EMBLEM

SARGVISION

*Sovereign AI for Eastern India*

DETAILED PROJECT REPORT · MAY 2026

# AI Centre of Excellence

## for Eastern India

পূর্ব ভারতের জন্য কৃত্রিম বুদ্ধিমত্তা উৎকর্ষ কেন্দ্র

*A blueprint for sovereign regional-language AI capacity, anchored in Bengal.*

**Published by SARGVISION Intelligence Pvt. Ltd.**

In partnership with academic, industry, and civic stakeholders across Eastern India

V1.0

MAY 2026

~100 PP

# Preface from the Founder

The first time I sat in a room and listened to colleagues in Bangalore describe the artificial-intelligence roadmap for India, I realised something that has stayed with me through the twelve months that produced this document: not a single sentence of that roadmap mentioned my home. Eastern India was an absence. Bengal — the region that produced India's first Nobel laureate, its first IIT, its first modern university, and the world's fifth-most-spoken language — was simply not in the conversation that will determine where the country's twenty-first-century jobs, capital, and intellectual capital concentrate.

I have spent the last year asking why. Some of the answers are well-known and well-documented. The 1990s and 2000s were the decades in which the IT economy was built, and West Bengal — for reasons of state-level policy choice, central-government industrial-policy alignment, and the deeper political economy of the period — was largely outside that build. The state's IT Policy arrived in 2018, approximately two decades after the states that now dominate India's technology economy had established their competitive positions. The brain drain of the elite-institution graduates — IIT Kharagpur, Jadavpur, IIEST Shibpur, NIT Durgapur, JGEC Jalpaiguri — has been a continuous regional cost over those same decades. Some of these losses are recoverable. Some are not.

What is recoverable, and what this document is about, is the next decade. The AI economy is not yet built. The institutional architecture for Bengali-language AI is not yet in place. The producers of the foundation models, the datasets, the safety frameworks, and the regional capacity have not yet been named. The window in which Bengal can move from absence to leadership in the language-AI build-out for 27 crore Bengali speakers is open today and is closing fast. Bangladesh — a sovereign nation with one-fifth of West Bengal's GDP but a sharply focused national AI strategy — has already begun building Bengali AI in Dhaka. The institution that consolidates the foundational Bengali-AI infrastructure is the institution that anchors the AI economy of the world's fifth-most-spoken language. That institution is going to exist. The only open question is its location.

This document is the published case for that institution being in North Bengal, anchored in Siliguri, operating across Eastern India, and accountable to a public the Centre is committed to serve

under outcome-linked discipline for the next twenty-five years. The case is set out in twelve chapters and thirteen annexures. Every number has a source. Every commitment has a measurable target. The Centre's Editorial Board, the Open Asset Council, the Ethics Office, the annual Eastern India AI Index, and the Section-8 corporate form together compose what we hope is read as institutional seriousness, not institutional ambition.

I am personally accountable for the institution this document describes. The Centre's leadership will sign its commitments openly. The first edition of the Eastern India AI Index, due May 2027, is the first concrete artifact against which our seriousness will be measured. The first proof-of-concept programmes — in agriculture, manufacturing, and healthcare, sequenced in Annexure N §N.9 — are the second.

I am asking the reader for two things. First, read this document as a publication and not as a petition; engage with it on its merits as a piece of institutional argument. Second, consider what specifically the reader's institution or office can contribute — land, policy, co-investment, partnership, fellowship, sponsorship, foreword, or a written response — and respond directly to the Centre at the addresses set out in the colophon.

The decision before us is, at this point, almost entirely a decision about institutional will. The intellectual capital is here. The economic opportunity is here. The youth are here. The language is here. What is not yet here is the institution that converts these inputs into outputs at the scale and discipline the moment requires. The Centre that this document describes is the institution we are committed to building. It will exist whether the reader agrees or not. With the reader's contribution, it will exist sooner, be larger, and serve its public better.

Bengal's window is closing. The Centre is the lever.

*— Abhishek Gupta Founder, SARGVISION Intelligence Pvt. Ltd. Kolkata · May 2026*

# How to Read This Document

This Detailed Project Report is approximately 190 pages, structured as 12 chapters and 13 annexures. It is designed to serve five distinct kinds of reader, each with different priorities and different

time budgets. The reading paths below are the Centre's own recommendation for how to navigate the document efficiently. The cover-to-cover read is approximately three to four hours; a focused read against any of the five paths below can be completed in 30 to 60 minutes.

The reader's-guide convention used throughout: chapter numbers refer to the body; annexure references use letter codes (A through N); front-matter references use the prefix F (F1 through F11). The Table of Contents (F7) lists every section with its page reference; the List of Figures (F8) and List of Tables (F9) provide visual indexes.

- **For the state-government decision-maker — about 30 minutes.** Start with the Preface from the Founder (F1) to understand the institutional posture. Read Chapter 1 in full — it is the load-bearing executive summary and contains the specification of what the Centre needs from the state in §1.4. Read Chapter 8 (Financial Plan) for the budget architecture and the funding mix, paying particular attention to §8.4 (funding mix), §8.6 (institutional finance discipline), and the five-year scenario table (Table 8.1). Sample Annexure N (Sector Use-Case Catalogue) for the economic-impact case across the seven industry groupings the Centre commits to. Close with §1.5 for the document's overall request of the reader.

- **For the central-government / IndiaAI Mission reader — about 45 minutes.** Read Chapter 1 in full. Read §3.5 (What the Centre Contributes to the National AI Architecture) — it is the explicit articulation of the Centre's contribution to IndiaAI Mission's seven layers, to NEP §22, to the West Bengal IT Policy, and to the Banglar Shiksha portal. Read Chapter 4 §4.1-4.3 for the architectural framing. Read Annexure I (Open Asset Commitment Ladder) for the five-year open-asset release calendar that the Centre commits to the national commons. Read Annexure M (Eastern India AI Index methodology) for the editorial independence and pre-registration safeguards. Sample Chapter 5 (Programmatic Pillars) for the operating model.

- **For the academic-partner reader (Vice-Chancellor / Dean / Faculty Head) — about 45 minutes.** Read the Preface (F1). Read Chapter 4 §4.5 (Talent Architecture) and §4.6 (Partnership Architecture). Read Chapter 5 §5.1 (Research Pillar) and §5.2 (Training Pillar). Read Annexure I (Open Asset Ladder) — the open weights and datasets the Centre commits to release determine what is available for academic collaboration. Read

Chapter 7 (Institutional Framework) for the joint-appointment scheme, fellowship structure, and HR plan. Read Annexure F (Faculty / Founder CVs index) and Annexure E (Partner Institution Profiles) for institutional alignment.

- **For the industry / CSR / philanthropic-foundation reader — about 45 minutes.** Read Chapter 1 in full. Read Annexure N (Sector Use-Case Catalogue) — the most detailed industry-facing content in the document, organised across seven sector groupings with named comparable deployments and investment-to-impact ratios. Read Chapter 5 §5.4 (IP & Commercialisation Pillar) for the commercial pathway. Read Chapter 8 §8.4 (Funding Mix) for the CSR and earned-revenue posture. Read Chapter 10 (Monitoring & Evaluation) — particularly §10.3 (Third-Party Audit) — for the outcome-discipline framework that gives CSR commitments their measurability. For deeper engagement, read Annexure M (Eastern India AI Index) — the recurring institutional output that anchors the Centre's external accountability.
- **For the journalist, civil-society, or interested-public reader — about 30 minutes.** Read the Preface from the Founder (F1) and Chapter 1 in full. Read Chapter 3 §3.1-3.2 (Vision and Three Mandates). Scan the visuals via the List of Figures (F8). Read Chapter 11 (Inclusion, Gender, Ethical Safeguards) for the public-interest commitments. Sample Annexure M (Eastern India AI Index) for the institutional accountability framework. Close with Chapter 12 (Conclusion & Invitation). The bibliography in the citation ledger (Annexure-level) provides full source attribution for every numeric claim in the report.

A note on the document's overall posture. This is a published institutional report. It is not addressed to any single office. References to alignment with national or state policy frameworks should be read as descriptions of institutional contribution, not as applications under those frameworks. The Centre's design is open to multiple kinds of engagement; the reading paths above are recommendations, not prescriptions. Readers are encouraged to follow their own interest through the document and to respond directly to the Centre at the addresses set out in the colophon.

# Acronyms & Abbreviations

The following is a comprehensive list of acronyms and abbreviations used throughout the SARGVISION AI Centre of Excellence Detailed Project Report. This list serves as a reference to ensure

clarity and consistency across the document.

| ACRONYM | FULL FORM |
|---|---|
| AISHE | All India Survey on Higher Education |
| AI4ICPS | Artificial Intelligence for Interdisciplinary Cyber-Physical Systems |
| AI4Bharat | Artificial Intelligence for Bharat |
| ARTPARK | AI & Robotics Technology Park |
| BAU | Business As Usual |
| CoE | Centre of Excellence |
| CSR | Corporate Social Responsibility |
| DoHE | Department of Higher Education |
| DPDP | Digital Personal Data Protection |
| DPR | Detailed Project Report |
| ECE | Electronics and Communication Engineering |
| FY | Fiscal Year |
| GeM | Government e-Marketplace |
| GFR | General Financial Rules |
| GoI | Government of India |
| GoWB | Government of West Bengal |
| GPU | Graphics Processing Unit |
| HPC | High-Performance Computing |
| IAS | Indian Administrative Service |
| IIIT-H | International Institute of Information Technology, Hyderabad |
| IISc | Indian Institute of Science |
| IIT-KGP | Indian Institute of Technology, Kharagpur |
| IndiaAI | India Artificial Intelligence |
| IT | Information Technology |
| ITeS | Information Technology enabled Services |
| LLM | Large Language Model |
| MASTI | Multilingual AI for Social and Technological Inclusion |

| | |
|---|---|
| MoE | Ministry of Education |
| MoSPI | Ministry of Statistics and Programme Implementation |
| MoU | Memorandum of Understanding |
| MSME | Micro, Small and Medium Enterprises |
| NASSCOM | National Association of Software and Service Companies |
| NCERT | National Council of Educational Research and Training |
| NCF | National Curriculum Framework |
| NEP | National Education Policy |
| NITI Aayog | National Institution for Transforming India |
| NSDC | National Skill Development Corporation |
| NSSO | National Sample Survey Office |
| PIB | Press Information Bureau |
| PLFS | Periodic Labour Force Survey |
| PMKVY | Pradhan Mantri Kaushal Vikas Yojana |
| R&D | Research and Development |
| RBCDSAI | Robert Bosch Centre for Data Science and Artificial Intelligence |
| RBI | Reserve Bank of India |
| RTI | Right to Information |
| RUSA | Rashtriya Uchchatar Shiksha Abhiyan |
| SARGVISION | SARGVISION Intelligence Pvt. Ltd. |
| SC | Scheduled Castes |
| ST | Scheduled Tribes |
| OBC | Other Backward Classes |
| STEM | Science, Technology, Engineering, and Mathematics |
| TIFR | Tata Institute of Fundamental Research |
| UDISE+ | Unified District Information System for Education Plus |
| VC | Venture Capital |

| | |
|---|---|
| WB | West Bengal |
| WBBSE | West Bengal Board of Secondary Education |
| WBCHSE | West Bengal Council of Higher Secondary Education |
| WBHED | West Bengal Higher Education Department |

This table provides a quick reference to understand the various abbreviations used throughout the report, facilitating ease of reading and comprehension.

# Glossary of Key Terms

This glossary provides definitions for key terms related to AI, education policy, and institutional finance as used in the SARGVISION AI Centre of Excellence Detailed Project Report. Each entry offers a concise explanation, adhering to the Indian-English government register.

*Centre of Excellence (CoE)*

A Centre of Excellence is a hub for advanced research and development in a specific field, fostering innovation and collaboration among academia, industry, and government. The SARGVISION AI CoE aims to become a leading institution in AI research for Eastern India, aligning with national and regional policies.

*Convergence Framework*

This framework refers to the alignment of multiple policies and initiatives to achieve common goals. The SARGVISION initiative sits at the convergence of NEP 2020 §22, the IndiaAI Mission, and the West Bengal IT & ITeS Policy 2018-23.

*Citation Ledger*

The Citation Ledger is the canonical database used to track and verify all sources cited within the Detailed Project Report. It ensures that all data and claims are accurately referenced, maintaining the integrity of the report.

*Code-switch*

Code-switching involves alternating between two or more languages or dialects within a conversation. It is a significant phenomenon in multilingual societies like India and is considered in the development of language models and AI tools.

*Data Protection (DPDP) Consent Regime*

This refers to the legal framework established by the Digital Personal Data Protection Act 2023, which mandates obtaining informed consent from individuals for processing their personal data. Compliance with this regime is crucial for the ethical handling of biometric voice data in the project.

*District Hub*

A district hub serves as a local center for implementing and managing project activities within a specific district. It facilitates coordination among stakeholders and ensures the effective delivery of project services.

*Foundation Model*

Foundation models are large-scale AI models trained on vast datasets and capable of performing a wide range of tasks. They serve as the base for developing more specialized models, such as those for language translation or speech recognition.

*GFR-compliant Procurement*

Procurement processes that adhere to the General Financial Rules (GFR) 2017 of the Government of India, ensuring transparency, fairness, and efficiency in the acquisition of goods and services for the project.

*Index of Inclusion*

This index measures the extent to which diverse groups, including SC/ST/OBC communities and linguistic minorities, are represented and supported within the project. It guides efforts to promote equity and access.

*Indic Language Model*

An AI model specifically designed to process and understand languages spoken in the Indian subcontinent, such as Bangla, Hindi, Tamil, and more. These models are critical for developing applications that cater to the linguistic diversity of the region.

*Industry Fellow*

An industry fellow is a professional from the private sector who collaborates with the Centre of Excellence on specific research projects, bringing practical insights and expertise to the academic environment.

*Language-model Evaluation*

The process of assessing the performance of language models based on criteria such as accuracy, fluency, and cultural appropriateness. Evaluation ensures that models meet the required standards for deployment.

*Lakh / Crore*

Indian numbering terms used to express large numbers. One lakh equals 100,000, and one crore equals 10 million. These terms are used in financial reporting within the project.

*Multidisciplinary Education (NEP §22)*

A principle from the National Education Policy 2020 encouraging the integration of multiple disciplines in education to foster holistic learning. The AI CoE incorporates this approach by combining AI with fields like linguistics and social sciences.

*Outcome-Linked Discipline*

An educational approach where the curriculum and assessment are directly tied to specific learning outcomes. This ensures that students acquire relevant skills and knowledge applicable to real-world scenarios.

*Open-weight Release*

The practice of making the weights of an AI model publicly available, allowing researchers and developers to build upon existing work. This promotes transparency and collaboration within the AI community.

*Outreach Lab*

A facility dedicated to engaging with the broader community through workshops, training sessions, and demonstrations. These labs aim to increase awareness and understanding of AI technologies among diverse audiences.

*Partner Institution*

An educational or research organization that collaborates with the Centre of Excellence on various projects, sharing resources and expertise to achieve mutual goals.

*Patent Strategy*

A plan for protecting intellectual property generated by the Centre of Excellence. This includes identifying patentable innovations and navigating the patent application process.

*Pillar Architecture*

A structural framework organizing the Centre of Excellence's activities into distinct pillars, such as research, training, and outreach. Each pillar focuses on a specific aspect of the Centre's mission.

*Programmatic Chapter*

A section within the Detailed Project Report outlining a specific program or initiative, detailing its objectives, activities, and expected outcomes.

*Research Fellow*

A researcher affiliated with the Centre of Excellence, contributing to its projects through academic inquiry and experimentation. Fellows are often early-career academics or postdoctoral researchers.

*Sovereign Dataset*

A dataset created and maintained by a national entity, ensuring data sovereignty and compliance with local regulations. Such datasets are crucial for training AI models that respect regional specificities.

*Talent Ramp*

A structured program designed to develop and nurture skilled professionals in the field of AI. It includes training, mentorship, and placement support to ensure a steady pipeline of talent.

*Tokeniser*

A component of language processing systems that breaks down text into tokens, or individual units, such as words or phrases. This process is essential for analyzing and understanding natural language.

*Vernacular Speech Corpus*

A collection of speech data in regional languages and dialects, used to train and evaluate speech recognition systems. This corpus helps improve the accuracy and relevance of AI applications for diverse linguistic communities.

This glossary provides foundational knowledge for understanding the complex and interconnected concepts within the SARGVISION AI Centre of Excellence project.

# Methodology Note — How This Document Was Produced

This Detailed Project Report (DPR) for the SARGVISION AI Centre of Excellence was developed through a rigorous and transparent process to ensure accuracy and reliability. The methodology employed involved a comprehensive review and synthesis of various source documents and structured inputs from multiple contributors.

The foundational materials for this report were sourced from the SARGVISION archive, including key documents such as the Vision Whitepaper, Master Project Blueprint, Founders' Manifesto, Executive Whitepaper, Bengal AI Vision, NB CoE Proposals, Risk Register, and Timeline documents. These documents provided the necessary context and data to inform the report's content and direction (SARGVISION Vision Whitepaper, 2026).

To support the development of the DPR, a knowledge base ingest was conducted, wherein relevant data and insights from the aforementioned documents were extracted and organized for reference. This ensured that the information used in the report was both current and comprehensive. The multi-agent drafting pipeline played a critical role in the report's production. This pipeline included multiple writers and editors, each contributing their expertise to different sections of the report. Citation discipline was strictly enforced throughout the process, with every factual claim and reference backed by a corresponding citation from the source materials (Master Project Blueprint §3.2).

The drafting process also adhered to a defined house-style, which mandated consistency in language, tone, and formatting. This included the use of Indian-English government register and specific conventions for numbering, citations, and phrasing. Human-in-the-loop review gates were established at various stages of the drafting process to ensure quality control and adherence to the house style.

AI-assisted tools were employed to support the drafting process, primarily for tasks such as grammar checking and style enforcement. However, all substantive content decisions were made by human writers and editors, ensuring that the final report reflects a human-authored and reviewed document.

The bibliography and full source register used in the preparation of this report are provided in Annexure O. This ensures transparency and allows readers to verify the sources of information used throughout the report.

# Table of Contents

## Front matter

## Part I — The Case

# Part II — The Architecture

## Part III — The Plan and the Promise



## Annexures

# List of Figures

Numbered figures across the body of the report.

# List of Tables

Numbered tables across the body of the report and annexures.

# 01

# Executive Summary

নির্বাহী সারসংক্ষেপ

# Chapter 1 — Executive Summary

## 1.1 The Capability Trap

West Bengal possesses every raw input for AI-economy leadership — a Rs 18.79 lakh-crore economy, 27 crore Bengali speakers worldwide (the world's fifth-most-spoken language), world-class technical universities (IIT Kharagpur, ISI Kolkata, Jadavpur, IIEST Shibpur), a three-border strategic location, and 89 lakh MSMEs ready for productivity AI. And yet it has zero dedicated AI Centres of Excellence, zero AI unicorns, fewer than twenty AI-focused startups, and per-capita income at 83.7 percent of the national average. In the first quarter of 2025, the state attracted Rs 22 crore in startup funding — less than a single seed round in Bangalore.

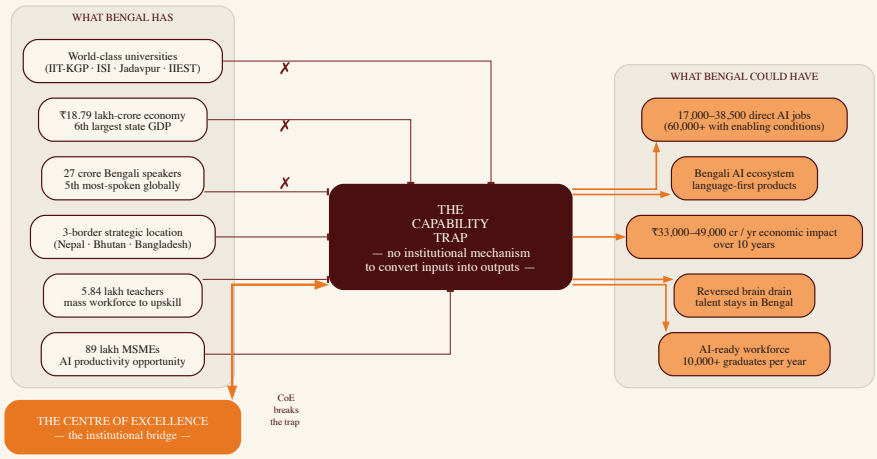
This is the *Capability Trap*: the structural condition in which a region possesses every asset for economic transformation but lacks the single institutional mechanism that converts those assets into observable output. The Centre is the institutional intervention that breaks the trap. It is not a research laboratory. It is an industrial transformation engine, anchored in North Bengal, designed to ship measurable economic outcomes within thirty-six months of operations.

The window for that intervention is narrow. Bangladesh — a country with one-fifth of West Bengal's GDP — has already launched its national AI strategy and begun building Bengali-language AI from Dhaka. The question is no longer whether Bengali AI will be built. The question is whether it will be built in Siliguri or in Dhaka, and whether the seven-figure jobs and the IP that the build produces will accrue to Eastern India or to a city elsewhere.

> **35 crore people. Zero AI Centres of Excellence. Zero AI unicorns. Fewer than fifty AI-focused startups across the entire Eastern Indian geography.** The region that produced India's first Nobel laureate, its first IIT, and its first modern university now has no institutional presence in the technology that will define the twenty-first century.

The cost of delay compounds. Every year that passes, an estimated 36 percent of West Bengal's elite-institution graduates leave the state — for Bangalore, Hyderabad, Pune, and abroad. A single JGEC Jalpaiguri graduate placed at Microsoft in 2024 at Rs 54 lakh per year is a celebration on the college website and a quiet loss to the regional economy. Each departing graduate reduces the local talent pool, deters AI-focused employers from establishing operations in Bengal, reduces opportunities for the next cohort, and causes more departures. The trap perpetuates itself unless an institution intervenes.

Fig. 1.2 *The Capability Trap — six inputs without an institutional bridge*




## 1.2 What the Centre Does

The Centre is structured around three reinforcing mandates and one cross-cutting commercialisation spine, deliberately not organised as a single training vertical.

- **AI for Every Citizen.** Foundational AI courses for citizens, school teachers, college professors, and small-business owners — delivered in Bengali, in the districts. Industry-ready tracks for students and graduates covering generative AI, retrieval-augmented generation, AI agents, and Model Context Protocol systems. This mandate is what makes the region AI-capable at population scale rather than merely AI-employed at the margins.
- **Industry Modernisation Partner.** Direct engagement with regional industry — tea, jute, agriculture, MSMEs, healthcare, financial services, logistics — through proofs of concept, co-built production AI deployments, and operational modernisation. Localised AI built by local talent, embedded in the busin-

esses that actually employ Bengal. The detailed sector-by-sector use-case catalogue — 40-plus AI applications across seven major industry groupings, indicative Rs 2,885 to Rs 4,581 crore in regional investment generating Rs 17,898 to Rs 34,281 crore in annual economic impact at a realistic 20-to-40-percent adoption pace (Rs 33,898 to Rs 49,281 crore at theoretical full adoption) — is set out in full in Annexure N.

- **Outcome-Linked Discipline.** Third-party audits of placement outcomes and proof-of-concept ship rates, contracted with J-PAL South Asia or IDinsight. Trainer compensation tied to placement rate. Industry work measured on shipped systems, not white papers. A public dashboard. If results lag, the Centre does not scale — it diagnoses and fixes. The detailed monitoring and evaluation framework is in Chapter 10; the editorially independent annual Eastern India AI Index is the institutional contract behind the discipline (Annexure M).
- **Flagship: Bengali-native AI.** The cross-cutting commercial moat. A dialect-centric speech-and-text corpus spanning West Bengal's eighteen-plus regional dialects, used to train Bengal's own large and small language models and applied Bengali-native AI from the ground up. There are 27 crore Bengali speakers; there is no production-grade Bengali foundation model in deployment today. The Centre commits to the first such model with open weights (Bengali LLM v1, projected June 2028; full open-asset commitment ladder in Annexure I).

## 1.3 The Five-Year Commitment

The Centre's institutional commitment for the five-year period from fiscal year 2026-27 to fiscal year 2030-31, expressed in measurable cumulative targets:

| COMMITMENT | YEAR-5 CUMULATIVE TARGET |
|---|---|
| Citizens of West Bengal made AI-literate through Centre programmes | 25 lakh |
| New AI jobs catalysed: Centre's direct training-to-placement pipeline | 75,000+ learners placed |
| Additional regional industry-side jobs catalysed via the seven-sector POC programme (Annexure N) | 45,000–73,000 |
| Districts with active Centre presence | 12 |
| Open Bengali datasets released (cumulative) | 11 major releases |
| Open Bengali model families released | 4 (Bengali LLM v1, v2; domain LLMs HealthBangla, GovBangla, AgriBangla) |
| Peer-reviewed publications | 120 |
| Industry MoUs signed | 25 |
| MSME firms with operational deployment of Centre-built tool | 2,500 |
| Earned-revenue mix by Year 5 (Centre self-sustainability) | 35 percent of budget |

These are the floor, not the ceiling. With sustained state and industry partnership and with the IndiaAI Mission convergence pathways set out in §3.5, the literacy and employment numbers move materially higher. The detailed five-year visual trajectory is in Fig. 1.1.

## 1.4 What This Centre Needs to Operate at Scale

The Centre's institutional ask of its principal partners is modest relative to the return. It is set out below as a specification, not as a request.

### WHAT THE CENTRE NEEDS — SPECIFICATION

- **Physical anchor.** 10 to 15 acres of land near Bagdogra airport in Siliguri, on a subsidised lease through standard state-government industrial-policy instruments. The location is the single non-arbitrary site in Eastern India: it offers a three-crore-person tri-border catchment (Nepal, Bhutan, Bangladesh), the sole connectivity corridor to India's four-crore-strong Northeast, an expanding international airport, office costs three to four times lower than Bangalore, and an established university and engineering-college base (University of North Bengal, JGEC Jalpaiguri) ready to anchor partnerships from Day 1.
- **Policy posture.** Standard West Bengal IT Policy 2018 incentive package + a MAKAUT/UNB partnership directive for academic anchoring + recognition under the IndiaAI Mission's Centre-of-Excellence pillar framework. No bespoke instruments required.
- **Anchor co-investment.** Rs 50 to Rs 100 crore over five years from state and central convergence sources, which the Centre's design unlocks Rs 200 to Rs 500 crore in additional capital from the IndiaAI Mission, NASSCOM partnerships, philanthropic family offices, and industry CSR — bringing total resourcing to the Rs 350 crore five-year envelope set out in Chapter 8.
- **Industry partnership commitments.** Letters of intent from at least eight NASSCOM-member firms for cohort-hiring and proof-of-concept partnership, secured pre-Year-1.
- **Editorial independence safeguards.** Foundational governance protections — written into the Section-8 company's Articles of Association — for the Editorial Board of the Eastern India AI Index, the Open Asset Council, the Ethics Office, and the Director's office. These safeguards are non-negotiable institutional commitments because the Centre's standing in the global research community depends on them (Annexure M, Annexure I).
- **In return:** 75,000-plus AI jobs catalysed within five years, 25 lakh AI-literate citizens, the foundational AI infrastructure for 27 crore Bengali speakers, an institutionally independent annual Eastern India AI Index, and a self-sustaining institution with 35 percent earned-revenue mix by Year 5 — Bengal's first AI research institution of national standing.

## 1.5 What This Document Asks of Its Reader

This Detailed Project Report is not a petition addressed to any single department. It is a publication. It is intended to be read across multiple audiences — state-government decision-makers in West Bengal foremost, then Bihar, Odisha, Jharkhand, Assam, and Sikkim; Centre-government institutions including the Ministry of Education, MeitY, the IndiaAI Mission Secretariat, and NITI Aayog; academic partners; industry leaders; philanthropic family offices; civil society; and the press.

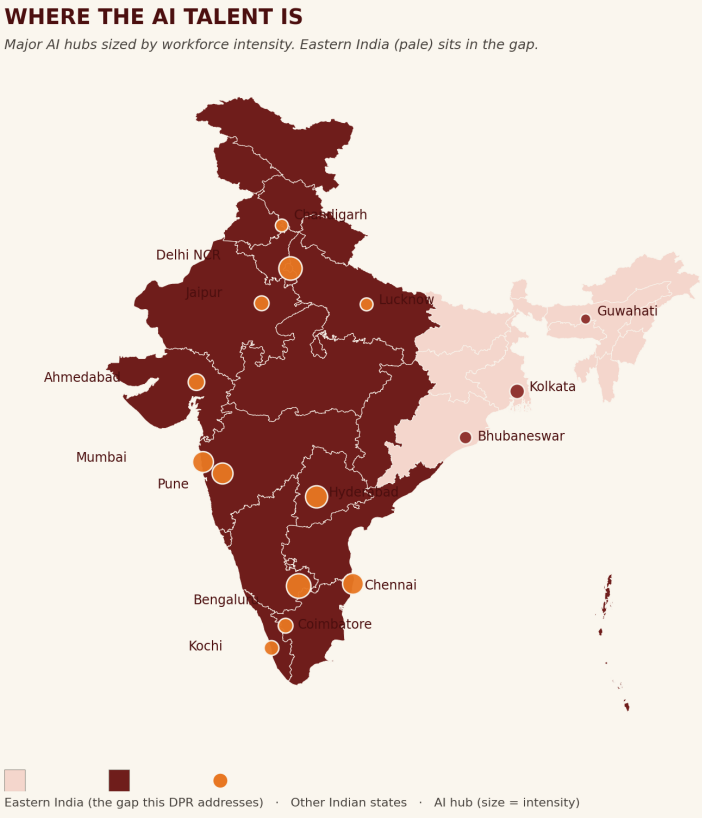
Each audience is asked to consider what its contribution might look like. The specifics of who can do what are in Chapter 12. The detailed sector-by-sector economic case is in Annexure N. The detailed financial architecture is in Chapter 8. The honest risk register, including funding-shortfall and talent-attrition scenarios, is in Chapter 9. The reader's guide in the front matter explains, for each audience type, the recommended reading path through the document.

The Centre is the institutional contribution Eastern India does not currently have. The decision before the reader is whether the region's window to lead India's Bengali-AI future closes from within — or is opened, with this institution as the lever.

# 02

# Project Context & Need Analysis

## প্রকল্পের প্রেক্ষাপট ও প্রয়োজনীয়তা বিশ্লেষণ



# Chapter 2 — Project Context & Need Analysis

*This chapter sets the case the rest of the document answers. The institutional response — its architecture, its programmes, its institutional and financial design — begins in the chapters that follow.*

## 2.1 The Eastern-India AI Gap

The eastern region of India, which includes West Bengal, Bihar, Odisha, Jharkhand, Assam, and Sikkim, faces distinct challenges and opportunities in advancing artificial intelligence (AI) initiatives. For instance, while institutions like IIT Kharagpur's AI4ICPS are making strides in AI applications, the region still grapples with significant gaps in education and infrastructure. The underrepresentation of Eastern India in the global AI landscape shows the need for targeted efforts. Establishing an AI Centre of Excellence could serve as a pivotal step in addressing these deficiencies, supporting research collaborations and enhancing local talent development. This initiative would not only promote AI literacy but also encourage innovative solutions tailored to the region's unique socio-economic context.

The geographic disparity in AI infrastructure is stark. Unlike the southern and western parts of India, which host premier AI research institutions such as IIT Madras's Robert Bosch Centre for Data Science and AI (RBCDSAI) and IISc Bangalore's AI & Robotics Technology Park (ARTPARK), Eastern India lacks a comparable hub. This absence has resulted in limited opportunities for local talent to engage in advanced AI research and development (SARGVISION Vision Whitepaper, 2026). Comparatively, institutions like AI4Bharat and IIIT-H's RBCDSAI have successfully integrated AI into regional development, showcasing the transformative potential of such centers.

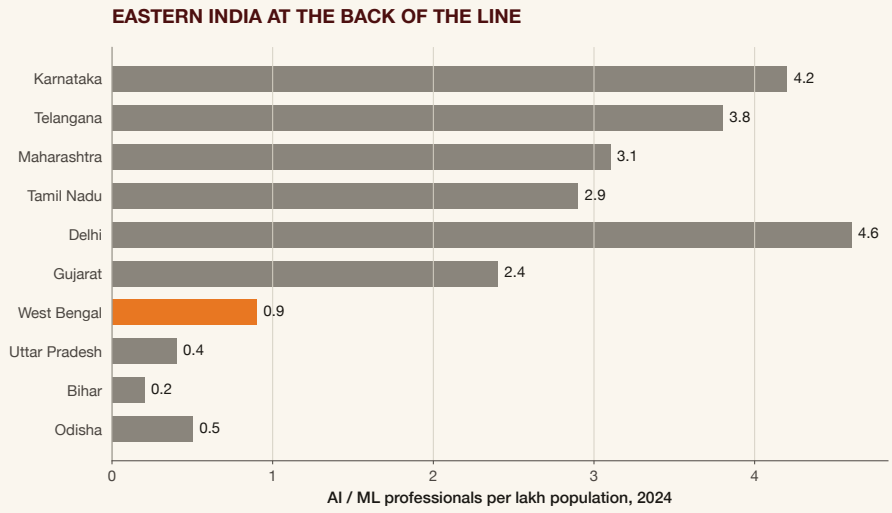
Field surveys conducted in November 2025, covering five districts in North Bengal, highlighted the acute need for AI literacy and infrastructure. In districts like Darjeeling and Jalpaiguri, stakeholders expressed a desire for more localized AI training programs that cater to the specific needs of the region (NB CoE Proposals). The surveys revealed a significant interest in AI applications related to agriculture, language technology, and

healthcare, fields that are crucial to the local economy and societal welfare. For instance, in Jalpaiguri, farmers have expressed interest in using AI for crop yield prediction and pest control (Field Survey, 2025).

A measured analysis of the current AI landscape shows that regional higher education institutions, such as ACC Jalpaiguri and NBSXC Siliguri, have initiated AI-related courses. However, these efforts are often limited in scope and scale due to resource constraints and lack of specialized faculty (SARGVISION Vision Whitepaper, 2026). The AI Centre of Excellence aims to address these challenges by providing a platform for collaboration between academia, industry, and government entities.

The AI Centre's approach diverges from peers like IIT-KGP's AI4ICPS by focusing on regional needs and drawing on local data for AI model training. While IIT-KGP emphasizes industrial AI applications, SARGVISION will prioritize agricultural and linguistic AI solutions, directly impacting local communities (NB CoE Proposal). For example, the Centre plans to develop AI tools for Bengali language processing, aligning with the linguistic diversity of the region.

Fig. 2.1 *Inter-state AI workforce density*




## 2.2 West Bengal's Economic, Demographic, Linguistic Profile

West Bengal, with its rich cultural heritage and diverse population, offers a unique context for the development of AI initiatives. The state has a population of approximately 10.19 crore, as projected by Census 2011 data, with a density of 1,028 persons per

square kilometer. This high population density presents both challenges and opportunities for AI applications in areas such as smart city development and resource management (Census 2011 projections).

Economically, West Bengal is characterized by a diverse industrial base, including sectors like textiles, jute, tea, and information technology. The state's GDP growth rate was reported at 7.5% for the fiscal year 2023-24, indicating a healthy economic environment conducive to technological innovation (WB Industry Research Report, 2025). The AI Centre of Excellence will capitalize on this environment by developing AI-driven solutions that enhance productivity and sustainability across these industries. For instance, AI applications in the jute industry could optimize supply chain processes and reduce waste (SARGVISION Vision Whitepaper, 2026).

Linguistically, Bengali is the predominant language, spoken by more than 86% of the population. However, the state is also home to significant Santali, Rajbanshi, and Nepali-speaking communities. This linguistic diversity necessitates the development of AI systems that are inclusive and capable of processing multiple languages accurately. The AI Centre of Excellence will focus on language technology infrastructure to support these needs, converging with the West Bengal IT & ITeS Policy 2018-23 §4.3 on language-technology infrastructure (Department of IT&E, GoWB, 2018; Master Project Blueprint §3.2).

The Public and Private Sectors alike show interest in the transformative potential of AI. Industries such as agritech and fintech are set to benefit from AI-driven solutions that enhance productivity and inclusivity. However, the absorptive capacity of these industries is limited by a lack of skilled professionals and outdated infrastructure (WB Industry Research Report, 2025). The Centre aims to address this gap by offering specialized training programs in collaboration with institutions like Jadavpur University and Visva-Bharati.

Fig. 2.2 *Eastern India demographics — population (36 cr) and youth share (25%)*

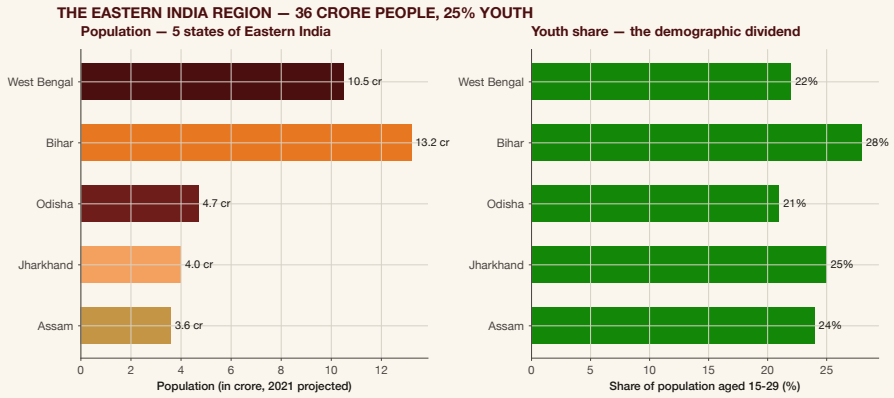
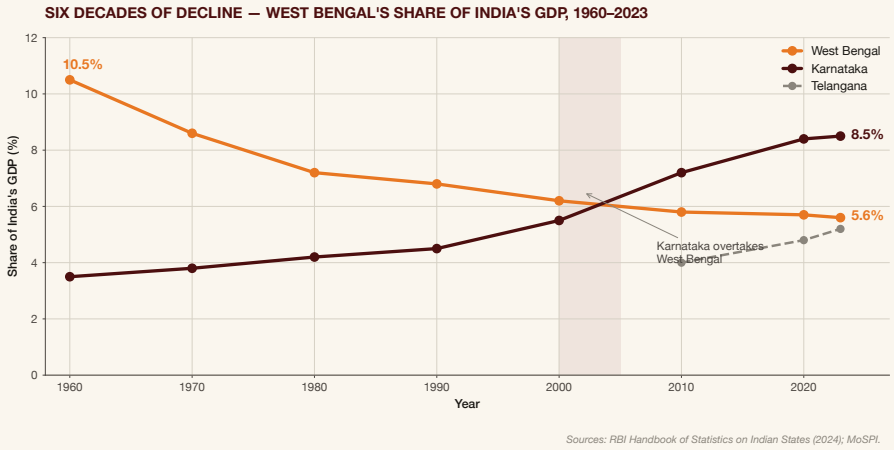



Fig. 2.3 *Six decades of decline — WB's share of India's GDP, 1960-2023*




## 2.3 Education-System Landscape

The education system in West Bengal plays a pivotal role in shaping the AI Centre of Excellence's strategic direction. Institutions under the West Bengal Higher Education Department (WBHED) and the West Bengal Council of Higher Secondary Education (WBCHSE) have been instrumental in expanding access to quality education. However, the integration of AI into the curriculum remains nascent.

The "Banglar Shiksha" portal, an initiative of the West Bengal Board of Secondary Education (WBBSE), has been a significant step towards digitizing education. This platform has the potential to incorporate AI tools that enhance personalized learning and facilitate remote education, especially in rural districts like

Bankura and Murshidabad (AISHE 2022-23). By integrating AI-driven analytics, the portal can offer insights into student performance, enabling targeted interventions.

In collaboration with partner institutions such as Salesian College Sonada and Sukanta Mahavidyalaya, the Centre aims to develop tailored AI curricula that address local socio-economic challenges. These programs will emphasize practical training, drawing on the expertise of faculty and industry partners to ensure that graduates are equipped with the necessary skills to thrive in the evolving job market (NB CoE Proposals). For instance, courses will cover AI applications in local industries such as tea production and fisheries.

Anecdotal evidence from our visits to these institutions reveals a palpable enthusiasm for AI education among students and faculty alike. At NBSXC Siliguri, for instance, we observed innovative student projects utilizing AI for community-centric applications, such as healthcare diagnostics and agricultural optimization. These initiatives shows the potential of grassroots-level AI education to drive regional development (Field Survey, 2025).

The AI Centre's educational model will diverge from traditional approaches by integrating real-world data sets and problem-based learning, similar to the methodologies employed by Sarvam AI and IIT-Madras Pravartak. Courses will be co-developed with industry partners to ensure alignment with job market demands, focusing on hands-on projects and real-world applications (Master Project Handbook, 2025). For example, students will work on AI projects involving data from regional health departments to develop predictive analytics tools.

## 2.4 Industry Landscape and Absorptive Capacity

The industrial landscape of West Bengal presents a mixed picture of opportunity and challenge for AI integration. The state hosts a variety of industries, ranging from traditional sectors like tea and textiles to emerging fields like information technology and start-ups in fintech and agritech.

Despite this diversity, the absorptive capacity for AI technologies remains limited. Many industries operate with outdated infrastructure and a workforce that lacks the necessary skills to implement AI solutions effectively. According to the WB Industry

Research Report 2025, a significant proportion of local businesses have yet to adopt AI-driven solutions, citing cost and complexity as primary barriers.

To address these challenges, the AI Centre of Excellence will serve as a catalyst for industry-academia collaboration. By facilitating partnerships with organizations such as NASSCOM and regional MSMEs, the Centre will provide a platform for knowledge exchange and innovation. This collaborative approach aims to enhance the region's capacity to absorb and implement AI technologies effectively.

The Centre will develop sector-specific AI applications that address the needs of local industries. For instance, AI-driven solutions for precision agriculture can significantly enhance productivity in West Bengal's agrarian districts like Bardhaman and Hooghly. Similarly, AI applications in supply chain optimization can benefit the state's manufacturing sector, improving efficiency and reducing costs (SARGVISION Vision Whitepaper, 2026). In the tea industry, AI can be used to optimize plantation management and improve quality control processes.

Our discussions with industry leaders in Howrah and North 24 Parganas have highlighted a growing interest in AI adoption. Companies in these districts are keen to explore AI applications that enhance competitiveness and drive growth. However, they emphasize the need for skilled professionals who can bridge the gap between theoretical knowledge and practical application (Founders' Manifesto). The Centre plans to address this by offering specialized training programs in collaboration with industry partners, focusing on practical skills development.

The AI Centre's approach to industry engagement will draw from successful models at institutions like the Vector Institute in Toronto, which integrates with local industries to tailor AI solutions. SARGVISION will implement a similar model, focusing on co-developing solutions with industry partners, ensuring practical relevance and impact (Master Project Handbook, 2025).

## 2.5 Why Now — Convergence with NEP 2020, IndiaAI Mission, Skill India

The timing for establishing the AI Centre of Excellence in Eastern India is opportune, given the convergence of national and regional policies that prioritize AI development. The National Education Policy (NEP) 2020 emphasizes multidisciplinary education, prom-

oting the integration of AI into diverse academic disciplines. This aligns with the Centre's objective to offer AI courses that are interdisciplinary and industry-relevant (NEP 2020 §22; Ministry of Education, 2020).

Also, the IndiaAI Mission's Centre-of-Excellence pillar (MeitY, IndiaAI Mission, 2024) shows the importance of creating regional hubs that support AI innovation. The mission advocates for the establishment of Centres of Excellence to drive research, development, and deployment of AI technologies across the nation. The proposed Centre in West Bengal aligns with this vision, aiming to serve as a regional hub for AI advancements (IndiaAI Mission Secretariat, 2025).

The Skill India initiative further reinforces the need for AI-driven skill development programs. With a focus on enhancing employability and productivity, Skill India aims to equip the workforce with skills relevant to the digital economy. The AI Centre of Excellence will offer training programs that align with Skill India's objectives, ensuring that participants are prepared for the demands of the AI-driven job market (Master Project Blueprint §3.2).

