

The AI-Data Skills Framework

A Playbook for Industry & Civil Society



Standardizing Skills and Scaffolds for
India's AI-Data Workforce

SABEENA MATHAYAS, MAHIMA TANEJA, VAIDEHI SAHASRABHOJANEE, DEEPTI

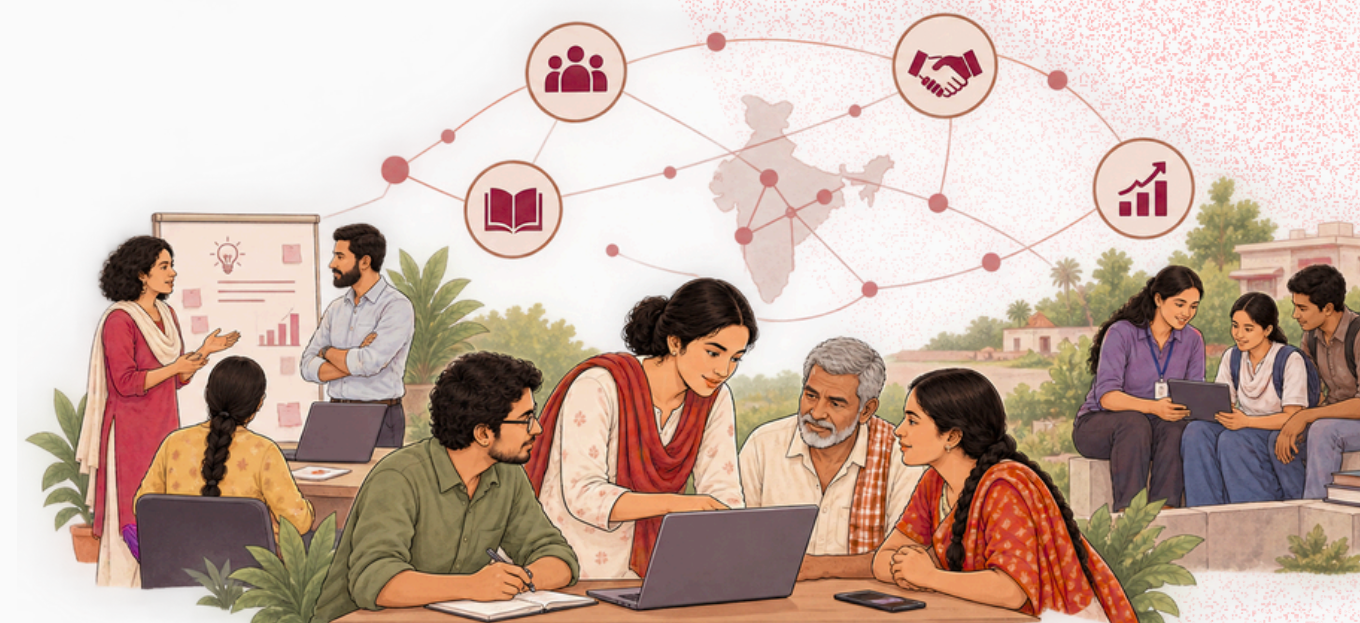




Table of Contents

Why AI-Data Skills Matter **01**

Who are AI-Data Workers? **03**

AISF Snapshot **05**

Notes on Reading the Framework **11**

Grounding the Framework in
Worker Experience **12**

Putting AISF into Practice **16**

Way Forward **24**

Related Resources **26**

References **27**

Acknowledgements **28**



Why AI-Data Skills Matter

As AI adoption grows, so does the range of human work involved in generating, preparing, enriching, validating and maintaining the data that AI systems rely on.

This work is typically outsourced to Business Process Outsourcing (BPO) firms and crowd-work platforms concentrated in Low and Middle-Income Countries, where labour costs are lower and English-language fluency and data literacy are higher (Tubaro et al., 2022). India has emerged as a major hub for this Humans-in-the-Loop (HITL) work. India is already a significant hub for this work: its data annotation market is projected to reach USD 7 billion and employ up to one million workers by 2030 (NASSCOM, 2021). This growth is taking place alongside wider investments in AI capabilities and skills, including an INR 10,300 crore commitment to the India AI Mission and Centres of Excellence for AI (Press Information Bureau, 2025) and a recent pledge to train one crore youth in AI skills within a year (PMINDIA, 2026).

AI-data work spans BPOs, online labour platforms, technology companies and social enterprises, including opportunities for workers in rural areas and Tier 2 and 3 towns.



INR 10,300 CRORE
committed over five years to the IndiaAI Mission and Centres of Excellence for AI



USD 1.7 TRILLION
AI could potentially contribute to India's economy by 2035



16%
India's share of the global AI talent pool, nearly one in six AI professionals





These roles involve different combinations of technical skills, contextual knowledge, human judgement and accountability. Yet the systems used to define roles, recognise existing capabilities and structure training and progression have not always kept pace with how the work is evolving.

