

Shaurya Singh

Builder & teacher — I find what hurts, then build at it.

Seattle ↔ New Delhi · spsjadon9@gmail.com · linkedin.com/in/spsthebest · github.com/spsthebest · shauryasingh.com

EDUCATION

University of Washington

Expected 2027

B.S., Computer Science & Data Science — Seattle, WA

- **Systems:** Hardware–Software Interface (x86-64, memory, the C/assembly boundary) · Software Design & Implementation · Data Structures & Algorithms.
- **ML & data:** Machine Learning · Computer Vision · Natural Language Processing · Statistical Analysis · Multivariable Calculus.

Modern School, Barakhamba Road

2023

High school diploma, New Delhi — calculus, data structures, database management, Boolean algebra.

EXPERIENCE

Undergraduate Research Assistant — DAIS Lab (Dr. Dong Si), University of Washington

May 2025 – Present

- DeepTracer, the cryo-EM → 3D-protein-structure pipeline: data preprocessing and model training in PyTorch and OpenCV on backbone trace alignment; PyMOL rendering and annotation.
- The unglamorous half is the real job: density maps arrive at different resolutions and contour levels, so the preprocessing has to normalise them without quietly deleting the signal the model is supposed to find.
- Renders are made to be argued with — a structural biologist has to be able to look at a trace and say where it is wrong, which means showing the uncertain regions rather than smoothing them into something confident and false.

AI Engineer (Internship) — Superpowr

2025 – Present

- The correspondent layer: a voice-and-email agent handling first-contact recruiting end to end, with a guardrail around every action it is allowed to take.
- Guardrail design, not prompt wishes: a confirmation gate that has to be staged and matched before any irreversible action fires; every tool call scoped to the acting user rather than trusting the model to stay in its lane; fail-closed defaults, so a missing caller identity refuses the call instead of guessing at one.
- Built an adversarial evaluation battery — tests that attack the system rather than confirm it — after a green regression suite masked a broken live path. Seeded known-good and known-bad controls into the battery: a run that flags zero controls is treated as void, not as clean.
- Worked the failure modes an agent system actually dies of: retries that duplicate side effects, a send path conflated with its persistence so a failed write reports "not sent" on a message that went out, and races where two concurrent turns both think they hold the gate.

Founder & AI Developer — TurtleShell, New Delhi

Nov 2022 – Oct 2023

- Tourist-safety startup. K-means (K=40, chosen by the Elbow Method) over 10,000+ LAPD crime datapoints to flag high-risk zones; PySpark for the clustering pass, Matplotlib to see what the clusters actually were.

- Swift / CoreLocation iOS app: real-time SOS, and routing that took intent as an input — the fastest route and the route you'd want as a tourist are not the same route, so the app asked which one you were after.
- The honest limit, stated then and now: clustering historical police reports finds where crime is *reported*, not where it is. That gap is a product decision, not a modelling detail — it drives who gets warned about which neighbourhood.
- Accepted into Microsoft for Startups; deliberately shut down in Oct 2023 rather than run it on a claim I couldn't stand behind.

Co-Founder & CTO — SkillTern

May 2022 – Aug 2023

- Education nonprofit teaching computer-science and AI literacy to 350+ students. Built the curriculum, website, and certification system; 4+ nonprofit partnerships; ran volunteer onboarding.
- Built it to outlive me on purpose: standardised the stack, wrote the volunteer training down instead of carrying it in my head, and made certification issuable by someone who had never met me. Then stepped off.

Dining Student Assistant — UW Housing & Food Services

Jan 2024 – Apr 2025

- Served 200+ students daily in a high-volume campus dining operation; trained and onboarded new team members. On the résumé because it is true and it overlapped every project above it.

