                            |

## Experience (continued)

|                                                                                                                                                                                                                                                                                                            |                            |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------|
| <b>Founder &amp; Lead Developer</b>   <i>UPAI Platform</i>                                                                                                                                                                                                                                                 | <b>May 2024 – Present</b>  |
| Remote                                                                                                                                                                                                                                                                                                     |                            |
| <ul style="list-style-type: none"><li>• Built an LLM orchestration platform spanning 100+ models with secure backend services, rate limiting, async jobs, retries, caching, and predictable API contracts.</li><li>• Built an LLM Orchestrator and SOTA benchmarking leaderboard for NVIDIA VIM.</li></ul> |                            |
| <b>AI Mentor &amp; Technical Lead</b>   <i>Edge Lab</i>                                                                                                                                                                                                                                                    | <b>Jun 2025 – Sep 2025</b> |
| Remote                                                                                                                                                                                                                                                                                                     |                            |
| <ul style="list-style-type: none"><li>• Led code reviews, testing practices, and technical mentorship to help teams ship production-ready projects with clean interfaces and deployable pipelines.</li></ul>                                                                                               |                            |

## Selected Research & Engineering Projects

|                                                                                                                                                                                                                                                                                                                                                                                  |             |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| <b>GS-DroneGym</b>   Python, Gymnasium, 3D Gaussian Splatting, VLA   GitHub                                                                                                                                                                                                                                                                                                      | <b>2025</b> |
| <ul style="list-style-type: none"><li>• Built a pip-installable 6-DOF quadrotor environment with photorealistic rendering, RGB/depth observations, safety constraints, language-conditioned navigation, and reproducible trajectories.</li><li>• Implemented a synthetic dataset factory and LeRobot/LIBERO adapters for multimodal training and evaluation workflows.</li></ul> |             |
| <b>LEMMA</b>   Rust, MCTS, neuro-symbolic reasoning   GitHub                                                                                                                                                                                                                                                                                                                     | <b>2026</b> |
| <ul style="list-style-type: none"><li>• Co-developed a mathematical reasoning engine combining a large symbolic transformation-rule library with transformer-guided MCTS, focused on explicit and verifiable reasoning trajectories.</li></ul>                                                                                                                                   |             |
| <b>Reachy Mini Talkie</b>   Reachy Mini SDK, MuJoCo, STT/LLM/TTS   GitHub                                                                                                                                                                                                                                                                                                        | <b>2026</b> |
| <ul style="list-style-type: none"><li>• Built on direct request from Hugging Face’s co-founder/CEO; integrated faster-whisper, Talkie-1930-13B, Kokoro TTS, rolling memory, synchronized robot motion, and simulator support.</li></ul>                                                                                                                                          |             |
| <b>Sieve / VRAMancer / heapwatch</b>   Rust, Zig, CLI/TUI, JSON automation                                                                                                                                                                                                                                                                                                       | <b>2026</b> |
| <ul style="list-style-type: none"><li>• Built developer-facing systems tooling for secret-leak prevention, VRAM inspection, and heap telemetry with automation-friendly interfaces, safe defaults, and repeatable workflows.</li></ul>                                                                                                                                           |             |

## Research, Open Source & Technical Communication

- **Publication:** Saxena, A., & Kharat, P. (2026). *Investigating Trajectory-Conditioned Semantics and Dynamic Computational Pathways in GPT-2-Small: A Mechanistic Interpretability Study*. doi:10.5281/zenodo.20098563
- **Open-source engineering:** Builds and maintains public tools including GS-DroneGym, Reachy Mini Talkie, Sieve, VRAMancer, heapwatch, LEMMA, Blackdrome Device Sync, WebShotify, and related developer/ML systems; also contributes across Bespoke Curator, OpenTelemetry Python, Pandera, Apache Superset, Apache Airflow (multiple PRs), Qiskit, Kornia, Litestar, OpenMed, toqito, OmniGent, TorchGeo Bench, Astral uv, Monocle, Eigent, Relic, InsForge Swift, ACoT-VLA, data-pipeline-env, and AIContract.
- **Contribution scope:** production fixes and features across ML/data infrastructure, observability, SDK/storage behavior, framework correctness, performance, packaging/dependencies, CLI/agent UX, RAG reliability, provider normalization, security-oriented tooling, and regression testing in Python, Rust, TypeScript/Bun, Swift, and mixed-stack repositories.
- **Technical communication:** Accepted 40-minute talk at Opportunity Open Source Conference 4.0, IIIT Allahabad - *From Paper to Working System: Reproducing AI Research When the Weights Aren’t There*; comfortable writing design notes, changelogs, contributor-facing docs, and explaining engineering tradeoffs to technical and non-technical stakeholders.
- **Community:** Founder, TechBlock Coding Club (IIT Madras); organized technical activities and mentored peers through project delivery.

## Education & Recognition

|                                                                                                                                                                                                                   |                       |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| <b>Indian Institute of Technology Madras</b>                                                                                                                                                                      | <b>Chennai, India</b> |
| <i>B.S. Data Science and Applications (ongoing)</i>                                                                                                                                                               |                       |
| <b>Recognition:</b> Techkriti Hackathon Finalist (IIT Kanpur)   Cyber Sentinel Winner (IIT Madras)   Top 250, Lovable Shipped Pre-Accelerator   Professional Diploma in Agile & Project Management, MTF Institute |                       |