This creates a practical challenge for employers and organisations working with AI-data workers:

- How do we define roles and skills consistently?
- How do we identify training needs? And how can workers better understand and demonstrate the expertise they acquire?

The AI-Data Skills Framework (AISF) provides a common language for addressing these questions. It maps AI-data work across five worker personas and the competencies associated with them. For industry, it can support role definition, recruitment, skills assessment and workforce development. For civil society organisations and social enterprises, it can help structure training, recognise skills developed through work, and build clearer links between community-based skilling and emerging opportunities in the AI-data economy.



Who are AI-Data Workers?

The AISF represents AI-data workers who **prepare, check, annotate, validate, and maintain the data** that AI systems are trained and run on, validate and evaluate AI outputs, handle edge cases and exceptions, and inject domain-specific human context so the AI stays accurate, safe, and reliable.¹



In practice, AISF covers people who:

- Collect or generate data for AI training. For example, recording voice samples, photographing products, or filming everyday tasks ("egocentric" data collection) for training datasets.
- Label, annotate, or tag data to make it machine-readable by drawing bounding boxes, transcribing audio, categorizing images or text, and so on.
- Review, audit, or moderate AI outputs for accuracy, safety, bias, or policy compliance.
- Handle exceptions and edge cases that automated systems cannot resolve on their own, escalating or resolving them using human judgment.
- Structure, validate, or maintain the datasets used to retrain and improve deployed models.



Defining the AI-Data Workforce: The AISF focuses on the AI-data workers as their knowledge, skills, and attributes are distinct enough from engineers' and people managers' to warrant a professional identity of their own. This recognition is the necessary first step before that identity can be built into India's broader economic and educational systems.

OUTSIDE OF SCOPE

Machine Learning Engineers, Data Scientists, and other AI specialists who focus on the engine: writing code, designing algorithms, building computational infrastructure, and optimizing models for raw capability, speed, and scale. The hand-offs between these roles are iterative and, at times, fluid, which makes strict boundaries hard to draw. However, both are essential to building AI, as they address two different sides of the same equation. AI data workers supply the foundational reality and ethical logic the architecture processes, relying on human cognitive and contextual judgment to translate reality into actionable logic for AI systems that ML engineers and Data Scientists build. High-level technical engineering roles depend as much on the human judgment, domain expertise, and data validation the AI-data workforce provides as they do on code.

1 Ferrone et al. (2026) describe AI work as organized into three tiers: (I) innovators, who build the underlying AI infrastructure and models; (II) facilitators, who deploy, integrate, and maintain AI capability for practical use; and (III) users, whose work is altered by applying AI. Facilitators are a broad family of roles which includes "data facilitators" i.e., data architects, engineers, annotators, curators, and data-quality specialists, as a specific sub-category responsible for making data usable for AI systems. AI-data workers, as defined in the AISF, are this sub-category.



AISF Snapshot

The AISF maps the AI-data value chain across five professional personas, each defined by a “Human Advantage” i.e., the capability that continues to make human labour valuable. Each persona is mapped to an NSQF/NCrF band (instead of a single level), because progression in AI-data work is layered and iterative, and not strictly linear.

Persona	Human Advantage in AI Data Value Chain Phase*	NSQF / NCrF	Employment type	Opportunity	‘So What’ for Industry and CSO Stakeholders: Utility & Persona-Specific Recommendations
Data Associate 	Collects or generates real-world data and establishes “Ground Truth” at the Dataset Generation stage.	Level 2.5–3.0 (Grade 10–12 / Early Vocational Certificate)	Gig / Platform Workers	This persona offers women a flexible, low-barrier entry into digitally-embedded livelihoods, but piece-rate gig work risks precarity and instability.	Industry: • Use this persona to structure large-scale, decentralised sourcing for high-volume data sets, • Use identified competencies to design screening evaluations and training tasks. • Shift compensation models from pure piece-rate volume toward quality-gated batch bonuses





Persona	Human Advantage in AI Data Value Chain Phase*	NSQF / NCrF	Employment type	Opportunity	'So What' for Industry and CSO Stakeholders: Utility & Persona-Specific Recommendations
 Data Associate	Collects or generates real-world data and establishes “Ground Truth” at the Dataset Generation stage.	Level 2.5–3.0 (Grade 10–12 / Early Vocational Certificate)	Gig / Platform Workers	This persona offers women a flexible, low-barrier entry into digitally-embedded livelihoods, but piece-rate gig work risks precarity and instability.	CSOs: • Introduce skill-equivalence programmes that recognise and build on the digital skills women already have to enable them to take on supplementary/gig work under this persona • Design training and peer-support programs focused on concrete, foundational digital and technical skills (secure logins, device troubleshooting, digital etiquette) instead of abstract AI theory or general digital literacy



