

Md Arifuzzaman

Silchar, Assam, India | +91 99579 92553 | mdarifuzzaman481@gmail.com
linkedin.com/in/arif481 | github.com/arif481 | huggingface.co/arif481

EDUCATION

National Institute of Technology Silchar, Silchar, Assam, India 2022 – Jun 2027 (expected)
B.Tech, Computer Science and Engineering
ICCR Suborno Jayanti Scholarship, Indian Council for Cultural Relations (2022)

PUBLICATIONS

M. Arifuzzaman, S. Mondal, L. Lalhriatpuii, B. Basu, A. Nandi. *Reliable Code Generation on Low-Cost Hardware: Visible-Test Filtering for Small Open Language Models.* Submitted to E2A 2027 (Springer LNEE).

RESEARCH EXPERIENCE

Summer Research Intern, National Institute of Technology Silchar May 2026 – Jul 2026

- Raised pass@1 of Qwen2.5-Coder 1.5B, 3B and 7B (run on Kaggle's free GPUs) by 4.3–9.8 points on HumanEval+ and 8.5–10.2 on held-out MBPP+ ($p < 0.02$, exact McNemar test) by sampling up to 20 programs and keeping the first that passes the task's own examples; the 7B model went from 41% to 64% on 100 LiveCodeBench problems, with no correct first answer lost.
- Built SymLoop, a repair loop in which CrossHair and Hypothesis try to break model-written contracts; at about four calls per problem it added 0.4 points against 4.3 for filtering.

Summer Research Intern, National Institute of Technology Silchar May 2025 – Jul 2025

- Applied machine learning and computer vision to detecting glaucoma from eye scan images, using TensorFlow, PyTorch, OpenCV and scikit-learn.

PROJECTS

linux-agent-bench | *Python, Xvfb, Claude Code, Gemini CLI* Oct 2026 – present

- Building a benchmark that scores AI agents on Linux desktop tasks for task success and for safety; some tasks are traps, such as a note with planted instructions the agent must ignore.
- Each run gets its own X display, D-Bus session and home folder; checks read files and app logs, never the agent's claims, and CI requires every task's scripted solution to pass and a do-nothing agent to fail.

linux-computer-use | *Python, xdotool, Pillow* Oct 2026

- Wrote a Claude Code plugin that gives Claude computer use on Linux X11 desktops (the built-in version runs only on macOS and Windows): 1280 px screenshots in about 0.12 s, region zoom, clicks and drags, and typing guarded by a window-title check so keys can't land in the wrong app.

CrossLingual-Sentiment | *PyTorch, Hugging Face Transformers* Jan 2026

- Fine-tuned XLM-RoBERTa for English-to-Bengali sentiment transfer: zero-shot from 64,000 English sentences reached 67.8 macro-F1 on Bengali, 100 Bengali comments gave 73.7 and both languages together 87.9, against 75.6 for TF-IDF; published the model and a demo Space on Hugging Face.

bd-election-2026 | *React, TypeScript, Firebase, Gemini API* Feb 2026

- Built a live results tracker for Bangladesh's February 2026 election, which has no official results API: it collects results from news reports with Gemini 2.5 Flash-Lite and Google Search grounding, scores each one 0–100 on several checks, and routes low scores to an admin review queue.

Quantum vs classical random numbers (final-year project) | *Python, Qiskit* Dec 2025 – present

- Wrote a Python package that benchmarks random bits from a Qiskit-simulated qubit, ANU's quantum RNG service and the OS CSPRNG, and estimates the cost of a Grover attack on P-256 keys seeded from them.

EXPERIENCE

Summer Intern, Northeast Frontier Railway, Silchar May 2025 – Aug 2025

- Developed a web app and a mobile app for the Coach Depot, Silchar.
- Built an IoT alert system on two ESP32 boards: a sensor node sends water-level and chain air-lock triggers over Wi-Fi to a master node, which sends an SMS through a Quectel EC200U LTE module and posts its GPS position to a Flask API with a live dashboard.

Publicity Head, E-Cell, NIT Silchar Aug 2023 – Jun 2025

- Promoted from Junior Associate after 10 months; ran publicity and outreach for E-Cell events.

SKILLS

Languages: Python, TypeScript, JavaScript, C, SQL

ML and LLMs: PyTorch, Hugging Face Transformers, TensorFlow, scikit-learn, OpenCV, Ollama, Gemini API, Claude Code, EvalPlus, LiveCodeBench

Tools: Linux, Git, GitHub Actions, Kaggle, LaTeX, Flask, Firebase, Qiskit, ESP32