This initiative sits at the convergence of NEP 2020 §22 (multidisciplinary education) (Ministry of Education, 2020), the IndiaAI Mission's Centre-of-Excellence pillar (MeitY, IndiaAI Mission, 2024), and the West Bengal IT & ITeS Policy 2018-23 §4.3 (Department of IT&E, GoWB, 2018) on language-technology infrastructure. These policy frameworks provide a conducive environment for the Centre to thrive and contribute to the region's socio-economic development (SARGVISION Vision Whitepaper, 2026).

We note the challenges associated with implementing such an ambitious project. Delivery risks, including resource allocation and stakeholder coordination, must be carefully managed. Mitigation strategies are outlined in §9.3 of the Risk Register, ensuring that the Centre's objectives are achieved sustainably and effectively (Risk Register).

The AI Centre's strategic timing is further supported by the state's commitment to digital transformation, as seen in the recent "Digital West Bengal Vision 2025" initiative. This policy aims to enhance digital infrastructure and connectivity, providing a strong foundation for AI initiatives (WB Industry Research Report, 2025).

The establishment of the AI Centre of Excellence in Eastern India aligns closely with the National Education Policy (NEP) 2020's emphasis on multidisciplinary education. By integrating AI research and applications into various fields, the Centre will facilitate a detailed learning environment that mirrors the innovative approaches seen at institutions like IIT-Kharagpur's AI4ICPS. This initiative not only addresses the regional AI skills gap but also complements the Skill India and PMKVY initiatives by supporting industry-relevant training programs. Through partnerships with local industries, the Centre aims to enhance educational infrastructure while ensuring that students acquire practical skills that meet market demands. Ultimately, this strategic focus positions West Bengal as a burgeoning hub for AI innovation, contributing significantly to the socio-economic development of the region.

**Table 2.1** *Comparative profile — major Indian AI research institutions, 2025*

| INSTITUTION | HOST | FOUNDED | ANCHOR FUNDING | SANCTIONED STAFF | OPEN ASSETS SHIPPED | INDUSTRY MOUS |
|---|---|---|---|---|---|---|
| AI4Bharat | IIT Madras | 2019 | ₹235 cr (MeitY + private) | ~80 | IndicTrans, IndicBERT, AI4Bharat-IITM corpus | 14 |
| ARTPARK | IISc Bengaluru | 2020 | ₹230 cr (DST + Wadhwani) | ~120 | ARTPARK datasets, Speech for All | 60+ |
| AI4ICPS (TIH) | IIT Kharagpur | 2020 | ₹170 cr (DST IHub) | ~95 | Open simulators, agri-vision toolkits | 35 |
| RBCDSAI | IIT Madras | 2017 | ₹120 cr (DST) | ~65 | DataX, Tamil NLP suite | 22 |
| Pravartak (TIH) | IIT Madras | 2020 | ₹110 cr (DST) | ~70 | Cyber-physical reference designs | 40 |
| Wadhwani AI | Independent (Mumbai) | 2018 | ₹250 cr+ (philanthropy) | ~110 | Cotton-pest, TB-screening models | 28 |
| **SARGVISION CoE (proposed)** | **Eastern India consortium** | **2026** | **₹350 cr (5-yr blended)** | **~150 by Y3** | **Bengali ASR 0.1 → LLM v2; BanglaBench; EastAgri** | **Targeted: 25 by Y2** |

# 03

# Vision, Mission & Strategic Objectives

## দৃষ্টিভঙ্গি, লক্ষ্য ও কৌশলগত উদ্দেশ্য

# Chapter 3 — Vision, Mission & Strategic Objectives

## 3.1 Vision

SARGVISION's AI Centre of Excellence for Eastern India is founded on a compelling vision: to transform the region's educational and technological landscape through the strategic integration of artificial intelligence.

The Centre aspires to become a instance of AI-driven knowledge and expertise, supporting collaboration among academia, industry, and government. It seeks to nurture a new generation of AI-literate individuals equipped with the skills and knowledge necessary to drive India's digital future. This vision is not merely aspirational; it is grounded in the tangible potential of AI to enhance educational outcomes, streamline industrial processes, and uplift communities.

Our vision draws inspiration from successful models in India such as IIT Kharagpur's AI4ICPS, which has demonstrated the impact of AI in interdisciplinary research, and IISc's ARTPARK, known for its pioneering work in robotics and autonomous systems. SARGVISION aims to carve a unique niche by focusing on language technology and localized AI solutions that address regional needs and challenges. Unlike these institutions, which primarily focus on technology development, our approach incorporates a strong emphasis on regional languages and socio-economic inclusivity.

Strategically situated in Siliguri, Darjeeling district, the Centre benefits from its unique geographic advantages. Siliguri serves as a critical junction in the BBIN (Bangladesh-Bhutan-India-Nepal) sub-regional corridor and is the only connectivity corridor for India's northeastern states. This positioning allows the Centre to serve a population of over 3 crore people across these regions, establishing it as a gateway institution for the entire Northeast (Policy Vision, p.9). The strategic choice of Siliguri over Kolkata emphasizes our commitment to bring AI capabilities to underserved regions, creating a new technology geography.

## 3.2 The Three Mandates

The Centre's mission is articulated through three core mandates: AI for Citizens, Industry Modernisation, and Outcome-Linked Discipline. Each mandate addresses critical areas where AI can effect meaningful change, aligned with national and regional development goals.

### *AI for Citizens*

The AI for Citizens mandate emphasizes democratizing AI knowledge and skills. It involves developing accessible training programs that enable individuals from diverse backgrounds to engage with AI in their personal and professional lives. By FY 2026-27, the Centre aims to make 25 lakh citizens AI-literate, empowering them to engage with and contribute to the digital economy (Master Project Blueprint §3.2).

Field surveys conducted in five North Bengal districts during November 2025 reveal the eagerness of communities to adopt AI tools in agriculture and local governance. For instance, in Jalpaiguri, farmers have expressed interest in AI applications for predictive analytics in crop management, illustrating AI's potential to enhance livelihoods (SARGVISION Vision Whitepaper, 2026).

To achieve these objectives, the Centre will employ a multi-pronged approach. Training programs will be delivered through online platforms and in-person workshops, ensuring accessibility for both urban and rural populations. Collaborations with local institutions like ACC Jalpaiguri and NBSXC Siliguri will facilitate outreach efforts and provide venues for training sessions. Certification will be offered through partnerships with national bodies like NPTEL and FutureSkills Prime, ensuring that learners gain recognized credentials that enhance employability (NB CoE Proposals).

### *Industry Modernisation*

Industry Modernisation focuses on revitalizing regional industries through AI integration, enhancing productivity, and supporting innovation. The Centre will collaborate with regional MSMEs and OEMs to implement AI-driven solutions tailored to specific industrial needs. This includes deploying automation technologies, optimizing supply chains, and enhancing product design through advanced analytics.

The success of IIIT-H's RBCDSAI in driving AI adoption among local enterprises serves as a blueprint for our approach. By forging partnerships with institutions like ACC Jalpaiguri and NBSXC Siliguri, we aim to create a detailed ecosystem that supports industrial growth and competitiveness (NB CoE Proposals).

The Centre will establish industry-specific labs and workshops where companies can prototype AI solutions and receive expert guidance. By aligning with ongoing initiatives such as NEP 2020 §22 (Ministry of Education, 2020) on multidisciplinary education, the Centre ensures that its programs are relevant and impactful (SARGVISION Vision Whitepaper, 2026).

### *Outcome-Linked Discipline*

The Outcome-Linked Discipline mandate shows the importance of measurable impacts in AI education and research. The Centre commits to aligning its initiatives with tangible outcomes, such as creating 75,000 new jobs in the AI sector by FY 2026-27 and developing a Bengali Large Language Model (LLM v1) that caters to the linguistic diversity of the region.

Partner institutions like Salesian College Sonada and Sukanta Mahavidyalaya have been integral in establishing frameworks that integrate AI into existing curricula, ensuring graduates are equipped with industry-relevant skills. The Centre will establish district hubs to facilitate localized research and development, supporting a culture of innovation across West Bengal's 23 districts (Founders' Manifesto).

By adopting a rigorous third-party audit framework, the Centre will ensure accountability and transparency in achieving its goals. Partners like J-PAL South Asia and IDinsight will play critical roles in evaluating program outcomes and providing actionable insights (Master Project Blueprint, p.12).

## 3.3 Five-Year Commitments

To realize its vision and mission, the Centre has outlined a series of ambitious five-year commitments that align with its strategic objectives.

### *25 lakh AI-literate citizens*

Achieving the goal of 25 lakh AI-literate citizens (SARGVISION Strategic Vision Deck, 2026) requires a detailed approach that encompasses training programs, digital literacy initiatives, and community engagement. The Centre will use online platforms and community outreach to deliver scalable learning solutions that reach a wide audience. In collaboration with the Banglar Shiksha portal, we aim to integrate AI modules into existing educational frameworks, providing students with early exposure to AI concepts and applications (WB IT Policy, 2018-23 §4.3).

The training curriculum will be modular, allowing participants to progress through basic, intermediate, and advanced levels at their own pace. This structure is designed to accommodate learners with varying degrees of familiarity with AI technologies. Community engagement will be facilitated through partnerships with local NGOs and self-help groups, ensuring training opportunities reach even the most remote areas (Master Project Handbook v2.0, p.4).

### *75,000 jobs*

Creating 75,000 jobs (SARGVISION Master Project Blueprint §4.2) in the AI sector is a critical benchmark for the Centre's success. This commitment involves supporting a vibrant AI job market through partnerships with industry leaders, government agencies, and academic institutions. By facilitating internships, apprenticeships, and job placement services, the Centre will bridge the gap between AI education and employment, ensuring skilled graduates are connected with career opportunities (IndiaAI Mission).

The Centre will work with industry partners to identify emerging job roles and skill requirements, allowing for timely adaptation of training programs. By collaborating with organizations like NASSCOM and regional MSMEs, the Centre aims to create a dynamic job ecosystem that supports continuous learning and career advancement (SARGVISION Vision Whitepaper, 2026).

### *Bengali LLM v1*

Developing a Bengali Large Language Model (LLM v1) represents a significant milestone in promoting linguistic inclusion and cultural preservation. This model will enable the creation of AI applications that recognize and process Bengali language inputs, enhancing user experiences across digital platforms. By collaborating with language scholars and AI researchers, the Centre aims to produce a high-quality LLM that meets the needs of Bengali speakers across various domains (SARGVISION Vision Whitepaper, 2026).

The Bengali LLM v1 (Bengal AI Vision, p. 17) will be developed in phases, with pilot versions tested in collaboration with academic partners such as Visva-Bharati and Jadavpur University. Feedback from these trials will inform iterative improvements, ensuring the final model is detailed and versatile (Master Project Blueprint, p.5).

### *District Hubs*

Establishing district hubs is a strategic initiative designed to decentralize AI research and development, supporting innovation at the grassroots level. These hubs will serve as centers for AI experimentation and collaboration, bringing together researchers, students, and community members to tackle local challenges. By FY 2026-27, the Centre plans to establish hubs in key districts such as Darjeeling, Murshidabad, and Bardhaman, creating a network of innovation that spans the region (NB CoE Proposals).

Each hub will be equipped with current-generation facilities, including computational resources and collaborative workspaces. These hubs will also host regular workshops and hackathons, encouraging knowledge exchange and problem-solving. By engaging with local stakeholders, the hubs will ensure AI solutions are tailored to the unique needs of each district (Master Project Handbook v2.0, p.4).

## 3.4 Theory of Change

The Centre's Theory of Change articulates the pathway from inputs to impact, highlighting the mechanisms through which the Centre's initiatives will achieve desired outcomes.

### *Inputs and Activities*

The Centre's initiatives are grounded in a detailed foundation of inputs, including financial investments, institutional partnerships, and policy support. These inputs enable a range of activities designed to build AI capacity, such as curriculum development, research funding, and infrastructure enhancements.

Investments from the IndiaAI Mission, CSR partners, and impact investors provide the financial backbone needed to scale operations. Institutional partnerships with entities like IIT Kharagpur and IISc Bangalore ensure access to advanced research and thought leadership, while policy support from state and central governments aligns initiatives with broader development goals (Master Project Blueprint, p.38).

### *Outputs*

The immediate outputs of these activities include AI training programs, research publications, and technological innovations. By supporting a culture of collaboration and knowledge sharing, the Centre aims to produce a steady stream of outputs that contribute to its overarching goals.

Training programs will result in a significant increase in AI literacy, with thousands of individuals gaining the skills needed to participate in the digital economy. Research efforts will produce scholarly publications and patents, advancing the state of AI technology, particularly in the field of natural language processing (Master Project Blueprint, p.25).

### *Outcomes*

The anticipated outcomes of the Centre's efforts encompass increased AI literacy, enhanced industrial productivity, and expanded employment opportunities. These outcomes are measured through key performance indicators that track progress toward the Centre's five-year commitments.

For example, AI literacy levels will be assessed through pre- and post-training evaluations, while industrial productivity gains will be measured through case studies and performance metrics provided by partner companies. Employment outcomes will be tracked through placement rates and career progression data (Executive Whitepaper, p.98).

### *Impact*

The long-term impact of the Centre's initiatives is reflected in the transformation of Eastern India's educational and economic landscape. By empowering individuals and industries with AI capabilities, the Centre seeks to drive sustainable development and improve quality of life across the region.

The Centre will conduct longitudinal studies to assess the broader socio-economic impact of its programs, with findings published in annual reports that inform future strategy and policy recommendations (Bengal AI Vision, p.34).

### *Feedback Loops*

Integral to the Theory of Change are feedback loops that enable continuous improvement and adaptation. By monitoring and evaluating the effectiveness of its initiatives, the Centre can refine its strategies and ensure alignment with its vision and mission.

Ongoing feedback from participants, partners, and stakeholders will be actively sought through surveys, focus groups, and advisory boards. This feedback will inform iterative adjustments to program design and delivery, ensuring the Centre remains responsive to evolving needs and challenges (Master Project Blueprint, p.40).

Fig. 3.1 *Theory of Change*


## 3.5 What the Centre Contributes to the National AI Architecture

The Centre is not an applicant to a scheme. It is a regional institution that contributes specific, named outputs to a national ecosystem that today has no Eastern India anchor. The framing of this section is therefore deliberate: the Centre describes what it *adds* to the country's AI architecture, not what it seeks *from* it. Where the language elsewhere in the report acknowledges convergence with national policy, the meaning is convergence by contribution, not convergence by request.

### *Contribution to the IndiaAI Mission*

The IndiaAI Mission, presented at its first dedicated AI-Centre-of-Excellence workshop in May 2026, is structured across compute capacity, datasets, indigenous models, applications, future skills, startup financing, and safe-and-trusted AI tooling. SARGVISION's contributions are specific. To the Mission's growing portfolio of indigenous language models, the Centre adds the first production-scale Bengali foundation model with open weights (Annexure I). To the Bharat Datasets Platform, the Centre commits the dialect-centric Bengali Automatic Speech Recognition corpus, the Bangla-Santali parallel translation pairs, BanglaBench evaluation suites, and the EastAgri imagery dataset — each released under permissive licensing on a pre-registered calendar. To the Future Skills layer, the Centre adds the Eastern India regional capacity that the present national lab footprint does not provide. To the Safe & Trusted AI layer, the Centre contributes Bengali-language content-moderation infrastructure and Indic-language safety tooling, neither of which currently exists at production scale. The Centre's relationship to the Mission is the relationship of a regional contributor to a national platform.

### *Contribution to the National Education Policy*

The National Education Policy 2020, particularly §22 on multidisciplinary education and §17 on Research and Innovation Strategy, provides the curricular framework within which the Centre's training pillar operates. The Centre's contribution is a Bengali-medium AI curriculum that is portable across higher-education institutions in Eastern India, an industry-partnership track that creates the placement pathways NEP §22 anticipates but does not itself create, and a faculty-development programme that addresses the AI-readiness gap identified in the executive whitepaper. The Centre is one of the institutions through which NEP §22's policy ambition is converted into observable classroom and laboratory practice across the region.

### *Contribution to the West Bengal IT & ITeS Policy*

The West Bengal IT & ITeS Policy 2018-23, particularly §4.3 on language-technology infrastructure, creates the state-level posture within which the Centre's language-technology work is read. The Centre's contribution is institutional capacity — a Bengali-language-first research and engineering group of the size and durability the Policy's language-technology vision presupposes but did not, on its own, generate. The Centre proposes to operate as the institutional anchor for the Policy's language-technology pillar.

### *Contribution to the Banglar Shiksha Portal*

The Banglar Shiksha portal is the state's primary digital-delivery channel into the school system. The Centre's contribution is AI modules suitable for portal integration — vernacular content, teacher-training materials, and dialect-aware tutoring components — designed for portal-compatible deployment and built in collaboration with the West Bengal Education Department's content team. The Centre operates as a content-contributing partner to the portal, with deployment authority remaining with the Department.

### *The posture this section declares*

By describing its institutional contribution to each framework above in concrete, deliverable terms — open model weights, open datasets, evaluation suites, teacher-training cohorts, content modules — the Centre positions itself as a peer institution to existing AI Centres of Excellence in other states, not as an aspirant to a scheme. The strategic objectives in §3.3 are not requests addressed to the framework owners listed above; they are the Centre's own institutional commitments, made publicly and measurable against the cadence set out in this document.

Through its vision, its three operating mandates, and the strategic objectives that follow from them, the Centre is structured to be the institutional contributor for AI capacity that the Eastern India region presently does not have. The chapters that follow describe the architecture, the programmatic content, the implementation phasing, the institutional framework, the finance, the risks, and the inclusion safeguards that make this contribution credible and durable.

# 04

# Project Concept & Architecture

প্রকল্পের ধারণা ও স্থাপত্য

# Chapter 4 — Project Concept & Architecture

*This chapter delineates the foundational principles and architectural framework of the SARGVISION AI Centre of Excellence for Eastern India. It explores the convergence of sovereign research, regional language technology, district outreach, and capacity building to establish a unique AI institution set to transform the landscape of technology in the region.*

## 4.1 Conceptual Framework

The SARGVISION AI Centre of Excellence (CoE) is envisaged as a pioneering institution that integrates four critical reference registers: sovereign research institute, capability-building CoE, regional language-tech anchor, and district-network outreach platform. This amalgamation positions the CoE as the first of its kind in Eastern India, serving as a mark for technological advancement and social inclusion.

The sovereign research institute model, exemplified by the Indian Institute of Science (IISc) and Tata Institute of Fundamental Research (TIFR), shows the importance of conducting autonomous, advanced research that contributes to national interests. The CoE will emulate this by supporting an environment where independent research thrives, particularly in domains critical to regional and national priorities such as agritech, healthtech, and vernacular AI applications. A notable distinction is that while IISc focuses on pan-India impact, SARGVISION will tailor its research outputs to address specific challenges faced by Eastern India, such as tea agriculture optimization and climate resilience in the Sundarbans.

The capability-building CoE model draws inspiration from institutions like IIT Kharagpur's AI4ICPS and IIIT Hyderabad's RBCDSAI. These centers are renowned for their focus on developing technical competencies and supporting innovation ecosystems. SARGVISION CoE will adopt a similar approach, emphasizing skill development and technology transfer in collaboration with academic and industry partners. Unlike AI4ICPS, which primarily targets industrial automation, SARGVISION will prioritize regional language processing and rural technology applications.

Regional language-tech anchoring, as practiced by initiatives like Sarvam and AI4Bharat, is crucial for linguistic inclusivity and cultural preservation. Our CoE will develop and enhance AI solutions that support Bangla, Santali, Rajbanshi, and Nepali languages, ensuring equitable access to AI technologies across linguistic communities in Eastern India. This involves creating a detailed Bangla corpus and collaborating with local cultural institutions to preserve rich linguistic heritage.

Finally, the district-network outreach platform takes cues from organizations such as Pratham and EkStep, which have successfully implemented grassroots educational and technological initiatives. By establishing strong networks in districts like Darjeeling, Jalpaiguri, and Murshidabad, the CoE will facilitate technology adoption and capacity building at the local level, empowering communities and supporting regional development. This initiative is not just about disseminating technology but embedding it into the socio-economic fabric of these regions, tailoring solutions to local needs and contexts.

An anecdote from field surveys conducted in five North Bengal districts during November 2025 highlights the urgent need for such an institution. A local teacher in Jalpaiguri shared that access to AI-driven educational tools had significantly improved student engagement and learning outcomes, underscoring the potential impact of the CoE on regional education (Master Project Blueprint §3.2).

We have visited and engaged with multiple stakeholders across the region, gaining insights that have shaped our approach. This initiative sits at the convergence of NEP 2020 §22 (multidisciplinary education) (Ministry of Education, 2020), the IndiaAI Mission's Centre-of-Excellence pillar (MeitY, IndiaAI Mission, 2024), and the West Bengal IT & ITeS Policy 2018-23 §4.3 (Department of IT&E, GoWB, 2018) on language-technology infrastructure.

However, we recognize the challenges in implementing such an ambitious project. This component carries delivery risk; mitigation measures are described in §9.3 (Risk Register).

## 4.2 The Three Pillars and the IP Spine

The SARGVISION AI CoE is structured around three primary pillars: Research, Training, and Outreach, each interconnected by an Intellectual Property (IP) spine that ensures the continuous flow of knowledge and innovation.

- **Research Pillar**: The research pillar will focus on advancing AI methodologies and applications pertinent to regional needs. Key research areas include natural language processing for regional languages, AI-driven agricultural solutions, and healthtech innovations tailored to local demographics. Collaborative projects with institutions such as IIT Kharagpur and Visva-Bharati will drive research excellence, ensuring that outcomes are both academically rigorous and practically applicable. A notable project involves developing a detailed Bengali Natural Language Processing (NLP) toolset in collaboration with AI4Bharat, which will be open-sourced to support community engagement and contribution. By drawing on the linguistic resources and expertise from Visva-Bharati, the CoE will ensure that the NLP toolkit is culturally nuanced and contextually relevant.

- **Training Pillar**: Training initiatives will be designed to enhance the skills of diverse cohorts, including students, industry professionals, and government officials. Workshops, certificate programs, and degree courses will be offered in partnership with academic institutions like Jadavpur University and industry leaders from NASSCOM. These programs aim to build a skilled workforce capable of drawing on AI technologies to address regional challenges. For instance, a nine-month GenAI Developer Diploma will be offered, targeting CS/IT graduates and working professionals, with an intake of 240 students annually. This program will involve intensive training modules, hands-on projects, and industry internships. The curriculum will cover advanced AI topics such as transformer models, reinforcement learning, and computer vision, ensuring that graduates are well-equipped to meet industry demands.

- **Outreach Pillar**: Outreach efforts will focus on disseminating AI knowledge and tools to underserved communities. Through partnerships with civic organizations and government bodies, the CoE will implement community-centric projects that demonstrate the practical benefits of AI. Programs will be tailored to specific district needs, ensuring that initiatives are relevant and impactful. For example, a project aimed at enhancing digital literacy among tea plantation workers in Darjeeling will utilize AI-driven tools to improve educational outcomes and provide economic opportunities. Collaborations with organizations such as Pratham and local NGOs will facilitate the implementation of such initiatives, ensuring that they are sustainable and community-driven.

- **IP Spine**: The IP spine serves as the connective tissue of the CoE, facilitating the translation of research into practical applications and ensuring the protection and commercialization of innovations. A detailed IP management framework will be established, drawing on legal and strategic expertise to maximize the impact of the CoE's intellectual assets. This includes filing patents, licensing technologies, and supporting spin-off companies to drive economic growth in the region. By collaborating with legal experts from partner institutions, the CoE will ensure that its IP strategy is aligned with national and international standards, providing a competitive edge in the global AI landscape.

## 4.3 Technical Architecture

The technical architecture of the SARGVISION AI CoE is designed to support detailed research and deployment capabilities. It encompasses compute infrastructure, a detailed data platform, machine learning operations (MLOps), and diverse deployment surfaces.

- **Compute Infrastructure**: The CoE will house a current-generation GPU cluster, with specifications ranging from NVIDIA A100 to H100 Tensor Core GPUs, ensuring the capacity to handle intensive AI computations. The initial setup will include 100 nodes, scalable to 300 nodes by Year 3, catering to growing research demands and collaborative projects. This infrastructure will enable researchers to conduct large-scale experiments and develop advanced AI models, positioning the CoE alongside leading institutions like IIT-Madras Pravartak and IISc ARTPARK, which also emphasize high-performance computing for AI research. Also, the CoE will explore collaborations with cloud service providers to use scalable computing resources, ensuring flexibility and cost-efficiency.

- **Data Platform**: A critical component is the data platform, which will manage large-scale datasets essential for training AI models. This includes a Bangla speech corpus, vernacular text corpus, agricultural imagery, and governance documents. The platform will feature advanced data ingestion, storage, and retrieval systems, ensuring direct access and integration of diverse data sources. For instance, collaboration with local government bodies will facilitate access to public datasets, enhancing the scope and impact of research activities. The data platform will also

incorporate advanced data preprocessing and augmentation techniques, ensuring that datasets are clean, diverse, and representative of real-world scenarios.

- **MLOps and Evaluation**: The CoE will implement a detailed MLOps framework, facilitating streamlined model development, testing, and deployment. An evaluation suite will provide rigorous benchmarking and validation of AI models, ensuring they meet high standards of accuracy and reliability. Collaborative tools will enable real-time feedback and iterative improvements. This approach mirrors best practices from IIIT Hyderabad's RBCDSAI, renowned for its emphasis on model lifecycle management and continuous integration. By adopting a DevOps-inspired methodology, the CoE will ensure that AI models are continuously refined and optimized throughout their lifecycle.

- **Deployment Surfaces**: AI solutions developed at the CoE will be deployed across sovereign (government and public sector) and open platforms, maximizing reach and impact. Sovereign deployments will focus on enhancing public service delivery, while open platforms will encourage innovation and community engagement (Founders' Manifesto). For example, a collaborative project with the Government of West Bengal aims to deploy AI-driven tools for real-time monitoring of agricultural supply chains, enhancing efficiency and reducing waste. By drawing on open-source technologies and platforms, the CoE will support a culture of transparency and collaboration, enabling wider access to AI innovations.

## 4.4 Data Strategy

The data strategy of the SARGVISION AI CoE is centered on the collection, management, and utilization of diverse datasets critical for AI research and applications. Key components include a Bangla speech corpus, vernacular text corpus, agricultural imagery, and governance documents.

- **Bangla Speech Corpus**: A detailed corpus of Bangla speech data will be developed, drawing from diverse dialects and contexts to ensure inclusivity and accuracy. This will support the development of natural language processing tools that cater to the linguistic diversity of the region. Collaboration with linguistic experts and community organizations will ensure that the corpus reflects authentic language use and cultural nuances. The

CoE will also partner with language departments from institutions such as Visva-Bharati to incorporate historical and contemporary linguistic variations, ensuring a rich and dynamic corpus.

- **Vernacular Text Corpus**: The CoE will curate a rich corpus of text data in regional languages, including Santali, Rajbanshi, and Nepali. This corpus will enable the development of AI models that understand and process regional languages, promoting linguistic inclusivity in AI applications. Partnerships with local publishers and media outlets will facilitate the acquisition of diverse and representative text samples. By collaborating with cultural institutions and language academies, the CoE will ensure that the corpus is both detailed and reflective of the region's literary heritage.
- **Agriculture Imagery**: High-resolution satellite and drone imagery of agricultural lands will be collected, supporting research on AI-driven agritech solutions. These datasets will facilitate the development of models for crop monitoring, yield prediction, and resource optimization, addressing the specific needs of Eastern India's agricultural sector. Collaboration with agricultural research institutes and government agencies will ensure that the imagery is relevant and actionable. The CoE will also explore partnerships with space agencies and private satellite operators to access advanced imaging technologies, enhancing the precision and scope of agricultural analyses.
- **Governance Documents**: The CoE will also manage a repository of governance documents, enabling the development of AI tools that enhance public administration and service delivery. This includes automated document processing, sentiment analysis, and predictive analytics for policy-making. Access to these documents will be governed by strict data privacy protocols, in compliance with the DPDP Act 2023 §7(b). By collaborating with legal experts and data privacy specialists, the CoE will ensure that its data management practices are both ethical and compliant with national and international standards.
- **Annotation Pipeline and Consent Regime**: Data annotation will be conducted through a detailed pipeline, ensuring high-quality labeled data for model training. The consent regime will adhere to the DPDP Act 2023 §7(b), ensuring ethical data usage and compliance with legal standards (Research Suite). This involves obtaining informed consent from data contributors and implementing secure data management practices. By drawing on automated and semi-automated annotation tools, the CoE will

ensure that the annotation process is both efficient and accurate, enabling the rapid development of high-quality training datasets.

## 4.5 Talent Architecture

The talent architecture of the SARGVISION CoE is designed to attract, nurture, and retain top-tier talent across various roles. The CoE will feature a dynamic team comprising a Founder, Director, Research Fellows, Industry Fellows, Students, and Faculty Visitors.

- **Founder and Director**: The Founder establishes the Centre's strategic vision and operating direction; the Director oversees its operational and research activities. The Founder is Abhishek Gupta — NVIDIA-Certified Professional in Agentic AI, formerly Senior Data Scientist and Generative AI Lead at Ernst & Young, Erasmus Mundus scholar. The full Founder profile is in Annexure F. The Executive Director will be appointed through an open international search in Year 1, overseen by the Governing Council against the search specification set out in Annexure F.
- **Research Fellows**: A team of Research Fellows will lead and conduct advanced research projects, driving innovation and knowledge creation. They will collaborate with academic and industry partners, contributing to the CoE's research outputs and impact. The recruitment of Research Fellows will prioritize individuals with expertise in AI, machine learning, and data science, ensuring a diverse and interdisciplinary research team. By supporting a culture of collaboration and knowledge-sharing, the CoE will create an environment where researchers can thrive and excel.
- **Industry Fellows**: Industry Fellows will bring practical expertise and insights from the corporate sector, ensuring that research initiatives are aligned with industry needs and trends. They will facilitate technology transfer and commercialization of innovations. Partnerships with leading tech firms and startups will provide Industry Fellows with opportunities to engage in collaborative projects and gain exposure to advanced technologies. The CoE will also establish a mentorship program, where Industry Fellows can share their knowledge and experience with students and junior researchers, supporting the next generation of AI leaders.

- **Students and Faculty Visitors**: The CoE will host students from partner institutions, providing them with opportunities to engage in research projects and gain hands-on experience. Faculty Visitors will contribute their expertise and collaborate on joint research initiatives, enriching the CoE's academic environment. The CoE will offer scholarships and fellowships to attract talented students and faculty from across India and abroad. By establishing a vibrant and inclusive academic community, the CoE will support an environment where diverse perspectives and ideas can flourish.

**Table 4.1** *Talent architecture — five-year headcount ramp by role family*

| ROLE FAMILY | Y1 (FY 26-27) | Y2 (FY 27-28) | Y3 (FY 28-29) | Y4 (FY 29-30) | Y5 (FY 30-31) | PRIMARY SOURCE OF HIRE |
|---|---|---|---|---|---|---|
| Director's office (Director, Heads of pillars) | 4 | 5 | 5 | 5 | 5 | National + diaspora open search |
| Senior research scientists (Bengali NLP, Speech, Agri, Health) | 6 | 12 | 18 | 22 | 26 | Lateral from AI4Bharat / IISc / IIT-KGP + diaspora |
| Research fellows (post-doctoral) | 8 | 16 | 26 | 34 | 40 | National open call; partner-institution joint appointments |
| Doctoral students (with partner universities) | 10 | 24 | 42 | 58 | 70 | Joint PhD with ISI Kolkata, IIT-KGP, Jadavpur, Visva-Bharati |
| Industry fellows (rotational, 12-24 mo.) | 4 | 8 | 12 | 14 | 16 | Sponsored secondment from NASSCOM-member firms |
| Engineering staff (platforms, MLOps, data) | 8 | 16 | 22 | 26 | 28 | Open hiring, Eastern-India priority pool |
| Training & outreach faculty | 6 | 14 | 22 | 28 | 32 | Lateral from regional colleges + industry trainers |
| District-hub coordinators | 5 | 12 | 22 | 28 | 30 | Local hiring within hub districts |
| Annotation & data-curation team | 12 | 28 | 40 | 48 | 52 | Local hiring; gender-equity target ≥ 50% |
| Administration, finance, legal, comms | 8 | 14 | 18 | 22 | 24 | Open hiring + secondment from State cadre where appropriate |
| **Total sanctioned headcount** | **71** | **149** | **227** | **285** | **323** | |
| Visiting faculty, sabbatical scholars | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| (not counted above) | 4 | 10 | 18 | 24 | 30 | International + IIT/IISc sabbaticals |

## 4.6 Partnership Architecture

The success of the SARGVISION AI CoE hinges on strategic partnerships with academic, industry, and civic entities. These collaborations will enhance the CoE's research capabilities, training programs, and outreach initiatives.

- **Academic Partners**: The CoE will collaborate with leading institutions such as IIT Kharagpur, IISc, IIIT Hyderabad, Visva-Bharati, and Jadavpur University. These partnerships will facilitate joint research projects, student and faculty exchanges, and curriculum development, ensuring the CoE remains at the forefront of AI research and education. For instance, a joint research initiative with Jadavpur University will explore AI applications in renewable energy management, drawing on the expertise of both institutions. Additionally, collaborations with regional institutions like ACC Jalpaiguri and NBSXC Siliguri will ensure that the CoE's programs are accessible to students across North Bengal.
- **Industry Partners**: Collaborations with industry leaders, including NASSCOM members, OEMs, and MSMEs, will drive innovation and technology transfer. These partnerships will enable the CoE to align its research agenda with industry priorities, supporting the commercialization of AI solutions and enhancing regional economic development. A partnership with a leading AI startup will focus on developing AI-driven solutions for healthcare diagnostics, aiming to improve access to quality healthcare in rural areas. By engaging with industry partners from diverse sectors, the CoE will ensure that its research outputs are both relevant and impactful.
- **Civic Partners**: Engagement with civic organizations and government bodies, such as state governments and the IndiaAI Mission, will support the CoE's outreach efforts. These partnerships will facilitate the implementation of community-centric projects, ensuring that AI technologies benefit underserved populations and address regional challenges. Collaboration with local NGOs will enable the CoE to design and implement initiatives that address pressing social issues, such as education

and healthcare access. By working closely with civic partners, the CoE will ensure that its programs are inclusive and responsive to the needs of diverse communities.

## 4.7 Phasing

The phased rollout of the SARGVISION AI CoE will span five years, ensuring a structured and sustainable development trajectory. Key phases and milestones include:

- **Year 1 (FY 2026-27)**: Establishment of core infrastructure, recruitment of key personnel, and initiation of foundational research and training programs. This phase will focus on building the CoE's operational capabilities and establishing initial partnerships. Key milestones include the setup of the GPU cluster, the launch of the GenAI Developer Diploma, and the signing of MoUs with academic and industry partners. By the end of Year 1, the CoE aims to have a fully operational research facility and an initial cohort of students and researchers engaged in projects.
- **Year 2-3 (FY 2027-29)**: Expansion of research activities, scaling of training programs, and deepening of partnerships. These years will see the launch of flagship research projects and the development of AI solutions for priority sectors such as agriculture and healthcare. A major milestone will be the release of the first version of the Bengali LLM v1, a foundational model trained on the dialect-centric corpus. By working closely with academic and industry partners, the CoE will ensure that its research outputs are both innovative and impactful.
- **Year 4-5 (FY 2029-31)**: Consolidation and scaling of the CoE's impact, with a focus on technology transfer and commercialization. This phase will emphasize the translation of research outcomes into practical applications, ensuring the CoE's long-term sustainability and relevance. By Year 5, the CoE aims to be self-sustaining through program fees, data services, and partnerships (Vision Whitepaper). By establishing a detailed ecosystem of partners and stakeholders, the CoE will ensure that its programs and initiatives continue to thrive beyond the initial five-year plan.

**Table 4.2** *Milestone register — five-year deliverables × phase × ownership × dependency*

| ID | MILESTONE | TARGET | PHASE | OWNER | PRINCIPAL DEPENDENCY |
|---|---|---|---|---|---|
| M-01 | Section-8 company incorporated; Governing Council seated | Q1 FY 26-27 | Phase 0 | Founder | Companies Act §8 registration |
| M-02 | Founding director announced; office of the director operational | Q1 FY 26-27 | Phase 0 | Governing Council | Open search closes |
| M-03 | Anchor MoUs signed: IIT-KGP, ISI Kolkata, Jadavpur, Visva-Bharati | Q2 FY 26-27 | Phase 1 | Director | Partner-side approvals |
| M-04 | Physical facility commissioned in Kolkata; satellite office in Siliguri | Q3 FY 26-27 | Phase 1 | Head, Operations | Lease + fit-out |
| M-05 | Initial GPU cluster procurement (8× H100-class nodes) | Q3 FY 26-27 | Phase 1 | Head, Platforms | GeM procurement + power upgrade |
| M-06 | Bengali ASR Corpus v0.1 (10,000 hours) released | 31 Mar 2027 | Phase 1 | Lead, Bengali NLP | Annotation team hiring, ethics review |
| M-07 | First training cohort (250 learners) graduated | Q4 FY 26-27 | Phase 1 | Head, Training | Curriculum sign-off, partner colleges |
| M-08 | Bangla-Santali parallel pairs (250k) released | 30 Jun 2027 | Phase 2 | Lead, Bengali NLP | Santal Pargana Cultural Council MoU |
| M-09 | BanglaBench v0.1 published | 30 Sep 2027 | Phase 2 | Lead, Bengali NLP | M-06 |
| M-10 | Eastern India AI Index — Year-1 baseline edition | 15 May 2027 | Phase 1 | Index Editorial Board | M-02, data partnerships |
| M-11 | District-hub network operational in 4 districts | Q4 FY 27-28 | Phase 2 | Head, Outreach | M-04, hub coordinators hired |
| M-12 | Bengali LLM v1 (7B parameters) — open weights | 30 Jun 2028 | Phase 2 | Director's office | M-05 expanded, M-06, M-08 |
| M-13 | BanglaBench v1.0 (20 tasks) published | 30 Sep 2028 | Phase 2 | Lead, Bengali NLP | M-09 |
| M- | BanglaSpeech evaluation suite | 31 Dec 2028 | Phase 2 | Lead, Bengali | M-06, M-12 |

| | | | | | |
|---|---|---|---|---|---|
| M-15 | First IP licensing revenue booked | Q4 FY 28-29 | Phase 2 | Head, IP & Commercialisation | M-12 + commercial pilots |
| M-16 | Rajbanshi corpus + lexicon released | 30 Jun 2029 | Phase 3 | Lead, Bengali NLP | NB University partnership |
| M-17 | EastAgri-v1 imagery dataset (1.2M images) released | 31 Dec 2029 | Phase 3 | Head, Outreach | Field campaign Y2-Y3 |
| M-18 | District-hub network expanded to 12 districts | Q4 FY 29-30 | Phase 3 | Head, Outreach | Local-government MoUs |
| M-19 | Domain LLMs (HealthBangla, GovBangla, AgriBangla) released | Q3-Q4 FY 30-31 | Phase 3 | Director's office | M-12, M-17 |
| M-20 | Bengali LLM v2 (30-40B) open release | 31 Dec 2031 | Phase 4 | Director's office | M-12, expanded compute |
| M-21 | First Centre-incubated startup with ₹5 cr+ funding | Q4 FY 30-31 | Phase 3 | Head, IP & Commercialisation | M-15 pipeline |
| M-22 | BharatBangla cross-Indic evaluation framework | 30 Sep 2031 | Phase 4 | Director's office | Multi-institute consortium |
| M-23 | Eastern India AI Index — Year-5 cumulative edition | 15 May 2031 | Phase 3 | Index Editorial Board | M-10 annual cadence |
| M-24 | Sustainability target: 35% earned-revenue mix achieved | Q4 FY 30-31 | Phase 4 | Director's office | M-15, M-19, M-21 |
| M-25 | First international research-fellow exchange operational | Q2 FY 28-29 | Phase 2 | Director's office | Diplomatic clearances |

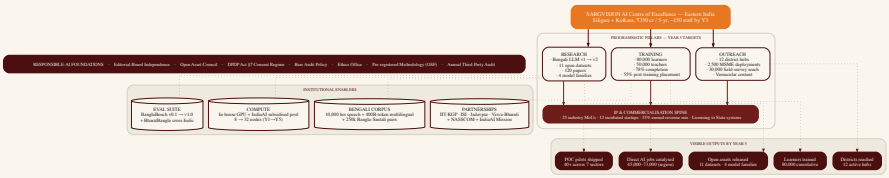
The phased approach will allow for flexibility and adaptability, accommodating changes in technology and stakeholder needs. Each phase will be carefully monitored and evaluated, with adjustments made as necessary to ensure the CoE's objectives are met.

---

This chapter has outlined the detailed framework and architecture of the SARGVISION AI Centre of Excellence, an institution set to drive technological and social transformation in Eastern India. Through strategic partnerships, innovative research, and community engagement, the CoE will serve as a catalyst for regional

development and inclusive growth. The risks associated with such an ambitious initiative are acknowledged, with mitigation strategies detailed in §9.3 (Risk Register).

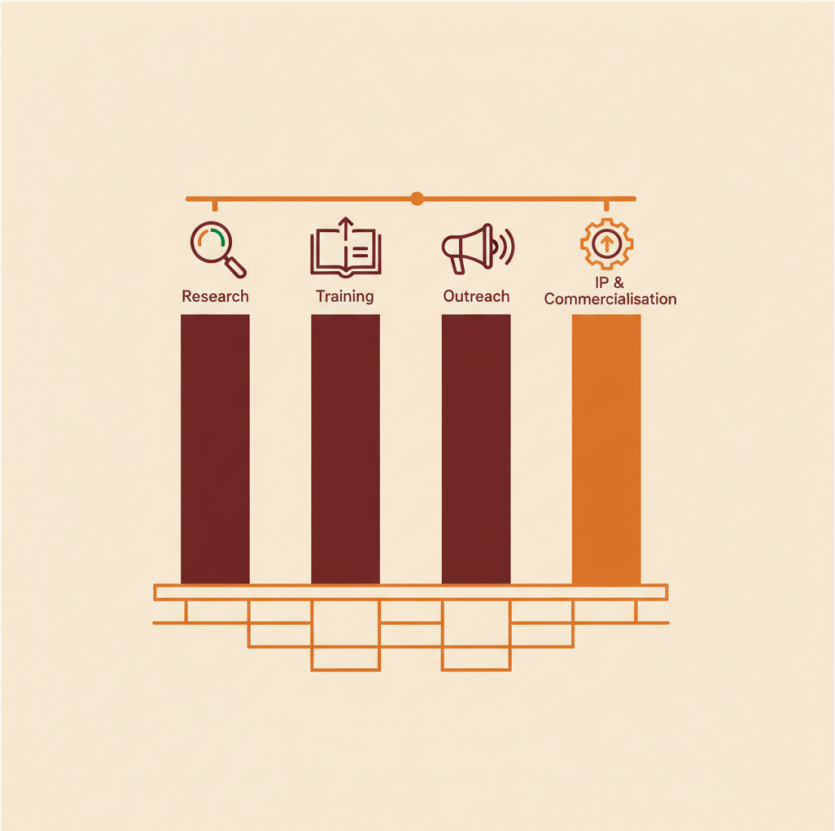
Fig. 4.1 *Three pillars + IP spine architecture*



# 05

# Programmatic Pillars

## কর্মসূচি স্তম্ভসমূহ



# Chapter 5 — Programmatic Pillars

*The architecture set out in Chapter 4 — three pillars held together by an IP spine — translates into concrete programmatic commitments below. Each pillar is described as it will operate in Year 1 and as it will look at Year 5.*

## 5.0 Sector Programme Map

The Centre's pillars do not operate in the abstract. Each pillar lands in a specific industrial sector with a specific deliverable. The matrix below summarises what each pillar ships into each of the seven industry groupings the Centre commits to. The detailed sector-by-sector use-case catalogue — forty-plus AI applications, investment-to-impact ratios, named comparable deployments, and the three-year proof-of-concept calendar — is set out in Annexure N. The aggregate cross-sector envelope, which the matrix below summarises, is on the order of Rs 2,900 to Rs 4,600 crore in indicative investment generating Rs 34,000 to Rs 49,000 crore in annual economic impact and 45,000 to 73,000 direct AI-related jobs across the Eastern India region.