Persona	Human Advantage in AI Data Value Chain Phase*	NSQF / NCrF	Employment type	Opportunity	'So What' for Industry and CSO Stakeholders: Utility & Persona-Specific Recommendations
Data Enricher 	Maps complex real-world information into precise, machine-readable taxonomies across the Dataset Generation → Model Training stages.	Level 4.0 (Higher Secondary Vocational / Undergraduate Entry)	Gig/Platform and/or Salaried/ Contracted	Commonly associated with the "Data Annotator" role, this category encompasses the spectrum of data labelling and annotation roles in AI training. Requires sustained focus but is often managed with high-turnover, low-touch methods. Also have to manage edge cases.	Industry: • Use this persona's competencies for worker training and certification. Standardize training on data modalities, technical skills, and transversal skills independent of processes and project-specifications. • Implement ambiguity-escalation protocols • Support ergonomic interventions against visual and physical fatigue; • Introduce tiered wages based on cognitive load. CSOs: Use the persona to identify the next tier of skills Data Associates can train toward, and support progression pathways.



Persona	Human Advantage in AI Data Value Chain Phase*	NSQF / NCrF	Employment type	Opportunity	'So What' for Industry and CSO Stakeholders: Utility & Persona-Specific Recommendations
Data Auditor 	Resolves ambiguity using empathy, sarcasm detection, and cultural nuance across the Dataset Generation → Model Training stages.	Level 4.5–5.0 (Advanced Diploma / B.Voc / Undergraduate Degree)	Salaried or Contracted	This persona is central to risk mitigation and is the frontline defence against AI hallucinations, toxicity, and systemic bias. Includes AI Data Quality Manager, Content Moderator, AI Quality Analyst etc. Most likely to progress along an enterprise's people-management track.	Industry: • Recognise critical thinking and comfort with complexity as a core transversal skill; • For content-moderation workflows, institutionalise mental-health debriefing and limit continuous exposure hours; • Train auditors to document and defend “grey-area” judgments to engineering teams; • Move workers from piece-rate to stable, contracted salaries commensurate with their expertise.





Persona	Human Advantage in AI Data Value Chain Phase*	NSQF / NCrF	Employment type	Opportunity	'So What' for Industry and CSO Stakeholders: Utility & Persona-Specific Recommendations
Data Analyst 	Identifies systemic logic failures and detects “drift” across the Model Training → Model Maintenance stages.	Level 5.5–6.0 (Professional Bachelor's Degree / Post-Graduate Diploma)	Salaried or Contracted	Operational bridge between the annotators and the ML engineers. Responsible for fixing dataset logic failures and throughput bottlenecks, but rarely hold the administrative authority to alter underlying engineering architecture.	Industry: - Align compensation with pipeline-health metrics — SLA adherence, IoU improvement, error reduction; - Train for strategic influence, building analysts' capacity to synthesise conflicting data and communicate root-cause problems persuasively to non-technical leadership. Professional development must pivot away from basic software execution toward abductive reasoning



Persona	Human Advantage in AI Data Value Chain Phase*	NSQF / NCrf	Employment type	Opportunity	'So What' for Industry and CSO Stakeholders: Utility & Persona-Specific Recommendations
Data Specialist 	Deconstructs professional wisdom into modular reasoning chains across the Model Maintenance → Dataset Refinement/Development → Model Training stages.	Level 6.5–7.0 (Master's Degree / Expert Professional)	Salaried or Contracted	Enterprises depend on this persona to deconstruct professional wisdom into modular reasoning chains that autonomous agents can emulate. Includes those (1) who guide the strategic direction of datasets for model development and maintenance and (2) domain experts who train the agent or model for targeted use cases, by breaking down their knowledge into rigid, sequential Chain of Thought logs.	Industry: • Offer competitive, consultancy-grade or equity-based compensation to retain this talent; • Focus training on prompt engineering and cognitive deconstruction — teaching experts to break their own expertise into machine-readable steps.

For a complete breakdown of the knowledge, technical, and behavioural competencies under each persona, and how they map onto the data value chain, see [AI-Data Skills Framework \(AISF\)](#).



Notes on Reading the Framework

BANDED MAPPING

Roles map to NSQF bands, instead of rigid single levels. For example, a Language Data Associate recording basic colloquial audio operates at a lower NSQF level than a Healthcare Data Associate recording responses to clinical queries, even though the core task of audio recording is similar.

SPECIALISED VS. LINEAR PROGRESSION

Progression in AI-data work rarely follows a linear ladder. Workers advance along two main tracks: a Domain Specialisation track and a People and Operations Management track (moving from individual contributor to leadership roles).

PSYCHOLOGICAL AND OPERATIONAL VALUE OF PERSONAS

Replacing transactional terms like “click-worker” with professional titles like “Data Associate” restores psychological legitimacy and self-efficacy, a meaningful lever for reducing attrition among marginalised workers. Personas also enable empathy-based training design: a facilitator can ask concretely, “What tools does the Data Auditor struggle with?” instead of addressing technical gaps in the abstract.