SELECTED PROJECTS

- **Automated Bug Detection** (Python, AST, regex) — static analysis for Python and JavaScript across 20+ bug and vulnerability classes, shipped as a VS Code extension; 89% on the repositories I tested it against. Regex finds the shape of a bug; the AST pass is what tells a real call from the same words inside a string or a comment. The two disagree often enough to be worth running both.
- **CiteGuard** — a verification rail for legal citations that sits *outside* the model and cannot be skipped (prototype). The design bet: a model asked to check its own citations will agree with itself, so the check has to live where the model has no vote.
- **“Harnessing Language Models to Mitigate AI-Driven Phishing”** — policy paper, co-authored with two others (July 2023). Not a detection model: an *incentive structure*. A shared flagged-sender database, maintained by academic institutions for course credit, funded by government, fed with training data by private companies who get the model plus loss reimbursement in return — with a cross-disciplinary oversight board so no single party defines what counts as phishing. Runner-up, WatGov Hackathon (University of Waterloo).
- **IKEA x TKS — GenZ and the metaverse** — led a strategy engagement for IKEA through The Knowledge Society: drove primary consumer research and practitioner interviews into a browser-based room-layout world where Gen-Z designs surface on in-store kiosks and top-voted creators earn a revenue share. The argument was participation over reach. The finding that shaped the whole deliverable: Gen-Z detects an inauthentic brand entrance instantly, so the deck had to answer “what does this give people that the store cannot?” before it was allowed to answer anything else.
- **Computer vision & ML build-outs** — real-time face detection (OpenCV Haar cascades, multi-scale, webcam), driver-drowsiness detection, hand-gesture recognition (MediaPipe), leaf-disease and plant-disease CNNs, speech-emotion recognition, and a fake-vs-real news classifier. The portfolio of a person learning the field by building one of everything in it.
- **H2OAquatics** — an autonomous underwater vehicle aimed at ocean acidification. The first time the reflex showed up.

HOW I BUILD WITH AGENTS

I do not hand work to a model and hope. I direct it the way you'd direct engineers you are responsible for — which means knowing the code well enough to catch it being wrong, and building the machinery that catches what I miss.

- **Orchestration.** Work is decomposed into bounded tasks with exclusive ownership of the files they touch, then run in parallel waves with a barrier between them. Models are matched to jobs by cost and by how expensive being wrong is — the cheap tier does mechanical sweeps, the strongest one does judgment and final review. Fan-out buys coverage, not speed.
- **Verification that tries to break things.** Reviewers are instructed to *refute* a claim, not confirm it — a finding that survives an attempt to disprove it is plausible, not proven. Every battery carries seeded controls, and a run that catches none of them is void rather than clean.
- **Measured or UNKNOWN.** A number ships with the method that produced it or it does not ship. This site is the working example: every figure on it is re-derived from the built artifact at build time, and the build fails if a printed number and the artifact disagree.
- **Harness truth is not wire truth.** A check that reads the build does not know what the edge actually serves. So the same site is verified twice — once against the artifact, once against the live origin — after a green local gate sat happily on top of production serving injected scripts that broke the page's own claims.
- **The irreversible wall.** Reversible work runs unattended; anything that cannot be taken back — publishing, credentials, destructive migrations — stops and waits for a human. That line is what makes the autonomy on the other side of it safe, and I hold it even when it is slower.
- **Steering the model itself.** Personas, hard rules, and output contracts, plus evals built to attack rather than flatter — because the failure mode of a helpful text predictor is agreeing with you.

WRITING

- Four self-taught neuroscience essays on Medium (@spsjadon9), 2022–23 — optogenetics, brain–computer interfaces, connectomics, and a CNN from scratch. 1,500+ views.

SKILLS

languages	Python · Java · C / C++ · TypeScript / JavaScript · Swift · SQL · HTML / CSS
ml & vision	PyTorch · TensorFlow / Keras · scikit-learn · OpenCV · MediaPipe · NLTK · NumPy · SciPy · Pandas · Matplotlib / Seaborn · PySpark · PyMOL
agent systems	Multi-agent orchestration · tool/function-call design · guardrails and confirmation gates · adversarial evaluation and control probes · prompt and context engineering · output contracts · retrieval and grounding
systems & tooling	Git · Linux / Bash · GDB · x86-64 and the C/assembly boundary · REST APIs · MySQL · JUnit · Gradle · Flask · Astro · Cloudflare Pages · static analysis (regex + AST)
data & viz	Power BI · Tableau · Plotly.js · Chart.js · statistical analysis · clustering (K-means, DBSCAN)

The narrative version — with the parts a résumé flattens — is at shauryasingh.com.