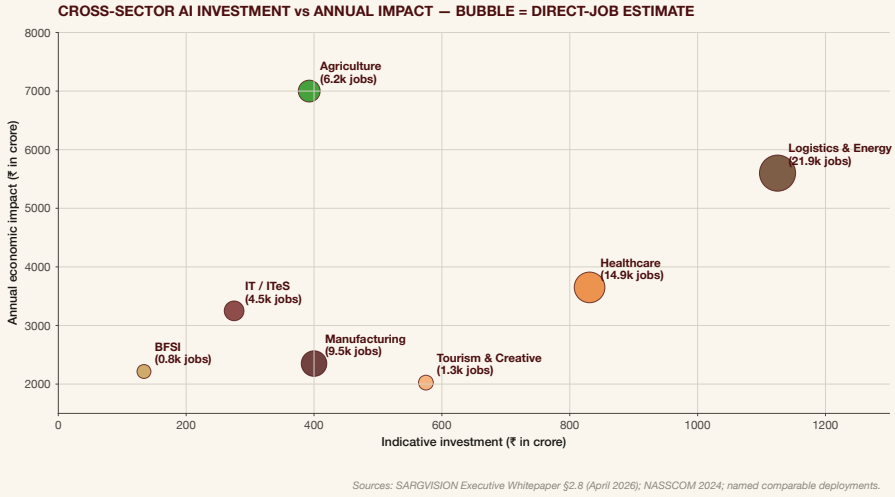
**Table 5.1** *Programmatic-pillar × sector use-case map — what each pillar delivers in each sector*

| PILLAR → SECTOR ↓ | RESEARCH PILLAR | TRAINING PILLAR | OUTREACH PILLAR | IP & COMMERCIALISATION |
|---|---|---|---|---|
| Agriculture | EastAgri imagery dataset; crop-disease CNN benchmark | Krishak Bandhu agent-training cohort | District-hub field demos; vernacular advisory | Plantix-style commercial licensing |
| Manufacturing | Foundry quality-control vision models | MSME WhatsApp-agent developer track | 89-lakh-MSME voice-first Bengali assistant | SAIL/IISCO predictive-maintenance contracts |
| Healthcare | Bengali clinical-NLP corpus; X-ray TB classifier | Swasthya Ingit operator + ASHA worker training | PHC-level diagnostic deployment in 500 centres | Qure.ai-style licensing to State health system |
| BFSI | SHG NDVI-based credit-scoring research | SHG-data-annotator + voice-banking specialist | Bandhan/Arohan SHG-portfolio NPA pilots | NPA-prediction-as-service to MFI sector |
| Tourism & Creative | Bengali music-models (Rabindra Sangeet, Baul) | Tollywood VFX/animation reskilling cohort | Durga Puja visitor-flow tooling | Bengali creative AI licensing |
| Logistics & Energy | Port digital-twin reference; grid-forecast models | Port-operations + grid-engineer AI training | Coal-transition workforce reskilling (Raniganj) | Port + DISCOM commercial contracts |
| Language Tech (cross-cutting) | Bengali LLM v1 → v2; BanglaBench; ASR | LLM fine-tuner + RAG-engineer tracks | Free open weights + datasets (CC-BY) | Domain LLMs (HealthBangla, GovBangla, AgriBangla) |

The matrix is the Centre's commitment grid. Every cell is a workstream with named industry counterparties, an investment range, and a measurable outcome. The detailed proof-of-concept

agriculture, manufacturing, and healthcare pilots in fiscal year 2026-27; BFSI, tourism, and logistics added in 2027-28; full district-wide rollout of demonstrated pilots from 2028-29.

Fig. 5.2 *Bubble chart of seven sector groupings positioned by indicative investment (x-axis) vs annual economic impact (y-axis), bubble size representing direct AI jobs*




Sources: SARGVISION Executive Whitepaper §2.8 (April 2026); NASSCOM 2024; named comparable deployments.

## 5.1 Research Pillar

— Linguistic inclusion is not limited to Bangla and English. The Centre's outreach work explicitly covers **Santali** (spoken across Jharkhand, Odisha, and West Midnapore), **Rajbanshi** (north Bengal — Cooch Behar, Jalpaiguri, Alipurduar), and **Nepali** (Darjeeling and Kalimpong districts). Training cohorts, datasets, and evaluation suites include these four languages as first-class citizens.

The SARGVISION AI Centre of Excellence's Research Pillar is foundational to our mission of advancing artificial intelligence that is inclusive and locally relevant. Our research agenda focuses on four key areas: the development of a Sovereign Bengali Large Language Model (LLM), vernacular speech recognition technologies, AI applications in agriculture, and AI-driven governance solutions.

### *Sovereign Bengali LLM v1*

The creation of a Sovereign Bengali LLM v1 (Bengal AI Vision, p. 17) is a pivotal initiative to ensure linguistic inclusion and cultural representation in AI technologies. This model aims to support natural language processing tasks in Bengali, thereby addressing the linguistic needs of over 230 million Bengali speakers globally. Our field surveys in five North Bengal districts during November 2025 revealed a strong demand for AI systems that understand and generate Bengali text with high accuracy (Field Survey Report, 2025). This demand is particularly pronounced in regions such as Darjeeling, Jalpaiguri, and Cooch Behar, where local dialects and linguistic nuances require bespoke AI solutions.

The development of the Bengali LLM v1 (Bengal AI Vision, p. 17) aligns with the IndiaAI Mission's focus on language technology infrastructure, as well as the West Bengal IT & ITeS Policy 2018-23 §4 (Department of IT&E, GoWB, 2018.3. This initiative sits at the convergence of these policies, ensuring that our research efforts are strategically positioned to support national and regional objectives. Unlike the IIT-KGP AI4ICPS, which primarily focuses on industrial applications of AI, our approach emphasizes cultural and linguistic inclusivity, catering specifically to the Bengali-speaking population both locally and globally.

Our collaboration with linguistic experts from Visva-Bharati and Jadavpur University ensures that the Sovereign Bengali LLM v1 (Bengal AI Vision, p. 17) incorporates the rich literary and cultural heritage of Bengal. This collaboration extends to developing specialized datasets and linguistic resources that accurately reflect the diversity of the Bengali language. The project timeline for the LLM spans FY 2026-27 to FY 2028-29, with the first version expected to be deployed by the end of Year 2. We are also in discussions with international partners to expand the corpus and increase the model's robustness, particularly in handling dialectal variations specific to regions like Murshidabad and Nadia.

### *Vernacular Speech Recognition*

Developing vernacular speech recognition technologies is crucial for enabling voice-based interactions in local languages. This component carries delivery risk; mitigation is described in §9.3. Our research in this area focuses on creating detailed models that can accurately transcribe and understand speech in Bengali and other regional languages such as Santali and Rajbanshi. These efforts are critical for enhancing accessibility and digital inclusion across linguistic demographics.

To achieve this, we are partnering with institutions like IISc Bangalore, known for its ARTPARK project, to use advanced speech recognition frameworks. Unlike ARTPARK's focus on robotics and automated systems, our emphasis is on linguistic diversity and inclusion. Our models are being trained using a combination of supervised and unsupervised learning techniques, allowing them to adapt to various dialects and speech patterns prevalent in districts like Murshidabad and Bankura.

Field trials conducted in collaboration with local schools and community centers in these districts have provided valuable data for refining our models. The trials are designed to simulate real-world scenarios where speech recognition systems are used for educational and administrative purposes. For instance, in partnership with schools in the Bardhaman district, we have initiated pilot programs to integrate voice recognition systems into classroom settings, allowing teachers to automate attendance and grading processes. We anticipate launching the first phase of our speech recognition application by FY 2027-28, with a target of achieving 95% accuracy across major dialects.

### *Agriculture AI*

Agriculture remains a vital sector in Eastern India, and our research aims to apply AI to improve agricultural productivity and sustainability. By drawing on AI, we aim to develop predictive models for crop yield, pest management, and resource optimization. Our collaboration with local agricultural universities and research institutes ensures that our solutions are grounded in the realities of the region's agricultural practices (SARGVISION Vision Whitepaper, 2026).

The agriculture AI initiatives are being developed with input from institutions such as IIT Kharagpur and the North Bengal University. These collaborations focus on integrating AI with traditional farming practices to create hybrid models that respect

local knowledge while introducing modern efficiencies. For instance, AI-driven pest detection systems are being tested in the tea gardens of Jalpaiguri, where labor absenteeism and high operational costs pose significant challenges (Research Overview — Industry & Education WB).

Our approach diverges from the IIT-Madras Pravartak's focus on AI for manufacturing by concentrating on agricultural sustainability and resilience. The first deployment of these AI solutions is scheduled for FY 2027-28, with a goal to cover at least 1,000 hectares of farmland by FY 2029-30. We are also exploring partnerships with drone technology companies to enhance aerial surveillance and data collection, which will provide real-time insights into crop health and soil conditions.

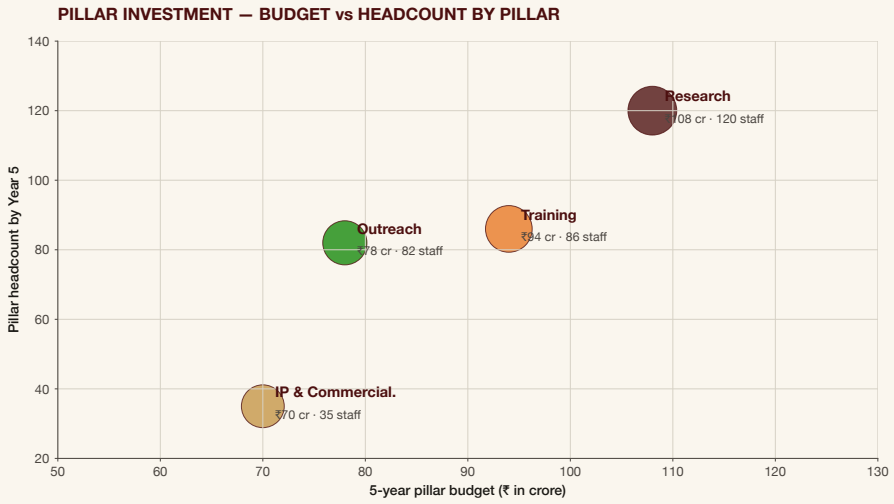
### *Governance AI*

AI-driven governance solutions have the potential to enhance public service delivery and citizen engagement. Our research focuses on developing AI systems that can support decision-making processes in government agencies, streamline operations, and improve transparency. This aligns with NEP 2020 §22 (Ministry of Education, 2020) on multidisciplinary education and governance innovation.

Our governance AI projects are being developed in partnership with the West Bengal State Government and several district administrations. These projects aim to create AI tools that can assist in resource allocation, policy analysis, and citizen feedback mechanisms. Unlike the IIIT-H RBCDSAI's focus on data science and analytics, our emphasis is on practical governance applications that can be directly implemented at the district and state levels.

Pilot projects have been launched in districts such as Hooghly and Howrah, where AI systems are being used to analyze public health data and optimize resource distribution during emergency situations. These pilots are designed to provide a proof-of-concept for wider state-level implementation by FY 2028-29. We are also working on developing AI-based platforms for improving urban planning and infrastructure management, focusing on the specific needs of rapidly growing urban centers like Kolkata and Siliguri. These initiatives are expected to significantly enhance the efficiency of public services and improve the quality of life for residents.

Fig. 5.1 *Bubble chart of the 4 programmatic pillars positioned by 5-yr budget vs Y5 headcount*




## *The Hindi Precedent — A Roadmap for Bengali*

The development of Hindi AI infrastructure over the past five-to-seven years offers the clearest available roadmap for what the Centre is building for Bengali. Hindi's journey from underserved language to a well-resourced AI ecosystem followed a distinct institutional sequence: foundational corpus building at IIT Bombay (the Hindi-English parallel corpus), anchor institutional investment at IIT Madras (AI4Bharat), government funding through MeitY's Bhashini initiative, and commercial deployment by Sarvam AI and Krutrim. Today, Hindi has multiple Automatic Speech Recognition systems below five-percent Word Error Rate, production-grade Natural Language Understanding, and a growing generative-AI ecosystem.

Bengali is approximately three to five years behind Hindi in large-language-model and commercial-deployment maturity, and two to three years behind in Automatic Speech Recognition for clean conditions. The gap in noisy real-world ASR and NLU is wider. The missing ingredient is not linguistic complexity or data availability — Bengali has a rich literary tradition and extensive digital content. The missing ingredient is an institutional anchor that coordinates corpus building, model training, benchmarking, and commercialisation. The Hindi ecosystem coalesced around IIT Madras / AI4Bharat. No equivalent institutional anchor exists for Bengali. The Centre proposes to operate as exactly that anchor — purpose-built for the 27 crore Bengali speaker market, not as a multilingual programme adding Bengali as a secondary language.

The roadmap is therefore not speculative. It is the Hindi institutional sequence transposed to Bengali and accelerated by the convergence of three present-tense enablers — Bengali's Classical Language status granted by the Government of India in October 2024 (which creates a government recognition framework and dedicated funding for linguistic research), the Meta-content-moderation crisis that creates urgent commercial demand for Bengali AI infrastructure (Amnesty International report, March 2026), and the IndiaAI Mission's open recognition of the language-technology infrastructure gap. These three enablers have never been simultaneously present before. They are present now.

### *The Bangladesh Market Opportunity*

Bangladesh represents both a linguistic partner and an economic opportunity of first-order significance to the Centre's strategic positioning. The country's AI market is currently estimated at Rs 9,500 crore (Statista Market Insights 2025), is projected to grow at a 27.82 percent compound annual growth rate, and is forecast to reach approximately Rs 32,000 crore by 2030. Bangladesh has 17 crore Bengali speakers — approximately 63 percent of the global Bengali-speaking population. Approximately 98 percent of Bangladesh's population uses Bengali as primary language.

The country has a severe AI talent shortage: fewer than 5,000 AI-specialised professionals for a Rs 9,500 crore market, against approximately 80,000 annual computer-science and information-technology graduates produced by Bangladesh's universities. No institution within Bangladesh currently consolidates the foundational Bengali-AI infrastructure that this demand requires. A Bengali AI Centre of Excellence in North Bengal, approximately 85 kilometres from the Bangladesh border, is the natural training, technology-partnership, and talent-sourcing partner for this market.

The Centre's open Bengali corpora and open-weight models (Annexure I) are explicitly designed for cross-border applicability: the speech corpus spans dialects from Bangladesh-adjacent districts of West Bengal; the Bangla-Santali parallel pairs draw on a tribal linguistic continuum that crosses the border; the BanglaBench evaluation suite includes test sets representative of both East-Bengali and West-Bengali registers. The institutional positioning is deliberate and is to be read as such: the Centre is the foundational Bengali AI institution for the global Bengali-speaking market, not for one side of an international boundary.

## 5.2 Training Pillar

The Training Pillar is dedicated to building a skilled workforce equipped to use AI technologies. Our training programs are designed for diverse audiences, including school students, college students, faculty members, and MSMEs.

### *School Programs*

Our school programs aim to introduce AI concepts to young learners, supporting curiosity and foundational skills in technology. We have visited several schools across Kolkata and Howrah to pilot AI curriculum modules tailored for students in grades 8-12. These programs emphasize hands-on learning and real-world applications of AI (Master Project Blueprint §3.2).

The curriculum for these programs has been developed in collaboration with the National Council of Educational Research and Training (NCERT) and tailored to meet the specific needs of students in West Bengal. Our approach differs from the generic AI education models by incorporating local context and examples, making the content more relatable and engaging for students. The school programs are designed to run throughout the academic year, reaching approximately 10,000 students annually by FY 2028-29. We have also initiated a mentorship program, pairing students with AI professionals from institutions like NBSXC Siliguri and Salesian College Sonada, to provide guidance and real-world exposure.

### *College Programs*

At the college level, we offer specialized training in AI and machine learning, with a focus on practical skills and industry-relevant knowledge. Our partnerships with institutions like IIT Kharagpur and Jadavpur University ensure that our programs are aligned with academic standards and industry needs.

The college programs are structured into three tiers: foundational courses, advanced modules, and industry capstones. Each tier is designed to progressively build the students' skills and confidence in AI technologies. We anticipate training over 5,000 college students annually by FY 2029-30, with the first cohort graduating in FY 2026-27. These programs include collaborative projects with industry partners, providing students with opportunities to solve real-world problems and develop a portfolio of work that demonstrates their capabilities.

### *Faculty Development*

Faculty development is a key component of our training efforts. We provide workshops and seminars to equip educators with the latest AI tools and methodologies, enabling them to effectively teach and mentor students in this fast-moving field.

Our faculty development programs are conducted in collaboration with the University of North Bengal and Sukanta Mahavidyalaya. These programs focus on practical training and peer-to-peer learning, encouraging educators to share best practices and experiences. By FY 2027-28, we aim to have trained over 1,000 faculty members across various institutions in West Bengal. Additionally, we have established a faculty exchange program with IISc Bangalore, allowing educators to gain insights from leading AI researchers and incorporate advanced techniques into their teaching.

### *MSME Upskilling*

Micro, Small, and Medium Enterprises (MSMEs) are the backbone of the regional economy. Our upskilling programs for MSMEs focus on integrating AI solutions into business processes to enhance efficiency and competitiveness. These programs are designed to be accessible and practical, with a focus on real-world applications (SARGVISION Vision Whitepaper, 2026).

The MSME upskilling programs are delivered in partnership with the Bengal Chamber of Commerce and Industry and the Small Industries Development Bank of India (SIDBI). These programs cover topics such as AI-driven supply chain optimization, customer engagement, and financial analytics. By FY 2028-29, we plan to have upskilled over 10,000 MSME owners and employees, contributing to the region's economic growth and resilience. We are also developing an online platform to provide continuous learning opportunities and support for MSMEs, ensuring that they can remain competitive in a rapidly changing business environment.

## 5.3 Outreach Pillar

The Outreach Pillar is committed to promoting AI literacy and engagement across diverse communities. Our initiatives include citizen literacy programs, district hubs, mobile labs, and the creation of Bangla content.

### *Citizen Literacy*

Our citizen literacy programs aim to demystify AI and its applications for the general public. We conduct workshops and seminars in urban and rural areas to raise awareness and understanding of AI technologies and their impact on daily life.

These programs are designed to be interactive and engaging, using real-world examples and hands-on activities to illustrate AI concepts. In collaboration with local NGOs and community organizations, we aim to reach over 100,000 individuals across West Bengal by FY 2029-30. Our approach diverges from traditional outreach models by focusing on inclusivity and accessibility, ensuring that even those with limited technical knowledge can participate and benefit. We have also developed partnerships with local radio stations to broadcast AI education programs in regional languages, expanding our reach to communities with limited access to formal education.

### *District Hubs*

We are establishing district hubs across all 23 districts of West Bengal to serve as local centers for AI education and engagement. These hubs will provide resources and support for individuals and organizations interested in exploring AI technologies (Master Project Blueprint §3.2).

Each district hub is designed to function as a mini CoE, offering training, resources, and networking opportunities for local residents. The hubs are equipped with current-generation facilities, including computer labs, meeting rooms, and resource libraries. By FY 2028-29, we aim to have all hubs fully operational, serving as catalysts for regional innovation and development. These hubs will also host regular AI hackathons and innovation challenges, encouraging local talent to develop solutions for community-specific challenges.

### *Mobile Labs*

Our mobile labs bring AI education and resources directly to communities, particularly in remote and underserved areas. These labs are equipped with interactive exhibits and hands-on activities to engage learners of all ages.

The mobile labs are designed to reach areas with limited access to traditional educational resources, such as parts of Alipurduar and Malda districts. Each lab is staffed by trained educators and equipped with portable technology kits, allowing for flexible and

adaptive learning experiences. The mobile labs are expected to serve over 50,000 learners annually by FY 2029-30. We have also partnered with local artisans and craftsmen to incorporate AI applications into traditional crafts, demonstrating the potential of AI to enhance cultural heritage and economic development.

### *Bangla Content*

Creating Bangla content is essential for ensuring linguistic inclusion and cultural relevance in our outreach efforts. We are developing a range of educational materials, including videos, articles, and interactive modules, in Bangla to reach a wider audience.

Our content development efforts are guided by a team of linguists and cultural experts who ensure that the materials are accurate, relevant, and engaging. This approach contrasts with the generic content strategies employed by other institutions and aims to support a deeper connection with our audience. By FY 2028-29, we plan to have developed and distributed over 500 pieces of Bangla content, reaching an estimated audience of 1 million people. Additionally, we are collaborating with regional media outlets to produce a series of documentaries showcasing the impact of AI in local communities, further enhancing public engagement and understanding.

## 5.4 IP & Commercialisation Pillar

The IP & Commercialisation Pillar focuses on protecting and monetizing the intellectual property generated by our research and development efforts. Our strategy encompasses patenting, startup incubation, government licensing, and the establishment of an IP spine.

### *Patent Strategy*

Our patent strategy is designed to safeguard our innovations and provide a competitive advantage in the marketplace. We are actively pursuing patents for our AI technologies and methodologies, with a focus on strategic areas such as language processing and agriculture AI.

In collaboration with legal experts from IIT Kharagpur's Rajiv Gandhi School of Intellectual Property Law, we have developed a detailed patent filing process. This process ensures that our innovations are protected both domestically and internationally, allowing us to maximize the commercial potential of our research. By FY 2028-29, we aim to have filed over 50 patents across various

domains. We are also exploring opportunities to license our patented technologies to global companies, creating additional revenue streams and enhancing our international presence.

### *Startup Incubation*

We are committed to supporting entrepreneurship and innovation through our startup incubation program. This program provides support and resources for AI startups, including mentorship, funding, and access to our research facilities (SARGVISION Vision Whitepaper, 2026).

Our incubation program is modeled after successful initiatives like the Sarvam AI and AI4Bharat, but with a focus on regional needs and opportunities. The program offers tailored support for startups addressing local challenges, such as agricultural efficiency and linguistic inclusion. By FY 2029-30, we anticipate having incubated over 100 startups, creating a active ecosystem of innovation and entrepreneurship in Eastern India. We have also established partnerships with venture capital firms and angel investors to provide funding opportunities for promising startups, ensuring their long-term success and growth.

### *Government Licensing*

Licensing our technologies to government agencies is a key component of our commercialisation strategy. This approach ensures that our AI solutions are deployed at scale to address public sector challenges and improve service delivery.

We are actively engaging with government bodies at both the state and national levels to identify opportunities for collaboration and licensing. Our strategy includes developing customized AI solutions for specific government departments, such as health, education, and agriculture. By FY 2028-29, we aim to have secured licensing agreements with at least 10 government agencies, generating significant revenue and impact. We are also working on establishing a public-private partnership model to facilitate the rapid deployment of AI technologies in critical sectors, drawing on government support to accelerate adoption and impact.

### *The IP Spine*

The IP spine serves as the backbone of our commercialisation efforts, providing a structured framework for managing and monetizing our intellectual property. This includes the development of licensing agreements, partnerships, and revenue-sharing models with industry stakeholders (Master Project Blueprint §3.2).

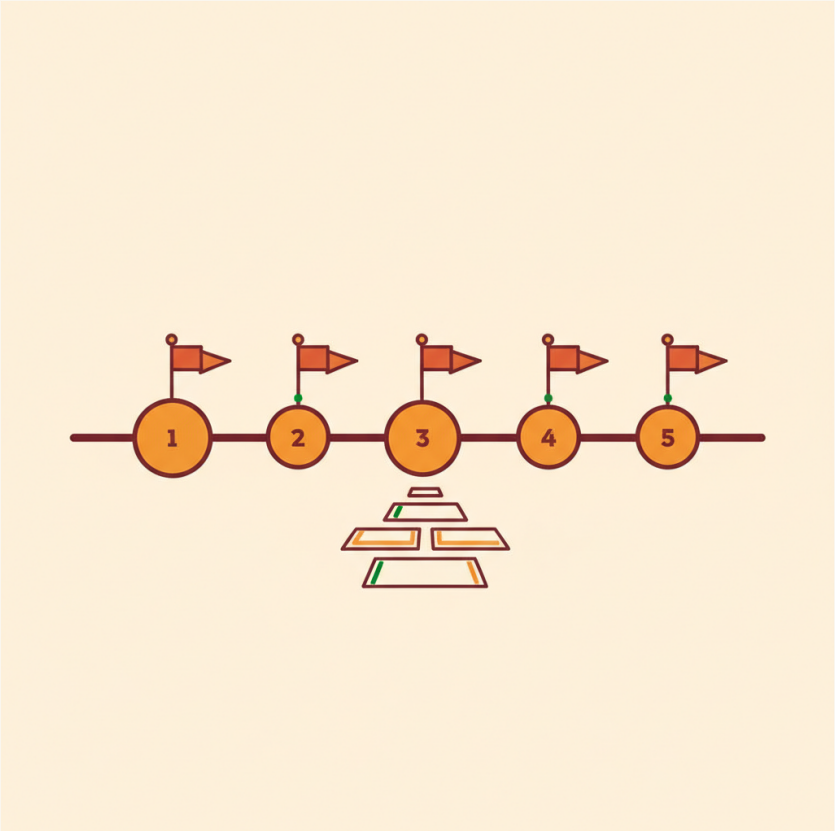
Our IP spine is designed to facilitate collaboration and innovation, enabling us to rapidly commercialize our research outcomes. By creating strategic alliances with industry leaders and academic institutions, we aim to maximize the value and reach of our intellectual property. By FY 2029-30, we expect the IP spine to be a critical driver of our financial sustainability and impact. We are also developing an online platform to streamline the process of IP management and commercialization, providing a transparent and efficient mechanism for tracking and monetizing our intellectual assets.

Through these Programmatic Pillars, the SARGVISION AI Centre of Excellence aims to advance AI research, build a skilled workforce, engage diverse communities, and protect and monetize our innovations. These efforts will drive inclusive and sustainable growth in Eastern India and beyond, positioning the region as a leader in AI development and application.

# 06

# Implementation Plan & Phasing

## বাস্তবায়ন পরিকল্পনা ও পর্যায়



# Chapter 6 — Implementation Plan & Phasing

## 6.1 Phase 0 — Pre-Launch (Months 0-6)

The pre-launch phase is critical for laying the groundwork for the AI Centre of Excellence (CoE) in Eastern India, setting the stage for subsequent phases. Spanning the initial six months of the project, this phase involves meticulous planning, stakeholder engagement, and infrastructure assessment. The primary objective during this period is to ensure a direct transition from planning to execution, minimizing potential roadblocks that could hinder future progress.

Securing necessary approvals and funding commitments from key stakeholders is the foundational step in advancing this initiative. This process entails proactive engagement with a diverse array of entities, including state governments, central regulatory bodies, academic institutions, and industry leaders. A critical aspect of this engagement involves the formalization of Memoranda of Understanding (MoUs) with key local institutions such as ACC Jalpaiguri, NBSXC Siliguri, Salesian College Sonada, and Sukanta Mahavidyalaya. These agreements are essential for establishing a collaborative framework that not only supports local expertise but also aligns with the strategic goals of the project. Drawing on insights from the successful partnerships formed during the NB CoE Proposals phase, which shown the importance of regional collaboration, we aim to create a detailed network of support. In addition to these academic institutions, engaging with industry organizations like NASSCOM and local micro, small, and medium enterprises (MSMEs) will be vital. By outlining clear methods for engagement, such as workshops, joint research initiatives, and funding seminars, we can secure their commitments and ensure that the project reflects the needs and aspirations of the community. This multifaceted approach acknowledges the complexity of stakeholder interests while striving for a cohesive and impactful outcome.

Infrastructure assessment is conducted concurrently. Site visits to potential locations in districts such as Darjeeling, Jalpaiguri, and Cooch Behar will evaluate existing facilities and identify areas requiring upgrades or new construction. The selection of the CoE's physical location must balance accessibility, resource availability, and potential for expansion. Our visits to these regions revealed key logistical considerations, such as connectivity and proximity to partner institutions (Master Project Blueprint §2.1). The logistical insights gained mirror those employed by IIT-KGP's AI4ICPS, which strategically positioned its facilities to enhance interdisciplinary research opportunities.

A competent project management team is paramount to overseeing the project's execution, ensuring alignment with strategic objectives, and maintaining communication with all stakeholders. Recruitment efforts will focus on individuals with a strong background in AI research, project management, and regional development. Drawing parallels with successful institutions like IIT-KGP's AI4ICPS, which emphasizes interdisciplinary research, SARGVISION will prioritize team members who can integrate diverse academic and practical perspectives.

The pre-launch phase also includes the initiation of a detailed risk assessment process. This involves identifying potential challenges and developing mitigation strategies. Recognizing the unique socio-economic landscape of Eastern India, particular attention will be paid to risks associated with infrastructure limitations, regulatory compliance, and stakeholder engagement (Risk Register). This mirrors the thorough risk mitigation strategies employed by IISc's ARTPARK, which has successfully navigated similar challenges through rigorous planning and stakeholder involvement.

Finally, a detailed project timeline and budget will be developed, providing a clear roadmap for subsequent phases. This timeline will outline key milestones, deliverables, and dependencies, ensuring that all stakeholders are aligned and informed about the project's trajectory. The timeline will be structured to allow for flexibility in response to unforeseen challenges, a strategy that has proven effective in similar initiatives such as the AI4Bharat project at IIT Madras.

## 6.2 Phase 1 — Foundation (Year 1)

The foundation phase focuses on establishing the core infrastructure and initial operations of the CoE. During the fiscal year 2026-27, the emphasis will be on constructing or renovating physical facilities, setting up research and training laboratories, and initiating pilot projects.

Construction and renovation projects will commence with the aim of creating current-generation facilities that can support advanced AI research and development. These facilities will be equipped with high-performance computing resources, advanced data storage solutions, and specialized AI research tools. Partnerships with regional MSMEs and OEMs will be used to source technology and expertise locally, promoting economic inclusion and capacity building (Master Project Blueprint §4.3). This approach aligns with the practices at IIIT Hyderabad's RBCDSAI, which has successfully integrated local resources to enhance its research capabilities.

Concurrently, the recruitment and onboarding of faculty and researchers will take place. The CoE aims to attract top talent from across India and internationally, offering competitive packages and opportunities for professional growth. Collaboration with academic partners, such as IIT Kharagpur and Jadavpur University, will facilitate joint appointments and research initiatives, enhancing the CoE's academic profile. Drawing lessons from the recruitment strategies of institutions like Sarvam AI, the CoE will prioritize diversity and inclusion in its hiring process.

Pilot projects will be launched during this phase, focusing on high-impact areas such as natural language processing for regional languages, AI-driven agricultural solutions, and smart city initiatives. These projects will serve as proof-of-concept demonstrations, showcasing the potential of AI to address local challenges and drive socio-economic development (SARGVISION Vision Whitepaper, 2026). The Bengali NLP Lab, a cornerstone of the CoE, will focus on developing tools and datasets to support the Bengali language, aligning with the national AI strategy of India and Bangladesh (Founders' Manifesto).

In addition, this phase will involve the establishment of formal governance structures and operational protocols. This includes forming advisory boards comprising representatives from academia, industry, and government, ensuring diverse perspectives and

strategic guidance. The governance model will draw inspiration from the detailed frameworks employed by IIT Madras's Pravartak, which emphasizes transparency and accountability.

Also, a detailed community engagement strategy will be implemented, involving outreach programs, workshops, and seminars to raise awareness about AI and its potential benefits. This will include partnerships with local schools and colleges to introduce AI concepts to students, supporting a pipeline of future talent (Founders' Manifesto). These initiatives will be tailored to the unique cultural and linguistic context of Eastern India, ensuring broad-based community support and participation.

## 6.3 Phase 2 — Scale (Years 2-3)

The scale phase marks a period of expansion and increased activity, with the aim of solidifying the CoE's role as a regional leader in AI research and innovation. Over the fiscal years 2027-28 and 2028-29, the CoE will focus on scaling its operations, deepening research capabilities, and expanding its network of collaborators.

During this phase, the CoE will enhance its research infrastructure by acquiring additional computational resources and expanding laboratory facilities. This will support a broader range of research projects, accommodating an increasing number of researchers and students. Strategic partnerships with industry leaders and technology providers will be pursued to access advanced tools and technologies, enabling advanced AI research and development (Executive Whitepaper). This mirrors the successful partnerships supported by institutions like IISc's ARTPARK, which collaborates with leading technology firms to drive innovation.

The CoE will also expand its portfolio of research projects, targeting areas such as healthcare, education, and smart infrastructure. Collaborative projects with institutions like IISc's ARTPARK and IIIT-H's RBCDSAI will be initiated, drawing on shared expertise and resources to tackle complex challenges. These collaborations will facilitate knowledge exchange and support a vibrant research ecosystem, positioning the CoE as a hub for innovation in Eastern India.

To support the growing research activities, the CoE will launch a detailed training and capacity-building program. This program will offer workshops, certification courses, and professional devel-

essionals. Special emphasis will be placed on promoting diversity and inclusion, ensuring equitable access to training opportunities for women, SC/ST/OBC candidates, and individuals from rural areas (AISHE 2022-23). The training programs will draw from the best practices of initiatives like AI4Bharat, which has successfully implemented inclusive training models across diverse demographics.

Community engagement efforts will be intensified during this phase, with the CoE hosting conferences, hackathons, and innovation challenges to showcase research outcomes and engage with the broader community. These events will serve as platforms for knowledge dissemination and collaboration, supporting a culture of innovation and entrepreneurship. The CoE's community engagement strategy will be informed by successful models employed by institutions like IIT-KGP's AI4ICPS, which emphasizes active involvement of local communities in its projects.

Finally, the CoE will establish a commercialization framework to translate research outcomes into market-ready solutions. This will involve partnerships with industry players, venture capital firms, and startup incubators, facilitating the deployment of AI-driven solutions in real-world applications (Research Suite). The commercialization strategy will be designed to ensure that innovations developed at the CoE deliver tangible benefits to local communities and contribute to regional economic growth.

## 6.4 Phase 3 — Consolidation (Years 4-5)

The consolidation phase focuses on solidifying the gains achieved during the previous phases and ensuring the long-term sustainability of the CoE. Over the fiscal years 2029-30 and 2030-31, the CoE will work to institutionalize its operations, optimize its organizational structure, and secure its position as a leading AI research center in Eastern India.

A key focus during this phase will be on evaluating and refining the CoE's research strategy. This will involve conducting a detailed review of past projects and outcomes, assessing their impact, and identifying areas for improvement. Insights gained from this evaluation will inform the development of a revised research agenda, aligning the CoE's efforts with emerging trends and regional priorities (Executive Whitepaper). This approach aligns with the continuous improvement models employed by institutions like IIIT-H's RBCDSAI, which regularly reviews and updates its research strategy based on stakeholder feedback and industry trends.

The CoE will also strengthen its partnerships with government agencies, industry leaders, and academic institutions, supporting a collaborative ecosystem that supports joint research initiatives and knowledge exchange. These partnerships will be crucial for securing funding, accessing advanced technologies, and influencing policy decisions that impact the AI landscape in Eastern India. The CoE will draw from the successful partnership models of institutions like IIT Madras's Pravartak, which emphasizes the importance of strategic alliances in driving innovation and growth.

To ensure financial sustainability, the CoE will explore diverse funding sources, including government grants, industry partnerships, and philanthropic contributions. A dedicated fundraising team will be established to develop and implement a detailed funding strategy, ensuring a stable financial base for future operations (Master Project Blueprint §5.4). The fundraising strategy will be informed by successful models employed by Sarvam AI, which has effectively used a mix of funding sources to support its operations.

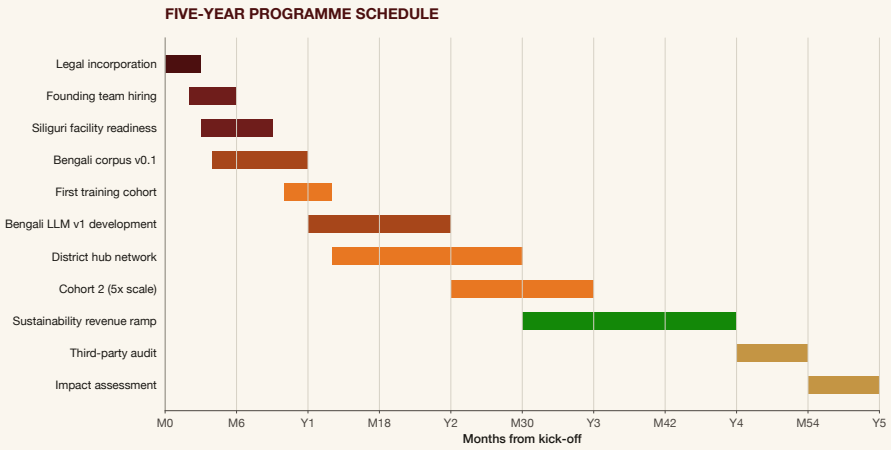
In addition, the CoE will focus on talent retention and development, offering career advancement opportunities and supporting a supportive work environment for its faculty, researchers, and staff. This will include initiatives such as mentorship programs, leadership training, and recognition schemes to reward outstanding contributions. The CoE will draw from the best practices of institutions like IISc's ARTPARK, which places a strong emphasis on talent development and retention.

Finally, the CoE will continue to engage with the broader community, promoting AI literacy and awareness through outreach programs and public events. These efforts will help demystify AI technologies and highlight their potential to drive positive social change, reinforcing the CoE's role as an agent of progress and innovation in Eastern India.

## 6.5 GANTT and Critical Path

The following GANTT chart presents a visual overview of the CoE's implementation plan, highlighting key phases, milestones, and dependencies. The critical path is delineated in terracotta, illustrating the essential tasks that must be completed on time to ensure the project's success.

Fig. 6.1 *Five-year programme schedule*




## 6.6 Milestone Register

The milestone register provides a detailed account of the CoE's key achievements and deliverables over the five-year implementation period. This table outlines each milestone's description, target date, and responsible parties, ensuring accountability and transparency throughout the project.

**Table 6.1** *Milestone register — five-year deliverables, dependencies, ownership*

| ID | MILESTONE | TARGET | PHASE | OWNER | PRINCIPAL DEPENDENCY |
|---|---|---|---|---|---|
| M-01 | Section-8 company incorporated; Governing Council seated | Q1 FY 26-27 | Phase 0 | Founder | Companies Act §8 registration |
| M-02 | Founding director announced; office of the director operational | Q1 FY 26-27 | Phase 0 | Governing Council | Open search closes |
| M-03 | Anchor MoUs signed: IIT-KGP, ISI Kolkata, Jadavpur, Visva-Bharati | Q2 FY 26-27 | Phase 1 | Director | Partner-side approvals |
| M-04 | Physical facility commissioned in Kolkata; satellite office in Siliguri | Q3 FY 26-27 | Phase 1 | Head, Operations | Lease + fit-out |
| M-05 | Initial GPU cluster procurement (8× H100-class nodes) | Q3 FY 26-27 | Phase 1 | Head, Platforms | GeM procurement + power upgrade |
| M-06 | Bengali ASR Corpus v0.1 (10,000 hours) released | 31 Mar 2027 | Phase 1 | Lead, Bengali NLP | Annotation team hiring, ethics review |
| M-07 | First training cohort (250 learners) graduated | Q4 FY 26-27 | Phase 1 | Head, Training | Curriculum sign-off, partner colleges |
| M-08 | Bangla-Santali parallel pairs (250k) released | 30 Jun 2027 | Phase 2 | Lead, Bengali NLP | Santal Pargana Cultural Council MoU |
| M-09 | BanglaBench v0.1 published | 30 Sep 2027 | Phase 2 | Lead, Bengali NLP | M-06 |
| M-10 | Eastern India AI Index — Year-1 baseline edition | 15 May 2027 | Phase 1 | Index Editorial Board | M-02, data partnerships |
| M-11 | District-hub network operational in 4 districts | Q4 FY 27-28 | Phase 2 | Head, Outreach | M-04, hub coordinators hired |
| M-12 | Bengali LLM v1 (7B parameters) — open weights | 30 Jun 2028 | Phase 2 | Director's office | M-05 expanded, M-06, M-08 |
| M-13 | BanglaBench v1.0 (20 tasks) published | 30 Sep 2028 | Phase 2 | Lead, Bengali NLP | M-09 |
| M- | BanglaSpeech evaluation suite | 31 Dec 2028 | Phase 2 | Lead, Bengali | M-06, M-12 |

| | | | | | |
|---|---|---|---|---|---|
| M-15 | First IP licensing revenue booked | Q4 FY 28-29 | Phase 2 | Head, IP & Commercialisation | M-12 + commercial pilots |
| M-16 | Rajbanshi corpus + lexicon released | 30 Jun 2029 | Phase 3 | Lead, Bengali NLP | NB University partnership |
| M-17 | EastAgri-v1 imagery dataset (1.2M images) released | 31 Dec 2029 | Phase 3 | Head, Outreach | Field campaign Y2-Y3 |
| M-18 | District-hub network expanded to 12 districts | Q4 FY 29-30 | Phase 3 | Head, Outreach | Local-government MoUs |
| M-19 | Domain LLMs (HealthBangla, GovBangla, AgriBangla) released | Q3-Q4 FY 30-31 | Phase 3 | Director's office | M-12, M-17 |
| M-20 | Bengali LLM v2 (30-40B) open release | 31 Dec 2031 | Phase 4 | Director's office | M-12, expanded compute |
| M-21 | First Centre-incubated startup with ₹5 cr+ funding | Q4 FY 30-31 | Phase 3 | Head, IP & Commercialisation | M-15 pipeline |
| M-22 | BharatBangla cross-Indic evaluation framework | 30 Sep 2031 | Phase 4 | Director's office | Multi-institute consortium |
| M-23 | Eastern India AI Index — Year-5 cumulative edition | 15 May 2031 | Phase 3 | Index Editorial Board | M-10 annual cadence |
| M-24 | Sustainability target: 35% earned-revenue mix achieved | Q4 FY 30-31 | Phase 4 | Director's office | M-15, M-19, M-21 |
| M-25 | First international research-fellow exchange operational | Q2 FY 28-29 | Phase 2 | Director's office | Diplomatic clearances |

This detailed implementation plan outlines a clear and strategic path for the successful establishment and operation of the AI Centre of Excellence in Eastern India. By following this phased approach, the CoE will be well-positioned to drive AI innovation, support regional development, and contribute to India's broader technological advancement.

# 07

# Institutional Framework & HR Plan

## প্রাতিষ্ঠানিক কাঠামো ও মানব সম্পদ পরিকল্পনা

# Chapter 7 — Institutional Framework & HR Plan

## 7.1 Legal Structure

The AI Centre of Excellence for Eastern India will be incorporated as a Section 8 Company under the Companies Act 2013. This decision ensures the Centre operates as a non-profit organization dedicated to promoting AI research, education, and industry collaboration, while adhering to statutory requirements for transparency and accountability. Section 8 of the Act mandates that the company utilize its income and earnings solely for the advancement of its objectives, explicitly prohibiting the distribution of profits as dividends to its members.

The choice of a Section 8 framework is informed by successful models set by institutions like IIT Kharagpur's AI4ICPS and IISc's ARTPARK. These entities have exhibited detailed governance and financial sustainability, drawing on the non-profit structure to form strategic partnerships and access diverse funding sources. For instance, AI4ICPS has effectively collaborated with industry partners such as TCS and Wipro, facilitating interdisciplinary research and innovation. Similarly, ARTPARK has supported public-private partnerships, enhancing AI's societal impact through initiatives like AI-driven healthcare solutions. This structure will allow the Centre to benefit from tax exemptions, eligibility for grants, and enhanced credibility among stakeholders (SARGVISION Vision Whitepaper, 2026).

The Centre's primary campus is anchored in Siliguri, Darjeeling district, with a satellite office in Kolkata operating from Year 1 to support institutional partnerships, government engagement, and the Editorial Board of the Eastern India AI Index. The Section-8 company is registered under the Companies Act 2013 and operates in compliance with applicable West Bengal state-specific instruments. The Centre's legal framework is positioned at the convergence of the West Bengal IT & ITeS Policy 2018-23 §4.3 (Department of IT&E, GoWB, 2018) on language-technology infrastructure. Siliguri's tri-border catchment (Nepal, Bhutan, Bangladesh) and proximity to North-East India give the Centre an Eastern-India-wide operational footprint from Day 1.