A detailed breakdown of the knowledge, technical, and behavioural skills for each persona is provided in the companion reference, [AI-Data Skills Framework \(AISF\)](#).





Grounding the Framework in Worker Experience

Organisational Scaffolds

Before turning to what industry and civil society should each do with the AISF, it helps to see the evidence those recommendations rest on. The ‘Women Workers in India's AI Data Economy’ report — a companion piece to the AISF based on in-depth interviews with workers, coordinators, and managers at Myna Mahila Foundation and NextWealth documents the organisational support structures that make women's participation and retention in AI-data work possible in practice, what workers identify as their most important unmet needs, and the structural gap that explains why some of those needs cannot be addressed through organisational action alone. The recommendations in this toolkit draw on this evidence base.



Grounding the Framework in Worker Experience

What support exists

Gig Workers – Myna Mahila

Community Mediator Support:

- Trusted, personally known community mobilisers;
- Intensive home visits by female facilitators for the first 10–12 days at entry, with active effort to prevent drop-out;
- Direct phone/WhatsApp access to coordinators at short notice for grievance redressal and troubleshooting;
- Support in opening bank accounts to receive payments

Infrastructure and Connectivity Support:

- Facilitators provide hotspot data from their own phones where home data is insufficient.
- Local community centres which gig workers can use to attend trainings or upload large data recordings.

Recognition:

- Verix certifications recognising workers' experience and contributions

BPO Workers – NextWealth

Onboarding and Training Support:

- Structured onboarding and a Process Knowledge Test framework for initial training;
- Continuous training and ramps for new processes

Infrastructure Support:

- Accessible, air-conditioned offices in Tier 2/3 towns; most workers live within 25km radius
- Shuttle services provided for those who need it

Supportive Management:

- Team leads as an accessible, approachable first point of contact; managers described as understanding of family emergencies;
- A collegial environment supporting skill and confidence building, especially for first-time formal workers.

Recognition:

- High-performance recognition through prizes, certificates, and public ceremonies.



Grounding the Framework in Worker Experience

	Gig Workers — Myna Mahila	BPO Workers — NextWealth
What workers asked for	1. Continuity of work above all else “Work should come regularly.” — Susan* “There should be continuous work.” — Jaya* 2. More training in digital and technical skills, with explicit interest in learning about AI. “These days people say that AI is growing a lot, so it would be good to learn about that. I have heard that there is a lot of work in it and that people’s skills are increasing through it.” — Priya* 3. Access to laptops to unlock a wider range of task types and higher-value work.	1. Upskilling — English communication training and advanced Excel, named as the clearest career bottleneck: “If I know English well, I can talk directly with others, including clients and management.” — Isha* 2. Infrastructural upgrades such as adequate washroom infrastructure relative to workforce size and staggered break windows. 3. Female supervisors to enable safe escalation of menstrual health and other needs. 4. Supportive policies such as menstrual leave, a five-day work week, and mitigation of eye strain and physical fatigue from prolonged screen time.

*Names of workers have been anonymised to protect confidentiality.





Grounding the Framework in Worker Experience

	Gig Workers — Myna Mahila	BPO Workers — NextWealth
Key structural gap	Continuity of work is a procurement and funding problem, not an organisational design problem — Myna Mahila cannot guarantee project availability without sustained client demand and grant funding.	Access to additional training, female supervisors, and infrastructural fixes are implementable by the organisation; however, the pay floor is a structural labour-market issue that internal scaffolding cannot address.

Thus, while internal scaffolding is essential, it operates under distinct market and structural constraints. For gig workers, where funding and project continuity cannot be guaranteed, organisations should move toward transparency and forecasting to help women anticipate dry periods and manage financial risk. BPO operations can directly resolve internal bottlenecks such as infrastructure and female leadership representation, but remain bound by industry pay floors, a structural labour-market issue that internal scaffolding alone cannot solve.





Putting AISF in Practice

Recommendations for Industry: Operations, ESG and Strategic Value

The private sector bears the primary operational and financial burden of mobilising, training, and sustaining the HITL workforce that builds the AI-data value chain. Because this workforce is largely informal and platform-mediated, however, industry's investment in human capital often goes unrecognised in official economic metrics, even though that investment is already substantial.

By adopting the AISF, private enterprises can translate this operational overhead into quantifiable, strategic assets.

THE PERSONA-LEVEL ACTIONS ALREADY SET OUT IN THE FRAMEWORK SNAPSHOT ABOVE ARE THE STARTING POINT.

For example, industry can use the persona-level list of knowledge, technical skills, and transversal skills to recruit, train, assess, certify, and upskill their workforce and institute persona-specific safeguards and guardrails (such as ambiguity escalation protocols, mental health support, critical thinking etc). The recommendations below build outward from individual roles to sector-wide strategy and government partnership.



