

Shubham Shinde

Research software and machine learning engineer · shubhamshinde.github.io

Bristol, UK shubham.edu.ms@gmail.com +44 7352 671872 github.com/ShubhamX57 linkedin.com/in/shubhams6

I build scientific and machine learning software, and I test whether its results can be trusted. MSc in Scientific Computing with Data Science from the University of Bristol, with three years of professional experience across a defence research programme, a machine learning company and an engineering firm, plus a year of robotics research. Recent work spans a silent numerical failure in the *kinisi* library, Bayesian neutron scattering analysis, a distributed Higgs analysis and grounded retrieval.

Education

2025 to 2026 Bristol, UK	MSc Scientific Computing with Data Science University of Bristol · Distinction expected · SARTHI Maharashtra Government Foreign Scholarship
2022 to 2023 Pune, India	PG Diploma in Data Science and Artificial Intelligence Savitribai Phule Pune University · GPA 8.5/10 · Thesis on building shadow detection from aerial imagery
2017 to 2020 Pune, India	BE Mechanical Engineering Savitribai Phule Pune University · University Merit Scholarship, 2018

Publications and talks

- Shinde, S. When a covariance matrix silently breaks your diffusion analysis. Lightning talk, RSECon26 (UK Research Software Engineering Conference), 2026.
- Shinde, S., et al. Building shadow detection using aerial imagery with attention-enhanced YOLOv4. JETIR 10(9), 2023. **Best Research Paper Award**, Savitribai Phule Pune University.

Technical skills

Scientific computing Python, NumPy, SciPy, HDF5 and h5py, MDAnalysis, emcee, Matplotlib	Numerical methods Bayesian MCMC, numerical linear algebra, Monte Carlo methods, uncertainty quantification, model validation	High-performance computing SLURM, MPI, OpenMP, Cython, Numba, PySpark, Dask, multi-GPU; Monte Carlo code 4.7× faster than its NumPy version
Machine learning PyTorch, scikit-learn, XGBoost, LightGBM, OpenCV, MLflow, DVC, RAG, cross-encoder reranking, vector search	Engineering practice FastAPI, PostgreSQL, Docker, Kubernetes, RabbitMQ, Prometheus, GitHub Actions CI/CD, pytest, mypy, Git	Languages Python, SQL, Bash, JavaScript and TypeScript, MATLAB

Experience

Jun 2024 to Aug 2025 Pune, India	Data scientist and project engineer ARDE research programme, Raj Security and Facility Management Pvt Ltd - Designed authenticated, versioned FastAPI microservices serving sensor data to dashboards at sub-100 ms on a restricted offline network; Docker CI/CD cut release time from hours to under ten minutes. - Built a parallelised PySpark and SLURM pipeline ingesting over 20 GB of sensor data a day, with automated failure recovery, schema validation and audit logging into PostgreSQL for full traceability. - Developed deep learning models for elevation-angle prediction from sensor and imaging data, and an XGBoost and LightGBM ensemble that cut prediction error by roughly 42% against the legacy baseline. - Gathered requirements from stakeholder teams, reported results to engineers and management, and ran on-call production support with root-cause analysis, at a mean time to resolution under four hours.
Jun 2023 to May 2024 Satara, India	Robotics research experience Rayat Science and Innovation Activity Centre Developed machine learning features for educational robotics: object-recognition models on robot sensor data and Python path-planning and navigation algorithms, demonstrated to researchers and students.

Feb 2023 to May 2023 Remote	Machine learning research engineer Spartifical Pvt Ltd – Developed a hybrid CNN and Transformer classifier reaching 89.7% accuracy on NASA lunar terrain imagery, integrating NASA dataset APIs with internal annotation systems. – Built a full-stack data service with FastAPI inference endpoints and a React monitoring dashboard, with pytest and Jest suites at over 90% coverage, packaged with Docker, DVC and CI for one-command reproduction.
Nov 2020 to May 2022 Satara, India	Design engineer Gurudatta Enterprises Pvt Ltd – Built 3D models and drawings in SolidWorks and CATIA V5 with tolerance stack-up analysis across 200+ prototype test cycles, and automated test-bench data logging in Python, cutting post-processing time by roughly 60%.

Research and open-source projects

	Numerical stability of covariance matrix inversion in kinisi MSc dissertation, University of Bristol, supervised by Dr Andrew McCluskey – kinisi carries the full MSD covariance through a least-squares fit; in some runs the matrix becomes non-positive-definite, silently degrading results. Built a testbed with known ground truth to separate artefact from signal. – Failure rate fell from 90% to 2% as sampling increased, isolating noise in long lag-time variance; benchmarked eigenvalue, ridge and shrinkage reconditioning and proposed an adaptive eigenvalue floor. ○ Validation across noise levels is ongoing; a consistency check is open upstream as pull request #222.	2025 to 2026
	gens: quasi-elastic neutron scattering analysis library Author, open source · gens.readthedocs.io – End-to-end toolbox for ISIS Mantid data, combining Fickian, Chudley–Elliott and Singwi–Sjölander diffusion models with measured resolution; joint-Q Bayesian MCMC with Gelman–Rubin diagnostics, validated on LET benzene data.	2025 to present
	Distributed ATLAS H → ZZ* → 4ℓ Higgs analysis Open source · github.com/ShubhamX57/HZZ-Distributed-Computing – Reproduces the ATLAS Higgs analysis at 4.5σ, distributing 40 ROOT files across horizontally scaled Docker workers through RabbitMQ, with exponential-backoff retries, a dead-letter queue, crash-recovery checkpointing and Prometheus metrics.	2026
	arxiv-rag: grounded retrieval over research papers Open source · github.com/ShubhamX57/arxiv-rag – Hybrid BM25 and BGE-M3 retrieval over 204 papers: reranking lifted MRR@10 from 0.801 to 0.955, and verbatim-quote verification gave zero false grounding. 224 tests.	2026
	Physics-informed machine learning for pore characterisation Open source · github.com/ShubhamX57/biscuit-dunking-analysis – Tested the Washburn capillary-flow equation on 3,000 measurements, then inverted it into a pore-radius feature: a random forest rose from 81% to 96% accuracy, and a hybrid physics and ML pipeline lifted R² from 0.79 to 0.985.	2026
	scisolve: agentic scientific computing with grounded answers Open source, MIT · github.com/ShubhamX57/scisolve – Solves ODE, optimisation, fitting and linear algebra problems by executing NumPy and SciPy code; every number must trace to a computed value and pass one of six independent checks. 122 tests. ○ In development: five of seven milestones complete, tested through a scripted harness and not yet run against a live model API.	2026

Earlier projects

2023 SPPU	Cervical cancer detection with machine learning Department of Technology, Savitribai Phule Pune University · Best Project Award Compared six classifiers with cross-validation, prioritising recall and ROC-AUC for screening.
2019 to 2020 BE capstone	CFD analysis of wind turbine performance Mesh- and convergence-checked simulations, processed in Python and MATLAB into power and C _p curves.
2018 to 2019 BE project	Shear-strength test fixture for cement-bonded tiles CAD design verified with parametric Python and MATLAB stress and safety-factor analysis.