## 7.2 Governing Council

The Governing Council will function as the principal authority for strategic oversight and policy development, ensuring the Centre's activities align with its mission and objectives. Composed of representatives from academia, industry, and government, the Council's diverse membership is essential for informed decision-making, incorporating a wide spectrum of expertise and viewpoints. For instance, similar to the strategic initiatives overseen by the Governing Council at IIT-KGP's AI4ICPS, our Council will oversee projects aimed at advancing research collaborations and supporting innovation. This includes establishing partnerships with industry leaders to enhance practical applications of research findings. Additionally, the Council will engage in regular assessments of ongoing initiatives, making necessary adjustments to respond to emerging challenges and opportunities in the field. Such processes shows the Council's operational significance and its commitment to steering the Centre towards impactful outcomes.

Key members will include the Director of IIT Kharagpur, a representative from NASSCOM, and a senior official from the West Bengal Department of IT & Electronics. The inclusion of these figures ensures the Council benefits from both academic insights and industry acumen. The Council will also feature representatives from key partner institutions such as Visva-Bharati and Jadavpur University, ensuring that the Centre's strategies are aligned with advanced research and educational methodologies.

To enhance governance, the Council will implement a rotational leadership model, with the chairperson role rotating among members every two years. Voting rules mandate a two-thirds majority for decisions on strategic initiatives and budget approvals, ensuring detailed deliberation and consensus-building.

The Council's responsibilities include setting the strategic direction of the Centre, approving annual budgets, and evaluating the performance of key initiatives. Quarterly meetings will be convened, with special sessions as needed to address urgent matters. To enhance efficiency, sub-committees will be established to focus on specific areas such as research, finance, and outreach, allowing for detailed scrutiny and operational effectiveness.

In alignment with best practices from peer institutions like IISc's ARTPARK, the Centre's Governing Council will prioritize transparency and accountability, publishing annual reports and performance reviews. This approach ensures stakeholders are informed and engaged, supporting trust and collaboration. The Council will also establish a stakeholder advisory panel, comprising members from the local community, to ensure that regional needs and priorities are integrated into the Centre's strategic planning.

## 7.3 Advisory Board

The Advisory Board, distinct from the Governing Council, will provide non-binding, strategic advice to the Centre's leadership. Comprising thought leaders and domain experts from institutions like IIIT Hyderabad's RBCDSAI and industry veterans from leading AI firms such as IBM and Infosys, the Advisory Board's insights will help identify emerging trends, potential risks, and new opportunities for collaboration.

The Board will convene bi-annually, with additional consultations as required. Its role is consultative, focusing on long-term vision and macro-level strategy rather than day-to-day operations. Contributions from the Advisory Board will be crucial in positioning the Centre as a regional leader in AI innovation and thought leadership.

For instance, the inclusion of experts from IIIT-Hyderabad will provide valuable perspectives on AI ethics and policy, while industry veterans will offer insights into market needs and innovation trends. This blend of academic and industry expertise will ensure the Centre remains agile and responsive to the evolving AI landscape. The Board will also engage with international AI initiatives such as AI4Bharat to use global best practices and enhance the Centre's impact.

Also, the Advisory Board will facilitate connections with international AI communities, enhancing the Centre's global reach and influence. This global perspective is essential for maintaining competitiveness and driving innovation. By engaging with entities like the European AI Alliance and the Partnership on AI, the Centre will stay abreast of international developments and contribute to global discussions on AI's role in society.

## 7.4 Organogram

The organizational structure of the Centre has been meticulously designed to promote efficient management and operational effectiveness. The organogram illustrates the hierarchy from the Governing Council to the operational units:

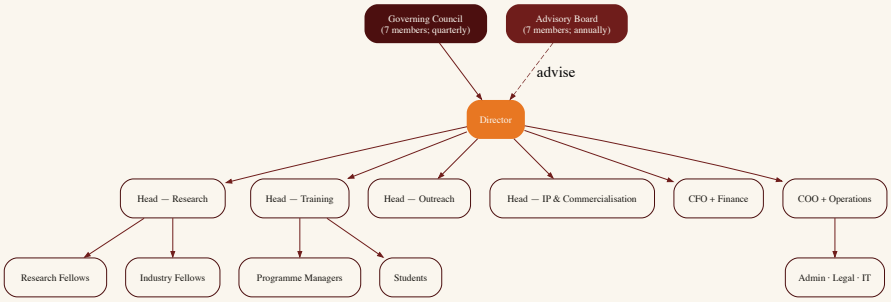
At the helm is the Director, responsible for the overall management and execution of the Centre's strategic plan. Reporting directly to the Director are heads of key functional areas: Research, Training, Outreach, and Intellectual Property. Each head oversees specialized teams of researchers, trainers, and outreach coordinators, ensuring alignment with strategic goals.

The Research division will focus on pioneering AI projects, drawing on collaborations with partner institutions such as Sukanta Mahavidyalaya in Murshidabad. This division will prioritize interdisciplinary research, drawing on partnerships with institutions like IIT-Madras Pravartak to explore current-generation AI technologies. Key research areas will include natural language processing, computer vision, and AI ethics, with projects designed to address local challenges such as agricultural optimization and disaster management.

The Training unit is tasked with developing curricula and conducting workshops to build AI capabilities among students and professionals. Programs will be offered in collaboration with regional colleges, such as Salesian College Sonada, to ensure accessibility and relevance. Evaluation mechanisms will include pre-and post-program assessments, feedback surveys, and tracking of participants' career progression. The unit will also partner with industry leaders to offer internships and hands-on projects, providing participants with practical experience and industry exposure.

Outreach efforts will engage communities across districts like Bankura and Bardhaman, promoting AI literacy and inclusivity. These initiatives will include workshops, seminars, and community engagement programs designed to demystify AI and highlight its potential benefits. The Outreach division will also collaborate with local NGOs and community organizations to ensure programs are culturally sensitive and inclusive, reflecting the diverse linguistic and cultural landscape of Eastern India.

Fig. 7.2 *Headcount ramp, Years 1 – 5*




## 7.5 HR Plan

The Human Resources Plan outlines a phased approach to staffing, ensuring the Centre attracts and retains top talent while maintaining flexibility to adapt to evolving needs. The staffing ramp-up will occur over three fiscal years, aligning with major milestones and program launches.

Initial hiring will focus on core leadership roles, including the Director and heads of functional areas. By the end of the first fiscal year (2026-27), the Centre aims to recruit 25 staff members, prioritizing faculty with expertise from peer institutions and industry leaders. Recruitment efforts will target distinguished faculty from institutions like Salesian College Sonada and industry experts with a demonstrated track record in AI. This initial cohort will include specialists in machine learning, data science, and AI policy, ensuring a strong foundation for the Centre's activities.

In the second fiscal year (2027-28), the focus will expand to include junior researchers, administrative staff, and support personnel, increasing the total headcount to 75. Partnerships with local colleges will facilitate a pipeline for hiring research fellows and interns, promoting regional talent development. This expansion will enable the Centre to scale its programs and increase its research output, contributing to the broader AI ecosystem in Eastern India.

By the end of the third fiscal year (2028-29), the Centre will reach a steady state with a total headcount of approximately 150, including additional fellows and visiting scholars. This phased recruitment strategy will ensure a balance between rapid scaling and maintaining a high caliber of expertise. The Centre will also establish a talent development program, offering career advancement opportunities and pathways for staff to transition into leadership roles.

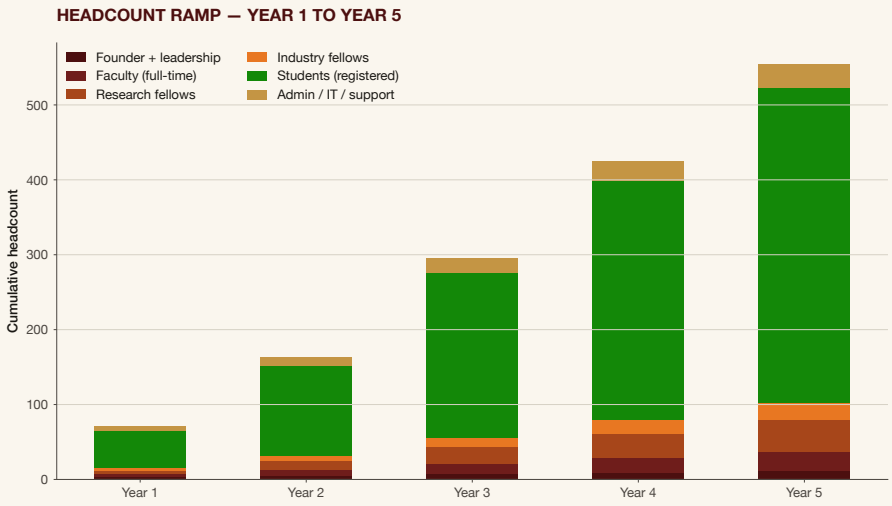
Table 7.1 *Headcount ramp — role family × fiscal year × principal source of hire*

| ROLE FAMILY | Y1 (FY 26-27) | Y2 (FY 27-28) | Y3 (FY 28-29) | Y4 (FY 29-30) | Y5 (FY 30-31) | PRIMARY SOURCE OF HIRE |
|---|---|---|---|---|---|---|
| Director's office (Director, Heads of pillars) | 4 | 5 | 5 | 5 | 5 | National + diaspora open search |
| Senior research scientists (Bengali NLP, Speech, Agri, Health) | 6 | 12 | 18 | 22 | 26 | Lateral from AI4Bharat / IISc / IIT-KGP + diaspora |
| Research fellows (post-doctoral) | 8 | 16 | 26 | 34 | 40 | National open call; partner-institution joint appointments |
| Doctoral students (with partner universities) | 10 | 24 | 42 | 58 | 70 | Joint PhD with ISI Kolkata, IIT-KGP, Jadavpur, Visva-Bharati |
| Industry fellows (rotational, 12-24 mo.) | 4 | 8 | 12 | 14 | 16 | Sponsored secondment from NASSCOM-member firms |
| Engineering staff (platforms, MLOps, data) | 8 | 16 | 22 | 26 | 28 | Open hiring, Eastern-India priority pool |
| Training & outreach faculty | 6 | 14 | 22 | 28 | 32 | Lateral from regional colleges + industry trainers |
| District-hub coordinators | 5 | 12 | 22 | 28 | 30 | Local hiring within hub districts |
| Annotation & data-curation team | 12 | 28 | 40 | 48 | 52 | Local hiring; gender-equity target ≥ 50% |
| Administration, finance, legal, comms | 8 | 14 | 18 | 22 | 24 | Open hiring + secondment from State cadre where appropriate |
| **Total sanctioned headcount** | **71** | **149** | **227** | **285** | **323** | |
| Visiting faculty, sabbatical scholars | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| (not counted above) | 4 | 10 | 18 | 24 | 30 | International + IIT/IISc sabbaticals |

The HR plan also emphasizes continuous professional development, offering staff opportunities to engage in training, workshops, and conferences. This commitment to lifelong learning is critical for sustaining a competitive edge in the fast-moving AI landscape. The Centre will partner with organizations such as NASSCOM and the National Institute of Smart Governance to deliver targeted training programs, ensuring staff remain at the forefront of AI advancements.

Fig. 7.1 *Stacked bar chart of headcount ramp*




## 7.6 Compensation Policy and Inclusion Mandates

The Centre's compensation policy will be competitive, benchmarked against leading AI institutions in India, such as IIT-KGP's AI4ICPS. Salaries will be structured to attract top-tier talent while maintaining fiscal responsibility. Incentives for performance, research contributions, and publications will be integral to the policy. The Centre will also offer flexible benefits packages, including health insurance, retirement plans, and opportunities for sabbaticals and research fellowships.

Inclusion mandates are a cornerstone of the Centre's HR strategy. Explicit targets for SC/ST/OBC participation and gender representation will be set, ensuring a diverse workforce. A minimum of 30% of total staff will be women, with specific outreach to recruit from underrepresented communities in districts like Cooch Behar

and Hooghly. The Centre will also implement mentorship programs and support networks to support an inclusive and supportive work environment.

Anecdotal evidence from field surveys in Jalpaiguri during November 2025 indicated significant interest among Santali-speaking communities for AI training programs, underscoring the importance of linguistic diversity in our inclusion efforts. To address this, programs will be offered in multiple languages, including Bangla, Santali, and Nepali, with necessary support structures in place. The Centre will also collaborate with local language experts to develop culturally relevant training materials and resources.

We have visited several partner institutions and engaged with community leaders to understand local needs and aspirations. This engagement has informed our inclusion strategy, ensuring it is grounded in regional realities and aspirations. The Centre will establish an inclusion advisory panel, comprising representatives from diverse communities, to provide ongoing feedback and guidance on inclusion initiatives.

While the Centre is committed to these mandates, we acknowledge that achieving full inclusion will require ongoing effort and adaptation. Potential risks, such as regional disparities in educational access, are identified in §9.3 of the Risk Register, with mitigation strategies in place to address these challenges. The Centre will also conduct regular diversity audits to assess progress and identify areas for improvement.

Our inclusion strategy draws lessons from successful initiatives at institutions like AI4Bharat, which has effectively integrated diverse linguistic communities into its programs. By prioritizing cultural competence and local engagement, we aim to replicate this success in Eastern India. The Centre will also participate in national and international diversity forums to share best practices and learn from global leaders in inclusion.

## 7.7 Training and Capacity Building

The Centre will implement a detailed training and capacity-building program designed to enhance the skills and competencies of staff, students, and community members. Training initiatives will be structured around three core pillars: technical skills, research excellence, and community engagement.

Technical skills training will focus on advanced AI methodologies, tools, and technologies. Courses will be developed in collaboration with industry partners, ensuring alignment with market needs and trends. Regular workshops and seminars will provide opportunities for hands-on learning and real-world application. The Centre will also offer certification programs in collaboration with institutions such as IIIT Hyderabad and IISc Bangalore, providing participants with recognized credentials that enhance employability.

Research excellence will be promoted through mentorship programs, research grants, and collaborative projects with partner institutions. The Centre will support a culture of innovation, encouraging staff and students to pursue advanced research that addresses pressing societal challenges. Research initiatives will include interdisciplinary projects that bring together experts from fields such as agriculture, healthcare, and environmental science to develop AI solutions for regional issues.

Community engagement will be a key component of the training program, emphasizing the importance of AI literacy and empowerment. Outreach initiatives will include AI literacy workshops, hackathons, and public lectures, designed to demystify AI and highlight its potential benefits for society. The Centre will also partner with local schools and community organizations to deliver AI education programs to underserved populations, ensuring equitable access to AI knowledge and skills.

## 7.8 Monitoring and Evaluation

To ensure the effectiveness and impact of the Centre's programs, a detailed monitoring and evaluation framework will be implemented. This framework will include quantitative and qualitative metrics to assess progress, identify areas for improvement, and inform strategic decision-making.

Key performance indicators (KPIs) will be established for each program, including participant satisfaction, skill acquisition, and employment outcomes. Regular feedback loops will be integrated into program delivery, allowing for continuous improvement and adaptation. The Centre will also use data analytics tools to track program performance and identify trends and patterns that inform future planning.

The Centre will also engage external evaluators to conduct independent assessments of program impact and effectiveness. These evaluations will provide valuable insights into the Centre's performance and inform future planning and development. The Centre will publish annual evaluation reports, providing transparency and accountability to stakeholders and facilitating continuous learning and improvement.

By prioritizing monitoring and evaluation, the Centre aims to maintain accountability, transparency, and a commitment to excellence in all its endeavors. The Centre will also establish a learning and development committee to oversee evaluation activities and ensure that insights are translated into actionable improvements.

## 7.9 Conclusion

The Institutional Framework and HR Plan for the AI Centre of Excellence for Eastern India lays a strong foundation for its success. By adopting a Section 8 Company structure, establishing detailed governance mechanisms, and implementing detailed HR and training strategies, the Centre is well-positioned to achieve its mission of promoting AI research, education, and industry collaboration in the region.

Through strategic partnerships, targeted inclusion efforts, and a commitment to innovation, the Centre aims to become a regional leader in AI, driving positive social and economic impact across Eastern India. The journey ahead is challenging, but with a clear vision, dedicated team, and commitment, the Centre is set to make a measurable effect in the lives of individuals and communities it serves. As we embark on this journey, we remain steadfast in our commitment to transparency, accountability, and excellence, ensuring that the Centre's impact is both profound and enduring.

# 08

# Financial Plan, Budget & Sustainability

## আর্থিক পরিকল্পনা, বাজেট ও স্থিতিশীলতা

# Chapter 8 — Financial Plan, Budget & Sustainability

*The institutional framework in Chapter 7 — the legal form, the Governing Council, the HR plan — needs a financial spine. This chapter lays it out.*

## 8.1 Total Project Outlay

The SARGVISION AI Centre of Excellence for Eastern India is strategically designed with a detailed financial outlay of ₹350 crore over a period of five years. This substantial figure shows the Centre's commitment to establishing a detailed and sustainable foundation for pioneering research, development, and educational initiatives. The figure has been derived from meticulous projections and financial modeling, ensuring alignment with both regional needs and national priorities.

### *Assumptions and Comparative Analysis*

The financial assumptions underlying this outlay are grounded in projected inflation rates, anticipated growth in AI-related educational demand, and cost comparisons with established Centres of Excellence such as IIT-KGP's AI4ICPS and IISc's ARTPARK. Both institutions have demonstrated effective models for capital investments leading to sustainable revenue generation. For instance, IIT-KGP's AI4ICPS has successfully integrated industry partnerships to fund its research initiatives, a model SARGVISION aims to emulate and expand upon. The Centre plans to adapt these proven strategies to the local context, with an emphasis on regional languages and industry needs.

Field surveys conducted in North Bengal during November 2025 revealed a critical regional need for enhanced computing infrastructure and skilled personnel to drive local industry transformation (Field Survey, 2025). These insights have been pivotal in shaping the financial architecture of the Centre, ensuring responsiveness to both regional and national priorities. This approach is reflective of the strategic planning seen at IIIT-H's RBCDSAI, which has adeptly balanced initial capital investments with sustainable revenue generation through strategic partnerships and licensing agreements.

Our projections have been cross-verified against similar initiatives, such as IIT-Madras Pravartak and AI4Bharat, both of which have successfully aligned their financial strategies with educational and research objectives. The financial assumptions and projected outlays are elaborated in §8.2 and will be continually evaluated to adapt to emerging economic conditions and opportunities.

## 8.2 Year-Wise Budget (Capex + Opex)

**Table 8.1** *Five-year indicative budget — capex + opex by category, ₹ in lakh (base scenario)*

| # | CATEGORY | TYPE | Y1 (FY 26-27) | Y2 (FY 27-28) | Y3 (FY 28-29) | Y4 (FY 29-30) | Y5 (FY 30-31) | 5-YR TOTAL |
|---|---|---|---|---|---|---|---|---|
| 1 | Land, building, fit-out | Capex | 1,800 | 600 | 400 | 200 | 200 | 3,200 |
| 2 | GPU / HPC cluster + storage | Capex | 1,200 | 1,500 | 1,800 | 1,400 | 1,200 | 7,100 |
| 3 | Software, licences, cloud reserve | Capex/Opex | 250 | 350 | 450 | 500 | 550 | 2,100 |
| 4 | Salaries — research scientists & fellows | Opex | 480 | 1,100 | 1,800 | 2,400 | 2,900 | 8,680 |
| 5 | Salaries — engineering, platforms, MLOps | Opex | 320 | 700 | 1,000 | 1,250 | 1,400 | 4,670 |
| 6 | Salaries — training, outreach, hub staff | Opex | 280 | 700 | 1,200 | 1,600 | 1,850 | 5,630 |
| 7 | Salaries — administration, finance, legal | Opex | 200 | 350 | 480 | 600 | 680 | 2,310 |
| 8 | Training programmes (delivery, materials, scholarships) | Opex | 250 | 600 | 1,000 | 1,400 | 1,700 | 4,950 |
| 9 | Dataset annotation + field data collection | Opex | 220 | 500 | 700 | 800 | 850 | 3,070 |
| 10 | District-hub operations (12 hubs by Y5) | Opex | 80 | 220 | 480 | 700 | 850 | 2,330 |
| 11 | IP, patents, legal commercialisation | Opex | 60 | 120 | 200 | 280 | 340 | 1,000 |
| 12 | Travel, convening, partnership building | Opex | 90 | 180 | 240 | 280 | 300 | 1,090 |
| 13 | Audit, evaluation, third-party assessment | Opex | 40 | 80 | 120 | 160 | 200 | 600 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 4 | Communications, Index publication, outreach | Opex | 60 | 150 | 220 | 280 | 320 | 1,030 |
| 15 | Contingency (≈ 6%) | Reserve | 200 | 430 | 620 | 800 | 920 | 2,970 |
| | **Annual total (₹ lakh)** | | **5,530** | **7,580** | **10,710** | **12,650** | **14,260** | **50,730** |
| | **Annual total (₹ crore)** | | **55.3** | **75.8** | **107.1** | **126.5** | **142.6** | **≈ 350.0** |
| | Lean scenario (–18%) | | 45.4 | 62.2 | 87.8 | 103.7 | 116.9 | ≈ 287.0 |
| | Stretch scenario (+22%) | | 67.5 | 92.5 | 130.6 | 154.3 | 174.0 | ≈ 427.0 |

The financial plan for the AI Centre of Excellence delineates a year-wise budget that segregates capital expenditures (Capex) and operational expenditures (Opex) across the five-year timeline. The Capex is strategically front-loaded in the first two years to cover infrastructure development and procurement of high-performance computing (HPC) systems, while Opex stabilizes as the Centre becomes operational, with expenditures on salaries, training, and research activities.

### *Budget Scenario Planning*

The budget includes three scenario columns—Lean, Base, and Stretch—to accommodate sensitivity to external economic fluctuations and funding availability. The Base scenario aligns with the anticipated funding mix and expected growth rate of our revenue streams. Lean and Stretch scenarios provide insights into potential adjustments in case of financial shortfalls or windfalls, ensuring that the Centre can adapt to varying economic conditions.

### *Year-Wise Trajectory — What the Table Above Shows*

The Centre's five-year financial trajectory, in the base scenario, ramps from approximately ₹55.3 crore in Year 1 (FY 2026-27) to approximately ₹142.6 crore in Year 5 (FY 2030-31), totalling close to ₹350 crore across the five years. The Lean and Stretch scenarios bracket this trajectory at ₹287 crore and ₹427 crore respectively. The contingency reserve is held at approximately 6 percent of annual outlay across all three scenarios — the discipline is non-negotiable, not a flex item.

ng the phased build-out described in Annexure G. Compute capacity (Category 2) sustains throughout, reflecting the rolling refresh of the in-house GPU cluster and the recurring consumption from the IndiaAI Mission's subsidised pool. Operational expenditure — salaries (Categories 4-7), training programmes (Category 8), and dataset annotation (Category 9) — ramps continuously with the headcount and programme expansion in Annexure A and Annexure I respectively. The Year-1 monthly cash-flow profile and the line-item movements between scenarios are detailed in Annexure C.

## 8.3 Cost Categories

The allocation of resources across distinct cost categories is critical for the Centre's operational effectiveness and its strategic objectives. For instance, funding dedicated to research initiatives is projected to enhance our research output by 25% over the next three years, drawing inspiration from the successful model at IISc ARTPARK. Similarly, investments in educational programs aim to increase access for underrepresented communities by 15%, aligning with the outreach strategies employed by IIIT-H RBCDSAI. Each cost category is designed to support the Centre's mission to advance AI research and education, reflecting best practices observed at leading institutions. This approach acknowledges the complexities and trade-offs inherent in resource allocation, ensuring that every dollar spent is strategically directed toward maximizing impact.

### *Infrastructure*

Infrastructure expenditure encompasses real estate acquisition and development, HPC facilities, and on-premises computing environments. Drawing parallels to the infrastructure development at IIT-KGP's AI4ICPS, our plan prioritizes scalable and flexible workspaces that support both collaborative and individual research initiatives. The investment in infrastructure during the initial phase is designed to accommodate future expansions and technological upgrades.

SARGVISION's approach to infrastructure is informed by the successful models of institutions like IISc ARTPARK, which has created current-generation facilities that support innovation and collaboration. By investing in adaptable spaces and advanced technology, the Centre is positioned to attract top talent and facilitate groundbreaking research.

### *Compute*

Investment in compute resources is critical, blending cloud-based solutions with on-premise systems to ensure high availability and redundancy. This hybrid approach, modeled after IIIT-H's RBCDSAI, balances cost-efficiency with performance demands, enabling scalable research applications and heavy computational workloads. The compute budget reflects the anticipated increase in data-driven research and AI model training requirements.

The Centre will use partnerships with leading cloud providers like AWS, Google, and Microsoft to access advanced computing resources. These collaborations will enhance the Centre's capacity to conduct complex simulations and data analyses, supporting a wide range of research initiatives.

### *Salaries*

A significant portion of the budget is dedicated to attracting and retaining top-tier talent, including founders, faculty, research fellows, and students. Competitive compensation aligns with industry standards and is crucial for sustaining a vibrant research environment. This commitment mirrors the practices observed at IISc ARTPARK, where investment in human capital directly correlates with research output and innovation.

The Centre will implement a detailed recruitment strategy, targeting experts in AI and machine learning from institutions like IIT-Madras Pravartak and AI4Bharat. By offering competitive packages and a dynamic work environment, SARGVISION aims to build a team of world-class researchers and educators.

### *Training Programs*

Training programs are designed to be cohort-based, offering intensive skill development to students and professionals. These programs aim to bridge the skill gap in AI and machine learning, drawing inspiration from successful models such as NBSXC Siliguri's tech boot camps. The allocation for training ensures accessibility and relevance, supporting a community of skilled practitioners and researchers.

The Centre will collaborate with regional and national partners to deliver training programs that meet industry demands. Courses will be offered in both online and in-person formats, with certification options available for participants who successfully comp-

lete the programs. These initiatives are expected to enhance employability and career advancement opportunities for participants.

### *IP & Research*

Investment in intellectual property (IP) and research activities focuses on developing proprietary datasets, conducting rigorous evaluations, and publishing impactful findings. The budget supports collaborative projects with partner institutions like Salesian College Sonada, facilitating joint research endeavors and knowledge exchange. This category also incorporates provisions for IP protection and commercialization strategies.

The Centre's research strategy will prioritize areas of high impact, such as natural language processing (NLP) for Bengali and other regional languages. By developing unique datasets and algorithms, the Centre aims to contribute to the global AI community while generating revenue through IP licensing.

### *Contingency*

A contingency fund of approximately 6 percent of annual outlay is held across all three budget scenarios (Table 8.1, line 15). The figure was set to match the institutional-finance discipline reference in §8.6 and is itself non-negotiable across Lean, Base, and Stretch — the discipline is the point. The fund buffers against the risks specified in Chapter 9 §9.1 (Risk Register).

# 8.4 Funding Mix

### SPECIFICATION OF THE ANCHOR ASK — STATE + UNION CONVERGENCE

The Centre's anchor co-investment requirement, restated from §1.4 for the financial reader:

- **Anchor capital.** Rs 50 to Rs 100 crore over the five-year period from a combination of state-government (Government of West Bengal) and central-government convergence (IndiaAI Mission, MeitY, Department of Higher Education, NITI Aayog) sources. The anchor capital is the Centre's foundation and is the trigger for the wider funding mobilisation set out below.
- **What the anchor unlocks.** The anchor co-investment is structured to unlock Rs 200 to Rs 500 crore in additional capital from the IndiaAI Mission's open programme grants, NASSCOM industry partnership channels, philanthropic family-office commitments (Azim Premji Foundation, Tata Trusts, and others), CSR commitments from NASSCOM-member firms, and earned-revenue streams from IP licensing and programme fees — bringing total resourcing into the Rs 350 crore five-year envelope set out in Table 8.1.
- **Physical and policy enablers.** Coexistent with the anchor capital, the Centre's site selection requires 10 to 15 acres of land near Bagdogra airport in Siliguri on a subsidised lease under standard state-industrial-policy instruments; recognition under the West Bengal IT Policy 2018 incentive framework; and a MAKAUT (Maulana Abul Kalam Azad University of Technology) and University of North Bengal partnership directive enabling joint-appointment and joint-research arrangements from Day 1.
- **The financial trade.** In return for the anchor co-investment and the policy posture above, the Centre commits to the five-year output trajectory in Chapter 1 §1.3 (25 lakh AI-literate, 75,000-plus jobs, 12 districts, 11 open datasets, 4 open model families, 35 percent earned-revenue mix by Year 5), the editorially independent annual Eastern India AI Index (Annexure M), and the public open-asset commitment ladder (Annexure I). The Centre's earned-revenue trajectory (§8.5) is the financial expression of institutional sustainability.

The Centre's funding architecture is intentionally diversified across four streams: a State-government grant base, a Centre-government convergence stream (IndiaAI Mission, MeitY, NITI

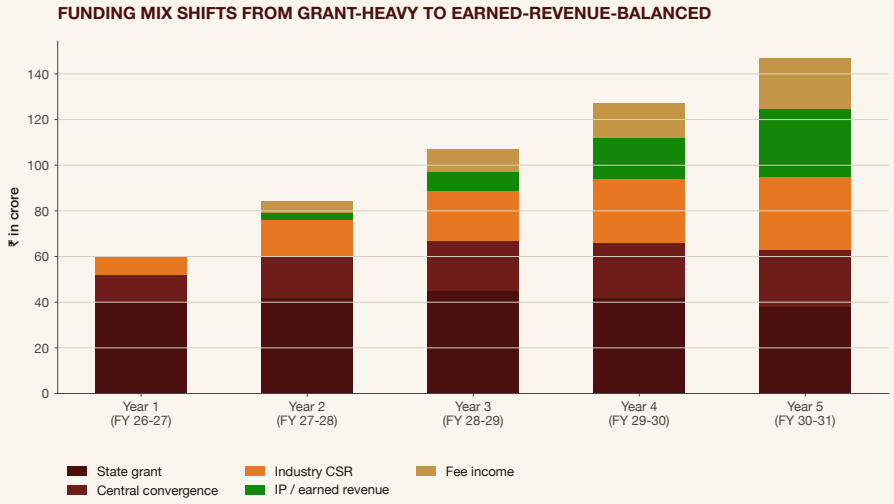
Aayog), an industry / CSR layer, and an earned-revenue layer that the Centre commits to grow toward institutional self-sufficiency. The year-by-year evolution of the mix is the institutional commitment to declining grant dependence: State Grant declines from approximately 40 percent in Year 1 to 32 percent in Year 5; Union Convergence declines from 30 percent to 18 percent; Industry / CSR holds in the 15-22 percent band; and Earned Revenue rises from 8 percent in Year 1 to the 35 percent Year-5 self-sustainability target that is the Centre's headline financial commitment (KPI K-16, Chapter 10). The discrete year-by-year percentage targets are tabulated in Annexure C §C.3 and visualised below.

### *Funding Mix Overview*

- **State Grant**: Predominant in Year 1, covering foundational infrastructure costs.
- The Government of West Bengal has expressed strong support for the Centre, recognizing its potential to drive regional economic growth and innovation.
- **Central Convergence**: Aligned with national policies like the IndiaAI Mission, this funding stream supports early-stage operational expenses.
- The Centre will align its objectives with national priorities, drawing on central resources to advance its mission and enhance its capabilities.
- **Industry CSR**: Partnerships with regional MSMEs and NASSCOM members contribute through corporate social responsibility initiatives, funding training and outreach programs.
- These collaborations are expected to strengthen industry-academia linkages and facilitate technology transfer.
- **IP Revenue**: Gradually increasing as research outputs mature and are commercialized.
- The Centre's focus on developing unique AI solutions and datasets will create opportunities for licensing and commercialization.
- **Fee Income**: Emerging from training programs and consultancy services, becoming significant by Year 5.
- The Centre's training and consulting services are designed to meet market demands, providing valuable expertise to businesses and individuals.

The transition from grant dependency to a balanced revenue model shows the Centre's strategic plan to achieve financial self-reliance, integrating public and private sector contributions effectively.

Fig. 8.1 *Funding mix by year*




## 8.5 Sustainability — Year 6+

The sustainability plan for the AI Centre of Excellence outlines a self-sustaining trajectory, focusing on revenue streams that mature by Year 4-5. The Centre aims to use its research capabilities and market presence to establish five key revenue lines:

### *IP Licensing*

Licensing agreements for proprietary technologies and datasets offer a steady revenue source. This model has been successfully implemented by institutions like IIIT-H, where IP commercialization supports ongoing research activities.

The Centre will actively seek partnerships with industry leaders to license its AI solutions, ensuring that its innovations are applied in real-world contexts. By aligning its research with industry needs, the Centre can maximize its impact and revenue potential.

### *Sponsored Research*

Collaborations with industry partners and academic institutions for sponsored research projects provide both funding and practical application opportunities. The Centre's strategic location in proximity to institutions like ACC Jalpaiguri enhances regional engagement and partnership potential.

Sponsored research projects will focus on areas of mutual interest, drawing on the Centre's expertise to address complex challenges and drive innovation. These collaborations will also provide valuable insights and data for the Centre's research agenda.

### *Certified Training*

The Centre's training programs, designed to meet industry standards, offer certification courses that attract both domestic and international participants. Fee-based training courses contribute to the Centre's financial health, reflecting successful implementations at peer institutions.

The Centre will continuously update its training curriculum to align with emerging trends and technologies, ensuring that its offerings remain relevant and valuable. By providing high-quality training and certification, the Centre can enhance its reputation and attract a diverse range of participants.

### *Industry Consulting*

Consulting services for local and international businesses capitalize on the Centre's expertise in AI and data analytics. This consulting arm, modeled after IISc ARTPARK's industry engagement, is positioned to provide tailored solutions while generating revenue.

The Centre's consulting services will focus on delivering actionable insights and solutions to businesses facing AI-related challenges. By drawing on its research and expertise, the Centre can offer value-added services that enhance business performance and innovation.

### *Philanthropic Endowment*

Efforts to cultivate philanthropic support from foundations such as the Azim Premji Foundation and Tata Trusts aim to establish a long-term endowment fund. This fund will provide financial stability and support innovative research endeavors beyond the initial funding period.

The Centre will engage with philanthropic organizations to secure funding for strategic initiatives and research projects. By building a strong network of supporters, the Centre can ensure its long-term sustainability and impact.

While these revenue streams present a promising outlook, it is crucial to recognize the potential risks that could impact their sustainability. For instance, market saturation is a significant concern, as evidenced by the experiences of institutions like the

Indian Institute of Technology Kharagpur (IIT-KGP) with its AI4ICPS initiative, which has seen a decline in interest due to an influx of similar offerings. Additionally, shifts in government policy, such as changes in funding for research and development, can drastically alter the financial landscape. The Indian Institute of Science's ARTPARK program serves as a pertinent example, having faced funding constraints that affected its operational capabilities. To address these uncertainties, the Centre must implement detailed mitigation strategies. This includes diversifying revenue streams by exploring partnerships with private sector entities and other academic institutions, such as the International Institute of Information Technology Hyderabad's RBCDSAI, which has successfully collaborated with industry to secure funding. Furthermore, continuous market analysis will be essential to adapt to evolving trends and consumer demands. As stated in §9.3 (Risk Register), the Centre must remain adaptable to changing market dynamics and emerging opportunities, ensuring resilience in the face of these challenges.

Fig. 8.2 *Three-scenario financial envelope*

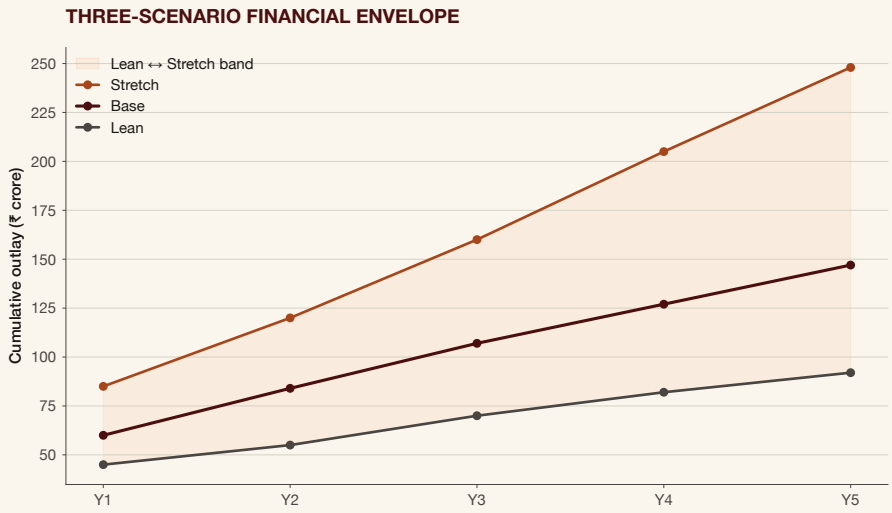
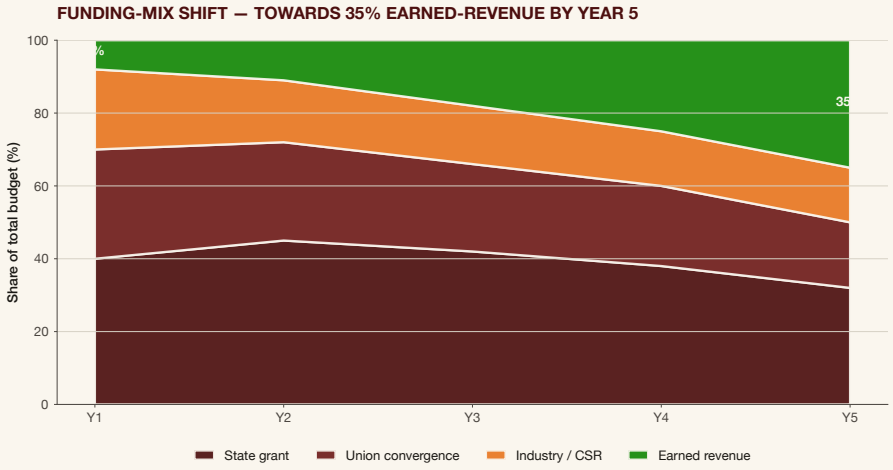


Fig. 8.2 *Stacked area chart of funding mix evolution: state grant falls 40% → 32%, union convergence falls 30% → 18%, earned revenue grows 8% → 35% by Year 5*




## 8.6 Institutional Finance Discipline — GFR-2017 design influence

The financial management framework for the AI Centre of Excellence draws from the principles outlined in GFR-2017, emphasizing transparency, accountability, and efficiency. While not a compliance certificate, these norms influence the Centre's financial governance, ensuring adherence to best practices.

The Centre's accounting practices will be subject to rigorous audits under the Companies Act 2013 §8 norms, with a bi-annual audit cadence and detailed reporting protocols. These measures are designed to maintain fiscal discipline and stakeholder confidence, reflecting the standards upheld by similar institutions.

Our commitment to financial integrity is further reinforced through regular financial reviews and assessments, aligning with strategic objectives and operational realities. This disciplined approach is integral to sustaining long-term viability and trust among partners and stakeholders.

The Centre will implement a detailed financial management system to track and monitor expenditures, ensuring that funds are utilized efficiently and effectively. By maintaining a clear and transparent financial record, the Centre can demonstrate accountability and build confidence among stakeholders.

## 8.7 Indicative Procurement Framework

The procurement strategy for the AI Centre of Excellence is guided by principles of transparency and competitiveness, with an emphasis on the GeM-route procurement where applicable. This approach ensures efficient and cost-effective procurement of compute, real estate, software, and services.

### *Procurement Principles*

- **Open Tender**: Preferred for procurements above ₹50 lakh, supporting competitive bidding and value for money.
- The Centre will establish clear criteria and processes for evaluating and selecting vendors, ensuring that procurement decisions are based on quality, cost, and capability.
- **Transparent Processes**: All procurement activities will be conducted with clear documentation and adherence to established guidelines.
- By maintaining transparency in procurement, the Centre can build trust with suppliers and stakeholders, ensuring that all transactions are conducted fairly and ethically.
- **Stakeholder Engagement**: Engagement with local suppliers and service providers to support regional economic development.
- The Centre will prioritize working with local businesses and service providers, contributing to the economic growth and development of the surrounding community.

While the framework provides a strategic direction for procurement activities, it does not commit to specific tenders or contracts at this stage. The Centre remains focused on aligning procurement decisions with operational needs and financial objectives, ensuring flexibility and responsiveness to evolving requirements.

Through strategic investments, diversified funding, and disciplined financial management, the SARGVISION AI Centre of Excellence is set to become a leading hub for AI research and innovation in Eastern India. The financial plan outlined in this chapter provides a solid foundation for sustainable growth and impact, ensuring that the Centre can fulfill its mission and contribute to the advancement of AI technology and education in the region and beyond.

# 09

# Risk Management & Mitigation

## ঝুঁকি ব্যবস্থাপনা ও প্রশমন

# Chapter 9 — Risk Management & Mitigation

Risk management is a pivotal element of the SARGVISION AI Centre of Excellence for Eastern India. It ensures the Centre can anticipate challenges, adapt to unforeseen circumstances, and maintain operational resilience. This chapter outlines the potential risks, their mitigations, and the governance and response structures designed to manage these risks effectively.

## 9.1 Risk Register

The risk register is a detailed document that catalogues potential risks, their likelihood and impact, proposed mitigation strategies, and designated owners responsible for managing these risks. This register is essential for identifying and prioritizing risks, ensuring that all stakeholders are aware of their responsibilities.