1. Formalising and Quantifying Grassroots Impact

To build effective AI data pipelines, industry partners need to reach deep into peri-urban and rural geographies to mobilise latent labour pools. This requires navigating socio-cultural barriers such as securing a woman's “permission to participate” from her family and establishing hyper-local delivery networks. The training these companies provide also extends beyond simple software navigation: it builds basic digital literacy, hardware management, and professional self-efficacy from the ground up.

THE AISF GIVES INDUSTRY A VOCABULARY TO QUANTIFY THIS ON-THE-GROUND EFFORT FORMALLY.

- **Translating Overhead into Asset Development**

At entry, a highly technical education background is not required for first three personas, and where participation of women is currently concentrated. The actual hiring threshold is basic computer fluency, not an engineering or computer science degree. Most technical skills are acquired on the job, developed through repetition and escalation. When guidelines do not work for a new file type, workers raise it with team leads, who take it to the client for validation. This is a structural asymmetry industry should recognise and act on directly: **workers are frequently overqualified for entry-level tasks by formal credential, yet under-recognised for the expertise they actually develop on the floor. This expertise is transferable but currently carries no portable credential**, requiring re-training at each new organisation despite identical skill requirements. By mapping this expertise to the AISF's personas, and certifying them, companies can evidence their asset development, and showcase how they are systematically moving workers along a standardised capability spectrum. Training a woman to draw bounding boxes is no longer just task-prep; it is the formal development of a **Data Enricher** equipped with "Structural Dexterity."



- **Evidencing ESG and Ethical Supply Chains**

Global technology clients and regulators are placing growing scrutiny on AI-data supply chains, seeking transparency on labour conditions and fair wages. Mapping workforce to AISF personas lets Indian data enterprises report Environmental, Social, and Governance (ESG) contributions with auditable evidence that they are building ethical human infrastructure and supporting upward social mobility. The organisational scaffolds observed among current partners are exactly this kind of evidence: at NextWealth, for instance, 84 percent of employees report feeling safe being employed locally, a result of the company's deliberate choice to place delivery centres within the towns where its largely female workforce already lives (NextWealth Social Impact Report, 2024).

2. Strategic Leverage through Accreditation and Government Partnerships

As the AI industry matures, the integration of internal company training with national accreditation frameworks is critical for long-term sustainability. The AISF serves as a bridge between dynamic private-sector needs and rigid public-sector compliance, utilizing a "Parallel Governance" strategy.

- **Expanding the Recognised Labour Force**

By aligning their training modules with the AISF, and partnering with recognised Awarding Bodies such as SSC-NASSCOM, companies can map AI-data workers to national skilling frameworks and formally move a fragmented workforce into a nationally recognised, credentialed talent pool. It ensures that the millions of workers contributing to the AI economy are officially counted, effectively expanding the recognized digital labor force of the country.



- **Positioning as a Strategic Investment Partner**

As enterprises formalise this workforce, their relationship with the government can shift from standard service vendor to strategic investment partner in national initiatives like the India AI Mission. By offering a standardized, highly skilled, and credentialed HITL workforce, the industry establishes a unique competitive moat.

- **Attracting Foreign Direct Investment (FDI)**

This formalization provides immense geopolitical and economic leverage. Global AI developers require massive, secure, and highly accurate datasets to train their models. An industry that can guarantee an ethically managed, structurally verified, and progressively skilled workforce—backed by a framework like the AISF—is a magnet for high-value international business. It positions the domestic industry to attract top-tier Foreign Direct Investment, cementing the market as the premier global destination for sovereign, high-quality AI development and deployment.



Recommendations for Civil Society: Scaffolding Empowerment and Training

Civil society organizations (CSOs), grassroots NGOs, and social enterprises serve as the critical bridge between the formal technological sector and the informal, marginalized workforce. While industry provides the infrastructure and government provides the regulatory framework, civil society provides the socio-emotional scaffolding required to make the AISF actionable on the ground. The following recommendations outline how civil society can leverage the AISF to drive genuine socio-economic empowerment.

1. Awareness, Advocacy and Community Relations

The transition into the digital economy is not merely a technical challenge for rural and semi-urban women; it is a profound socio-cultural shift. Fieldwork consistently demonstrates that community support and localized learning opportunities act as the primary gateway for women to seek, secure, and retain any form of employment, let alone highly novel AI-data work (Worker Experience Report, forthcoming). In this context, CSOs can:

- **Facilitate Peer Networks and SHGs:** Civil society actors can facilitate community-level support, mediation, peer mentoring, and local digital collectives. For women navigating gendered household and community norms, these groups provide a safe, culturally sanctioned space to build digital confidence, secure household buy-in, and reduce early dropout.
- **Mediate Industry Relationships:** Civil society can act as a buffer between marginalised gig workers and corporate employers, using the AISF's professional vocabulary to advocate for the dignity of labour and to keep platform-mediated work from being treated as commoditised, invisible piece-meal work.



