Position: Software Developer / Pipeline Engineer (Flexible – Freelance to Full-time)

Location: Flexible (Mumbai, Bangalore, remote – deliverables-focused)

We are seeking a **software developer** to help build and scale data-processing pipelines for large volumes of astronomical data from a range of **satellite and ground-based observatories**. The role involves transforming research-level machine learning models into robust, scalable systems that can handle diverse astrophysical datasets efficiently.

This position is open in scope: from freelance/contract-based engagements to a longer-term full-time hire, depending on mutual fit and interest.

Key Responsibilities:

- Develop and maintain scalable data-processing pipelines for astronomical datasets.
- Translate machine learning research prototypes into production-ready code.
- Optimise algorithms for performance, scalability, and reproducibility.
- Use of high-end GPUs capable of real-time detection, classification, and analysis of PetaBytes of astronomical transients data across a large electromagnetic spectrum (radio, optical, X-ray, gamma-ray observations)
- Collaborate with scientists and students to integrate workflows into broader research frameworks.
- Maintain version control, testing practices, and documentation for codebases.

Qualifications:

- Strong programming skills in Python (experience with C++/CUDA is a plus).
- Familiarity with ML frameworks (e.g., PyTorch, TensorFlow), scientific computing libraries (NumPy, SciPy, pandas, etc.) and more generic, scalableTransformer-based models.
- Experience building data pipelines, APIs, or distributed workflows.
- Knowledge of AI/ML concepts (not necessarily deep expertise, but comfort with implementation).
- Strong problem-solving and communication skills.
- Ability to work independently and deliver reliably against milestones.

Nice-to-have:

- Comfortable using generative-AI algorithms
- Experience with cloud computing platforms (Azure, AWS, GCP) or HPC clusters.
- Familiarity with containerization (Docker, Singularity) and workflow orchestration tools (Airflow, Nextflow, etc.).
- Background in astrophysics, physics, or related sciences (not required, but a plus).

What we offer:

- Flexible engagement (freelancer, part-time, or full-time).
- Opportunity to apply coding skills to large-scale scientific data challenges.
- Deliverables-driven environment with location flexibility.
- Collaboration with international teams at the intersection of science and AI.
- Salaries start from 1.5 Lakh Rupees per month (including allowances) and are negotiable; pay is commensurate with experience and capability of the applicant. For part-time Depending on the number of hours committed, part-time applicants will be compensated on a pro-rata basis.
- This position can be extended for up to 3 years subject to positive and successful annual performance reviews.

If you are a strong coder interested in scaling ML-driven pipelines for cutting-edge science, we'd love to hear from you. Applications should be addressed to hanasoge@tifr.res.in with the subject line "Application for pipeline developer"

Required Materials

1. **Cover Letter**
    - A short letter explaining your motivation, prior experience in software engineering, and interest in working on scientific/AI pipelines.

2. **Curriculum Vitae (CV)**
    - Detailed work experience, coding skills, relevant tools and frameworks (Python, ML libraries, data engineering, cloud platforms, etc.).
    - Should include past projects (academic or industrial) that involved large datasets, pipelines, or ML integration.

3. **Portfolio / Code Samples**
    - GitHub, GitLab, or Bitbucket repositories (or links to equivalent code samples).
    - If proprietary work cannot be shared, a description of contributions to past software projects.
    - Sample Project Documentation (if available) — e.g., user manuals, technical writeups, or pipeline documentation.

4. **Technical Statement (1–2 pages)**
    - Outline of experience in software design, ML pipeline integration, database management, or distributed computing.
    - Emphasis on scaling code for large datasets.

5. **References**
    - Contact details for 2–3 professional referees (mentors, supervisors, collaborators).