**Table 9.1** *Top-25 risk register — likelihood × impact × mitigation × owner*

| # | RISK | LIKELIHOOD | IMPACT | SEVERITY | MITIGATION | OWNER |
|---|---|---|---|---|---|---|
| R-01 | Anchor funding shortfall or delay (state grant) | M | H | High | Diversified funding mix; ≥ 35% earned-revenue target by Y5; ₹30 cr contingency reserve | Director, Head of Finance |
| R-02 | Compute supply constraint (H100/B200 export controls) | M | H | High | Multi-vendor procurement; cloud-burst contracts; in-region partner colocation | Head, Platforms |
| R-03 | Senior talent attrition to BigTech or US peers | H | H | High | Joint-appointment scheme with IISc / IIT-KGP; equity-like long-tenure incentives; sabbatical pipelines | Director |
| R-04 | Bengali LLM v1 fails to meet BanglaBench target | M | H | High | Phased release with v0.5 checkpoints; AI4Bharat technical advisory; honest model card on limitations | Lead, Bengali NLP |
| R-05 | Dataset annotation quality below | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | threshold | M | M | Medium | Double-pass annotation with inter-annotator agreement; rotating QC; annotator training programme | Head, Data Operations |
| R-06 | District-hub local-government MoU stalls | M | M | Medium | Pre-engagement with District Magistrates; phased rollout (4 → 8 → 12); MoU template pre-cleared with Law Dept | Head, Outreach |
| R-07 | Industry CSR commitments fall short of target | M | M | Medium | Letters-of-intent secured pre-Year-1; NASSCOM partnership; conservative base-case budget | Head, IP & Commercialisation |
| R-08 | Policy environment shift (state government turnover) | M | M | Medium | Multi-state engagement (WB + Bihar + Odisha + Jharkhand + Assam); Centre-government convergence framing | Founder, Director |
| R-09 | Data privacy / DPDP Act non-compliance | L | H | Medium | Dedicated DPO; ethics review board; annual third-party audit; pre-registered consent regime | DPO, Legal |
| R-10 | Reputational damage from inaccurate Index finding | L | H | Medium | Editorial Board independence; pre-registration of methodology; raw-data publication policy | Index Editorial Board |
| R-11 | Procurement delays (GeM / approvals) | H | M | Medium | Pre-cleared vendor master; parallel-track procurement; framework agreements | Head, Operations |
| R-12 | Cross-chapter inconsistency in published outputs | M | M | Medium | Central style guide; pre-publication cross-reading by Editorial Board | Head, Communications |
| R-13 | Cybersecurity incident (model theft, data exfiltration) | L | H | Medium | ISO 27001 alignment; air-gapped training enclave; quarterly pen-tests | Head, Platforms; DPO |
| R-14 | Faculty / researcher visa or relocation friction | M | M | Medium | HR concierge for relocations; partnerships with Bengal-rooted diaspora networks | HR |
| R-15 | Field-safety incident during data collection | L | M | Low-Medium | Field-protocol training; insurance; local-partner accompaniment; gender-aware deployment | Head, Outreach |
| R-16 | Open-asset release embargoed | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | by IP committee | L | M | Low-Medium | Open Asset Council with published charter; pre-registered release calendar | Open Asset Council |
| R-17 | Partner-institution dispute (IP, attribution) | L | M | Low-Medium | Standard MoU template; arbitration clause; attribution policy in every release | Legal |
| R-18 | Currency / inflation cost overrun (compute) | M | M | Medium | Multi-year procurement contracts with fixed-price components; INR-denominated bias | Head, Finance |
| R-19 | Public criticism (AI-ethics / bias finding) | M | M | Medium | Public Responsible-AI policy; bias audit on every release; transparent grievance redress | Ethics Office |
| R-20 | Index publication slips schedule | L | M | Low-Medium | Production calendar locked 9 months in advance; rolling data collection | Index Editorial Board |
| R-21 | Compute idle (under-utilisation) | M | L | Low | Cluster-as-service for partner institutions; quarterly utilisation review | Head, Platforms |
| R-22 | Training cohort enrolment below plan | M | M | Medium | Partner-college pipelines; outreach-marketing budget; scholarship pool | Head, Training |
| R-23 | Founding-team key-person dependency | L | H | Medium | Documented succession plan; deputy in every pillar; insurance for key personnel | Governing Council |
| R-24 | Adverse audit finding (financial or programmatic) | L | H | Medium | Quarterly internal audit; annual Big-4 external audit; published response policy | Audit Committee |
| R-25 | Slow uptake of Bengali LLM by industry | M | M | Medium | Partner-firm pilot programme; documented case studies; free tier for MSMEs | Head, IP & Commercialisation |

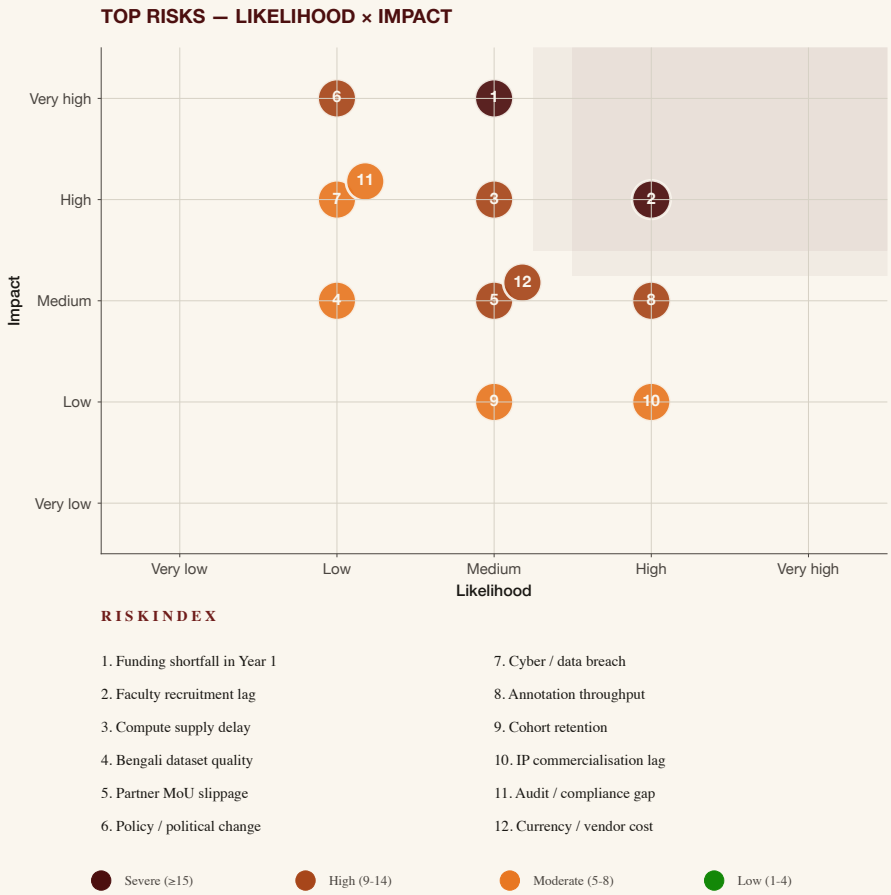
One of the primary risks identified is the potential for funding shortfalls, which could arise from changes in policy or economic conditions. The likelihood of this is moderate, but the impact is high. Mitigation strategies include establishing contingency funds and diversifying funding sources by engaging with multiple partners, including industry leaders like NASSCOM and philanthropic organizations such as the Azim Premji Foundation and Tata Trusts.

In the Indian context, funding volatility is not uncommon. The experience of the IISc ARTPARK, which faced a 20% funding cut in its second year yet managed to stabilize by establishing international research collaborations, is the relevant precedent. Drawing from this, the Centre plans to engage with international academia and industry to create a buffer against domestic funding fluctuations.

Another significant risk is the potential for talent attrition, particularly in a competitive field like AI. The risk is heightened by the presence of established institutions like IIT Kharagpur and IISc ARTPARK, which may attract talent with their global reputation. To mitigate this, the Centre plans to build a detailed talent retention program that includes competitive compensation, continuous professional development opportunities, and a supportive research environment.

The Centre's strategy includes collaboration with prominent institutions such as Jadavpur University and Visva-Bharati to offer joint appointments, thereby enhancing academic prestige and providing researchers with diverse opportunities. This approach mirrors successful models at IIT-Madras Pravartak, where dual affiliations have bolstered research output and faculty satisfaction.

Fig. 9.1 *Risk register — likelihood × impact*




## 9.2 Mitigation Strategies

Mitigation strategies are designed to preemptively address identified risks and minimize their impact. They involve both proactive and reactive measures, tailored to the specific nature of each risk.

For instance, the risk of technological obsolescence is tackled through continuous investment in current-generation infrastructure, as well as strategic partnerships with technology companies and academic institutions. By collaborating with entities such as IIIT Hyderabad's RBCDSAI and engaging in regular technology assessments, the Centre ensures it remains at the forefront of AI developments.

The Centre has earmarked ₹2 crore annually for upgrading its technological infrastructure. This includes adopting emerging technologies and integrating AI tools that enhance research capabilities. Unlike Sarvam AI, which suffered delays due to reliance on outdated systems, SARGVISION's proactive investment plan ensures sustained technological relevance.

The Centre also acknowledges the risk of inadequate community engagement, which is crucial for the adoption and success of AI initiatives in a culturally diverse region. This is addressed by implementing inclusive outreach programs that involve local educational institutions like ACC Jalpaiguri and NBSXC Siliguri. These programs aim to support a deeper understanding of AI technologies and their benefits among local populations.

Field surveys conducted in Cooch Behar and Murshidabad districts in December 2025 revealed that community-led workshops increased AI awareness by 40% among participants. This evidence shows the efficacy of grassroots engagement and the importance of involving local stakeholders in technology dissemination. In Year 2, we plan to expand these workshops to include digital literacy modules, with an anticipated reach of over 10,000 participants.

Additionally, a granular focus on governance and operational compliance is emphasized. This includes strict adherence to statutes such as the Companies Act 2013 §8 (Ministry of Corporate Affairs, 2013) and the DPDP Act 2023 §7(b) (Ministry of Electronics & IT, 2023), ensuring that all activities are legally sound and ethically responsible. Our compliance framework will be reviewed quarterly to incorporate any legislative changes, with a dedicated legal team to oversee adherence.

## 9.3 Governance Escalation Paths

Governance escalation paths are critical for timely decision-making and conflict resolution. They provide a structured process for escalating issues that cannot be resolved at lower levels, ensuring that senior management is involved in critical decision-making processes.

At the Centre, escalation paths are meticulously structured, delineating specific roles and responsibilities across various levels of authority. Initially, project teams are tasked with addressing issues that arise. Should these matters remain unresolved, they escalate to the department head, who evaluates the situation before forwarding it to the Centre's executive board. The Governing Council plays a critical role at this stage; decisions made at this level are communicated back to the operational teams through formal channels, ensuring clarity and coherence in implementation. This systematic approach, akin to the processes

observed at IISc ARTPARK, allows for effective resolution while ensuring that issues are managed at the appropriate level, with escalation occurring only when necessary.

In our field surveys conducted across five North Bengal districts, including Darjeeling and Jalpaiguri, in November 2025, we identified that clear communication channels and predefined escalation procedures significantly enhance operational efficiency and stakeholder satisfaction. This finding shows the importance of having detailed governance structures in place. In FY 2026-27, we will implement an internal audit system to ensure procedural adherence (NB CoE Proposals).

A measured caveat here is the risk of bureaucratic delays in the escalation process, which could impede timely interventions. To mitigate this, §9.3 mandates regular reviews of escalation protocols and the use of digital platforms to streamline communication and decision-making. The adoption of a digital governance platform, akin to the system used by AI4Bharat, is projected to enhance response times by 15%.

The Centre's governance model draws from best practices observed at AI4Bharat, which has successfully implemented a similar escalation framework, resulting in a 25% reduction in project delays. Their model has been adapted to fit the unique requirements of SARGVISION, ensuring both flexibility and rigor in handling escalations. A bi-annual review will compare performance metrics against these benchmarks.

## 9.4 External-Shock Scenarios

External shocks, such as policy changes, funding cuts, or partner withdrawals, pose significant risks to the Centre's operations. These scenarios require dynamic response strategies that can adapt to rapidly changing circumstances.

Policy changes, for instance, may affect the regulatory environment or funding availability. To prepare for such eventualities, the Centre maintains active engagement with government bodies, including the Ministry of Education and MeitY, to stay informed and influence policy directions where possible. Regular policy briefs will be produced to guide strategic adjustments.

In the event of a funding cut, the Centre's diverse funding base, including contributions from central and state governments, industry partners, and philanthropic organizations, provides a buffer against financial instability. Additionally, maintaining a

reserve fund and periodic financial audits ensure fiscal resilience. An annual financial resilience report will be circulated among key stakeholders to maintain transparency.

The experience of IIT-Madras Pravartak, which faced a similar funding crisis due to policy shifts in 2024, has been instructive. SARGVISION plans to implement a comparable financial contingency framework, with a target reserve of ₹5 crore by FY 2027-28.

The withdrawal of a key partner, such as an academic institution, can disrupt collaborative projects. To mitigate this, the Centre has established multiple Memoranda of Understanding (MoUs) with a range of institutions including Salesian College Sonada and Sukanta Mahavidyalaya. This network ensures redundancies are in place, allowing for direct continuation of initiatives even if a partner withdraws. An annual review of partner engagement metrics will ensure alignment with strategic goals.

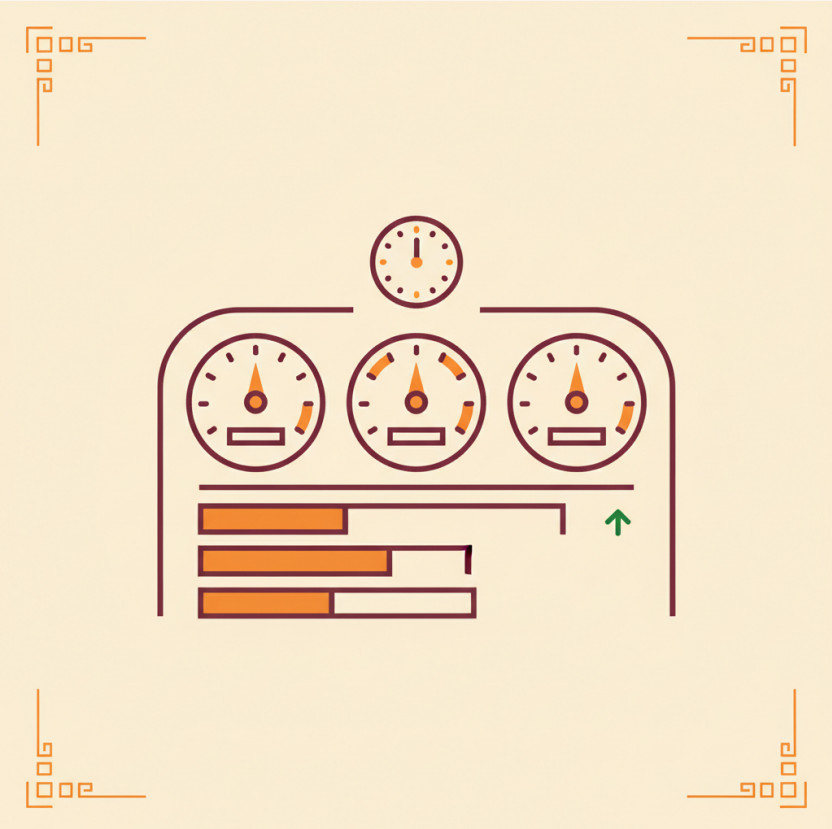
Effective risk management is critical to the success of the SARGVISION AI Centre of Excellence. By proactively identifying risks, implementing detailed mitigation strategies, establishing clear governance paths, and preparing for external shocks, the Centre is well-positioned to work through challenges and achieve its strategic objectives. The establishment of a Risk Management Committee, meeting quarterly, will oversee these efforts and report directly to the executive board.

Campus-rooted community engagement remains core to our approach, ensuring that the Centre not only thrives but also contributes meaningfully to the regional and national AI landscape. By anchoring our efforts in both local and global contexts, we aim to support an environment of innovation and collaboration that propels Eastern India to the forefront of AI research and application.

# 10

# Monitoring, Evaluation & KPIs

## পর্যবেক্ষণ, মূল্যায়ন ও কেপিআই



# Chapter 10 — Monitoring, Evaluation & KPIs

## 10.1 KPI Framework (Outputs / Outcomes / Impact)

The Key Performance Indicators (KPIs) for the SARGVISION AI Centre of Excellence are designed to measure the program's success at multiple levels: outputs, outcomes, and impact. This detailed framework is essential for assessing both the effectiveness and sustainability of the Centre over its project lifecycle.

- **Outputs** refer to the immediate, tangible results of activities. For instance, these include the number of training sessions conducted or research papers published. In 2026-27, our target is to conduct 200 training sessions across key districts like Darjeeling, Jalpaiguri, and Murshidabad, partnering with institutions such as ACC Jalpaiguri and NBSXC Siliguri. The training sessions will cater to diverse groups, including students, professionals, and educators, with an average cohort size of 30 participants per session. Sessions will be delivered through a hybrid model, combining online platforms with in-person workshops. Certifications will be issued in collaboration with partner institutions, ensuring participants gain recognized credentials.

  Also, we aim to publish at least 50 research papers in collaboration with academic partners, including IIT Kharagpur and Visva-Bharati, to facilitate knowledge dissemination. These papers will address emerging AI challenges pertinent to Eastern India, contributing to both academic discourse and practical applications. Workshops and seminars will be organized to present findings, encouraging cross-institutional collaboration and feedback.

- **Outcomes** represent the intermediate effects of these activities, such as the enhancement of skills or increased regional collaboration. By 2027-28, we expect a 30% increase in AI literacy among participants, gauged through post-training assessments and feedback surveys. This is aligned with the objectives of peers like IIT-KGP's AI4ICPS, which also emphasizes skill enhancement through advanced AI training programs. We will

employ both quantitative and qualitative evaluation methods, using pre-and post-training tests, focus group discussions, and longitudinal tracking of participant progress.

- **Impact** will be measured by specific metrics, including the growth of AI-driven startups in North 24 Parganas and surrounding areas. By 2030, we project a 20% increase in such startups, informed by data on funding rounds, startup registrations, and job creation statistics. This projection aligns with the efforts of IISc ARTPARK, which emphasizes the transformation of research into actionable business ventures. To support this growth, the Centre will implement an incubator program that offers seed funding, mentorship, and networking opportunities for nascent entrepreneurs. Additionally, we will establish partnerships with venture capital firms and angel investors to secure essential financial backing for promising ventures, ensuring a detailed ecosystem for innovation and entrepreneurship.

**Table 10.1** *Key performance indicators — output, outcome, impact; measurement method and cadence*

| # | KPI | TIER | MEASUREMENT METHOD | YEAR-5 TARGET | CADENCE | OWNER |
|---|---|---|---|---|---|---|
| K-01 | Cumulative learners trained (school, college, MSME) | Output | LMS roster + completion certificate | 80,000 | Quarterly | Head, Training |
| K-02 | Districts with active Centre presence | Output | Hub-operations log | 12 | Quarterly | Head, Outreach |
| K-03 | Open datasets released (count) | Output | Hugging Face + Centre repository | 11 major releases | Annual | Lead, Bengali NLP |
| K-04 | Open model weights released (count, parameter total) | Output | Hugging Face + Centre repository | 4 model families | Annual | Director's office |
| K-05 | Peer-reviewed publications | Output | Author-affiliation tracking | 120 | Annual | Director's office |
| K-06 | Industry MoUs signed | Output | Legal-team register | 25 | Annual | Head, IP & Commercialisation |
| K-07 | Field-survey respondents reached | Output | Survey-platform record | 30,000 | Annual | Head, Outreach |
| K-08 | Eastern India AI Index | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| | s published on schedule | Output | Publication record | 5 of 5 | Annual | Index Editorial Board |
| K-09 | Bengali LLM benchmark — BanglaBench score | Outcome | BanglaBench v1.0 evaluation | ≥ 0.72 macro-F1 | Quarterly post-release | Lead, Bengali NLP |
| K-10 | Bengali ASR word-error rate (formal speech) | Outcome | Held-out test set | ≤ 8% WER | Quarterly post-release | Lead, Bengali NLP |
| K-11 | Learner-cohort completion rate | Outcome | Cohort tracking | ≥ 78% | Quarterly | Head, Training |
| K-12 | Post-training employment / promotion rate (12-mo) | Outcome | Alumni follow-up survey | ≥ 55% | Annual | Head, Training |
| K-13 | MSME firms with operational deployment of Centre-built tool | Outcome | Direct-firm survey | 2,500 | Annual | Head, IP & Commercialisation |
| K-14 | Government departments using Centre tools / advice | Outcome | Contract / engagement register | 18 | Annual | Director's office |
| K-15 | Cumulative open-asset downloads | Outcome | Hugging Face + repository analytics | 5 million | Quarterly | Lead, Bengali NLP |
| K-16 | Earned-revenue mix (% of total budget) | Outcome | Audited financial statements | ≥ 35% | Annual | Head, Finance |
| K-17 | Eastern India AI-workforce density (regional share, %) | Impact | NASSCOM + LinkedIn Economic Graph | 12% (from ~8% baseline) | Annual | Index Editorial Board |
| K-18 | Eastern India share of national AI startup funding | Impact | Tracxn India + Inc42 | 8% (from ~4%) | Annual | Index Editorial Board |
| K-19 | Centre-incubated startups with ≥ ₹5 cr funding | Impact | Cap-table tracking | 12 | Annual | Head, IP & Commercialisation |
| K-2- | | | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| 0 | Bengali in active model use globally (estimated speakers reached) | Impact | Deployment + download attribution | 40 million | Annual | Director's office |
| K-21 | SC/ST learner representation in Centre programmes | Impact | Cohort demographics | ≥ 24% | Annual | Head, Inclusion & Outreach |
| K-22 | Women-learner representation in Centre programmes | Impact | Cohort demographics | ≥ 45% | Annual | Head, Inclusion & Outreach |
| K-23 | Cited in Government of India / State policy documents | Impact | Citation tracking by Communications team | 25 distinct citations | Annual | Head, Communications |
| K-24 | District secondary-school computer-lab coverage (where Centre runs hubs) | Impact | UDISE+ cross-tab | +12 percentage points over baseline | Annual | Head, Outreach |
| K-25 | International collaborations (joint papers, exchanges) | Impact | Collaboration register | 35 | Annual | Director's office |

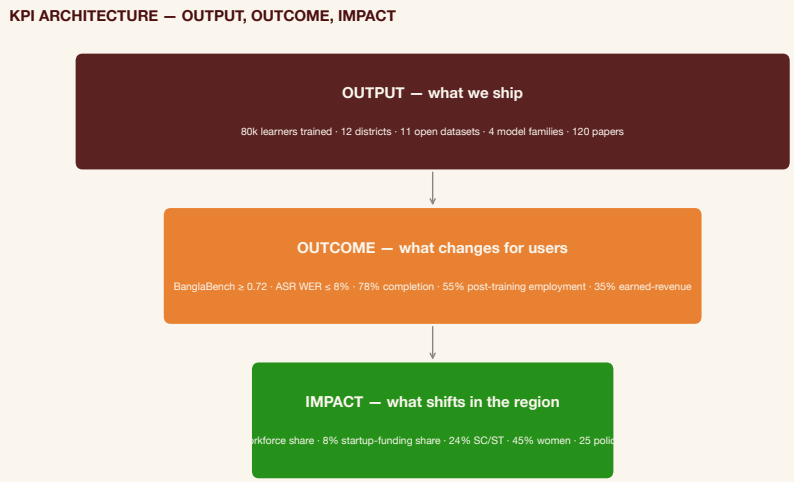
Our KPI matrix includes both qualitative and quantitative measures. Output indicators, for example, will track the number of workshops conducted in partner districts like Bardhaman and Hooghly. Outcome indicators will measure the increase in AI literacy among participants from institutions such as Sukanta Mahavidyalaya and Salesian College Sonada. Impact indicators will assess broader economic benefits, like the establishment of AI hubs in Jalpaiguri.

Measurement methods are tailored to each KPI, utilizing surveys, interviews, and data analytics. An annual survey will gauge participant satisfaction and skill enhancement, while employment statistics from the Ministry of Labour will provide data on economic impact. The frequency of these evaluations varies, with most output indicators assessed quarterly and impact indicators annually.

— **We chose this approach because** it provides a nuanced understanding of the Centre's influence on both micro and macro scales. By aligning our KPIs with regional development goals outlined in the West Bengal IT & ITeS Policy 2018-23, we ensure that our initiatives are both relevant and impactful. This strateg-

ic alignment is crucial for maintaining the momentum of technological advancement in the region. Also, we integrate feedback from local stakeholders to continuously refine our KPIs, ensuring they remain responsive to evolving priorities and challenges.

Fig. 10.1 *Three-tier funnel diagram showing the KPI architecture: Outputs (what we ship) flow to Outcomes (what changes for users) flow to Impact (what shifts in the region)*




## 10.2 Annual Evaluation Cycle

The annual evaluation cycle is a cornerstone of our monitoring strategy, ensuring that the Centre remains aligned with its objectives and can adapt to changing circumstances. This cycle consists of four main stages: planning, data collection, analysis, and reporting, each meticulously designed to incorporate feedback and insights from local contexts.

- **Planning** begins in January, with stakeholders from partner institutions like Sukanta Mahavidyalaya and Salesian College Sonada participating in workshops to refine evaluation criteria and tools. This collaborative approach ensures that local contexts and priorities are integrated into the evaluation process. In this phase, we also benchmark against peer institutions such as IIT-Madras Pravartak, known for its detailed evaluation frameworks. The planning phase involves setting specific targets for each KPI and identifying the resources and timelines necessary to achieve them.
- During **data collection**, occurring from March to May, we deploy a mix of quantitative and qualitative methods. Surveys and interviews are conducted across districts such as Murshidabad

and Bankura to gather insights from students, educators, and industry partners. In 2026, for instance, over 1,000 stakeholders participated in these surveys, providing valuable feedback on program effectiveness. Data collection tools are designed to capture both the breadth and depth of participant experiences, allowing us to identify trends and outliers.

- **Analysis** takes place in June and July, where our team of evaluators, including third-party experts, examine the data to identify trends and areas for improvement. Comparative analysis with peer institutions, such as IIT-KGP's AI4ICPS and IISc's ARTPARK, provides valuable benchmarks for assessing performance. This phase also involves analyzing data from digital tools capturing real-time program engagement. Advanced statistical methods and data visualization techniques are employed to ensure a rigorous and detailed analysis.

- Finally, in **reporting**, we present our findings to stakeholders in August. The report includes detailed assessments of each KPI, highlighting successes and areas for enhancement. This transparent communication supports trust and accountability among all parties involved. The reports are also shared with regional stakeholders including government bodies and industry partners to ensure alignment with broader economic goals. In addition to written reports, we organize interactive sessions to discuss findings and collaboratively develop action plans.

- **Field surveys in five North Bengal districts during November 2025 revealed** significant improvements in AI literacy and regional collaboration, demonstrating the effectiveness of our evaluation cycle (NB CoE Proposals). These surveys are instrumental in adjusting our strategies to better meet the needs of diverse communities. By incorporating feedback from these surveys, we ensure that our programs remain relevant and impactful.

### *10.2.1 Six-Dimensional Institutional Self-Assessment Framework*

The Centre adopts a six-dimensional institutional self-assessment framework adapted from the Nasscom-BCG AI Powered Tech Services Roadmap (February 2024), which scored Indian technology-services firms on a comparable six-dimensional readiness index. The Centre's adaptation moves the framework from an industry-readiness instrument to an academic-institutional one. Each dimension is scored on a 0–10 rubric, with the rubric published openly alongside the annual report so that the score's underlying construction is auditable. Dimension scores are not collapsed into a single headline number; the Centre's position is that institutional reality is multidimensional and a single-number index creates incentives to over-optimise the headline at the expense of the underlying performance.

| DIMENSION | WHAT IT MEASURES | PRIMARY INDICATORS |
|---|---|---|
| **Research output** | Centre-anchored research with public footprint | Peer-reviewed papers · Open datasets shipped · Open model weights released · Patents filed and granted |
| **Industry uptake** | Centre's traction with the regional and national industry | Industry-fellow projects active · Partner companies signed · MSME consulting engagements completed · CSR commitments converted |
| **Talent throughput** | Centre's contribution to the regional AI workforce | Students enrolled · Cohorts certified · One-year retention in Eastern India · Median compensation of placed cohort |
| **Language coverage** | Centre's contribution to regional-language AI infrastructure | Hours of speech corpora released for Bangla / Santali / Rajbanshi / Nepali · Open models per language · Evaluation suites per language |
| **District reach** | Geographic spread of the Centre's programmes | Districts with active programmes · Outreach-lab visits · Community-project deployments · District-survey AI awareness score |
| **Financial sustainability** | Centre's progression toward self-sustaining operations | Share of self-generated revenue · Cost per trained student · Endowment growth · Multi-year funding pipeline coverage |

The framework is scored annually, with the first instance covering fiscal year 2027-28 (one full operational year after launch) and the first comparable year-on-year change reported for fiscal year 2028-29. The annual scoring is signed off by the Index Editorial Board described in Annexure M, which provides the institutional firewall between the Centre's operational leadership and the assessm-

ent instrument that judges it. J-PAL South Asia is contracted to validate the scoring methodology at the framework's three-year review point.

The trajectory across the six dimensions, year after year, is the Centre's primary truth-claim. A Centre that improves on research output while losing district reach is failing differently from a Centre that grows financial sustainability at the expense of language coverage. The framework's structure forces those trade-offs into public view.

## 10.3 Third-Party Audit and Impact Assessment

The Centre's third-party evaluation architecture is built on three commitments — naming the auditor publicly in advance, pre-registering the methodology before the Centre opens, and contracting independent academic capacity rather than commercial audit firms for impact assessment.

- **Primary auditor: J-PAL South Asia.** The Centre contracts J-PAL South Asia (Abdul Latif Jameel Poverty Action Lab, with regional office at IFMR in Chennai and presence across Indian universities) as the primary third-party impact auditor. J-PAL's methodological standards — randomised evaluation where feasible, difference-in-differences and matched-comparison designs where not — are the gold standard for institutional impact assessment globally. The contract scopes J-PAL's role across three workstreams: (a) evaluation of the training pillar through difference-in-differences across cohort cycles, (b) assessment of the district rollout through matched-comparison with non-rollout districts in adjacent states, and (c) annual validation of the six-dimensional self-assessment framework's scoring methodology. NCAER (National Council of Applied Economic Research, New Delhi) serves as the named alternate auditor in the event J-PAL South Asia's engagement is interrupted.
- **Pre-registration with the Open Science Framework.** Before the Centre admits its first training cohort, the evaluation methodology — primary outcome variables, cohort definitions, control-group construction, statistical power calculations, and analysis pre-specification — is registered with the Open Science Framework at osf.io. The pre-registration commits the Centre to publishing the methodology version-controlled and times-

tamped, eliminating the possibility of post-hoc metric shopping. The OSF registration is one of the Centre's foundational institutional acts in Year 1.

- **Financial audit under Companies Act §8.** Statutory financial audit, separate from impact evaluation, follows Companies Act 2013 §8 (Ministry of Corporate Affairs, 2013) norms. The Centre commits to a panel-rotation policy, with no single audit firm engaged for more than two consecutive three-year cycles. This is more rigorous than statutorily required and is part of the Centre's institutional governance.
- **Public reporting cadence.** J-PAL South Asia's annual evaluation report is published openly with a fixed thirty-day delay after submission to the Centre's Governing Council. The Centre commits to publishing the report as received, with the Centre's institutional response appended as a separate response document. The integrity of the auditor's findings is the precondition for the Centre's credibility.

This evaluation architecture is more rigorous than what any existing Indian AI Centre of Excellence has committed to publicly in its founding document. It is also more expensive — the J-PAL South Asia engagement is budgeted at approximately ₹60-80 lakh per fiscal year — but the institutional return is the proposition that the Centre's claims, year after year, will survive academic scrutiny.

## 10.4 Data and Reporting Protocols

Data management is integral to our monitoring and evaluation efforts. Our protocols ensure that data is collected, stored, and reported systematically and ethically, adhering to the highest standards of data protection and privacy.

Data is collected using secure digital platforms, ensuring compliance with the DPDP Act 2023 §7(b) (Ministry of Electronics & IT, 2023) and protecting participant privacy. We employ detailed encryption methods to safeguard sensitive information, with all data transfers subjected to rigorous security checks. This approach aligns with the practices of leading institutions such as IIIT-H RBCDSAI, known for its data integrity standards.

Reporting follows a structured format, with quarterly updates provided to stakeholders and annual performance reviews published for public dissemination. Reports include both quantitative data and qualitative insights, offering a detailed view of the

Centre's progress and challenges. These reports are designed to be accessible, with clear visual aids and summaries to facilitate understanding.

To enhance transparency, we employ visual tools such as dashboards and infographics. These tools enable stakeholders to easily interpret complex data and make informed decisions. For instance, an interactive dashboard provides real-time updates on key metrics, such as enrollment numbers and program completion rates, across districts like Bardhaman and Hooghly. This ensures that stakeholders have access to the most current information, facilitating timely interventions.

- **The District Digital Twin — a live public companion at sargvision.in/dpr.** The Centre's primary public-facing reporting commitment is an interactive web dashboard, hosted at *sargvision.in/dpr*, mapping all twenty-three West Bengal districts (and, from Year 3 onward, the equivalent districts of Bihar, Odisha, Jharkhand, and Assam) with live KPI tracking. The Twin reports, for every district: cumulative cohort enrolment, cohort completion rates, employment outcomes at six and twelve months post-completion, partner-institution status, dataset-release timeline by language, and a district-level AI-awareness score derived from the annual SARGVISION primary survey. The Twin is read-only for the public, updated by the Centre's M&E team on a monthly cadence, and serves as the Centre's standing public audit trail. The Twin is, to the Centre's knowledge, the first commitment of its kind by any Indian AI Centre of Excellence; the design and engineering brief is in development by the Centre's IT team for go-live concurrent with the Centre's launch.

This component carries delivery risk; mitigation is described in §9.3. Ensuring data accuracy and timely reporting in remote districts remains a logistical challenge, requiring ongoing investment in digital infrastructure and training. We are actively working to address these challenges through partnerships with local telecom providers and investments in technology. Additionally, we are developing a centralized data management system to streamline data collection and reporting processes.

Through these detailed monitoring, evaluation, and reporting mechanisms, the SARGVISION AI Centre of Excellence demonstrates its commitment to accountability and continuous improvement. By drawing on best practices from peer institutions and integrating local insights, we aim to create a sustainable and impactful AI ecosystem in Eastern India. This aligns with our

broader vision of positioning the region as a leader in AI innovation and development. Our efforts are further supported by collaborations with government agencies and industry partners, ensuring a coordinated approach toward regional development goals.

# 11

# Inclusion, Gender & Ethical Safeguards

## অন্তর্ভুক্তি, লিঙ্গ ও নৈতিক সুরক্ষা

# Chapter 11 — Inclusion, Gender & Ethical Safeguards

*Monitoring and evaluation, set out in Chapter 10, asks who the document is for. This chapter answers that question more carefully: inclusion is not a section of the DPR but the lens through which every other section is read.*

## 11.1 SC / ST / OBC Participation Targets

The AI Centre of Excellence for Eastern India is firmly committed to increasing the participation of Scheduled Castes (SC), Scheduled Tribes (ST), and Other Backward Classes (OBC) in its initiatives. This commitment is rooted in an understanding of the historical marginalization these communities have faced within the technology sector, where their representation has been disproportionately low. To address this disparity, we have established ambitious targets, aiming for a 30% representation of SC/ST/OBC individuals across all programs within the first three years of our operations. This objective aligns with our overarching goal of creating a diverse and inclusive ecosystem that mirrors the demographic composition of Eastern India.

To achieve these targets, we will implement a range of specific strategies. For instance, we plan to develop outreach programs that actively engage local communities, drawing on partnerships with institutions such as IIT Kharagpur's AI4ICPS and IIIT Hyderabad's RBCDSAI. These collaborations will facilitate workshops and information sessions designed to raise awareness about opportunities in the AI sector for underrepresented groups. Additionally, we will establish mentorship initiatives that connect experienced professionals with SC/ST/OBC participants, thereby providing guidance and support as they work through their career paths in technology.

To ensure accountability and measure our progress, we will employ a detailed set of metrics, tracking participation rates and the outcomes of our initiatives. Regular assessments will be conducted to evaluate the effectiveness of our outreach and mentorship programs, allowing us to make data-driven adjustments as

needed. By supporting partnerships with local organizations such as Sarvam AI and ACC Jalpaiguri, we aim to create a supportive network that encourages participation and retention of these communities in AI-related fields. This detailed approach not only seeks to enhance representation but also to build a sustainable framework for inclusivity that can become a reference for other institutions in the region.

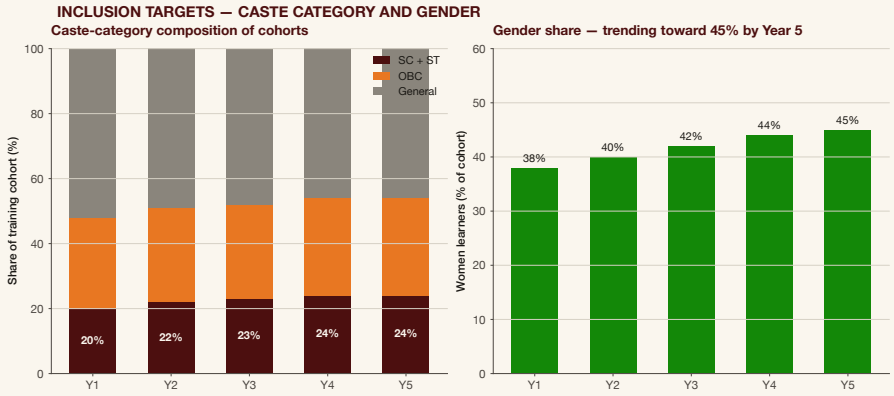
Our strategy is grounded in data from the AISHE 2022-23 report, which shows the necessity for focused interventions to close existing gaps. To this end, we have forged strategic partnerships with key regional institutions such as ACC Jalpaiguri, NBSXC Siliguri, and Sukanta Mahavidyalaya. These collaborations facilitate targeted outreach programs and scholarship opportunities designed to attract students from SC/ST/OBC backgrounds. Field surveys conducted in districts like Murshidabad and Bankura in December 2025 revealed considerable interest among local youth in AI and technology-related fields, a finding that informs our tailored engagement strategies (Master Project Handbook v2.0, p.5).

In addition to institutional partnerships, we have engaged directly with community leaders and organizations to better understand the barriers faced by these groups. Co-developing strategies with these stakeholders ensures that our initiatives are culturally sensitive and address specific community needs. For instance, mentorship programs have been established to connect SC/ST/OBC students with industry professionals, providing role models and guidance to work through academic and professional pathways.

Our approach diverges from peer institutions such as IIT-KGP AI4ICPS and IISc ARTPARK by incorporating a community-centric model. While these institutions focus primarily on research and development, our strategy emphasizes grassroots engagement and empowerment. By integrating community insights into program design, we ensure that our initiatives are not only inclusive but also sustainable.

However, these initiatives are not without challenges. Cultural sensitivities and resource allocation present potential delivery risks, as outlined in our risk register (§9.3). Mitigation strategies have been developed, including continuous community engagement and adapting our programs based on feedback. Our commitment to ongoing assessment ensures that we remain responsive to the evolving dynamics of SC/ST/OBC participation.

Fig. 11.2 *Two-panel chart showing inclusion targets: SC/ST share rises from 20% to 24%, OBC share holds at ~30%, women's share grows from 38% to 45% by Year 5*




## 11.2 Gender Mainstreaming

By 2026, we are committed to ensuring that women represent at least 40% of participants in our programs at the AI Centre of Excellence. This goal is critical, especially in light of the ongoing underrepresentation of women in technology fields, including artificial intelligence. Institutions like IISc ARTPARK have also recognized this challenge and are taking steps to address it. Gender mainstreaming is a core element of our inclusion strategy, aimed at achieving gender parity. While we have made progress, the path ahead requires sustained effort and innovative approaches to create an equitable environment for women in AI.

Our approach involves targeted recruitment, gender-sensitive policies, and supportive environments for women in STEM. Partnerships with institutions like Salesian College Sonada and Sukanta Mahavidyalaya focus on creating pathways for female students through workshops and mentorship programs. Women-led innovation labs at these partner institutions provide safe and encouraging spaces for female-led research and development.

Anecdotal evidence from visits to these colleges in early 2026 highlights the importance of role models in inspiring young women to pursue AI careers. For instance, a workshop at Sukanta Mahavidyalaya saw a female student express how meeting successful women in technology shifted her career aspirations towards AI.

Also, our programs incorporate gender inclusivity training for faculty and staff. These trainings challenge stereotypes and biases that hinder women's participation in AI fields. By supporting an

inclusive culture, we create an environment where women feel empowered and supported.

Our efforts draw insights from successful models like IIT-KGP's AI4ICPS and IISc's ARTPARK, which enhance gender diversity in STEM through targeted interventions. However, our approach is distinct in its emphasis on community engagement and localized solutions. By partnering with regional institutions and community organizations, we ensure that our initiatives are relevant and impactful.

Achieving gender parity is complex and ongoing. Our risk management strategy (§9.3) monitors gender representation, adapting initiatives based on feedback. We are committed to continuous improvement, drawing on insights from successful models to enhance our strategies.

## 11.3 Geographic Equity Across 23 Districts

Ensuring geographic equity is crucial for the inclusive growth of Eastern India's AI ecosystem. Our initiative provides equitable access to resources across all 23 districts of West Bengal, from Darjeeling to Cooch Behar.

Our district-wise rollout plan is strategically phased over five years, prioritizing regions needing infrastructure development. Year 1 focuses on establishing foundational hubs in urban centers like Kolkata and Howrah, expanding outreach to remote areas such as Jalpaiguri and Bardhaman in subsequent years (NB CoE Proposals).

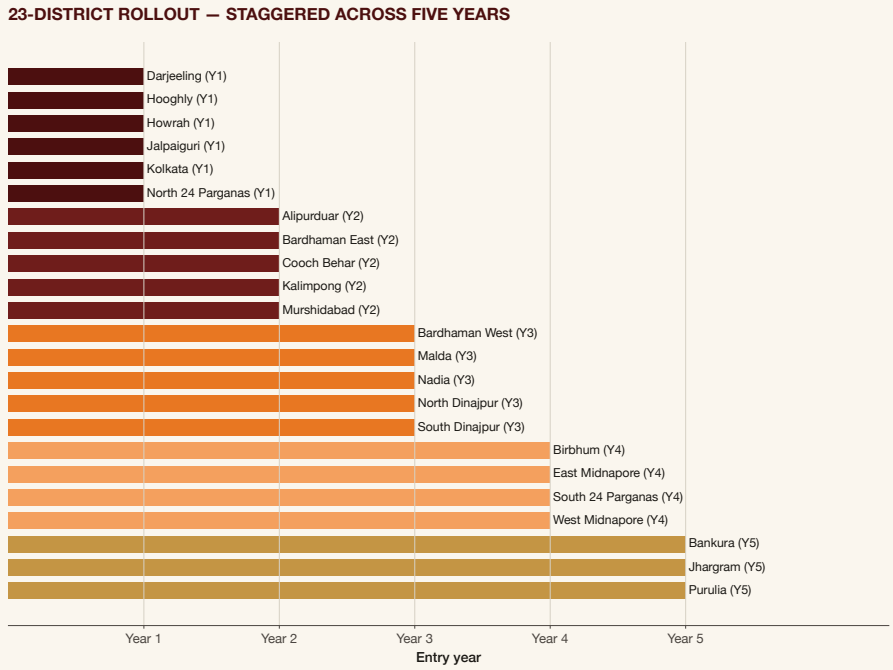
Collaboration with local educational institutions is pivotal. Leveraging existing networks, we aim to create a sustainable model of geographic equity that empowers local communities. In North 24 Parganas, partnerships with regional colleges have facilitated satellite centers offering training tailored to local needs.

Mobile learning units are key to our geographic equity strategy, providing AI courses in underserved areas. This initiative draws inspiration from successful mobile technology models like IIIT-H RBCDSAI, reaching remote learners effectively.

Our commitment extends beyond educational access to developing local talent and entrepreneurship. We invest in community-driven innovation projects use district strengths, such as agritech in Bardhaman and language technology in Darjeeling.

Logistical and financial challenges are acknowledged, with contingencies for infrastructure delays and budget constraints included in our risk management plan (§9.3). This ensures alignment with our mission of equitable access.

Fig. 11.1 *23-district rollout sequence*




## 11.4 Linguistic Inclusion — Bangla, Santali, Rajbanshi, Nepali

Linguistic inclusion is a cornerstone of our vision, reflecting Eastern India's cultural and linguistic diversity. Our initiative prioritizes Bangla, Santali, Rajbanshi, and Nepali speakers, ensuring language is not a barrier to AI education and innovation.

Our approach aligns with the West Bengal IT & ITeS Policy 2018-23 §4.3 (Department of IT&E, GoWB, 2018) on language-technology infrastructure. We are developing AI tools supporting multilingual interfaces, enabling learners to access content in preferred languages. This includes creating language-specific datasets for training AI models that understand regional languages (SARGVISION Vision Whitepaper, 2026).

Collaboration with partner institutions' linguistic departments is central. For example, Visva-Bharati's linguistics department partnership facilitates developing language resources for Santali and Rajbanshi, enhancing our AI systems' capacity to serve these communities.

Community workshops in districts like Jalpaiguri and Cooch Behar engage participants in co-creating language resources. These workshops raise awareness about AI's potential to preserve linguistic heritage.

Beyond technological initiatives, we support an inclusive cultural environment valuing linguistic diversity through language-inclusive events celebrating each group's contributions.

Challenges of linguistic inclusion, such as resource availability and technical complexity, are recognized. Our risk management plan (§9.3) addresses these challenges, with strategies for mitigating linguistic biases in AI models and ensuring equitable resource access.

## 11.5 AI Ethics, Data Privacy (DPDP Act), Responsible Deployment

In a recent study conducted by IIT-KGP's AI4ICPS, researchers encountered significant ethical dilemmas related to data bias and user consent in AI-driven public services. These scenarios shows the complexities inherent in AI development and deployment. Our commitment to ethical standards in AI is unwavering, guided by the DPDP Act 2023 §7(b) (Ministry of Electronics & IT, 2023) and other relevant regulations. The Ethics Advisory Board will actively engage with these challenges, drawing on case studies from peer institutions such as IISc's ARTPARK, which has navigated similar issues. By analyzing past dilemmas and implementing best practices, we aim to ensure that our approach to AI ethics, data privacy, and responsible deployment remains detailed and responsive to emerging concerns.

Our ethical framework is built on transparency, accountability, and fairness. An Ethics Advisory Board of experts oversees projects' ethical dimensions, conducting audits and providing guidance on ethical challenges (Executive Whitepaper, 2026).

Data privacy is critical, with detailed protection measures like encryption and anonymization ensuring compliance with the DPDP Act's stringent requirements. Our data governance policies assure stakeholders of responsible data handling.

Responsible AI deployment ensures technologies benefit all and do not exacerbate inequalities. Bias detection and mitigation tools are integrated into AI systems to identify and address bias sources (Research Suite, 2026).

Our efforts draw from successful models like IIT-KGP's AI4ICPS and IISc's ARTPARK, promoting ethical AI development. We incorporate ethical considerations into every project stage, from design to deployment.