- 
- **Negotiate for Flexibility and gender-friendly workplace policies:** A recurring reality for women is the burden of domestic responsibilities and care work. CSOs can support in designing flexible and asynchronous work models, so that gig platforms and industry accommodate domestic realities without penalising a worker's earning potential or progression along AISF pathways. Myna Mahila's own model illustrates what this looks like in practice: after research found that fragmented participation was common, it set up temporary centres within a five-minute walk of workers' homes, let women come in during school hours, work in short stretches from home, and reschedule recording calls around care schedules, accommodating the structure of gig workers' days instead of expecting them to reorganise it.

2. Community-Centred, Holistic Program Design

To effectively transition workers from basic digital literacy to advanced AI data synthesis, program designers, trainers, and facilitators must abandon generic "computer classes" in favour of holistic, empathy-based program design.

THE AISF PROVIDES THE EXACT PEDAGOGICAL BLUEPRINT REQUIRED TO BUILD THESE TARGETED INTERVENTIONS.

- **Progressive Mapping of Digital Skills:** Trainers should use the AISF to scaffold technical education. Programs must start by addressing the immediate realities of the Data Associate (Level 1)—focusing on basic device familiarity, secure login practices, troubleshooting, and strict adherence to standard operating procedures (SOPs). Some of this is currently learned informally, from peers or from workers' own children, instead of a curriculum, a pathway the AISF does not capture but which program design should. Further, rather than overwhelming entry-level workers with complex coding concepts, curricula should progressively map to the Data Enricher (Level 2), teaching them how to navigate hierarchical menus, manage overlapping digital windows, and utilize specific platform hotkeys to increase their piece-rate efficiency.



- 
- **Cultivating Critical Thinking:** A holistic program must actively train the cognitive transitions outlined in the framework. Instructional designers and facilitators could design scenario-based learning modules that move workers from foundational compliance (following the rules) to applied critical thinking (choosing between rules based on context). For workers advancing to *Data Auditor* (Level 3) roles, training must simulate resolving "grey area" policy disputes, detecting cultural nuances, and evaluating the factual accuracy of AI outputs, — mirroring the escalation pathway already used at NextWealth, where workers raise novel edge cases with team leads, who take them to the client for validation.
 - **Embedding Behavioural and Transferable Skills:** Because technical tools in the AI economy depreciate quickly, behavioural attributes are the most durable asset a worker can develop — professional digital etiquette, managing the cognitive load of high-precision screen work, and ethical handling of Personally Identifiable Information (PII). By focusing on these transferable skills, civil society ensures that even if a specific AI platform becomes obsolete, the women are equipped with the critical reasoning, adaptability, and professional resilience required to thrive elsewhere in the modern digital economy. This should further extend to workers' own AI-value-chain awareness - understanding not how models work, but where their own contribution sits in the chain. Workers who do not know where their data goes cannot easily articulate the value of their own contribution; workers who do describe their role with a confidence CSOs should be actively cultivating.



- **Designing programs keeping in mind women’s entry level pre-conditions:** Other than knowledge and technical skills already identified under the personas, women working as AI-data workers in either gig or full-time roles, additionally might need:

Infrastructure and devices	Smartphone (with adequate access in terms of time and quality) or laptop (based on nature of work)
	Stable and adequate Internet connectivity (to upload data sets)
Enabling environment	Mobility support or freedom to commute
	Family/care work support
	Quiet environment/space at home (for gig workers)
	Informal and formal learning and troubleshooting pathways (from children, peers, community mediators)

The framework on its own does not capture how workers actually acquire competencies, navigate normative constraints, or what the pre-conditions of their entry into this work are such as device fluency, ecosystem support, and informal learning pathways. These entry-level realities are the starting conditions program design has to work with.



Way Forward: Implications for Practice

PULLED TOGETHER, THE EVIDENCE AND RECOMMENDATIONS ABOVE POINT TO FOUR IMPLICATIONS FOR HOW INDUSTRY AND CIVIL SOCIETY CAN APPLY THE AISF ON THE GROUND.

1. The AISF is necessary but not sufficient on its own, for three reasons:

- a. Workers need enabling conditions such as device access and fluency, formal and informal pathways to learn and troubleshoot, and supportive household and social norms to enter this work and to participate and progress in it over time. These conditions are gendered, and need organizational and civil society support to ensure equitable participation.
- b. Workers need portable credentials: The AISF names and recognises the skills workers already build on the job, but that recognition falls short without portability. A credential (such as a skills passport) only has value if it travels with the worker, if regulators and employers recognise it, and if workers have the bargaining power to use it elsewhere. This is why the accreditation and government-partnership recommendations matter as much as the framework itself.
- c. Third, workers should understand where their own work sits in the AI-data value chain, for the AISF to hold meaning for the workers themselves, and unlock progression pathways. AI skilling for data roles should include a disclosure and transparency layer, so workers know what their data trains and how it is used, not only how to do the task.



