

Action Plan

Asks from data governing bodies:

Government: Allocate funding for open datasets tailored to India. Set standards for data collection. Provide anonymized governmental data.

Private Sector: Contribute industry-specific datasets adhering to ethical sourcing principles. Participate in data challenges/hackathons. Provide compute resources.

Academia: Curate multilingual datasets. Host competitions to expand datasets. Ensure datasets support diversity. Guide ethical data practices.

Non-profits: Aggregate public domain crowd-sourced data. Campaign for ethical, inclusive data. Advocate for accountability in datasets.

International Agencies: Support with global best practices on data. Facilitate cross-border data sharing within norms. Fund data inclusion efforts.

Roadmap

Phase 1

- Form working groups for data, models, hardware, applications etc.
- Define standards and guidelines for each track
- Develop initial open datasets for Indian languages
- Experiment with data and model architectures

Phase 2

- Release open source LLM and diffusion models trained on Indian data
- Expand datasets iteratively covering new domains
- Develop open APIs and integration tools
- Pilot custom hardware and infrastructure optimizations
- Validate applications in key sectors like agriculture, education etc.

Phase 3

- Establish federated infrastructure for sharing compute, data etc.
- Mature open source models to enterprise-grade performance
- Commercialize custom chips and hardware
- Scale application deployments across sectors
- Expand to more regional languages and datasets
- Drive widespread literacy and adoption of GenAI

The charter provides a principled framework. Collaborative execution of this plan and roadmap can make India a leader in inclusive AI advancement. This will require sustained commitment from stakeholders.