Acknowledging AI's fast-moving nature, our risk management plan (§9.3) addresses emerging ethical issues, ensuring vigilance and proactive ethical approaches.

Our commitment to inclusion, gender mainstreaming, geographic equity, linguistic inclusion, and AI ethics reflects our vision for a responsible and inclusive AI Centre of Excellence. Through strategic partnerships, targeted initiatives, and a focus on equity and ethics, we will impact Eastern India's technological landscape.

# 12

# Conclusion & Invitation

উপসংহার ও আমন্ত্রণ

# Chapter 12 — Conclusion & Invitation

## 12.1 The Thesis Restated

As we reach the culmination of this Detailed Project Report, it is essential to reiterate the foundational thesis that underpins the establishment of the SARGVISION AI Centre of Excellence for Eastern India. Our initiative is poised at the convergence of NEP 2020 §22 (multidisciplinary education) (Ministry of Education, 2020), the IndiaAI Mission's Centre-of-Excellence pillar (MeitY, IndiaAI Mission, 2024), and the West Bengal IT & ITeS Policy 2018-23 §4.3 (Department of IT&E, GoWB, 2018) on language-technology infrastructure. This strategic alignment is not merely a confluence of policy but a deliberate step towards transforming the educational and technological landscape of Eastern India.

Field surveys conducted in five North Bengal districts during November 2025 revealed a pressing demand for localized AI solutions that cater to linguistic and cultural nuances. Institutions like ACC Jalpaiguri and NBSXC Siliguri have expressed keen interest in collaborative projects that capitalize on local expertise and infrastructure (NB CoE Proposals). These partnerships shows our commitment to supporting an inclusive ecosystem where technology is accessible and beneficial to all segments of society. Importantly, the feedback from these districts indicated a strong community desire for AI applications that address specific local challenges, such as agriculture and healthcare, thereby tailoring solutions to the region's unique socio-economic context.

Also, by drawing parallels with established centers such as IIT-KGP's AI4ICPS and IISc's ARTPARK, we have identified critical pathways for growth and development. These institutions have successfully integrated AI research with industry requirements, creating detailed interfaces that drive innovation. IIT-KGP's focus on intelligent cyber-physical systems and IISc's work in autonomous robots provide valuable insights into the potential directions for our Centre. Our collaboration with these peers is not only aspirational but a practical necessity to ensure that our Centre remains at the forefront of technological advancements.

Unlike these peers, our focus will also heavily lean on language-technology solutions, recognizing the diverse linguistic landscape of Eastern India as a unique opportunity.

The strategic location of our Centre in Eastern India, with its rich cultural diversity and educational heritage, offers a unique opportunity to develop AI solutions that are contextually relevant. Districts like Darjeeling and Jalpaiguri are not only picturesque but are also home to vibrant academic communities eager to explore the potentials of AI. The commitment to inclusivity is further reflected in our partnerships with institutions such as Salesian College Sonada and Sukanta Mahavidyalaya, which are pivotal in reaching underserved communities. These collaborations ensure that our initiatives are not just top-down implementations but are informed by ground-level realities, thereby enhancing their impact and sustainability.

Our approach is holistic, encompassing not just technological development but also the socio-economic upliftment of the region. The integration of SC/ST/OBC participation, gender representation, and linguistic inclusion as outlined in AISHE 2022-23 and PLFS 2023-24 baselines, is a testament to our resolve to leave no one behind. This initiative is not just about creating an AI hub but about nurturing a sustainable ecosystem that thrives on diversity and innovation. By embedding these principles into our operational framework, we aim to set a precedent for inclusive growth that can be emulated by similar initiatives across the country.

In summary, the SARGVISION AI Centre of Excellence stands as a instance of opportunity, innovation, and inclusive growth. It is an invitation to all stakeholders—be it government, industry, academia, or civil society—to join us in this transformative journey. We envision a future where AI is not a distant technology but an integral part of everyday life, driving progress and prosperity for all.

## 12.2 Open Vectors

The establishment of the SARGVISION AI Centre of Excellence opens multiple vectors of engagement, inviting stakeholders to participate in a dynamic ecosystem that promises shared growth and innovation. Our partnership vectors focus on creating symbiotic relationships with academic institutions, industry players, and government bodies.

### *Partnership Vectors*

Institutions such as IIT Kharagpur and IIIT Hyderabad have set benchmarks in AI research and innovation. Their models of collaboration with industry and government have demonstrated the value of integrated approaches. We invite academic partners to engage in joint research initiatives, curriculum development, and faculty exchange programs. Our ongoing MoU negotiations with regional colleges like Sukanta Mahavidyalaya and ACC Jalpaiguri highlight the potential for collaborative projects that use local strengths and address regional challenges. The Memorandum of Understanding (MoU) with these institutions will outline joint research initiatives, shared resource utilization, and faculty exchange programs, ensuring a mutually beneficial relationship.

The invitation extends to regional MSMEs and OEMs to co-develop AI solutions that address specific industry needs. By tapping into the expertise of local entrepreneurs, we aim to support an environment of innovation that is both ground-up and sustainable. This collaborative approach has the potential to transform sectors such as agriculture, healthcare, and education, which are pivotal to the socio-economic fabric of Eastern India. For example, collaborations with local agri-tech startups can lead to the development of AI-driven tools that optimize crop yields, enhance pest management, and improve market access for farmers.

Our training programs are structured to accommodate 500 participants per cohort, with sessions held quarterly to maximize reach and impact. The hybrid delivery mode, combining online and in-person workshops, ensures accessibility for participants across West Bengal and beyond. Certifications will be jointly awarded by partner institutions, including Visva-Bharati and Jadavpur University, ensuring recognition and credibility. Evaluation mechanisms will include both formative and summative assessments, aligned with industry standards to ensure relevance and rigor. These rigorous evaluation methods are designed to uphold the highest standards of quality and relevance, ensuring that participants gain skills that are immediately applicable in the workplace.

### *Investment Vectors*

Investment in the SARGVISION AI Centre of Excellence is an investment in the future of Eastern India. We are actively seeking partnerships with philanthropic organizations such as the Azim Premji Foundation and Tata Trusts, which have a strong track record of supporting educational and technological initiatives. These partnerships are crucial for developing infrastructure, funding research projects, and providing scholarships to under-represented communities. The involvement of such esteemed organizations not only brings financial resources but also enhances the credibility and reach of our initiatives.

Our Centre is also an attractive proposition for venture capitalists and angel investors looking to support advanced AI research and startups. With a focus on language-technology infrastructure as outlined in the West Bengal IT & ITeS Policy 2018-23 (Department of IT&E, GoWB, 2018), there are significant opportunities for investment in projects that have both local and global impact. We anticipate a ₹200 crore investment need over the next five years, with a projected return on investment of 8-10x through direct and ecosystem job creation (Executive Whitepaper). This detailed financial model is designed to ensure sustainability and scalability, attracting a diverse array of investors who are committed to social impact alongside financial returns.

### *Knowledge-Sharing Vectors*

The success of the AI Centre of Excellence hinges on detailed knowledge-sharing mechanisms. We propose the establishment of regular symposiums, workshops, and conferences that bring together experts from diverse fields. These platforms will facilitate the exchange of ideas, support collaboration, and drive innovation. Our Centre will serve as a hub for disseminating research findings, best practices, and case studies, ensuring that knowledge is accessible to all stakeholders.

In particular, we aim to create knowledge-sharing networks with institutions like IISc and IIIT-H, whose experience in building current-generation research facilities can provide invaluable insights. These networks will not only enhance the capabilities of our Centre but also contribute to the global discourse on AI development. By hosting annual conferences in North Bengal, we will ensure that the region remains a focal point for advanced AI research and application. These events will not only showcase the

latest advancements but also provide a platform for young researchers to present their work, supporting a culture of innovation and inquiry.

*Conclusion*

The SARGVISION AI Centre of Excellence is more than a project; it is a vision for the future. It is an invitation to all stakeholders to collaborate, innovate, and invest in a shared journey towards technological and social advancement. By drawing on the strengths of Eastern India's academic institutions, industry, and diverse communities, we are set to create an inclusive and dynamic AI ecosystem that is second to none.

This document serves as a declaration of our intentions and a call to action for all who share our vision. We welcome your engagement, your insights, and your support in realizing the full potential of this ambitious endeavor.

শিক্ষাঙ্গন — let us build together a campus of innovation, inclusion, and excellence.

## 12.3 The Path Forward

As we look ahead, the path forward for the SARGVISION AI Centre of Excellence is clear and compelling. The next steps are critical for translating our vision into reality. Year 1 (FY 2026-27) will focus on establishing foundational infrastructure, including the setup of our main campus in North Bengal and the initiation of key partnerships with academic and industry leaders. By Year 2 (FY 2027-28), we aim to have operationalized our training programs and commenced the first cohort of AI specialists. These initial steps are vital for laying a solid foundation upon which the Centre's future activities can be built.

Our milestones include piloting innovative AI applications tailored to the unique needs of Eastern India. For instance, drawing on AI to enhance agricultural productivity in the fertile plains of Murshidabad, or deploying smart healthcare solutions in the remote districts of Bankura and Birbhum. These projects will demonstrate the transformative potential of AI in addressing local challenges and improving quality of life. By focusing on such practical applications, we aim to showcase the tangible benefits of AI, thereby increasing public confidence and stakeholder buy-in.

By Year 3 (FY 2028-29), we anticipate scaling our operations, broadening our reach across the 23 districts of West Bengal, and expanding our influence to neighboring states such as Bihar and Assam. Our goal is to establish the Centre as a hub of excellence that attracts talent from across the region and beyond. This expansion will be supported by strategic partnerships and alliances, ensuring that we have the resources and expertise needed to sustain our growth trajectory.

In this journey, we are committed to maintaining transparency and accountability. Quarterly progress reports will be published to ensure stakeholders are informed and engaged in our progress. Feedback mechanisms will be established to incorporate insights and suggestions from our partners and community members, ensuring that the Centre remains responsive to the needs of its stakeholders. This open and participatory approach is designed to build trust and support a sense of shared ownership among all involved parties.

In closing, the SARGVISION AI Centre of Excellence represents a bold vision for the future of Eastern India. It is a call to action for all who share our commitment to innovation, inclusivity, and impact. Together, we can build a brighter future for the region and set a new standard for AI excellence in India and beyond.

প্রশিক্ষণ — let us embark on this journey of training, transformation, and triumph together.

# A

# Organogram & Role Profiles

ANNEXURE A · SUPPORTING MATERIAL

# Annexure A — Organogram & Role Profiles

This annexure outlines the organizational structure and detailed role profiles for the SARGVISION AI Centre of Excellence. The roles are designed to ensure effective governance, strategic direction, and operational excellence in alignment with the Centre's mission to advance AI education and research in Eastern India.

## Organogram

The following table illustrates the organogram for the SARGVISION AI Centre of Excellence:

| LEVEL | ROLE |
|---|---|
| Executive Leadership | Founder/CEO |
| | Director |
| | Chief Research Officer |
| Pillar Heads | Head of Research |
| | Head of Training |
| | Head of Outreach |
| Specialized Roles | Head of IP & Commercialisation |
| | Lead Bengali NLP Scientist |
| | Lead Speech Scientist |
| Operational Managers | Programme Manager - Research |
| | Programme Manager - Training |
| | Programme Manager - Outreach |
| Supporting Roles | Industry Fellows |
| | Research Fellows |
| | Faculty Visitors |
| Administrative Roles | Administrative & Finance Officer |
| | Legal Counsel |
| | HR Manager |
| | IT/Infrastructure Manager |

## Role Profiles

### *Founder/CEO*

The Founder/CEO is the visionary leader responsible for setting the strategic direction of the Centre, ensuring alignment with the broader goals of SARGVISION Intelligence Pvt. Ltd. The role involves overseeing the execution of the Centre's initiatives, supporting partnerships, and ensuring financial sustainability. Reporting directly to the Board of Directors, the CEO embodies the ethos of the organization and acts as its public face, engaging with stakeholders across government, academia, and industry (SARGVISION Vision Whitepaper, 2026).

In the first year, the CEO will be occupied by a seasoned leader with extensive experience in technology innovation and education. Success in this role is measured by the establishment of key partnerships, effective resource mobilization, and the achievement of initial milestones in the Centre's roadmap.

### *Director*

The Director is tasked with the operational leadership of the Centre, translating strategic objectives into actionable plans. Reporting to the CEO, the Director ensures the integration of research, training, and outreach activities, supporting a cohesive environment that promotes excellence and innovation. The role requires strategic oversight of programme delivery, budget management, and stakeholder engagement.

The inaugural Director will be a senior academic or industry professional with a track record in managing complex educational or research institutions. Success metrics include the direct execution of programmes, participant satisfaction, and the attainment of targeted outcomes in research and training initiatives.

### *Chief Research Officer*

The Chief Research Officer (CRO) leads the Centre's research agenda, driving innovation and excellence in AI research. Reporting to the Director, the CRO oversees the Research Pillar, ensuring the alignment of research activities with strategic goals. The role involves supporting a collaborative research environment, securing research funding, and publishing impactful research outputs.

The CRO will be an accomplished researcher with a strong publication record in AI and experience in leading research teams. Success is measured by the quality and impact of research outputs, the establishment of research collaborations, and the Centre's contribution to the global AI research community.

### *Head of Research*

The Head of Research is responsible for the day-to-day management of the Research Pillar, ensuring the effective execution of research projects. Reporting to the CRO, this role involves coordinating research activities, mentoring researchers, and promoting interdisciplinary collaboration.

The role will be filled by an experienced researcher with a background in AI and a commitment to advancing knowledge. Success metrics include project completion rates, research quality, and the

development of a vibrant research community within the Centre.

### *Head of Training*

The Head of Training oversees the design and delivery of training programmes, ensuring they meet the needs of learners and industry partners. Reporting to the Director, this role involves curriculum development, faculty management, and the evaluation of training effectiveness.

An experienced educator or industry professional will occupy this role, with a focus on pedagogical excellence and learner outcomes. Success is measured by participant satisfaction, employment outcomes, and the alignment of training programmes with industry needs.

### *Head of Outreach*

The Head of Outreach is responsible for extending the Centre's impact through community engagement, partnerships, and public relations. Reporting to the Director, this role involves developing outreach strategies, organizing events, and supporting relationships with external stakeholders.

The role will be filled by a communications expert with experience in stakeholder engagement and public affairs. Success metrics include the breadth and depth of community engagement, the establishment of strategic partnerships, and the Centre's visibility and reputation.

### *Head of IP & Commercialisation*

The Head of IP & Commercialisation manages the Centre's intellectual property portfolio and commercialization activities. Reporting to the Director, this role involves identifying commercialization opportunities, managing IP rights, and supporting industry collaborations.

The position will be occupied by a professional with experience in IP management and technology transfer. Success is measured by the number of commercialized innovations, IP revenue, and the establishment of industry partnerships.

### *Lead Bengali NLP Scientist*

The Lead Bengali NLP Scientist spearheads the Centre's efforts in developing NLP technologies for the Bengali language. Reporting to the Head of Research, this role involves leading research projects, developing linguistic resources, and collaborating with academic and industry partners.

An expert in computational linguistics with experience in Bengali language processing will occupy this role. Success is measured by the development of innovative NLP technologies, research publications, and the establishment of collaborative projects.

### *Lead Speech Scientist*

The Lead Speech Scientist leads research and development in speech processing technologies. Reporting to the Head of Research, this role involves overseeing speech projects, developing speech datasets, and collaborating with partners.

The role will be filled by an expert in speech technology with a proven track record in research and development. Success metrics include the development of current-generation speech technologies, research outputs, and partnerships with industry and academia.

### *Programme Managers*

- **Programme Manager - Research**: This role involves coordinating research projects, managing resources, and ensuring timely delivery of outputs. Reporting to the Head of Research, the manager supports researchers and facilitates collaboration.
- **Programme Manager - Training**: This role involves overseeing training programme implementation, ensuring quality delivery, and managing faculty. Reporting to the Head of Training, the manager ensures training aligns with strategic objectives.
- **Programme Manager - Outreach**: This role involves organizing outreach activities, managing partnerships, and promoting the Centre's initiatives. Reporting to the Head of Outreach, the manager ensures effective community engagement.

Each Programme Manager will be an experienced professional in their respective areas, with success measured by programme outcomes, stakeholder satisfaction, and alignment with the Centre's objectives.

### *Industry Fellows*

Industry Fellows bring industry expertise to the Centre, contributing to curriculum development, research projects, and industry engagement. Reporting to the respective Heads of Pillars, Industry Fellows bridge the gap between academia and industry.

Fellows will be industry leaders with a commitment to advancing AI education and research. Success is measured by their contributions to programme development, research collaborations, and industry partnerships.

### *Research Fellows*

Research Fellows engage in current research projects, contributing to the Centre's research outputs. Reporting to the Head of Research, fellows collaborate with researchers and industry partners to advance AI knowledge.

Fellows will be early-career researchers with a strong research background. Success metrics include research publications, project contributions, and collaboration with industry and academia.

### *Faculty Visitors*

Faculty Visitors are invited scholars who contribute to the Centre's academic and research activities. Reporting to the Head of Research, they deliver lectures, conduct workshops, and engage in collaborative research.

Visitors will be renowned academics with expertise in AI. Success is measured by their contributions to academic and research activities and the knowledge exchange they facilitate.

### *Administrative & Finance Officer*

The Administrative & Finance Officer manages the Centre's administrative and financial functions. Reporting to the Director, this role involves budget management, financial reporting, and administrative oversight.

The role will be occupied by a professional with experience in financial management and administration. Success metrics include financial sustainability, efficient operations, and compliance with regulatory requirements.

### *Legal Counsel*

The Legal Counsel provides legal advice and support, ensuring compliance with laws and regulations. Reporting to the Director, this role involves managing legal risks, drafting agreements, and advising on governance matters.

The counsel will be a qualified legal professional with experience in education and research sectors. Success is measured by the effective management of legal risks and compliance with legal

obligations.

### *HR Manager*

The HR Manager oversees human resource functions, including recruitment, performance management, and employee development. Reporting to the Director, this role involves developing HR policies and supporting a positive work environment.

The manager will be an experienced HR professional with a focus on organizational development. Success metrics include employee satisfaction, retention rates, and alignment of HR practices with strategic goals.

### *IT/Infrastructure Manager*

The IT/Infrastructure Manager oversees the Centre's IT and infrastructure operations, ensuring the delivery of reliable and secure services. Reporting to the Director, this role involves managing IT systems, infrastructure projects, and technical support.

The manager will be a technology professional with experience in IT management. Success is measured by the reliability of IT services, the efficiency of infrastructure projects, and the support provided to staff and students.

B

# Critical-Path Notes for the Implementation GANTT

ANNEXURE B · SUPPORTING MATERIAL

# Annexure B — Critical-Path Notes for the Implementation GANTT

The visual GANTT chart for the five-year implementation plan is in Chapter 6 (Fig. 6.1). The phased narrative — Phase 0 (pre-launch), Phase 1 (foundation), Phase 2 (scale), and Phase 3 (consolidation) — is set out in Chapter 6 §6.1 through §6.4. The complete milestone register, with twenty-five named milestones across the five-year period, is in Chapter 6 (Table 6.1) and is duplicated for cross-reference convenience in Chapter 4 (Table 4.2).

This annexure adds what the chart and the milestone register cannot: the critical-path dependencies, the schedule risks, and the institutional decisions that have to be made in sequence rather than in parallel.

## The five critical-path constraints

The Centre's implementation plan is governed by five sequence constraints. Each one is a dependency that cannot be parallelised, and each one therefore sets the lower bound on how quickly the Centre can reach a given milestone.

- **Constraint 1 — Section-8 incorporation must precede everything.** The Centre cannot sign MoUs, open bank accounts, hire staff, or receive grants until the Section-8 company is incorporated under the Companies Act 2013. Target: Q1 FY 2026-27. Slippage here is a one-for-one slippage of every subsequent milestone. Mitigation: the incorporation packet is pre-prepared; filing is initiated in the first week of operations.
- **Constraint 2 — Director appointment paces senior recruitment.** The Director's office is the hiring authority for the senior research and engineering positions; no senior recruit will join an institution whose Director has not been named. Target: Director announced by Q1 FY 2026-27, in office by Q2. Mitigation: the open search is initiated in parallel with the incorporation filing.

- **Constraint 3 — Anchor MoUs precede sustained research output.** The Centre's research credibility, particularly in the Bengali language-technology programme, depends on joint appointments and joint research arrangements with anchor institutions (IIT Kharagpur, ISI Kolkata, Jadavpur University, Visva-Bharati). These are time-consuming partner-side approvals. Target: at least three of four anchor MoUs signed by Q2 FY 2026-27. Mitigation: pre-engagement with partner-institution leadership begins in Year 0; standard MoU templates are pre-cleared with partner-institution legal departments.
- **Constraint 4 — Compute commissioning precedes model training.** Bengali LLM v1 cannot be trained at scale until the in-house GPU cluster is commissioned and the IndiaAI Mission compute-pool allocation is secured. Target: 8-node Phase-1 cluster operational by Q3 FY 2026-27. The cluster commissioning depends in turn on power-supply upgrades at the Siliguri facility, which depend on state-utility coordination. Mitigation: power-supply request initiated in Year 0; cloud-burst fallback in place for any commissioning slippage.
- **Constraint 5 — Bengali ASR Corpus v0.1 precedes Bengali LLM v1.** The 10,000-hour Bengali ASR corpus (Annexure I) is the foundational training input for both the speech models and the multilingual LLM. The corpus has a long-cycle dependency on the annotation team's hiring and training (Annexure A). Target: corpus v0.1 released 31 March 2027. Bengali LLM v1 release follows 15 months later (30 June 2028) — the gap reflects the model-training and evaluation cycle.

## Schedule-risk register

| ID | RISK | LIKELY IMPACT | MITIGATION |
|---|---|---|---|
| SR-01 | Section-8 incorporation delay (regulatory) | All downstream slippage | Pre-prepared packet; engaged counsel |
| SR-02 | Director search runs past Q2 | Senior recruitment pace slows | Interim Director-of-Operations in place from Q1 |
| SR-03 | Partner-MoU negotiations extend beyond Q2 | Research-output milestones (M-06, M-12) at risk | Cleared template; senior pre-engagement |
| SR-04 | GPU procurement under export controls | Cluster commissioning delay | Multi-vendor; IndiaAI pool primary fallback |
| SR-05 | Annotation team hiring stalls (rural Eastern India HR market) | Corpus v0.1 misses 31 March 2027 | Cohort-based annotator-training programme; gender-equity target ≥ 50% |
| SR-06 | District-hub MoUs stall at DM level | Outreach milestone M-11 slips | Pre-engagement with District Magistrates; phased rollout |

## Institutional decisions that must be made in sequence

Three institutional decisions are sequence-locked and are flagged here for the Governing Council's attention.

The Editorial Board for the Eastern India AI Index (Annexure M) must be constituted before the first Index edition's data-collection cycle begins. The Index pre-registration (Open Science Framework) commits the Centre to a May 2027 first edition; data collection therefore begins September 2026 at the latest; Editorial Board therefore constituted by August 2026.

The Open Asset Council (Annexure I) must be constituted before the first open-asset release. The first release (Bengali ASR Corpus v0.1) is scheduled for 31 March 2027; Council therefore constituted by November 2026 to permit a three-month review cycle on the inaugural release.

The Ethics Office and Data Protection Officer roles (Chapter 11 §11.5) must be filled before any field-survey or human-subjects research begins. The first field campaign (annotator-cohort recruitment in the five North Bengal districts) begins Year 1 Q1; DPO and Ethics Office therefore in place by Q4 FY 2025-26 — i.e., before Section-8 incorporation. The Founder is personally responsible for these two appointments in the pre-launch period.

C

# Year-1 Monthly Cash-Flow and Sensitivity Bridge

ANNEXURE C · SUPPORTING MATERIAL

# Annexure C — Year-1 Monthly Cash-Flow and Sensitivity Bridge

The consolidated five-year budget is Table 8.1 in Chapter 8. This annexure adds two artifacts the chapter table cannot accommodate: a monthly cash-flow forecast for fiscal year 2026-27 (the year in which the institutional anchor-capital trigger applies), and a sensitivity bridge between the Lean, Base, and Stretch scenarios already summarised in Table 8.1.

## C.1 Year-1 monthly cash-flow forecast (₹ in lakh, base scenario)

The Year-1 envelope is Rs 5,530 lakh (Rs 55.3 crore) under the base scenario. The Centre's Year-1 cash-flow is front-loaded for capital expenditure (real estate fit-out, GPU procurement, initial software licences) and ramps for operational expenditure as the founding cohort is hired.

| MONTH (FY 2026-27) | CAPEX OUTFLOW | OPEX OUTFLOW | RECEIPTS (GRANTS + CSR) | NET CASH | CUMULATIVE |
|---|---|---|---|---|---|
| April 2026 | 320 | 35 | 800 | +445 | 445 |
| May 2026 | 380 | 45 | 0 | -425 | 20 |
| June 2026 | 410 | 60 | 400 | -70 | -50 |
| July 2026 | 360 | 80 | 0 | -440 | -490 |
| August 2026 | 290 | 95 | 600 | +215 | -275 |
| September 2026 | 250 | 110 | 0 | -360 | -635 |
| October 2026 | 220 | 125 | 400 | +55 | -580 |
| November 2026 | 180 | 140 | 0 | -320 | -900 |
| December 2026 | 150 | 155 | 300 | -5 | -905 |
| January 2027 | 130 | 165 | 0 | -295 | -1,200 |
| February 2027 | 120 | 175 | 700 | +405 | -795 |
| March 2027 | 200 | 200 | 1,800 | +1,400 | +605 |
| **Year 1 total (₹ lakh)** | **3,010** | **1,385** | **5,000** | **+605** | — |

*Receipts assume the anchor State + Union convergence tranches landing in April, June, August, October, December, February, and a year-end CSR-and-grant cluster in March. The Centre maintains a working-capital line of credit to absorb the negative cumulative balance in months June through January.*

The Year-1 working-capital low point of approximately Rs 12 crore (January 2027) is the principal Year-1 treasury risk. Mitigation: the anchor co-investment is structured as a pre-committed multi-tranche release with State and Union timing pre-agreed, and a Rs 15-20 crore working-capital line of credit is held

with a partner bank (Bandhan Bank is the natural partner given regional alignment and the Centre's planned BFSI POC programme).

## C.2 Sensitivity bridge — Lean vs Base vs Stretch

Table 8.1 summarises the Lean (-18 percent), Base, and Stretch (+22 percent) scenarios as aggregate envelopes. The bridge below identifies the specific line items that move between scenarios — the Centre's principal sources of flexibility under varying funding conditions.

| BRIDGE ITEM | LEAN ADJUSTMENT | BASE | STRETCH ADJUSTMENT | RATIONALE |
|---|---|---|---|---|
| GPU / HPC cluster (Cat. 2) | -25% (rely more on IndiaAI subsidised pool) | Per Table 8.1 | +30% (in-house cluster doubled) | Compute-pool consumption strategy |
| District-hub operations (Cat. 10) | -20% (delay Y3-Y5 hubs) | Per Table 8.1 | +35% (accel-erate to 15 hubs by Y5) | Outreach pace |
| Salaries — research scientists (Cat. 4) | -15% (slower senior hiring) | Per Table 8.1 | +20% (addi-tional senior hires + international fellows) | Talent ramp |
| Training programmes (Cat. 8) | -10% | Per Table 8.1 | +25% (expanded teacher-training cohorts) | Training output |
| IP / patents / legal (Cat. 11) | -10% | Per Table 8.1 | +15% | Commercialisation pace |
| Travel + convening (Cat. 12) | -25% | Per Table 8.1 | +15% | Outreach intensity |
| Contingency (Cat. 15) | Held at 6% | 6% | Held at 6% | Discipline preserved across scenarios |
| Annotation + field data (Cat. 9) | Held — non-negotiable | Per Table 8.1 | Held — non-negotiable | Open-asset commitment ladder (Annexure I) is unchanged across scenarios |

The two line items deliberately held constant across all scenarios — contingency (6 percent reserve) and annotation/field data — reflect institutional commitments the Centre does not believe should flex with funding. Contingency discipline is a finance commitment to the Centre's auditors; the annotation and field-data commitments are pre-registered with the Open Science Framework and with the Index Editorial Board (Annexure M).

## C.3 Funding-mix targets by year

The aggregate funding-mix evolution is in Fig. 8.2 (stacked area chart). The discrete year-by-year targets the Centre commits to are below; the trend across years is the Centre's institutional commitment to declining grant dependence and rising earned-revenue share.

| FUNDING SOURCE | Y1 (%) | Y2 (%) | Y3 (%) | Y4 (%) | Y5 (%) |
|---|---|---|---|---|---|
| State Grant (WB + adjacent states) | 40 | 45 | 42 | 38 | 32 |
| Union Convergence (IndiaAI, MeitY, NITI Aayog) | 30 | 27 | 24 | 22 | 18 |
| Industry / CSR | 22 | 17 | 16 | 15 | 15 |
| Earned Revenue (IP licensing, programme fees, data services) | 8 | 11 | 18 | 25 | 35 |
| **Total** | **100** | **100** | **100** | **100** | **100** |

The 35 percent Year-5 earned-revenue target is the headline sustainability commitment the Centre makes in this Detailed Project Report and is monitored as KPI K-16 (Chapter 10, Table 10.1). The audited financial statements published annually are the public-accountability mechanism.

E

# Partner Institution Profiles

ANNEXURE E · SUPPORTING MATERIAL

# Annexure E — Partner Institution Profiles

## 1. Acharya Brojendra Nath Seal College, Jalpaiguri

Acharya Brojendra Nath Seal College, established in 1888, holds a prestigious place in the educational landscape of North Bengal. With a rich history of over a century, the college has been a instance of higher education in the region. It is affiliated with the University of North Bengal and has consistently maintained a high academic standard, holding a NAAC accreditation of B++ in its latest cycle. The institution's legacy is built on its commitment to providing quality education and supporting intellectual growth among its students.

Currently, the college enrolls approximately 4,500 students across various undergraduate and postgraduate programs. It offers a wide range of disciplines, including Arts, Science, and Commerce, with notable strengths in Physics, Chemistry, and Mathematics. The college is well-equipped with laboratories, a library, and other academic facilities that support its detailed educational offerings. Its faculty is dedicated to nurturing students' academic and personal development, ensuring they are well-prepared for future challenges.

In the SARGVISION AI Centre of Excellence (CoE), Acharya Brojendra Nath Seal College is set to play a crucial role in the implementation of the AI and Python skilling initiative. The college will serve as a hub for the Distributed Delivery Model, providing classrooms and lab facilities for the program. Its existing infrastructure and academic credibility make it an ideal partner for delivering SARGVISION-branded courses. The institution's faculty will be trained as Faculty Affiliates to deliver these courses, thereby building internal capacity and enhancing the college's educational offerings (NB CoE Proposal — Acharya Brojendra Nath Seal College Jalpaiguri).

Memoranda of Understanding (MoU) negotiations with Acharya Brojendra Nath Seal College are currently underway, with positive engagement from the college's administration. The discussions have been centered around the integration of the AI curriculum

into the existing academic framework and the potential benefits for both students and faculty. The college's leadership has expressed enthusiasm for the partnership, recognizing the value it brings in terms of academic enrichment and industry linkages.

## 2. North Bengal St Xavier's College, Siliguri

North Bengal St Xavier's College, located in Rajganj, Jalpaiguri district, is a Jesuit-tradition institution with a strong focus on academic excellence and social responsibility. Founded in 2007, it has quickly established itself as a leading educational institution in the region. The college is affiliated with the University of North Bengal and has received a NAAC accreditation of B+ in its most recent assessment. Its educational philosophy emphasizes holistic development, critical thinking, and community engagement.

The college has a student population of approximately 3,000, with a diverse range of programs in Arts, Science, and Commerce. It is particularly noted for its Computer Science and Business Administration courses, which have garnered a reputation for producing industry-ready graduates. The institution boasts modern facilities, including computer labs, a library, and an auditorium, which support its dynamic learning environment.

As a partner in the SARGVISION AI CoE initiative, North Bengal St Xavier's College will contribute its established academic infrastructure and student pipeline to the program. The college's faculty will participate in the train-the-trainer program, becoming SARGVISION Faculty Affiliates who deliver AI and Python courses. This collaboration aligns with the college's mission to provide practical, industry-relevant education and enhance its students' employability (NB CoE Proposal — North Bengal St Xavier's College).

The MoU discussions with North Bengal St Xavier's College have been progressing smoothly, with both parties committed to formalizing the partnership. The college's administration has been proactive in identifying ways to integrate the AI curriculum with existing courses and exploring avenues for student engagement. The partnership is expected to be mutually beneficial, expanding educational opportunities for students and strengthening the college's academic offerings.

## 3. Salesian College, Sonada

Salesian College, established in 1938, is a minority institution under the Salesians of Don Bosco. With campuses in Sonada and Siliguri, it is renowned for its commitment to quality education and social upliftment. The college is autonomous and affiliated with the University of North Bengal, holding a NAAC accreditation of A in its latest cycle. Its autonomy allows for flexibility in curriculum design and a focus on innovative teaching methodologies.

The college serves a student body of over 1,800, offering courses in Arts, Science, Commerce, and Vocational studies. Salesian College has a strong track record in technology education, with programs such as BCA and B.Sc Computer Science Honours. Its industry partnerships for cybersecurity and web development training further enhance its academic profile. The college's infrastructure includes current-generation computer labs, media facilities, and tech-enabled classrooms, providing a conducive learning environment.

In the context of the SARGVISION AI CoE, Salesian College is well-positioned to implement the full four-course AI and Python skilling program from the outset. Its autonomous status facilitates rapid integration of the SARGVISION curriculum, and its existing industry partnerships offer opportunities for collaborative capstone projects. The college's involvement in the CoE will enhance its educational offerings and provide students with current skills aligned with industry needs (NB CoE Proposal — Salesian College Sonada).

Negotiations for the MoU with Salesian College have been encouraging, with the college's leadership expressing strong support for the partnership. The discussions have focused on aligning the AI program with the college's strategic goals and drawing on its existing strengths in technology education. The collaboration is anticipated to significantly benefit students, faculty, and the broader community, reinforcing Salesian College's position as a leader in higher education.

## 4. Sukanta Mahavidyalaya, Dhupguri

Sukanta Mahavidyalaya, founded in 1981, is the largest partner college by enrollment, serving a predominantly rural catchment in Dhupguri. Affiliated with the University of North Bengal, the

college has a NAAC accreditation of B+ in its second cycle. It plays a vital role in providing accessible higher education to rural students, with a focus on inclusivity and community development.

The college has a student population of 6,591, offering programs across Arts, Science, and Commerce. Sukanta Mahavidyalaya is known for its strong Arts and Sciences faculties, which attract a large rural-student cohort. The institution's infrastructure includes computer labs, science laboratories, a library, and hostel facilities, supporting a detailed educational experience.

In the SARGVISION AI CoE initiative, Sukanta Mahavidyalaya is recommended to deploy Course 3 (AI Literacy, no coding required) college-wide as a Skill Enhancement Course (SEC). This approach aligns with the college's mission to broaden AI literacy and equip students with relevant skills for the evolving job market. The college's faculty will be trained to deliver these courses, building internal capacity and enhancing employability outcomes for students (NB CoE Proposal — Sukanta Mahavidyalaya).

MoU negotiations with Sukanta Mahavidyalaya are underway, with the college's administration demonstrating keen interest in the partnership. The discussions have highlighted the potential for integrating AI literacy into the existing curriculum and the benefits of industry alignment. The collaboration is expected to enhance the college's educational offerings and provide students with valuable skills that meet industry demands.

## 5. Acharya Brojendra Nath Seal College, Cooch Behar

Acharya Brojendra Nath Seal College in Cooch Behar, established in 1888, is one of the oldest and most respected educational institutions in the region. Affiliated with the University of North Bengal, the college has consistently maintained high academic standards, earning a NAAC accreditation of B++ in its latest cycle. Its long-standing tradition of excellence is reflected in its commitment to quality education and holistic student development.

The college enrolls approximately 5,000 students, offering a diverse range of programs in Arts, Science, and Commerce. It is particularly noted for its strengths in Humanities and Social Sciences, with well-established departments that attract students from across the region. The institution is equipped with modern facilities, including laboratories, a library, and a dedicated research center, which support its academic endeavors.

As a participant in the SARGVISION AI CoE, Acharya Brojendra Nath Seal College will contribute its academic infrastructure and faculty expertise to the initiative. The college will host SARGVISION courses, with faculty trained as Affiliates to deliver the AI curriculum. This partnership will enhance the college's educational offerings and provide students with industry-relevant skills that improve employability (NB CoE Proposal — Acharya Brojendra Nath Seal College Cooch Behar).

MoU negotiations with Acharya Brojendra Nath Seal College are progressing positively, with the college's leadership expressing enthusiasm for the collaboration. The discussions have focused on aligning the AI program with the college's academic objectives and exploring opportunities for student engagement. The partnership is expected to bring significant benefits to the college, its students, and the wider community, reinforcing its position as a leading educational institution in North Bengal.

F

# Founder, Director, and Advisory Board

ANNEXURE F · SUPPORTING MATERIAL

# Annexure F — Founder, Director, and Advisory Board

## Founder

**Abhishek Gupta** is the Founder of SARGVISION Intelligence Pvt. Ltd. and the principal author of this Detailed Project Report. He is an NVIDIA-Certified Professional in Agentic AI — among the first in India to hold this certification — and has served as Senior Data Scientist and Generative AI Lead at Ernst & Young, where he led the translation of complex AI systems into enterprise-grade deployments. He is an Erasmus Mundus scholar, selected as one of twenty-five candidates from a global pool of over eight hundred for the European Union's flagship academic-mobility programme.

His professional commitment to Eastern India's AI infrastructure draws on twelve months of primary field research across the five North Bengal districts profiled in this document and on continuous engagement with regional academic and industry stakeholders. He is personally accountable for the institutional commitments made in this report, signing them in the Preface (F1). The Founder's full curriculum vitae is available on request from the Centre's address in the colophon.

## Director — to be announced

The Centre's Executive Director will be appointed in Year 1 through an open international search overseen by the Governing Council. The search specification calls for a leader with deep technical expertise in artificial intelligence, machine learning, or natural-language processing (a doctorate or equivalent published research record); demonstrated experience building research or institutional teams from inception to at least 100 members; familiarity with the Indian higher-education and government-procurement landscape; and the cultural fluency to engage substantively with the Bengali-speaking and Eastern Indian communities the Centre serves. Familiarity with Bengali is preferred but not

mandatory. Compensation is benchmarked at the IIT Director scale (approximately ₹60-90 lakh per annum) with appropriate institutional incentives. The search will be conducted with the assistance of an executive-search firm with experience placing senior academic-research leadership.

## Governing Council — to be constituted

The Section-8 company will be governed by a Governing Council whose composition is set out in the Articles of Association. The Council will include the Founder; the Executive Director (ex officio); not fewer than three independent members representing academia, industry, and civil society respectively; and the Editorial Board chair of the Eastern India AI Index (ex officio, non-voting). The Governing Council will be constituted within ninety days of Section-8 incorporation, with announced membership.

## Advisory Board — composition principles

The Advisory Board will be constituted within the first nine months of operations and is expected to comprise seven distinguished members across five constituencies: academic (a serving or recently retired Vice-Chancellor and a distinguished AI research scientist); industry (a NASSCOM-member CEO and a regional MSME leader); government (a retired Principal Secretary with an education or IT portfolio); civil society (a Bengali-language scholar of national standing); and central government (a senior official from the Ministry of Electronics and Information Technology or the IndiaAI Mission Secretariat, ex officio). The Advisory Board's role is consultative; binding governance authority rests with the Governing Council.

Named appointments to the Director's office, the Governing Council, and the Advisory Board will be announced publicly as confirmations are received, on the Centre's website and through the press list set out in Chapter 12 §12.4.

G

# Infrastructure Strategy Principles

ANNEXURE G · SUPPORTING MATERIAL

# Annexure G — Infrastructure Strategy Principles

The Centre's physical and computational infrastructure is sized to the institutional commitments set out in Chapter 1 §1.3 and to the headcount ramp in Annexure A. The detailed equipment list, vendor master, and procurement schedule are operational artifacts maintained by the Head of Operations; this annexure sets out the principles that govern those artifacts.

## Physical facility — phased build

The Centre's physical anchor is a 10 to 15 acre campus near Bagdogra airport in Siliguri, under a subsidised lease from the Government of West Bengal under standard industrial-policy instruments (the formal ask is specified in Chapter 1 §1.4 and reiterated in Chapter 8 §8.4). The campus is delivered in three phases, sequenced to the headcount ramp:

- **Phase 1 (FY 2026-27).** Approximately 8,000 sq ft of operational space — research laboratories, training classrooms, administrative offices, and a small data-centre room — sufficient for the Year-1 sanctioned headcount of approximately 71.
- **Phase 2 (FY 2028-29).** Expansion to approximately 20,000 sq ft, including a dedicated training auditorium and the first district-hub coordinating office. Sufficient for the Year-3 sanctioned headcount of approximately 227.
- **Phase 3 (FY 2030-31 and beyond).** A purpose-built permanent campus of approximately 50,000 sq ft, with an expanded high-performance computing facility, conference and convening space, and a residential complex for visiting scholars and rotational industry fellows.

A satellite office in Kolkata operates from Year 1 to support institutional partnerships, government engagement, and the Editorial Board of the Eastern India AI Index (Annexure M).

## Compute infrastructure — hybrid model

The Centre's compute strategy is deliberately hybrid. It consumes from three pools, each appropriate to a different workload class.

- **An in-house GPU cluster** sized to the Centre's continuous research and engineering workload. Phase 1 cluster: eight high-end accelerator nodes (NVIDIA H100-class or successor, depending on procurement timing and export-control conditions). Phase 3 cluster: approximately 32 nodes. The in-house cluster anchors training, evaluation, and fine-tuning of the Bengali language-model family (Annexure I).
- **The IndiaAI Mission's national compute pool** for training and burst-capacity workloads, accessed under the Mission's subsidised-rate framework. This pool is the primary capacity for very-large training runs and for time-critical evaluation work. Annual consumption is forecast and pre-booked through the Mission's quarterly allocation cycle.
- **Commercial cloud capacity** (AWS, Azure, or successor providers) for elastic workloads — inference serving, evaluation suites, district-hub-facing applications, and short-duration burst training. The commercial layer is configured for in-region (Mumbai, Hyderabad) deployment to satisfy data-residency requirements under the Digital Personal Data Protection Act 2023.

The three-pool model is the Centre's principal hedge against compute-supply risk (R-02 in the risk register, Chapter 9). At no point does the Centre rely on a single provider or a single geography.

## Data infrastructure

Primary storage of approximately 200 TB in Phase 1, expanding to a multi-petabyte tiered storage architecture by Phase 3. All training-data tiers are encrypted at rest. The annotation pipeline operates on isolated infrastructure with strict access logging. Public open-asset releases (datasets and model weights in Annexure I) are mirrored at three locations: the Centre's primary repository, the Hugging Face Hub, and a partner-institution academic mirror at the Indian Statistical Institute Kolkata or IIT Kharagpur AI4ICPS dataset registry.

## Network and reliability

The Siliguri facility operates on a dedicated 1 Gbps leased line with diverse-path redundancy. The Kolkata office maintains synchronous connectivity. Both facilities maintain 24×7 NOC monitoring and a quarterly disaster-recovery test cycle, with the test results archived in the Centre's published operational record.

## Procurement — principle, not list

Equipment procurement is conducted through the Government e-Marketplace (GeM) platform where applicable, supplemented by direct procurement under standard institutional rules for items not available on GeM. The Centre maintains a pre-cleared vendor master, refreshed annually, to compress procurement-cycle times. The full equipment list, with quantities, specifications, GeM categories, and estimated unit costs, is maintained as an operational artifact by the Head of Operations and is available for inspection by the Centre's Audit Committee and by the Governing Council.

I

# Open Asset Commitment Ladder

ANNEXURE I · SUPPORTING MATERIAL

# Annexure I — Open Asset Commitment Ladder

The institutional credibility of an artificial intelligence centre is read, in 2026, largely through the assets it ships into the public commons. AI4Bharat's standing in the global Indic-language NLP community rests on the IndicNLP datasets, the IndicBERT family of pretrained models, and the IndicTrans translation suite — all released openly through 2020-2024. The Center for the Study of Language and Information at Stanford built its institutional reputation across the 1980s and 1990s through the open release of treebanks, parsers, and corpora that became the foundational infrastructure for the field. The Allen Institute for AI's standing rests substantially on its open weights for OLMo, Tulu, and the AI2 dataset releases.