- 
2. **Organisational scaffolding is necessary but structurally limited.** Non-profits like Myna Mahila cannot guarantee project continuity without sustained funding, and enterprises like NextWealth cannot raise the pay floor through internal design alone. Expecting social-purpose organisations to compensate for income precarity and the absence of portable credentials through program design alone is not sustainable. Procurement standards, labour protections, and regulatory pay floors need an ecosystem response that sits above any one industry or civil-society actor.
3. **Gender-intentional design is not optional.** The full lifecycle of an AI-data worker is gendered: women enter this work for care-compatible reasons, stay for internal progression and workplace community, and leave for marriage or relocation. Proximity and flexibility are not preferences, but are structural requirements imposed by care burdens and mobility constraints. Any framework or policy that treats this workforce as gender-neutral will reproduce the exclusions it claims to address.
4. **Recognition matters.** A credential is valued first for what it signals within a worker's household and community, not only for what it unlocks in the job market. Measurement and evaluation frameworks that assess impact only through income and employment outcomes will undercount what this work does for workers' self-efficacy, dignity, and standing, and recognition programmes aimed only at employers will miss the relational value that makes recognition meaningful to this workforce in the first place. National measurement frameworks need to catch up too: official surveys should capture the distinct, micro-tasked nature of AI-data refinement directly, rather than folding it into generic clerical categories or leaving it uncounted altogether.

None of these implications can be resolved by one actor alone. The AISF depends on industry, civil society, and government each doing the part only they can do, and doing it in coordination rather than in sequence. What determines whether India's AI-data workforce moves from informal to recognised is not the framework itself, but whether that coordination actually happens.



Related Resources

1. [Webpage](#)
2. [Toolkit | September 2026](#)
3. [Worker Experience Report \(Forthcoming\)](#)

APPROACH | HOW AISF WAS DEVELOPED

The AISF is the outcome of an intensive, voluntary consultation with AI-data enterprises and their workers in India, conducted through DACUM (Developing A Curriculum) workshops, Key Informant Interviews, and In-Depth Interviews with 43 participants across three organisations: **NextWealth Entrepreneurs Pvt. Ltd.** (a BPO/KPO with salaried employment contracts), **Myna Mahila Foundation** (a non-profit with a crowdsourcing structure for gig and micro-work), and **Karya** (DAIA Tech Pvt. Ltd., a platform-based social enterprise sourcing workers through community organisations and self-help groups). The research was anchored by LEAD's Gender x Digital (GxD) hub and approved by IFMR's ethics committee.





References

Fragiadakis, G., Diou, C., Kousiouris, G., & Nikolaidou, M. (2024). Evaluating human-AI collaboration: A review and methodological framework. arXiv preprint arXiv:2407.19098.

IEEE. (2025). P7015 – Draft Standard for Data and Artificial Intelligence (AI) Literacy, Skills, and Readiness. New York: The Institute of Electrical and Electronics Engineers, Inc.

International Organization for Standardization (ISO). (2023). Development of a competency framework for standards development professionals. Geneva, Switzerland: ISO.

NASSCOM. (2021). Data Annotation: Billion Dollar Potential Driving the AI Revolution. NASSCOM.

NextWealth. (2024). Social Impact Report.

Norton, Robert E. (1997). DACUM Handbook. Second Edition. Leadership Training Series No. 67. Columbus, Ohio: Center on Education and Training for Employment, College of Education, Ohio State University.

Press Information Bureau Research Unit. (2025). Transforming India with AI. 30 December.

PMINDIA. (2026). PM announces AI skilling of 1 crore youth in one year. 15 August. https://www.pmindia.gov.in/en/news_updates/pm-announces-ai-skilling-of-1-crore-youth-in-one-year/.

Prosus, BCG, and MeitY. (2026). “AI for All: Catalysing Jobs, Growth, and Opportunity.” Prosus News & Insights. 29 May. <https://www.prosus.com/news-insights/2026/ai-for-all-catalysing-jobs-growth-and-opportunity>.

Sarangi, A., Sahasrabhojane, V., Taneja, M. & Jhurani, Y. (2025). Decoded: Women and the Future of Digital Work in India. GxD hub, LEAD at Krea University, IFMR.

Tubaro, P., Coville, M., Le Ludec, C., & Casilli, A. A. (2022). Hidden inequalities: the gendered labour of women on micro-tasking platforms. Internet Policy Review, 11(1), 1-26.





Acknowledgements

This framework builds on the foundational work of The/Nudge Institute, whose earlier iteration (the Common Digital Skills Framework - CDSF) anchored this research from the outset. We are grateful to the management and teams at NextWealth, Myna Mahila Foundation, and Karya (DAIA Tech Pvt. Ltd.) for their sustained engagement — opening their processes and platforms to us, facilitating worker interviews, and offering critical validation of the framework.