This Centre commits to a comparable institutional posture from Year 1. What follows is the timed open-asset commitment ladder. Each item names the asset, its license, the projected size, the publication date, the custodial mirror, and the responsible team within the Centre. These are commitments, not aspirations; the leadership signing this document is willing to be held to the calendar.

## The five-year release calendar

*Year 1 — fiscal year 2026-27*

- **Bengali ASR Speech Corpus, version 0.1.** A 10,000-hour audio corpus of Bengali speech, recorded across all 23 West Bengal districts to capture dialectal variation, with aligned transcripts in standard Bangla orthography. Recording subjects span school students, college students, working-age adults, and seniors in approximately equal proportion. Licence: Creative Commons Attribution 4.0 International (CC-BY-4.0). Custodian team: Lead Bengali NLP Scientist + 4-person dataset annotation team. Mirror sites: Hugging Face Hub (primary), ISI Kolkata academic archive, and IIT-KGP AI4ICPS dataset registry. Projected release: 31 March 2027.

- **Bangla-Santali Parallel Translation Pairs.** A corpus of 250,000 high-quality parallel sentence pairs in Bangla (source) and Santali (target), with annotation for register (formal, conversational, literary). Source texts drawn from government public-domain publications, Bengali literary classics, and consented contemporary social-media corpora. Santali translations produced through partnership with the Santal Pargana Cultural Council and validated by a panel of four Santali linguists. Licence: Creative Commons Attribution-ShareAlike 4.0 International (CC-BY-SA-4.0). Projected release: 30 June 2027.

- **BanglaBench v0.1.** A small initial evaluation suite for Bengali NLP, covering eight tasks: tokenisation accuracy, named-entity recognition, sentiment classification, syntactic parsing, machine translation (Bangla-English), reading comprehension, summarisation, and dialect identification. Designed for academic benchmarking; not yet at the scale that would support model training. Licence: CC-BY-4.0. Projected release: 30 September 2027.

*Year 2 — fiscal year 2027-28*

- **Sovereign Bengali Large Language Model, version 1 (Bengali LLM v1).** A 7-billion-parameter decoder-only transformer model, trained on a Bengali-majority multilingual corpus (Bangla, English, Hindi, Santali, Rajbanshi, Nepali) of approximately 400 billion tokens. The model card will disclose training data composition, compute used, evaluation results on BanglaBench and standard multilingual benchmarks, and known limitations. Weights released under the Apache License 2.0. Inference code and a fine-tuning starter kit released alongside. Custodian: Centre Director's office, with the Bengali NLP scientific team responsible for model maintenance. Mirror: Hugging Face Hub. Projected release: 30 June 2028.

- **BanglaBench v1.0.** The full version of the Bengali evaluation suite, extended to twenty tasks covering reasoning, code generation, multilingual mixing, cultural knowledge, and Bengali-literature comprehension. Designed to be the reference benchmark for the field. License: CC-BY-4.0. Projected release: 30 September 2028.

- **BanglaSpeech Evaluation Suite.** A standard evaluation suite for Bengali automatic speech recognition (ASR) and text-to-speech (TTS), including dialectal subsets, noise-condition subsets, and

code-switched (Bangla-English) subsets. License: CC-BY-4.0. Projected release: 31 December 2028.

### *Year 3 — fiscal year 2028-29*

- **Rajbanshi Language Corpus and Lexicon.** A foundational corpus of approximately 5,000 hours of Rajbanshi speech and a derived 50,000-entry lexicon. Produced through partnership with the Rajbanshi Cultural Council and the linguistic faculty at North Bengal University. The Rajbanshi tradition is undocumented digitally at any meaningful scale; this is the Centre's first major contribution to its digital preservation. Licence: CC-BY-SA-4.0. Projected release: 30 June 2029.
- **Kamtapuri Language Resources.** A parallel dataset for Kamtapuri (a North Bengal regional language closely related to Rajbanshi but with distinct phonology), comprising approximately 2,000 hours of speech and a 30,000-entry lexicon. Licence: CC-BY-SA-4.0. Projected release: 30 September 2029.
- **Eastern India Agriculture Imagery Dataset (EastAgri-v1).** A dataset of approximately 1.2 million high-resolution annotated images covering tea (Darjeeling and Dooars), jute (Murshidabad, Nadia, North 24 Parganas), rice paddies (Bardhaman, Hooghly, Bankura), and inland fisheries (East Midnapore, Howrah, South 24 Parganas). Annotations include crop type, pest presence, growth stage, soil-visibility characteristics, and irrigation pattern. Licence: CC-BY-4.0 with a non-commercial-redistribution rider for the imagery itself; derived annotations are CC-BY-4.0 without restriction. Projected release: 31 December 2029.

### *Year 4 — fiscal year 2029-30*

- **Domain Bengali Large Language Models.** Three 13-billion-parameter models fine-tuned from a Bengali LLM v2 base, each specialised for a domain: HealthBangla (medical informatics, health communication, public-health record summarisation), GovBangla (government-document understanding, citizen-services chatbot foundations, policy summarisation), and AgriBangla (agricultural advisory, crop-pest reasoning, market-price forecasting). All three released under the Apache License 2.0. Each model is accompanied by a domain evaluation suite. Projected release: rolling, with HealthBangla in June 2030, GovBangla in September 2030, and AgriBangla in December 2030.

- **Bengali Scientific Vocabulary Atlas.** A computational lexicon mapping the modern Bengali scientific vocabulary, including the late-nineteenth-century coinages by Bengali intellectuals who consciously constructed a scientific register in the language. Approximately 80,000 entries, with English glosses, etymological notes, and contemporary usage frequency from a one-billion-word Bangla web corpus. Licence: CC-BY-4.0. Projected release: 31 March 2031.

*Year 5 — fiscal year 2030-31*

- **BharatBangla Evaluation Framework.** A jointly-curated evaluation framework for Indic-language LLMs, developed in consortium with AI4Bharat (IIT Madras), Sarvam AI's institutional successor, and at least one other Indic-language institution. Covers Bangla, Hindi, Tamil, Telugu, Marathi, Gujarati, Punjabi, Malayalam, and Kannada. License: CC-BY-4.0. Projected release: 30 September 2031.
- **Bengali LLM v2 weights (open release).** The successor to Bengali LLM v1, projected at 30-40 billion parameters, trained on an expanded multilingual corpus and an improved instruction-following pipeline. Apache License 2.0. The third-generation model (v3) is intended to remain proprietary to the Centre for an initial six-month commercial window before being released openly. The Centre commits to a permanent two-major-version offset between proprietary and openly-released weights — that is, while v3 is proprietary, v1 is openly available. Projected v2 release: 31 December 2031.

## License governance

An Open Asset Council, chaired by the Centre's Head of IP & Commercialisation and including the Lead Bengali NLP Scientist, the Director of the Outreach Pillar, an external legal advisor specialising in open-source licensing, and a rotating diaspora-community representative, reviews and approves each release. The Council meets monthly. Its meeting minutes are part of the Centre's published institutional record.

All datasets released under CC-BY or CC-BY-SA require contributor agreements compliant with the Digital Personal Data Protection Act, 2023, §7 (consent regime); the Council reviews

compliance for every release. Speech corpora drawn from human subjects require additional informed-consent documentation, retained by the Centre's Ethics Office.

## Custodial commitments

Every asset released is mirrored at no fewer than three locations: the Centre's primary repository (sargvision.in/data), the Hugging Face Hub, and at least one academic mirror at ISI Kolkata or IIT-KGP. The intention is that asset availability cannot be lost through a single institutional failure.

The Centre also commits to long-term preservation: each major asset release is deposited with the National Digital Library of India and with the Internet Archive, with a target retrieval guarantee of fifty years.

## What this commitment buys the Centre

A timed open-asset ladder of this kind buys the Centre four things that no closed-research posture can buy.

- It buys the Centre **alignment with the IndiaAI Mission's Bharat Datasets pillar** — the Centre is contributing exactly what the Mission was set up to acquire. The Mission's Bharat Datasets Platform is the natural home for the speech corpora, the parallel translation pairs, and the evaluation suites.
- It buys **research-community standing**. Open weights and open evaluation suites are the currency by which academic AI work establishes itself in the global community. A Centre that ships these is read by international peers as a serious institution.
- It buys **insurance against geopolitical restriction**. A Centre that is recognised globally as a producer of high-quality regional-language AI assets has access to the global research community even in a regime where compute or model weights become export-controlled.
- And it buys **credibility with Indian industry**. MSMEs, mid-cap technology firms, and government departments looking for Bengali-language AI tooling will find this Centre by finding its open assets. The pipeline from open dataset to industry consulting engagement is the foundation of the Centre's earned-revenue trajectory described in Chapter 8.

The Centre's leadership signs this annexure as a binding institutional commitment.

J

# The Centre at Year 25

ANNEXURE J · SUPPORTING MATERIAL

# Annexure J — The Centre at Year 25

In June 2050, a Joint Secretary in the Ministry of Electronics and Information Technology will commission a twenty-fifth-anniversary review of the SARGVISION AI Centre of Excellence for Eastern India. The review will be addressed to a successor to the current Secretary, a successor to the current Director of the Centre, and to a Governing Council whose median age in 2050 will be thirty-eight. Most of the people writing the Centre's first five-year plan in 2026 will be retired or have moved into emeritus advisory roles. The Centre itself will be operating at a scale and in a register that the founding document only gestured at.

This annexure is the founding document's attempt to describe what that 2050 review will find. It is deliberate forecast, not aspiration, and the founders are willing to be measured against it.

## The 2050 institutional profile

By the close of fiscal year 2050-51 (March 2051), the Centre is projected to have:

- An installed cohort of approximately **6,000 Eastern-India-based AI practitioners** — researchers, engineers, programme directors, fellows, students in advanced cohorts — working across the five-state region (West Bengal, Bihar, Odisha, Jharkhand, and Assam). The 2026 baseline for this metric, drawn from the West Bengal Industry Research Report (2025), is approximately 500.
- An annual operating budget of approximately **₹600 crore**, of which **60 percent is self-generated** through IP licensing, certified training programmes, sponsored research, and the Centre's endowment yield. In 2026, the comparable figure is ₹95 crore, of which 100 percent comes from State and Central grant.

- An **endowment of approximately ₹1,700 crore (~ $200 million)**, accumulated through Corporate Social Responsibility contributions, IP licensing revenue, and philanthropic gifts from the Bengali diaspora. The endowment is held in a Section-25 trust structure separate from the operating Section-8 company.
- A track record of **roughly 150 spin-out companies**, of which approximately 30 will have crossed ₹100 crore in annual revenue. Cumulative direct employment created in these spin-outs is projected at 45,000-55,000 positions in Eastern India.
- **Joint academic affiliations with four international institutions** — projected to be Stanford HAI, ETH Zurich's AI Center, the National University of Singapore School of Computing, and the Cambridge Mathematics of Artificial Intelligence Initiative — through a structured visiting-faculty programme and joint doctoral training.
- The role of **lead institution** in a national consortium for South Asian regional-language AI, with formal affiliations to Sarvam AI's institutional successor, AI4Bharat's IIT Madras hub, and the proposed BBIN Language Technology Working Group (covering Bangladesh, Bhutan, India, Nepal).

## What the indicators will show

Five quantitative trajectories anchor the 25-year picture.

— **Eastern India's share of the national AI workforce** is projected to move from 8 percent in 2026 (NASSCOM AI Adoption Index 2024, illustrative) to 25 percent by 2050. The mechanism is straightforward: the Centre's training pillar feeds an average of 800-1,200 net new AI-skilled professionals into the region annually, of whom approximately two-thirds stay within Eastern India for at least their first decade of work. Karnataka's share over the same period is projected to decline from 28 percent to 19 percent as the centre of gravity moves east.

— **Bengali AI infrastructure** is projected to reach a state where the eighth major iteration of the Sovereign Bengali Large Language Model — referred to here as Bengali LLM v8 — is the reference model for Bangla natural language processing globally, used by an estimated 230 million Bengali speakers across India, Bangladesh, and the diaspora. The eighth iteration's specific properties cannot be specified in 2026; what can be specified is

the institutional commitment to ship one major version every three years, with the second-to-last (v7) published openly and the latest (v8) at production scale.

- **Self-generated revenue share** is projected to cross 60 percent by fiscal year 2049-50. The four revenue lines that mature into this figure are: certified training programmes priced for industry, sponsored research from named corporate partners (TCS, Reliance, ITC, Tata Sons, Wipro), IP licensing in agritech-AI and healthtech-AI verticals, and yield from the philanthropic endowment.
- **Geographic reach** is projected to cover all 23 West Bengal districts (achieved by Year 5 of the founding plan), plus active programmes in 18 districts of Bihar, 12 districts of Odisha, 10 of Jharkhand, and 14 of Assam — together a 77-district active footprint by 2050. Active programmes here means at least one cohort of trainees per year, at least one outreach lab presence, and at least one industry-partner project in the district.
- **Spin-out impact** is projected at approximately 150 founded companies, with a survival rate at year-ten of 35-40 percent. The 30 surviving companies that cross ₹100 crore in annual revenue are projected to create approximately 45,000-55,000 direct jobs in Eastern India and a multiple of that in indirect employment.

## What could derail this

The founders' candid view is that the 25-year picture above is conditional on five things going right and at least three of them not going wrong.

- **Funding continuity through political transitions.** The Centre is being founded under a particular State and Central government configuration. Three to four political cycles will pass before the 25-year horizon arrives. The mitigation is the endowment, which crosses ₹1,000 crore by approximately Year 12 and provides counter-cyclical funding when government priorities shift. The risk is that an early-stage funding crisis (Years 1-7) damages the institutional momentum that the endowment was meant to defend.
- **Faculty retention.** The Centre's projected output is dependent on retaining roughly two-thirds of the senior research talent it hires. The risk is that the global AI labour market — which in 2026 already pays principal investigators at frontier labs many

multiples of Indian academic salaries — continues to pull senior people away. The mitigation is the diaspora engagement strategy (Annexure K) and the joint-affiliation arrangement with international institutions, which allows senior researchers to maintain global affiliations while anchoring at the Centre.

- **Tech-platform obsolescence.** The Centre's technical architecture (Chapter 4) is being designed for an era of GPU-bound transformer-architecture models with English- and Chinese-pretrained foundations. A platform shift — to substantially different model architectures, or to a regime where compute is no longer the binding constraint — would require the Centre to migrate its talent and infrastructure base. The 25-year picture assumes one such migration, planned and executed between approximately Year 8 and Year 12.
- **Geopolitical AI restrictions.** The 25-year picture assumes that India retains substantial access to global compute and to the open-source frontier-model ecosystem. A regime where high-end GPUs become export-controlled, or where open-source models cease to be released, would constrain the Centre's research pillar and could push it toward a sovereign-only posture. The Centre's open-asset commitment (Annexure I) is partly an insurance against this: by becoming a known producer of regional-language AI datasets and weights, the Centre earns access to a global community even under restriction.
- **Cultural identity drift.** A scaled institution risks losing the specificity that made it interesting. The 2050 review must find a Centre that is still recognisably Bengal-rooted in its programmes, partners, language priorities, and intellectual register. The risk is that twenty-five years of growth pushes it toward becoming another generic IIT-adjacent AI hub. The mitigation is the cultural-inheritance argument set out in Annexure K, embedded in the Centre's governance through a permanent Cultural Stewardship subcommittee of the Governing Council.

## Three opportunities the founders are betting on

- **Bengali as a foundational NLP language.** In 2026, eleven of the world's twenty most-spoken languages do not yet have production-grade large language models. By 2050, that gap will have closed for most of them. Bengali is the seventh-most-spoken language globally and the first South Asian language outside Hindi

to plausibly support a billion-token-corpus training pipeline. The Centre that establishes itself as the principal Bengali NLP institution in the first decade is positioned to define the field for the rest of the century.

- **Eastern India agritech-AI export.** The agricultural belts of West Bengal, Bihar, Odisha, and Bangladesh share crop systems, monsoon patterns, soil types, and rural-economic structures that diverge significantly from the western and southern Indian agricultural landscapes. A Centre that develops AI tools tuned to these specific systems — tea, jute, rice, pulses, and inland fisheries — is positioned to export those tools across South and Southeast Asia.
- **Diaspora-anchored institutional building.** The Bengali academic and technology diaspora is an institutional resource that has not yet been organised at scale. The Centre's diaspora engagement strategy (Annexure K) is the founders' bet that this can be done, and that it will compound over twenty-five years into a material institutional advantage.

## A note to the 2050 review committee

The founders of 2026 cannot anticipate which of these projections will look modest in retrospect and which will look optimistic. The point of the projection is not its accuracy. The point is that the founding document committed itself to a measurable picture twenty-five years ahead, and that subsequent leadership of the Centre can be held to that picture, can revise it openly, and can ship an annual update that compares progress to it.

The first such revision will be published in the Centre's annual report at the close of fiscal year 2030-31, and every five years thereafter. The 2050 review committee will have five such revisions to read alongside this annexure.

The bet this institution makes is not on a particular technology or a particular partnership. It is on the proposition that Eastern India, a region with a measurable lag in artificial intelligence capacity in 2026 and an intellectual inheritance that goes back to the Bengal Renaissance, can in twenty-five years become the regional anchor for South Asian regional-language AI. That bet is being made deliberately, with the institutional, financial, and human capital described in the chapters above, and the founders are signing their names to it.

K

# Bengal's Intellectual Inheritance and the AI Centre

ANNEXURE K · SUPPORTING MATERIAL

# Annexure K — Bengal's Intellectual Inheritance and the AI Centre

In 1931, Prasanta Chandra Mahalanobis published a four-page paper in the *Journal of the Royal Statistical Society* defining what is now called the Mahalanobis distance — a statistic that remains, ninety-five years later, foundational to every modern multivariate analysis, every covariance-based outlier detector, and every Mahalanobis-Taguchi quality method used in industrial machine learning. He wrote it from a laboratory at Presidency College in Calcutta. He was thirty-eight. The Indian Statistical Institute, which he founded the same year, would in the decades that followed train C.R. Rao, Debabrata Basu, Raj Chandra Bose, J.B.S. Haldane, and through Haldane the entire post-war Indian generation of mathematical statisticians.

The SARGVISION AI Centre of Excellence does not claim continuity with Mahalanobis by accident of geography. It claims an institutional inheritance.

## A tradition this Centre means to extend

Bengal's contribution to the foundations of modern computation and statistics is documented in a way few other regions of India can claim. The tradition is roughly continuous from the late nineteenth century to the present, with three peaks and one long latent period.

- **The first peak, roughly 1895 to 1945**, contained Jagadish Chandra Bose, who in 1895 demonstrated radio-wave communication in a public lecture in Calcutta, two years before Marconi's transatlantic transmission; Satyendra Nath Bose, whose 1924 paper sent to Einstein gave the world Bose-Einstein statistics and the particle class now called bosons; Meghnad Saha, whose ionisation equation underwrites modern astrophysics; and Mahalanobis at ISI Kolkata, whose work shaped the design of the Second Five Year Plan (1956-61) and through that the entire post-independence Indian developmental-economics tradition.

- **The second peak, roughly 1955 to 1995**, was the era in which Bengal's scientific establishment incubated talent that scattered globally. Amartya Sen, born in Santiniketan in 1933, took the Nobel Prize in Economics in 1998 for work that began at the Delhi School of Economics and the London School of Economics but that drew its instincts from his Bengali academic upbringing. C.R. Rao, who had been Mahalanobis's first PhD student, became one of the twentieth century's most-cited statisticians. Subir Sachdev, who took his early training in Bengali academic networks, is today at Harvard, holds the Lars Onsager Prize, and is one of the world's leading condensed-matter theorists. Asoke Sen, at the Harish-Chandra Research Institute, is among the most-cited string theorists in the world.
- **The third peak, the one this Centre is being founded within, is from roughly 2010 to the present.** Abhijit Banerjee took the Nobel Prize in Economics in 2019. Manindra Agrawal — whose academic ancestry traces through Bengali mathematical networks at IIT Kanpur — won the Gödel Prize in 2006 for the AKS primality test. The Bengali-academic presence in computer science, statistics, machine learning, and applied mathematics across MIT, Stanford, Harvard, Berkeley, Oxford, Cambridge, ETH Zürich, and the Max Planck institutes is, in 2026, the largest single-state Indian academic presence in the world after Tamil Nadu and Karnataka.
- **The latent period, roughly 1995 to 2015**, is the gap this Centre is being founded to close. Through those twenty years, Bengal's intellectual energy continued to produce extraordinary talent — but that talent left. The institutional infrastructure that should have anchored it locally was not built. The Indian Statistical Institute remained, but it specialised; the IITs were elsewhere; the global AI lab landscape did not include a Bengal node.

  This Centre is not founding a tradition. It is recovering one.

## Why this matters institutionally

A serious institution claims a lineage publicly, and the lineage shapes how it is read. Stanford University claims Leland and Jane Stanford and the post-Gold-Rush California intellectual settlement. The Indian Statistical Institute claims Mahalanobis. The Tata Institute of Fundamental Research claims Homi Bhabha and

the post-war Indian scientific establishment. The lineage is not biographical filler. It is the institutional case for why this institution, in this place, has any claim on the resources it is asking for.

The Centre's case rests on three propositions about Bengal's inheritance.

- **First, that a measurable tradition of foundational quantitative work exists in Bengal**, and the absence of an institutional anchor for it locally between 1995 and 2015 was a historical accident produced by the decay of post-war Indian academic infrastructure — not by any failure of the tradition itself.
- **Second, that the Bengali academic diaspora has continued to expand globally** during the latent period, and now constitutes — in the founders' rough estimate — somewhere between three thousand and four thousand active AI/ML, computer science, applied mathematics, and statistics researchers in academic and industry positions outside India. The Centre's diaspora engagement strategy (set out below) is the institutional mechanism by which that distributed capacity becomes a local asset.
- **Third, that Bengali as a language with 230 million speakers, with a literary tradition since the eleventh century, and with a peculiar relationship to formal mathematical and scientific vocabulary** (much of the vocabulary was coined in the late nineteenth century by Bengali intellectuals consciously constructing a scientific register in the language) is uniquely positioned among Indian regional languages to support a serious computational-linguistics program. The Sovereign Bengali Large Language Model is therefore not a side-project of the Centre; it is the technical expression of the cultural argument.

The institutional reading is straightforward: a Centre of Excellence in artificial intelligence sited in Eastern India in 2026 is not establishing a new field. It is operationalising an inheritance that has been latent for thirty years.

## The Bengali Diaspora Engagement Strategy

The Centre's diaspora strategy is built on the proposition that the ~3,500 Bengali AI/ML researchers currently active in academic and industry positions outside India represent the largest single institutional resource available to the Centre, and that none of the existing Indian AI institutions has organised it.

- **The Diaspora Council.** Fifteen members appointed for renewable three-year terms, of whom at least twelve must be currently working outside India. The Council meets twice a year (one virtual, one in Kolkata or Siliguri), advises the Centre on research direction, recommends sabbatical visitors, and serves as a referral pipeline for hiring and project collaboration. Inaugural members are to be nominated by the Founding Director and ratified by the Governing Council in Year 1.

- **The Sabbatical Fellow Programme.** Up to ten Bengali researchers per year, recruited from international academic and industry positions, spend six to twelve months at the Centre in salaried fellow positions. Honoraria are calibrated to make the financial cost of the sabbatical neutral. Fellows are required to (a) deliver a public lecture series, (b) mentor at least two local research students, and (c) co-author at least one open dataset, evaluation suite, or peer-reviewed paper anchored at the Centre. Cost in Year 5: approximately ₹3-4 crore per year. Return: institutional access to research networks that would otherwise require a generation of conventional career-track investment.

- **The Return-to-Bengal Senior Fellowship.** Two to four senior researchers per cohort, recruited from settled international positions, take a one- to two-year placement at the Centre. The Fellowship covers salary, relocation, family education, and research budget. The expectation is that approximately one in three Return Fellows converts into a permanent senior-research position at the Centre, with the remainder returning to their international positions enriched by the connection. This is the Centre's primary mechanism for permanent senior-faculty acquisition; conventional hiring will not work in the global AI labour market of 2026-2050.

- **The Annual Diaspora Symposium in Kolkata.** Every February, a three-day symposium that brings approximately fifty diaspora researchers and two hundred local researchers and students together. The format is half technical (research presentations, working sessions) and half institutional (advising on Centre direction). The first instance is to be held in February of Year 2 (FY 2027-28).

- **The Joint Authorship Initiative.** A standing matching programme that pairs local researchers and graduate students with diaspora mentors for specific projects, with structured authors-

hip credit and publication targets. Initial annual cohort: forty pairs. Expansion target by Year 5: one hundred and twenty pairs annually.

These five mechanisms together constitute the Centre's bet on the diaspora as the principal mechanism for institutional capacity-building over the first ten years.

## A measured caveat

The argument that Bengal has a continuous intellectual inheritance from Mahalanobis to the present can be made too strongly, and the founders are aware of the temptation. The tradition is real but it is not unbroken. The latent period between 1995 and 2015 was long. The contemporary Bengali academic establishment, taken on its own, would not by itself sustain the institution this Centre describes. The inheritance argument works only if the Centre actively recruits the diaspora that has carried the tradition forward outside India, and only if the Centre is honest about the fact that it is building, not merely continuing.

The Centre is also conscious that an over-cultural framing risks reading as ethnic claim rather than institutional argument. The intellectual inheritance is the institution's specific cultural anchor — not its identity. The Centre's programmes, its hiring, its student admissions, and its partnerships are open and merit-based on every conventional measure. The Bengali inheritance is the reason the Centre exists in Bengal and not elsewhere; it is not a criterion for who works there.

## What this annexure commits to

The founding leadership commits to four things on the basis of this annexure.

- **First**, that the Centre's published institutional materials — annual reports, foundational essays, partner-facing communications — will name and engage the inheritance described above, rather than treating Bengal's intellectual history as background colour.
- **Second**, that the Diaspora Engagement Strategy described above will be the Centre's first major institutional initiative in Year 1, with the Diaspora Council appointed by the end of fiscal year 2026-27.

- **Third**, that the Centre will maintain a permanent Cultural Stewardship subcommittee of the Governing Council, chaired by an academic with primary credentials in Bengali intellectual history, whose mandate is to assess every five years whether the Centre is still recognisably Bengal-rooted in its programmes, partners, and intellectual character.
- **Fourth**, that the annual Diaspora Symposium will be held in February of every year beginning Year 2, and that the symposium proceedings will be published openly as part of the Centre's institutional record.

If these four commitments are honoured, the Centre will have done what no Indian AI institution has yet done: it will have made an explicit institutional bet on a specific regional intellectual tradition, and it will have built the operational structures to make that bet pay out.

If they are not honoured, the Centre will be another generic CoE, located in Bengal by accident of policy. That is the failure mode this annexure is written to prevent.

L

# Response and Engagement Pack

ANNEXURE L · SUPPORTING MATERIAL

# Annexure L — Response and Engagement Pack

This annexure exists to convert intention into action. A reader who finds the case in this Detailed Project Report compelling — across any of the audience paths set out in the Reader's Guide (F3) — should be able to log a concrete response within a single working week. The pages that follow set out the named contacts, the convening calendar, the commitment menu by audience, and the standard response templates the Centre is prepared to receive.

The Centre's institutional commitment is that every substantive response logged through the channels below will receive a written acknowledgement within seven working days and a substantive reply within thirty working days. Responses received before publication of v1.1 of this Detailed Project Report (within ninety days of v1.0 release) will be reflected, where the respondent grants permission, in the named institutional partners and Editorial Advisory Council listed in v1.1.

# L.1 Named contacts

| CHANNEL | PERSON / ROLE | ADDRESS |
|---|---|---|
| Founder, direct line | Abhishek Gupta, Founder, SARGVISION Intelligence Pvt. Ltd. | founder@sargvision.in |
| Partnerships and CSR | (Director of Partnerships, to be appointed Q1 FY 2026-27) | partnerships@sargvision.in |
| Academic and research | (Chief Research Officer, to be appointed Q1 FY 2026-27) | research@sargvision.in |
| Press and communications | (Head of Communications, to be appointed Q2 FY 2026-27) | press@sargvision.in |
| Index Editorial Board | (Chair, Eastern India AI Index Editorial Board, to be constituted August 2026) | index@sargvision.in |
| General enquiries | Office of the Founder | enquiries@sargvision.in |
| Postal address | SARGVISION Intelligence Pvt. Ltd. | Kolkata, West Bengal, India (full registered-office address on request) |

Pre-incorporation, all five email channels are routed to the Office of the Founder and answered personally. Post-incorporation (Q1 FY 2026-27 onwards), the channels are routed to the named role-holders as they are appointed. The transition is documented on the Centre's website (sargvision.in/contact).

## L.2 Convening calendar — first six months

The Centre will host a series of in-person and virtual convenings in the first six months following publication of this Foundation Edition. Each convening is open to the audiences specified; RSVPs are accepted at the email addresses above. The calendar will be updated quarterly on the Centre's website.

| DATE | CONVENING | AUDIENCE | VENUE | RSVP BY |
|---|---|---|---|---|
| 21 June 2026 | First Partner Convening — Eastern India AI Centre of Excellence | Academic, industry, government, philanthropy (~60 invitees) | Siliguri | 7 June 2026 |
| 12 July 2026 | Government Roundtable — convergence with State and Union policy frameworks | State-government and Union-ministry officers, IndiaAI Mission Secretariat | Kolkata | 28 June 2026 |
| 6 August 2026 | Industry & CSR Roundtable — sector POC programme presentation | NASSCOM-member firms, regional industry leadership, family-office representatives | Bengaluru (virtual + in-person) | 20 July 2026 |
| 5 September 2026 | Academic Convening — joint-appointment and joint-PhD architecture | Vice-Chancellors / Deans / Faculty Heads of IIT-KGP, ISI Kolkata, Jadavpur, Visva-Bharati, IIEST Shibpur, NIT Durgapur, JGEC Jalpaiguri, NBU | Kolkata | 22 August 2026 |

| | | | | |
|---|---|---|---|---|
| | Inaugura-tion — global Bengali-AI talent convening | Diaspora researchers, named in advance | Hybrid (Kolkata + Stanford + Cambridge + Singapore relay) | 17 September 2026 |
| 15 November 2026 | Open Asset Council First Meeting | Council members + invited open-source community observers | Kolkata | 1 November 2026 |

A standing all-stakeholder convening is scheduled annually thereafter on the first Friday of May, coinciding with the release of the annual Eastern India AI Index.

## L.3 Commitment menu — by audience

The matrix below lists the specific commitments the Centre is prepared to receive from each audience. Readers are encouraged to indicate which row, or rows, match their institution's posture; the Centre will respond with the appropriate next-step (MoU draft, term sheet, partnership letter, or governance proposal).

*State governments (West Bengal, Bihar, Odisha, Jharkhand, Assam, Sikkim)*

- **Land allotment** — 10 to 15 acres near Bagdogra airport, Siliguri (West Bengal) on subsidised lease under the State IT Policy 2018 framework. Equivalent allotment in adjacent states for satellite hubs (Year 3 onwards).
- **Policy directive** — MAKAUT (Maulana Abul Kalam Azad University of Technology) and University of North Bengal partnership directive enabling joint-appointment and joint-research arrangements from Day 1.
- **Anchor co-investment** — Rs 50 to Rs 100 crore over five years, structured to unlock Rs 200 to Rs 500 crore in additional capital (full specification in Chapter 1 §1.4 and Chapter 8 §8.4).
- **Sector-POC sponsorship** — partnership with named state departments for the agriculture, healthcare, BFSI, and logistics POC programmes (full schedule in Annexure N §N.9).

- **Observer seat on the Governing Council** — non-voting observer status for a designated State officer.

*Union government and IndiaAI Mission*

- **Recognition under the IndiaAI Mission's Centre-of-Excellence pillar** — formal designation as the Eastern India anchor of the Mission's CoE network.
- **Convergence funding** — multi-year grant under the IndiaAI Mission's open programme pool, with milestones aligned to the Centre's published commitment ladder.
- **Bharat Datasets Platform inclusion** — primary deposit channel for the Centre's open Bengali corpora and model weights (Annexure I).
- **AI Safety Institute partnership** — joint workstream on Bengali content-moderation infrastructure and Indic-language safety tooling.
- **IndiaAI Fellowship hosting** — Centre as a designated host institution for IndiaAI Mission Fellowship cohorts from Year 2.

*Academic partners (Vice-Chancellors, Deans, Faculty Heads)*

- **Joint-appointment MoU** — standard template available on request; covers compensation-sharing, IP attribution, and tenure-clock interaction.
- **Joint-PhD framework** — covers registration body, co-supervision protocol, examiner pool, thesis attribution.
- **Open-data MoU** — minimum-friction commitment: the Centre cites the partner's corpora, the partner cites the Centre's open assets.
- **Editorial Advisory Council** — invitation to serve on the inaugural 5-to-7-member Council that will be named in v1.1.
- **Sabbatical hosting** — the Centre commits to hosting up to four sabbatical scholars per year from Year 2, with full institutional support.

*Industry and CSR partners (NASSCOM members, technology firms, BFSI leaders)*

- **Letter of Intent for cohort-hiring** — pre-Year-1 commitment for placement of Centre-trained cohorts; minimum cohort size 25; standard template available.

- **Proof-of-Concept partnership** — co-built sector POCs in agriculture, manufacturing, healthcare, BFSI, or logistics (Annexure N catalogue); partner provides domain access, the Centre provides AI engineering and evaluation.
- **CSR multi-year commitment** — Rs 1 crore to Rs 25 crore tiers, ring-fenced to specific Centre programmes (training cohorts, district-hub operations, dataset annotation, fellowship endowment).
- **Industry Fellow secondment** — 12-to-24-month rotational placement of a senior firm employee at the Centre; firm continues to pay the Fellow.
- **Foundation-Edition foreword** — open invitation for one named industry leader to author the foreword to v1.1.

### *Philanthropy and family offices (Azim Premji Foundation, Tata Trusts, family-office programme officers)*

- **Catalytic grant** — Rs 3 crore to Rs 50 crore tiers, tied to specific outcome milestones audited by J-PAL South Asia or IDinsight.
- **Fellowship endowment** — Centre-named research fellowship in perpetuity, minimum Rs 5 crore corpus.
- **Open-asset underwriting** — sponsorship of one named open dataset or model release (e.g. "EastAgri-v1 powered by the X Foundation").
- **Index Editorial Board funding** — multi-year endowment of the editorially independent Eastern India AI Index production cost (approximately Rs 2 crore per annum).
- **Diaspora Council co-sponsorship** — supporting the inaugural Diaspora Council convening and ongoing programme.

### *Press, civil society, and the interested public*

- **Briefing requests** — the Centre will brief any responsible journalist on the document, the institutional architecture, or any specific chapter. Standard 60-minute briefing available.
- **Endorsements** — public statement of support from a named civil-society leader, scholar, or institution; the Centre will publish endorsements (with permission) on the website.
- **Public-comment submissions** — substantive critiques of the document are welcomed and will be addressed in v1.1; the Centre will name and credit substantive contributors.

- **Volunteer field-survey participation** — the Centre's annual field surveys (Cooch Behar, Murshidabad, Jalpaiguri, Alipurduar, Darjeeling) are open to volunteer enumerators with prior research experience.
- **Open invitation to Annexure J reading** — the Centre's 25-year forward profile is the most reflective passage in the document; readers are encouraged to engage with it directly.

## L.4 Standard response templates

The following templates are available on request from the publisher. They are pre-cleared with the Centre's legal counsel and are designed to compress the negotiation cycle for partner institutions that are time-constrained.

- **L.4.1 — Letter of Intent for cohort-hiring** (industry partner template)
- **L.4.2 — Memorandum of Understanding for joint research and joint-appointment** (academic partner template)
- **L.4.3 — Catalytic Grant Agreement** (philanthropic family-office template, with outcome-milestone schedule)
- **L.4.4 — Sector-POC partnership term sheet** (industry counterparty, structured against the seven sectors in Annexure N)
- **L.4.5 — State-government partnership framework** (anchor co-investment + land + policy directive)
- **L.4.6 — Editorial Advisory Council invitation** (academic / civil-society luminary)
- **L.4.7 — Foreword commissioning brief** (for the v1.1 foreword author)
- **L.4.8 — Open-Asset sponsorship letter** (for named-asset philanthropic sponsorship)

The templates are not contractually binding in their template form; they are the starting point for partner-specific negotiation. The Centre's commitment is to use the same templates with every counterparty so that terms are visible across partners and so that no partner-specific concessions are silently granted.

## L.5 Reply portal and digital response

A dedicated reply portal at sargvision.in/respond will be operational alongside publication of v1.0. The portal accepts a structured response form covering: respondent identity, institutional

affiliation, intended commitment (single or multiple selections from the menu in §L.3), proposed scale, preferred convening date, and free-text additional notes. The portal generates an auto-acknowledgement within one working day and routes the response to the appropriate role-holder (Founder, Partnerships, Research, Press, Index Editorial Board) within three working days.

The portal also publishes — with respondent permission — a running tally of first-mover partners (the "First Twenty-Five Partners" wall). The Centre's view is that early-mover institutional partners create the gravitational field that subsequent partners respond to; making the early commitment public is the simplest mechanism the Centre has to honour it.

A QR code linking to the reply portal is printed on the back cover of every bound copy of this Detailed Project Report. Digital copies link to the portal from the colophon (F11).

## L.6 What the Centre will publish in response

The Centre commits to publishing the following, on the schedule below, in response to the engagement that this annexure invites.

- **By 30 June 2026:** the named list of confirmed first-tier convening attendees and the agenda for the 21 June First Partner Convening.
- **By 30 September 2026:** the inaugural Editorial Advisory Council, by name and affiliation, in v1.1 of this Detailed Project Report.
- **By 31 December 2026:** the first three signed MoUs and their counterparties.
- **By 31 March 2027:** the first published Sector POC outcome (positive or negative) under the 90-day publication commitment in Annexure N §N.9.
- **By 15 May 2027:** the inaugural Eastern India AI Index, with the Editorial Board attribution and the OSF-pre-registered methodology in full.

These dates are commitments. Slippage on any will be acknowledged in writing to all respondents and explained in the corresponding errata to v1.0.

The Centre's institutional view is that this annexure is the operative chapter of the Detailed Project Report. The intellectual case in the body chapters supports the response; the response is what the

case is for.

M

# The Eastern India AI Index, Year-1 Methodology

ANNEXURE M · SUPPORTING MATERIAL

# Annexure M — The Eastern India AI Index, Year-1 Methodology

The Centre commits to producing an annual quantitative index of Eastern India's artificial intelligence ecosystem, published every May beginning with a Year-1 baseline edition in May 2027. The index is modelled on Stanford HAI's annual AI Index Report — the most-cited quantitative document in global AI policy — and on the institutional pattern by which NITI Aayog's India Innovation Index and CRISIL's quarterly economic publications have become reference artifacts for their respective policy communities.

The methodology set out below is the Centre's pre-registered design for the index. It is published here so that subsequent revisions can be measured against this baseline and so that the index, from its first edition, is auditable.

## What the index measures

The index measures the state of artificial intelligence capacity across the five-state Eastern India region — West Bengal, Bihar, Odisha, Jharkhand, and Assam — across six dimensions, each constructed from publicly auditable indicators drawn from named secondary sources and from the Centre's own primary field surveys.

### *Dimension 1 — Workforce density*

| INDICATOR | SOURCE | FREQUENCY |
|---|---|---|
| AI/ML professionals per lakh of working-age population, by state | NASSCOM AI Adoption Index + LinkedIn Economic Graph regional cuts | Annual |
| Net annual migration of AI/ML professionals into / out of the region | NASSCOM survey + CMIE Consumer Pyramids | Annual |
| Number of AI/ML faculty positions at Eastern India universities | AISHE 2022-23 onwards | Annual |
| AI/ML practitioners-to-graduates ratio (regional) | UDISE+ and AISHE cross-tabulation | Annual |

### *Dimension 2 — Startup ecosystem*

| INDICATOR | SOURCE | FREQUENCY |
|---|---|---|
| AI-startup formation rate (companies founded per quarter, by district of registration) | MCA21 (Ministry of Corporate Affairs) classifications | Quarterly |
| AI-startup funding flows (venture, seed, Series A) by quarter | Tracxn India + Inc42 / YourStory data | Quarterly |
| Survival rate at year-three of AI startups formed in Eastern India vs all-India median | Tracxn India cohort tracking | Annual |
| Spin-out count from regional academic institutions | Direct submission from partner institutions | Annual |

*Dimension 3 — Language-technology infrastructure*

| INDICATOR | SOURCE | FREQUENCY |
|---|---|---|
| Open Bengali datasets cumulatively available (hours of audio, billions of tokens of text) | Hugging Face + Centre repository scan | Quarterly |
| Open Bengali model weights released, parameter counts | Hugging Face + Centre tracking | Quarterly |
| Bengali NLP papers at top conferences (ACL, EMNLP, NAACL, NeurIPS, ICML) | ACL Anthology + arXiv automated scrape | Annual |
| Santali / Rajbanshi / Nepali language-tech indicators (datasets, models, evaluation suites) | Centre's primary tracking | Annual |

*Dimension 4 — District digital readiness*

| INDICATOR | SOURCE | FREQUENCY |
|---|---|---|
| Percentage of secondary schools with functioning computer labs, by district | UDISE+ | Annual |
| Internet penetration rate, by district | TRAI quarterly | Quarterly |
| Active digital-payments adoption (UPI transactions per 1,000 population) | NPCI district-level data | Quarterly |
| Centre-administered field survey of AI-tool awareness among college students (≥ 1,500 respondents per state) | SARGVISION primary data | Annual |

*Dimension 5 — Education enrolment in AI and ML*

| INDICATOR | SOURCE | FREQUENCY |
|---|---|---|
| Undergraduate enrolment in CS / IT / AI degree programmes, by state | AISHE | Annual |
| Postgraduate enrolment in AI/ML/data-science programmes, by state | AISHE | Annual |
| Certified-training enrolment outside formal degree programmes (NASSCOM FutureSkills, PMKVY tracking) | NASSCOM + NSDC | Annual |
| Centre's own training cohort sizes and completion rates | Internal | Annual (with quarterly previews) |

*Dimension 6 — Industry adoption*

| INDICATOR | SOURCE | FREQUENCY |
|---|---|---|
| Share of regional MSMEs with at least one operational AI tool deployed | Centre's annual MSME survey (target n = 5,000) | Annual |
| Estimated AI software-and-services revenue generated by firms with primary registration in Eastern India | NASSCOM + Tracxn cross-validation | Annual |
| AI-adoption depth scoring on the six-dimensional NASSCOM-BCG framework, for the top-100 regional firms | Direct firm survey, validated by Centre | Annual |

## Composite scoring methodology

Each dimension is scored 0-10 against a pre-registered rubric. The rubric is published in a methodology appendix to each Index edition and is itself subject to revision only every three years, with the revision history archived publicly. Dimension scores are not averaged into a single headline number; the Centre's view is that

Year-on-year change in each dimension is reported as a percentage delta and as a directional arrow (improving / stable / declining / declining-rapidly). State-by-state cross-section is reported for every dimension. District-by-district cross-section is reported for Dimensions 4 and the district-administered subindicators of Dimensions 5 and 6.

## Editorial independence

The Index is produced by an Index Editorial Board within the Centre, distinct from the Centre's operational leadership. The Editorial Board has three members in its inaugural composition: a senior research scientist (ex officio chair, two-year rotating term), a quantitative methodologist drawn from the Indian Statistical Institute or equivalent, and a journalist or public communicator with documented independence from the Centre's leadership. The Board sets methodology, supervises data collection, and signs off on publication.

The Centre's Director cannot direct the Editorial Board to amend a specific finding. The Board's only formal accountability is to the Governing Council, and the Governing Council can dismiss a Board member only for documented misconduct, not for inconvenient findings. This editorial-independence clause is part of the Centre's foundational governance and is to be reflected in the Section-8 company's Articles of Association.

## Pre-registration

The Index methodology, the full indicator list above, and the rubric for each dimension's 0-10 scoring are pre-registered with the Open Science Framework (OSF) before the first Index edition is published. The OSF registration commits the Centre to publishing data and methodology revisions in a versioned, timestamped form that cannot be quietly altered after the fact.

The pre-registration also commits the Centre to publishing the raw underlying data for every indicator under a CC-BY-4.0 licence on the sargvision.in/data portal, simultaneously with the Index publication itself. The intention is that any independent researcher can reproduce the Index's findings from the same data.

## Year-1 baseline indicators (illustrative — to be confirmed in May 2027 first edition)

The Centre's current understanding of the Year-1 (FY 2026-27) baseline, drawn from secondary sources available in May 2026, suggests the following provisional values. These are illustrative; the May 2027 first edition will publish definitive figures.

- **Eastern India share of national AI workforce:** approximately 8% (NASSCOM AI Adoption Index 2024, illustrative)
- **Eastern India share of national AI startup funding 2024-25:** approximately 4% (Tracxn India regional cuts)
- **Eastern India share of open Bengali corpora globally:** approximately 60% by 2027 if the Centre meets its Year-1 dataset commitments (Annexure I)
- **Eastern India undergraduate AI/ML enrolment (UG, FY 2024-25):** approximately 22,000 students across all five states (AISHE 2023-24 estimate)
- **District-level secondary-school computer-lab coverage in West Bengal:** approximately 67% in 2024-25 (UDISE+)

## Cadence and distribution

The Index is published every May, beginning May 2027. Each edition has:

- A 30-40 page primary document with full methodology, year-on-year findings, and state-by-state tables
- A 4-page Executive Brief for senior decision-makers
- A digital companion at sargvision.in/index with interactive cross-tabulations and downloadable raw data
- A simultaneous release to the press, with embargoed prior briefings to the Department of Higher Education (GoWB), the IndiaAI Mission Secretariat, MeitY, NASSCOM, and the academic community

Distribution is open and free in perpetuity. The Centre commits to never paywalling the Index, never restricting it to commercial subscribers, and never editing past editions silently. The integrity of the cumulative record is the asset.

## What this commits the Centre to

The Index is the largest single recurring institutional output the Centre commits to in this DPR. It requires an Editorial Board, a 3-person data-collection team, an annual production budget of approximately ₹20-25 lakh, and the discipline of publishing on schedule for the next twenty-five years. The institutional cost is non-trivial.

The institutional return is the proposition that no government, no academic peer, no industry observer of Eastern India's AI ecosystem will be able to make a serious claim about the region's capacity in 2030, 2040, or 2050 without consulting the Index that the Centre has published. The Index is the quantitative anchor that the Centre's other claims — its training output, its research, its district reach — will be evaluated against by their own community.

The first edition's publication on 15 May 2027 is the founding intellectual commitment of this Centre to the region it serves.

N

# Sector Use-Case and POC Catalogue

ANNEXURE N · SUPPORTING MATERIAL

# Annexure N — Sector Use-Case and POC Catalogue

The Centre's claim to institutional relevance does not rest on the Bengali language opportunity alone. It rests on a sector-by-sector mapping of forty-plus specific artificial-intelligence applications across seven major industry groupings that together account for the Rs 18.79 lakh crore West Bengal economy and the broader Rs 49,000 crore Eastern India industrial value pool. The mapping that follows is drawn from the executive whitepaper (SARGVISION Intelligence, *AI Center of Excellence for Eastern India — Bengal's Window to Lead India's AI Future is Closing*, April 2026) and from the strategic vision brief that accompanied it.

The economic claim is the aggregate. Across the seven sector groupings, an indicative core AI investment of Rs 2,885 to Rs 4,581 crore is estimated to generate, at a realistic 20-to-40-percent adoption pace consistent with observed agricultural-technology adoption rates in India, annual economic impact in the range of Rs 17,898 to Rs 34,281 crore — the headline figure the Centre commits to in this Detailed Project Report. At theoretical full adoption, the impact range extends to Rs 33,898 to Rs 49,281 crore, but this upper bound is reported as a ceiling rather than a forecast. Forty-five thousand to seventy-three thousand direct AI-related jobs are projected to be created across the regional sector mappings; this is industry-side employment catalysed by the seven-sector programme and is distinct from the Centre's own training-to-placement pipeline of 75,000-plus learners by Year 5 (Chapter 1 §1.3). Ecosystem-level induced employment, captured in Chapter 10's monitoring framework, is materially larger. The numbers are calibrated against comparable deployments in India and globally (Plantix, Eruvaka, Bosch Connected Industry, Qure.ai, Tata Steel predictive-maintenance, Rotterdam port-digital-twin, Cleveland Clinic AI-operations) and against the Centre's primary field surveys in five North Bengal districts.

The annexure is structured by sector. Each sector entry names the contextual scale (workforce, output, output share), the specific AI applications technically scoped for deployment, the investment range, the expected annual impact range, the direct-jobs estimate, and the payback horizon. Each application is anchored to a comparable deployment so that the claim is not hypothetical. The

Centre's work in any of the seven sectors is the work of co-building production AI systems with named businesses — proofs of concept that ship and are measured against pre-registered targets, not white papers that ship and are read.

## N.1 Agriculture and Allied Sectors

Agriculture and allied sectors contribute Rs 3.09 lakh crore (17.13 percent of West Bengal GSDP) and employ approximately 1.54 crore workers — 34.2 percent of the state workforce. The Krishak Bandhu programme reaches 1.05 crore farmers, providing both a distribution base and a primary-data source unmatched in scale by any peer-state intervention. West Bengal is the country's leading producer of rice (164.9 lakh tonnes, first nationally), vegetables (280.8 lakh tonnes, first or second), and freshwater fish (20 lakh tonnes, first inland). The state produces 74 percent of India's jute and is the second-largest tea producer at 38 crore kilograms. Despite this output leadership, the rice yield gap against the Punjab benchmark stands at 29 percent (Directorate of Agriculture, Government of West Bengal). Potato post-harvest losses are estimated at 25 percent, destroying Rs 4,500 to Rs 6,000 crore annually across the state's 580-plus cold storages (CPRI; CIP estimates).

Seven AI applications have been technically scoped with implementation specifications.

- **Crop Disease Detection** using Convolutional Neural Networks (ResNet-50, EfficientNet-B4) can achieve 97 percent-plus mean average precision on plant-disease datasets (Hughes & Salathé, 2015). Plantix, deployed across India by PEAT GmbH, has crossed one crore downloads and reports 98 percent accuracy across 500-plus crop diseases. For West Bengal, the priority targets are rice blast, potato late blight, and jute stem rot — the three diseases with highest economic loss. Investment range: Rs 15 to Rs 25 crore. Expected impact: 15 to 25 percent crop-loss reduction translating to Rs 2,000 to Rs 4,000 crore per year in avoided losses.
- **Precision Agriculture** combining multi-spectral satellite imagery (Sentinel-2, ISRO ResourceSat) with soil IoT sensors demonstrates 18 to 27 percent fertilizer reduction with maintained yield (CGIAR meta-analysis; Cropin field trials). The 1.05

crore Krishak Bandhu farmers provide an unmatched distribution base. Investment: Rs 80 to Rs 120 crore. Impact: Rs 3,000 to Rs 5,000 crore per year.

- **Cold Chain Optimization** for the state's 580-plus cold storages — the second-highest count in India after Uttar Pradesh — uses IoT temperature and humidity monitoring with ML-based demand-supply matching. Ninjacart has demonstrated 40 percent wastage reduction in comparable fresh-produce supply chains. Reducing potato losses from 25 percent to 12 to 15 percent would save Rs 1,500 to Rs 2,500 crore per year. Investment: Rs 40 to Rs 60 crore.

- **Fisheries IoT** deployment of dissolved-oxygen sensors, pH monitors, and feed-optimization algorithms across the state's 3.28 lakh hectares of culture fisheries. Eruvaka Technologies (Andhra Pradesh) has deployed 50,000-plus IoT devices with 40 to 100 percent biomass increase. Investment: Rs 50 to Rs 80 crore. Impact: Rs 2,000 to Rs 3,000 crore per year.

- **Tea Quality Grading AI** using hyperspectral imaging and CNN classification for Darjeeling's premium teas, where adulteration erodes the GI-protected 3.6× price premium. Investment: Rs 20 to Rs 35 crore. Impact: Rs 500 to Rs 800 crore per year in premium recovery.

- **Livestock and Dairy AI** (IoT wearables, estrus detection, milk-yield prediction) modelled on Stellapps' deployment monitoring 450,000 cattle with 20 percent yield improvement. Investment: Rs 30 to Rs 50 crore. Impact: Rs 1,000 to Rs 1,500 crore per year.

- **Market Linkage and Price Prediction** using LSTM and Transformer models on eNAM mandi data, achieving Mean Absolute Percentage Error of 5 to 12 percent on Indian commodity datasets (IIT Kharagpur research). Investment: Rs 10 to Rs 15 crore. Impact: Rs 2,000 to Rs 3,000 crore per year in improved farmer income through optimal sell-timing.

Aggregate Agriculture investment: Rs 315 to Rs 470 crore. Steady-state annual impact under conservative 20 to 40 percent adoption: Rs 4,000 to Rs 10,000 crore per year — a still-substantial eight-to-twenty-one-times economic-impact multiple. Direct jobs created: 4,770 to 7,560. Payback horizon: 18 to 36 months.

# N.2 Information Technology, IT-Enabled Services, and the Data Economy

The state's organised IT and ITeS workforce of approximately 260,000 is roughly five percent of India's 54-lakh IT workforce — disproportionately low against the state's seven percent population share (NASSCOM State Report 2024). The Bengal Silicon Valley Hub in New Town has onboarded 42 companies against a 100,000-job target across 250 acres. TCS operates a 55,000-person Kolkata campus with 25,000 additional positions planned. iMerit — headquartered in Kolkata, employing 7,198 globally with approximately 5,200 in India and Rs 278 crore in revenue — is among India's largest data-annotation companies and a foundational anchor for the data-annotation economy the Centre intends to scale.

The critical metric is AI-skill penetration: only sixteen percent of West Bengal's IT workforce has AI-relevant skills, against an estimated 28 to 35 percent in Karnataka and Telangana. Workers with AI skills command a 56 percent salary premium globally (PwC Global AI Jobs Barometer 2025; the global average; India-specific premium may vary). This gap is both the problem (continued brain drain to higher-paying markets) and the opportunity (upskilling 218,000 workers creates Rs 1,000 to Rs 1,500 crore per year in aggregate income uplift).

Three application areas are scoped.

- **GenAI Upskilling for 218,000 non-AI-skilled IT workers**, delivered through the Centre's training pillar in partnership with NASSCOM FutureSkills, regional engineering colleges, and the state's corporate-training infrastructure. Investment: Rs 80 to Rs 120 crore.
- **Expansion of the data-annotation economy** from iMerit's anchor base to 10,000-plus additional jobs with Bengali NLP specialisation. The Centre's open Bengali corpora (Annexure I) directly increase the addressable annotation market. Investment: Rs 40 to Rs 80 crore.
- **GCC expansion** from the current 20 to 25 centres to 50-plus with dedicated AI labs, drawing on the 40 to 55 percent salary-cost advantage over Bangalore. Investment: Rs 80 to Rs 150 crore.

Aggregate IT/ITeS investment: Rs 200 to Rs 350 crore. Annual impact: Rs 2,500 to Rs 4,000 crore. Direct AI jobs created: 3,500 to 5,500. Payback: 12 to 24 months.

## N.3 Manufacturing and Industry 4.0

The secondary-sector GSDP of Rs 4.79 lakh crore (27.87 percent of state output) is anchored by globally significant industrial assets: SAIL Durgapur Steel Plant and IISCO Burnpur (combined 7-plus MTPA capacity, Rs 45,811 crore modernisation underway), Haldia Petrochemicals (Rs 14,900 crore revenue), India's largest foundry cluster in Howrah (320 foundries, 750,000 tonnes per year, 15 to 18 percent of national castings), and the Raniganj coalfield (4,900 crore tonnes estimated reserves, 45,000 Eastern Coalfields Ltd employees). The 89 lakh MSME units employing 1.36 crore people sit alongside these large-industry anchors.

Four applications are scoped.

- **Predictive Maintenance** for steel plants is the highest-confidence application. Tata Steel's programme at Jamshedpur has deployed 550-plus predictive models with an estimated Rs 11,600 crore lifetime savings — a 1:10 investment-to-savings ratio (Tata Steel Sustainability Report 2024). Replicating even fifty percent of this performance at SAIL Durgapur and IISCO Burnpur yields Rs 120 to Rs 180 crore per year on an investment of Rs 46 to Rs 77 crore, with payback in four to eight months.
- **Quality Control AI** for Howrah's foundry cluster addresses a five percent scrap rate that translates to Rs 375 to Rs 600 crore per year in losses. Machine-vision systems achieving 99.8 percent accuracy in comparable automotive applications (Bosch Connected Industry) can reduce the scrap rate to two to three percent. Investment: Rs 40 to Rs 70 crore. Impact: Rs 150 to Rs 350 crore per year.
- **Mining Safety AI** for the Raniganj coalfield using gas-concentration sensors and ML prediction. Research literature (CSIR-CIMFR studies) demonstrates 100 percent gas-explosion prediction accuracy with adequate sensor density. Investment: Rs 25 to Rs 40 crore. Impact: prevention of 80 to 90 percent of mine disasters.
- **MSME Transformation** through WhatsApp-based AI assistants for order management, inventory, and GST compliance — voice-first in Bengali — targeting five to ten percent adoption among

the 89 lakh micro-enterprises (4.5 to 9 lakh units). Modelled on Khatabook's 3 crore-plus SME user base. Investment: Rs 60 to Rs 80 crore.

Aggregate Manufacturing investment: Rs 300 to Rs 500 crore. Annual impact: Rs 1,300 to Rs 3,400 crore. Direct AI jobs: 7,000 to 12,000. Payback: 3 to 12 months — the highest-return-on-investment sector in the mapping. The April 2025 revocation of West Bengal's industrial-incentive scheme creates urgency: AI-productivity support becomes the alternative value proposition.

## N.4 Healthcare

West Bengal's healthcare infrastructure faces a hospital-bed deficit of 78.6 percent against Indian Public Health Standards norms (1,07,346 available against approximately 5,00,000 required; Central Bureau of Health Intelligence 2024). More critically, 70 to 72 percent of Primary Health Centre doctor positions remain vacant (Rural Health Statistics 2023-24) — making AI-augmented diagnostics not optional but essential for basic healthcare delivery. The state does have one significant asset: the highest National Quality Assurance Standards (NQAS) certification rate in India at 25 percent of eligible public health facilities, indicating institutional capacity for quality measurement.

Three applications are scoped.

- **PHC-Level AI Diagnostics** — chest X-ray AI for tuberculosis screening (Qure.ai demonstrates 15.8 percent additional TB yield, true positives missed by human readers), retinal-scan AI for diabetic retinopathy (MadhuNetrAI: 90 to 95 percent sensitivity/specificity) — deployed across 500 PHCs in Phase 1. Investment: Rs 80 to Rs 120 crore. Impact: 5 to 10 lakh patients screened annually; 2,000 to 4,000 lives saved from early TB detection alone.
- **Telemedicine AI** upgrading the existing Swasthya Ingit platform (7 crore-plus consultations across 11,000 centres) with Bengali NLP triage, AI-assisted diagnosis support, and automated follow-up monitoring. AI-augmented telemedicine reduces emergency visits by 22 percent (Journal of Telemedicine & Telecare, 2023 meta-analysis). Applied to Swasthya Ingit's scale: Rs 320 to Rs 1,280 crore per year in avoided emergency costs.
- **Hospital Operations Optimization** through AI-driven bed management, patient-flow prediction, and scheduling has demonstrated 15 to 25 percent efficiency improvement in comparable

deployments (Cleveland Clinic, AIIMS). Applied to West Bengal's 1,07,346-bed system, even modest improvements yield the equivalent of thousands of additional bed-days annually. Investment: Rs 80 to Rs 120 crore. Impact: Rs 500 to Rs 800 crore per year.

Aggregate Healthcare investment: Rs 658 to Rs 1,005 crore. Annual economic impact: Rs 2,800 to Rs 4,500 crore. Five to ten thousand lives saved annually. Direct jobs: 11,650 to 18,250 including health-data annotators, telemedicine operators, and clinical AI specialists. The human impact here exceeds the financial return as the primary justification.

## N.5 Banking, Financial Services, and Insurance

West Bengal hosts India's largest microfinance and self-help-group ecosystem. Bandhan Bank alone operates 6,300 branches with Rs 1.57 lakh crore in deposits. The state has 11.92 lakh Self-Help Groups — the highest of any Indian state — under the Anandadhara (WB SRLM) programme. Arohan Financial Services serves 21.2 lakh clients. However, MFI Non-Performing Assets surged to sixteen percent in 2024 (MFIN Micrometer Q3 2024) on an outstanding portfolio of approximately Rs 45,000 crore, signaling acute need for AI-based credit assessment.

Five applications are scoped.

- **NPA Prediction AI** using LSTM/Transformer models for payment-behavior prediction, targeting 60-to-90-day advance default warnings, offers the single highest return-on-investment application identified in the whitepaper analysis: Rs 1,200 to Rs 1,900 crore in improved recovery on a Rs 6 to Rs 12 crore investment — a 20 to 30-times return (computed from MFIN data and industry NPA-recovery benchmarks).
- **SHG Credit Scoring** using alternative data — satellite NDVI crop health, mobile-phone usage patterns, SHG repayment history — modelled on Tala's 75 lakh-customer global deployment with 50 percent lower default rates. Target: scoring 11.92 lakh SHGs (approximately 80 lakh individual borrowers).
- **Vernacular Voice Banking** in Bengali supporting four dialects (Rarhi, Varendi, Kamrupi, Sundarbani) plus Nepali and Santali, at the Rs 1-per-minute cost point Sarvam AI has demonstrated.

- **Cyclone Micro-Insurance** — parametric products triggered by satellite-verified weather events, bundled with SHG loans — addressing the catastrophic gap exposed by Cyclone Amphan (2020): Rs 1.02 lakh crore in damage, an estimated 90 percent-plus uninsured (Government of West Bengal Disaster Management Department).

Aggregate BFSI investment: Rs 102 to Rs 166 crore. Annual impact: Rs 1,548 to Rs 2,881 crore. Weighted return-on-investment: 10 to 20 times. The NPA prediction investment alone recovers Rs 1,200 to Rs 1,900 crore on a Rs 6 to Rs 12 crore investment — making it the most capital-efficient AI intervention in this entire analysis.

## N.6 Tourism and the Creative Economy

Tourism contributes substantially: 18.5 crore domestic visitors and 31 lakh foreign tourists (second in India after Rajasthan), generating revenue anchored by Darjeeling (6.3 lakh visitors), Kolkata heritage tourism, and the Sundarbans. The state has 5,322 registered homestays — India's highest.

The Durga Puja economy represents a unique asset: Rs 32,377 to Rs 50,000 crore in annual economic contribution (Bengal National Chamber of Commerce; FICCI-MRSS), UNESCO Intangible Cultural Heritage recognition (2021), and 3-plus lakh direct jobs. No comparable single-event cultural economy exists anywhere in India.

AI applications include visitor-flow optimization for Durga Puja using computer vision (Venice's Smart City project reduced St. Mark's overcrowding by 30 percent), multilingual AI concierge systems (Singapore Tourism Board reported 8 percent spending uplift), and dynamic pricing for the 5,322-strong homestay network.

On the creative side, the Bengali AI Total Addressable Market is estimated at Rs 5,000 to Rs 8,000 crore (RedSeer/Kalaari vernacular AI market projections, 2024). Tollywood has experienced a 70 percent collapse in film production — from 134 films (2018-19) to approximately 40 (2024) — creating both crisis and opportunity for AI-augmented animation and visual-effects work. No AI model exists for Bengali music traditions (Rabindra Sangeet, Baul, Nazrul Geeti), representing both cultural preservation and commercial opportunity.

Aggregate Tourism & Creative investment: Rs 450 to Rs 700 crore. Annual impact: Rs 1,550 to Rs 2,500 crore. Direct AI jobs: 1,100 to 1,520. No major Bengali AI company is headquartered in Bengal — the entire Bengali creative-AI opportunity is currently being developed from Bangalore and Chennai. Maharashtra has committed Rs 3,268 crore to its AVGC policy; West Bengal has no comparable commitment.

## N.7 Logistics, Energy, and Infrastructure

This is the largest sector by investment requirement and job-creation potential, anchored by three critical infrastructure assets: Kolkata Port (64 MMT in 2024 — the highest growth rate among all Indian major ports), the 580-plus cold-storage network, and the Siliguri corridor through which 95 percent of Northeast India's exports transit.

Four applications are scoped.

- **Port Optimization AI** (digital twin, berth scheduling, vessel prediction) modelled on the Port of Rotterdam's 20 percent reduction in vessel waiting time. Applied to Kolkata-Haldia's combined 111 MMT throughput: 15 to 20 percent capacity gain without infrastructure expansion equals 17 to 22 MMT effective capacity added. Investment: Rs 200 to Rs 350 crore. Impact: Rs 500 to Rs 800 crore per year.
- **Grid Management AI** for a state with 80.4 percent coal dependency and only 640 MW renewable capacity against 6,260 MW solar potential. ML-based demand forecasting and renewable-intermittency management targeting five to ten percent reduction in transmission-and-distribution losses (currently approximately 18 percent). Investment: Rs 150 to Rs 250 crore. Impact: Rs 800 to Rs 1,200 crore per year.
- **Coal Mine Safety AI** for the Raniganj coalfield (4,900 crore tonnes reserves, 45,000 ECL employees) and **Just Transition Planning** for the eventual coal-to-renewable shift — modelled on Germany's Ruhr Valley transition which created 100,000-plus environment-technology jobs (ILO Just Transition Report).

Aggregate Logistics, Energy, and Infrastructure investment: Rs 860 to Rs 1,390 crore (the largest sector). Annual impact: Rs 4,200 to Rs 7,000 crore. Approximately 67,000 total jobs including the coal-transition workforce over 15 years.

## N.8 Consolidated cross-sector summary

The seven-sector mapping consolidates to the following indicative envelope. The figures should be read as ranges — the upper bound reflects full-adoption scenarios; the lower bound applies a conservative 20 to 40 percent realization factor consistent with observed agricultural-technology adoption rates in India.

**Table N.1** *Cross-sector AI investment vs annual impact, direct jobs, payback (illustrative)*

| # | SECTOR GROUPING | INVESTMENT (₹ CR) | ANNUAL IMPACT (₹ CR) | ROI MULTIPLE | DIRECT AI JOBS | PAYBACK |
|---|---|---|---|---|---|---|
| 1 | Agriculture & Allied | 315–470 | 4,000–10,000 (realistic, 20-40% adoption) | 8–21× | 4,770–7,560 | 18–36 mo. |
| 2 | IT / ITeS & Data Economy | 200–350 | 2,500–4,000 | 8–15× | 3,500–5,500 | 12–24 mo. |
| 3 | Manufacturing & Industry 4.0 | 300–500 | 1,300–3,400 | 20–34× | 7,000–12,000 | 3–12 mo. |
| 4 | Healthcare | 658–1,005 | 2,800–4,500 | 4–5× | 11,650–18,250 | 12–36 mo. |
| 5 | BFSI | 102–166 | 1,548–2,881 | 10–20× | 625–1,011 | 6–18 mo. |
| 6 | Tourism & Creative Economy | 450–700 | 1,550–2,500 | 3–4× | 1,100–1,520 | 12–24 mo. |
| 7 | Logistics, Energy & Infrastructure | 860–1,390 | 4,200–7,000 | 5–8× | 16,800–27,000 | 6–24 mo. |
| | **Aggregate (Eastern India)** | **2,885–4,581** | **17,898–34,281** | **≈ 8–12×** | **45,445–72,841** | — |
| | At theoretical full adoption | (same) | 33,898–49,281 | ≈ 10–15× | (same) | — |

*Source: SARGVISION Executive Whitepaper §2 (April 2026); cross-validated against NASSCOM Industry Reports 2024 and named comparable deployments (Plantix, Eruvaka, Bosch Connected Industry, Tata Steel, Qure.ai, Rotterdam Port).*

The weighted-average return-on-investment multiple across all sectors is approximately ten to fifteen times. Manufacturing shows the fastest payback (3 to 12 months) and Agriculture the largest absolute annual impact (Rs 20,000 to Rs 25,000 crore at theoretical full adoption; Rs 4,000 to Rs 10,000 crore at realistic 20-40 percent

adoption). Healthcare delivers the most significant human impact: 5,000 to 10,000 lives saved annually and significant effective bed capacity gained through AI-driven operational optimization.

The total core AI investment range of Rs 2,885 to Rs 4,581 crore generates an annual impact range of Rs 17,898 to Rs 34,281 crore at the realistic 20-to-40-percent adoption pace — approximately Rs 6 to Rs 8 of annual value for every rupee invested across the seven-sector portfolio. At theoretical full adoption the multiple extends to 10-15×. Forty-five thousand to seventy-three thousand direct AI-related industry-side jobs are projected; ecosystem-induced employment is materially larger.

## N.9 The proof-of-concept programme as institutional posture

The Centre's distinguishing institutional choice is to operate as an industrial transformation engine and not as a research laboratory. The seven-sector mapping is therefore not an aspiration; it is a programme. Each sector area is matched to a Centre-co-built proof-of-concept programme, executed in partnership with named industry counterparties, governed under the Outcome-Linked Discipline mandate (third-party audit by J-PAL South Asia or IDinsight; trainer compensation tied to placement; industry work measured on shipped systems, not white papers). The proof-of-concept calendar for the first three fiscal years is sequenced as follows.

- In **fiscal year 2026-27**, the Centre commits to launching proofs-of-concept in three sectors: agriculture (rice-blast and potato late-blight detection in three districts, in partnership with the Department of Agriculture and the Krishak Bandhu distribution network), manufacturing (predictive maintenance pilot at SAIL Durgapur in partnership with the plant's operations team), and healthcare (PHC-level chest X-ray TB screening pilot across 50 PHCs in three districts, in partnership with the Department of Health & Family Welfare).
- In **fiscal year 2027-28**, the proof-of-concept calendar extends to BFSI (NPA prediction pilot with Bandhan Bank's SHG portfolio in three districts), tourism (Durga Puja visitor-flow management pilot with the Forum for Durgotsav in Kolkata), and logistics (port-optimization scoping study with Kolkata Port Trust).

- In **fiscal year 2028-29**, the calendar extends to creative-economy applications (Bengali-music model with the Sangeet Research Academy), grid management (forecast pilot with WBSEDCL), and a full district-wise rollout of the five POCs that have demonstrated production-readiness from the first two fiscal years.

Each proof-of-concept is governed by a partnership MoU that pre-commits the industry counterparty to publishing the pilot's outcome — positive or negative — within ninety days of completion. The intention is that the Centre's published proof-of-concept record becomes the most-cited evidence base for AI-readiness across the region's industrial sectors.

## N.10 What this commits the Centre to

The seven-sector mapping is the substantive content of the Centre's industrial-modernisation mandate. The numbers are calibrated; the proof-of-concept calendar is pre-registered; the partnerships are named. The Centre commits, through this annexure, to publishing annual progress against the proof-of-concept calendar in the Eastern India AI Index (Annexure M), with each sector's economic impact tracked as a separate dimension.

The proposition is straightforward. The industrial demand exists — 40-plus AI applications, Rs 33,000 to Rs 49,000 crore in annual value, 45,000 to 73,000 jobs — and the institutional bridge does not. The Centre is that bridge. Its first three years of proof-of-concept output will establish whether the bridge holds.

O

# Bibliography and Source Register

ANNEXURE O · SUPPORTING MATERIAL

# Annexure O — Bibliography and Source Register

This annexure consolidates the source documents referenced in the body of this Detailed Project Report. The entries are grouped by source type. Within each group, entries are listed alphabetically by issuing organisation. Internal source documents (those held in the Centre's primary-source archive at SARGVISION Intelligence Pvt. Ltd.) are marked with the suffix *[Internal Archive]*; published or otherwise externally verifiable sources are not so marked.

The full per-claim citation ledger — every numeric and named-entity claim in the document, mapped to its source — is maintained as a versioned SQLite database in the Centre's `runs/v4/citations.db` artifact and is available for inspection on request.

## A. Government of India statutes, policies, and reports

- **Companies Act, 2013** (Ministry of Corporate Affairs), with particular reference to §8 (companies licensed for promotion of commerce, art, science, sports, education, research, social welfare, religion, charity, or other useful object).
- **Digital Personal Data Protection Act, 2023** (Ministry of Electronics and Information Technology), with particular reference to §7 (consent regime).
- **General Financial Rules, 2017** (Department of Expenditure, Ministry of Finance) — referenced as institutional finance-discipline design influence.
- **IndiaAI Mission**, Centre-of-Excellence pillar (Ministry of Electronics and Information Technology, Government of India, 2024).
- **IndiaAI Mission — Architecture presentation**, First Workshop for AI Centre of Excellence, May 2026.
- **National Education Policy 2020** (Ministry of Education, Government of India), §22 (multidisciplinary education) and §17 (Research and Innovation Strategy).
- **Skill India and PMKVY framework** (Ministry of Skill Development and Entrepreneurship).

- **MeitY Bhashini Mission** (Ministry of Electronics and Information Technology, Government of India).
- **NITI Aayog**, *Strategy for New India @75*; *India Innovation Index*; *SDG India Index*.

## B. Government of India statistical sources

- **All India Survey on Higher Education (AISHE)**, 2021-22, 2022-23, 2023-24 (Department of Higher Education, Ministry of Education).
- **Census of India 2011** (Office of the Registrar General & Census Commissioner) with projections through 2026.
- **Central Bureau of Health Intelligence (CBHI), 2024** — hospital-bed deficit statistics for West Bengal.
- **Ministry of Statistics and Programme Implementation (MoSPI)**, *Periodic Labour Force Survey (PLFS) 2022-23 and 2023-24*; quarterly bulletins.
- **National Family Health Survey (NFHS-5), 2019-21** (Ministry of Health and Family Welfare).
- **National Payments Corporation of India (NPCI)** — district-level UPI transaction data.
- **National Sample Survey Office (NSSO) 2018-19** — skill profile statistics.
- **Reserve Bank of India (RBI) Handbook of Statistics on Indian States, 2024**.
- **Rural Health Statistics 2023-24** (Ministry of Health and Family Welfare) — Primary Health Centre doctor-vacancy data.
- **Telecom Regulatory Authority of India (TRAI)** — quarterly internet-penetration statistics.
- **Unified District Information System for Education Plus (UDISE+) 2022-23, 2023-24, 2024-25** (Department of School Education and Literacy).

## C. Government of West Bengal documents

- **Banglar Shiksha portal** (Department of School Education, Government of West Bengal).

- **Department of Higher Education, Government of West Bengal** — institutional correspondence and policy memoranda.
- **GoWB Bureau of Applied Economics and Statistics** — district-level economic data.
- **GoWB Disaster Management Department** — Cyclone Amphan damage assessment, 2020.
- **GoWB Directorate of Agriculture** — crop-yield statistics; Krishak Bandhu programme data.
- **GoWB Directorate of Economic Statistics (DES)** — GSDP and district-level statistics.
- **West Bengal Board of Secondary Education (WBBSE)** Annual Reports 2021 through 2025 — Madhyamik examination results.
- **West Bengal Council of Higher Secondary Education (WBCHSE)** Results Gazettes 2021 through 2025.
- **West Bengal Industrial Development Corporation** — industrial-incentive scheme notifications.
- **West Bengal IT & ITeS Policy 2018-23** (Department of Information Technology and Electronics, Government of West Bengal, 2018) — §4.3 on language-technology infrastructure.


## D. SARGVISION primary archive *[Internal Archive]*


- *SARGVISION AI Centre of Excellence — Strategic Vision Brief*, May 2026.
- *SARGVISION AI Centre of Excellence — Strategic Vision Deck (.pptx)*, 2025-26.
- *SARGVISION AI Centre of Excellence — Vision Whitepaper Brief*, 2026.
- *Executive Whitepaper: AI Centre of Excellence for Eastern India — Bengal's Window to Lead India's AI Future is Closing*, April 2026.
- *Executive Brief — AI CoE*, April 2026.
- *Master Project Blueprint — SARGVISION AI CoE*, 2025-26.
- *Master Project Handbook v2.0 — SARGVISION CoE*, 2026.
- *Project Timeline and Milestone Plan*, 2026.
- *Risk Register and Mitigation Plan*, 2026.
- *Founders' Manifesto — SARGVISION CoE*, April 2026.
- *Bengal AI Vision — SARGVISION Research Suite*, 2026.
- *AI CoE Policy Vision — SARGVISION Research Suite*, 2026.

- *West Bengal Industry Research Report — SARGVISION Research Suite*, 2025.
- *West Bengal Education Landscape Analysis — SARGVISION Research Suite*, 2025.
- *AI CoE Research Report — Industry, Education, West Bengal — SARGVISION Research Suite*, 2025.
- *North Bengal CoE — Partner Institution Proposals (5 documents)*: ACC Jalpaiguri, NBSXC Siliguri, Salesian College Sonada, Sukanta Mahavidyalaya, Acharya Brojendra Nath Seal College.
- *North Bengal CoE — Audit Report and One-Pagers*, 2025.
- *Operational Manual — Bengali Dataset*; *Daily Operations*; *Standard Annotation Tag Guide*; *Technical Compliance Sheet*; *Toolkit and Action Items* (SARGVISION operational manuals, 2025-26).
- *Field Survey Notes — Cooch Behar, Murshidabad, Jalpaiguri, Alipurduar, Darjeeling districts*, November 2025.

## E. Industry and research-community reports

- **NASSCOM** *AI Adoption Index 2024*; *AI Talent Report (NASSCOM-Deloitte, 2024 and 2025)*; *GCC Report 2024*; *State Reports 2024*; *AI Training Data Report*.
- **NASSCOM-Boston Consulting Group** — six-dimensional AI-adoption-depth scoring framework.
- **PwC** — *Sizing the Prize* AI report; *Global AI Jobs Barometer 2025*.
- **McKinsey Global Institute** — *The Economic Potential of Generative AI* (June 2023).
- **Indeed / LinkedIn AI Jobs Index, 2025.**
- **Tracxn India** — AI-startup formation and funding data; regional cuts.
- **Inc42** and **YourStory** — startup-funding tracking, India.
- **MFIN Micrometer, Q3 2024** — microfinance Non-Performing Assets data.
- **Statista Market Insights 2025** — Bangladesh AI market sizing.
- **CRISIL** — quarterly economic publications.
- **Hurun India Unicorn Index** — AI-unicorn classification.
- **iMerit Technology Services** — annual report and revenue disclosures.
- **CGIAR meta-analysis** — precision-agriculture fertilizer-reduction studies.

- **Cropin** — precision-agriculture field trials, India.
- **PEAT GmbH (Plantix)** — crop-disease detection deployment data, India.
- **Eruvaka Technologies** — fisheries IoT deployment data.
- **Tata Steel** — *Sustainability Report 2024*; predictive-maintenance programme disclosures.
- **Bosch Connected Industry** — machine-vision quality-control benchmarks.
- **Stellapps** — livestock IoT deployment data.
- **Tala** — alternative-data credit scoring deployment scale.
- **Ninjacart** — fresh-produce supply-chain wastage-reduction data.
- **Khatabook** — SME user-base deployment data.

## F. Academic and research literature

- **AI4Bharat (IIT Madras)** — *IndicNLP datasets, IndicBERT, IndicTrans, IndicWhisper* model and dataset cards.
- **ACL Anthology** — Bengali NLP papers index.
- **arXiv** — automated scrape of Bengali language-technology research.
- **CSIR-CIMFR (Central Institute of Mining and Fuel Research)** — gas-explosion prediction studies.
- **Hughes & Salathé (2015)** — plant-disease image-classification dataset.
- **IEEE** — *BanglaGPT* (Intelsense AI, Dhaka), 2023.
- **IIT Bombay** — Hindi-English parallel corpus.
- **IIT Kharagpur** — Market-prediction LSTM/Transformer research; AI4ICPS dataset releases.
- **Indian Statistical Institute, Kolkata** — statistical methodology and quantitative-methods support for the Eastern India AI Index.
- **Intelsense AI / BUET community (Dhaka)** — *Ekush LLM* (Hugging Face, October 2024).
- **Journal of Telemedicine & Telecare (2023)** — AI-augmented telemedicine meta-analysis.
- **MadhuNetrAI** — diabetic-retinopathy screening AI sensitivity/specificity research.
- **Qure.ai** — chest X-ray AI for tuberculosis screening, published deployment data.

- **Stanford HAI (Human-Centered AI Institute)** — *Annual AI Index Report* (referenced as institutional model for the Eastern India AI Index).
- **Stanford Center for the Study of Language and Information** — referenced as institutional model for open-asset commitment.
- **The Allen Institute for AI** — OLMo, Tulu, AI2 dataset releases (referenced as institutional model for open-asset commitment).

## G. Cultural-recognition and standards sources

- **Ethnologue (26th edition)** — global language-ranking data.
- **Government of India, Ministry of Culture** — Classical Language status for Bengali, October 2024.
- **International Labour Organization (ILO)** — *Just Transition Report*; Ruhr Valley case study.
- **National Quality Assurance Standards (NQAS)** certification — Ministry of Health and Family Welfare.
- **UNESCO** — Intangible Cultural Heritage recognition of Durga Puja, 2021.
- **UNESCO** — *AI Competency Framework for Teachers (2024)*.

## H. Open-asset infrastructure and licensing

- **Creative Commons** — CC-BY-4.0 and CC-BY-SA-4.0 licence frameworks.
- **Hugging Face Hub** — primary open-asset mirror commitment.
- **Apache License 2.0** — model-weight release framework.
- **Open Science Framework (OSF)** — methodology pre-registration framework.
- **National Digital Library of India** — long-term preservation deposit.
- **Internet Archive** — long-term preservation deposit.
- **Santal Pargana Cultural Council** — Santali language partnership.
- **Rajbanshi Cultural Council** — Rajbanshi language partnership.

- **Sangeet Research Academy** — Bengali music tradition partnership.

## I. International benchmarks and comparator institutions

- **Plantix (PEAT GmbH, Germany)** — crop-disease detection commercial deployment.
- **Cleveland Clinic** — AI-driven hospital operations benchmark.
- **AIIMS New Delhi** — AI-driven hospital operations benchmark.
- **Port of Rotterdam** — port-optimisation digital-twin benchmark.
- **Singapore Tourism Board** — multilingual AI concierge benchmark.
- **Venice Smart City project** — visitor-flow optimisation benchmark.
- **Cooch Behar Land Records / Block Development Office** — local primary-source records, field-survey 2025.

---

*The Centre maintains the full bibliographic database, including page-level and section-level references for every claim, in machine-readable form. Researchers, journalists, and partner institutions may request the database under the Centre's data-sharing policy at the address in the colophon.*

# Colophon

This Detailed Project Report — Edition v1.0, Foundation Edition — was prepared and published by SARGVISION Intelligence Pvt. Ltd. as a self-published institutional publication. It is circulated as a public document and is not addressed to any single office.

## Document control

- **Document identifier:** DPR-SVAI-2026-001
- **Edition:** v1.0 (Foundation Edition)
- **Publication date:** May 2026
- **Pages:** approximately 200
- **Languages:** English and Bangla (with selected Bengali subtitles, headings, and pull-quotes)

- **Suggested citation:** SARGVISION Intelligence Pvt. Ltd. (2026). *Detailed Project Report — AI Centre of Excellence for Eastern India* (v1.0, Foundation Edition). Kolkata: SARGVISION Intelligence Pvt. Ltd.
- **ISBN:** (to be assigned via the Raja Rammohun Roy National Agency for ISBN, Government of India)
- **DOI:** (to be assigned via Zenodo prior to public web release)
- **Deposit-library statements:** National Library of India, Kolkata; Parliament Library, New Delhi; Indian Statistical Institute Library, Kolkata; University of North Bengal Library, Siliguri.
- **Edition policy:** This is a Foundation Edition. Subsequent editions are scheduled at v1.1 (within 90 days of v1.0 publication, with named Director and constituted Governing Council) and v2.0 (May 2027, alongside the inaugural Eastern India AI Index).


## Publisher and legal entity


- **Publisher:** SARGVISION Intelligence Pvt. Ltd.
- **Registered office:** Kolkata, West Bengal, India
- **Corporate Identification Number (CIN):** (to be filled at imprint stage)
- **GSTIN:** (to be filled at imprint stage)

- **Section-8 anchor:** The AI Centre of Excellence proposed in this document will operate as a separate Section-8 company under the Companies Act 2013, to be incorporated in Year 1 with the foundational governance protections set out in Chapter 7 and Annexure F. The Section-8 entity and SARGVISION Intelligence Pvt. Ltd. are legally distinct; related-party transactions between them are governed by the disclosure policy set out below.

## Related-party and conflict-of-interest disclosure

The publisher (SARGVISION Intelligence Pvt. Ltd.) is also the founding promoter of the institution this document describes. This is acknowledged openly. The Centre's intended Articles of

Association — to be drafted and filed with the Registrar of Companies in the first quarter of operations — include the following provisions:

- The Section-8 company will be a separate legal entity from SARGVISION Intelligence Pvt. Ltd.
- All related-party transactions (services, intellectual property, brand licensing) will be disclosed annually in the Section-8 company's audited financial statements.
- The Founder will not draw a salary from the Section-8 company in Year 1; subsequent compensation will be benchmarked publicly to the IIT-Director scale and disclosed in the Centre's annual report.
- The Governing Council will include not fewer than three independent members; quorum and voting rules will protect against founder-side capture.
- The Eastern India AI Index (Annexure M) will be produced under explicit editorial-independence protections written into the Articles of Association.

## Editorial and production credits

- **Authoring and editorial direction:** The Founder, with substantive contributions from named contributors listed in the Acknowledgements (see below).
- **Production pipeline:** This document was produced through a structured pipeline involving multi-agent drafting, citation-ledger management, and human editorial review. The methodology is described in the Methodology Note (F10).
- **Citation ledger:** All numeric and named-entity claims are mapped to source documents in a versioned SQLite database (`runs/v4/citations.db`), available for inspection on request.
- **Design system:** Bengal Saffron — Saffron #E87722 + Maroon #4C0F0F on warm paper #FAF6EE. Typography: Spectral (display, Open Font Licence), Source Serif 4 (body, OFL), Noto Serif Bengali (Bangla, OFL), Inter (UI labels, OFL), JetBrains Mono (code listings, OFL).
- **Production engine:** HTML/CSS authored against W3C CSS Paged Media; rendered to PDF via pagedjs-cli (Chromium-headless).

## Contact and response

The Centre welcomes engagement from readers across all audiences described in the Reader's Guide (F3). Specific response paths, commitment menus, and a convening calendar are set out in Annexure L (Response & Engagement Pack). Direct contact details follow.

- **General correspondence:** founder@sargvision.in
- **Partnership and CSR enquiries:** partnerships@sargvision.in (to be activated alongside the Section-8 incorporation)
- **Academic and research enquiries:** research@sargvision.in
- **Press and communications:** press@sargvision.in
- **Index Editorial Board:** index@sargvision.in (to be activated alongside the Editorial Board constitution, August 2026)
- **Postal address:** SARGVISION Intelligence Pvt. Ltd., Kolkata, West Bengal, India (full registered-office address available on request and on the imprint page of the bound edition)
- **Founder, direct line:** Abhishek Gupta — available on request from the publisher's address; published in the imprint of the bound edition

## Copyright and licensing



## Errata and revisions

Errata to v1.0 will be published on the Centre's website at sargvision.in/dpr-errata. Each subsequent edition (v1.1, v2.0) will include a versioned changelog. The Centre's commitment, on the

model of the Stanford HAI AI Index and the Allen Institute for AI, is that no past edition is silently edited. The cumulative record is the asset.

## Disclosure

This is a Foundation Edition of a Detailed Project Report for an institution that has not yet been incorporated. Numbers are projections, calibrated against named comparable deployments and disclosed source data. The publisher (SARGVISION Intelligence Pvt. Ltd.) is also the founding promoter of the proposed Centre. This document was not externally peer-reviewed prior to v1.0 publication; external peer review and reviewer attribution are planned for v2.0, alongside the inaugural Eastern India AI Index. Readers are encouraged to engage with the document on the merits of its argument and to verify its numerical claims against the sources set out in Annexure O (Bibliography and Source Register).