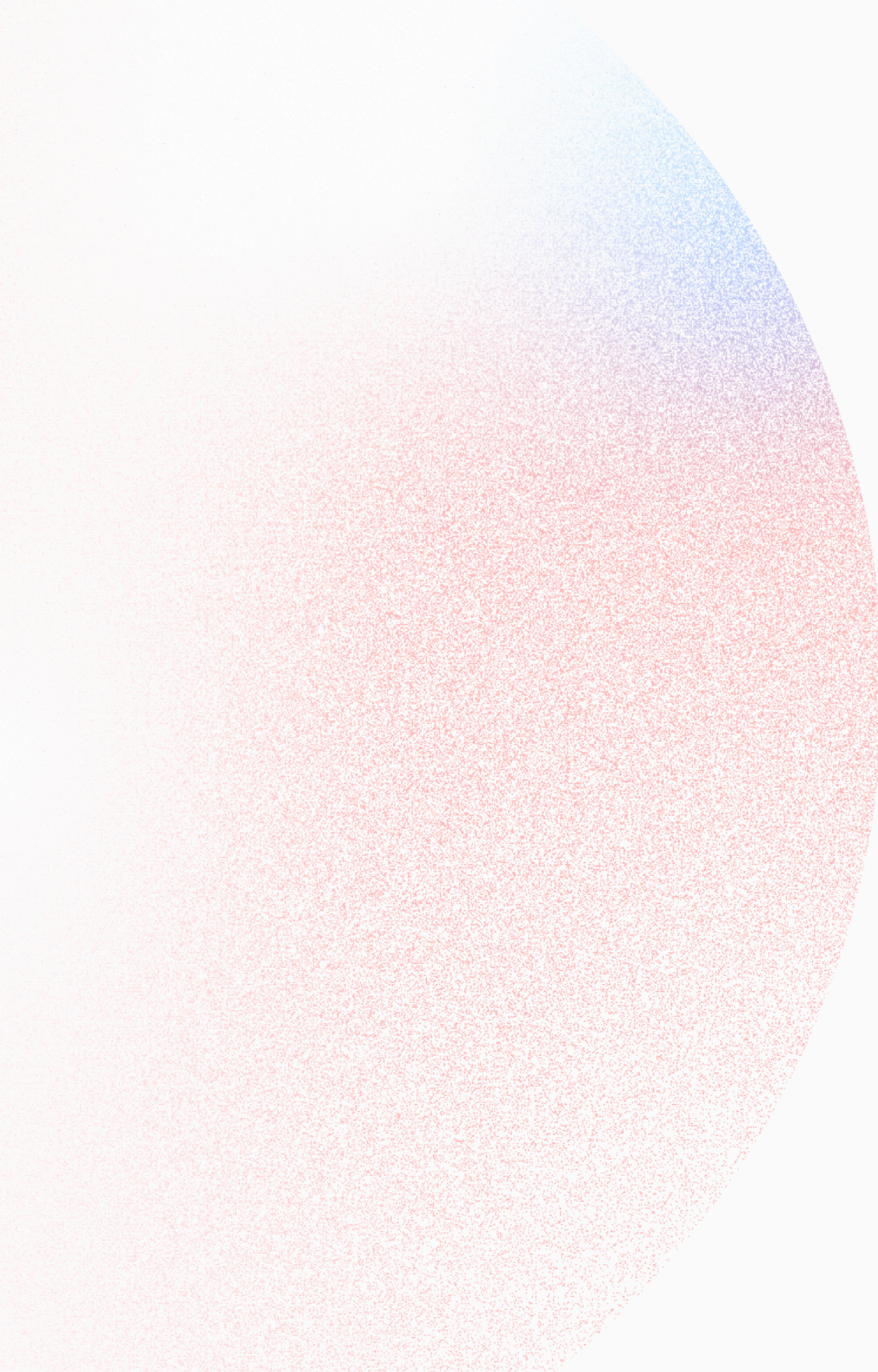
Most importantly, we thank the AI-data workers across NextWealth and Myna Mahila whose knowledge and experience this framework seeks to make visible and credential.

SABEENA MATHAYAS, MAHIMA TANEJA, VAIDEHI SAHASRABHOJANEE, DEEPTI

© LEAD at Krea University (IFMR), 2026.

Illustrations used were developed using AI-assisted tools, with human input and refinement.





The AI-Data Skills Framework (AISF) builds on The/Nudge's Common Digital Skills Framework (CDSF) and is part of Sanmati 2.0, a collaboration between the Gender x Digital (GxD) hub and The/Nudge exploring women's participation and opportunities across India's evolving digital value chain. GxD hub is an initiative of LEAD at Krea University (IFMR).



www.gxdhub.org



www.thenudge.org

GO TO COVER

