

AI-Data Skills Framework (AISF)

Defining a Common Skills Language for India's Future-Ready Workforce

September 2026



Our Collaborators



Foreword

AI could transform 38 million jobs in India by 2030, as new forms of work emerge across the AI-data value chain. This creates an opportunity to widen access to digital employment and shape the quality of these pathways early. At LEAD, we are keen to understand where these opportunities are emerging across sectors and geographies, who can access them, and the enabling conditions—skills, credentials, market linkages, and institutional support—that can support progression into more complex and higher-value work.

We hope the AISF provides a starting point for industry, skilling bodies and practitioners to build this understanding and strengthen the pathways that develop as the market grows.



Sharon Buteau
Executive Director
LEAD at Krea University



Kanishka Chatterjee
Managing Director, the^delta
Ecosystem Advisor, The/Nudge

Frameworks rarely excite people. But as AI changes the nature of work, we need ways to understand where people's existing abilities fit, what they can build on, and which opportunities they can access. AISF is our effort to make these connections visible across rural and urban India. For women in particular, participation in the AI economy requires more than training. It needs recognition of their abilities, credible pathways into paid work, and support that reflects their realities. This is the work of democratising opportunity: helping people become visible for the value they can create, and enabling employers to recognise that value.

Purpose of the Toolkit

This toolkit provides an actionable, multi-stakeholder framework to advance the AI-data workforce driving India's growing AI economy. As AI-related services and employment opportunities in India expand, there is a growing need to better understand the roles emerging across the AI-data value chain, the competencies they require, and the pathways they can create for workers.

By defining these roles and competencies, the framework provides a common language that can inform labour-market measurement and workforce planning, align skilling with emerging demand, support recognition and formal credentialing of existing skills, and create clearer pathways for mobility and progression.

For women and other groups who may enter the digital economy through these emerging forms of work, strengthening this skills infrastructure can also help identify where opportunities are concentrated and what conditions can support sustained participation and access to better-quality employment as the sector evolves.

The toolkit is designed for four key stakeholder groups, with clear applications across the AI-data workforce ecosystem:



Government & Regulators

Serves as a technical blueprint to map AI-data roles to NSQF/NCrF levels, becoming a basis for curriculum design and to NCO-2015 codes, capturing this workforce in national statistics



Industry

Delivers a standardized taxonomy to streamline skilling, build transparent career pathways, and provide auditable metrics for global clients.



Civil Society Organizations

Provides a tool for designing community-centered training, negotiating flexible work arrangements, and advocating for worker dignity.



AI Data Workers

Grants a formal occupational identity, and the basis for portable, recognized credentials that have the potential to unlock long-term job security.

Note: This technical framework is complemented by toolkits for policy makers and the industry, and a companion narrative report that captures the lived experiences and gendered barriers that women workers face in India's AI data economy.

How to Navigate this Toolkit

Page 08-11

Introduction & Methodology

- **Context & Gap:** Situates the AISF within India's rapidly expanding AI economy and identifies gaps in the recognition, classification and measurement of the emerging AI-data workforce.
- **Defining the AISF:** Introduces the framework as a standardised competency mapping of AI-data roles and skills to national skills and qualification systems, including the NSQF and NCrF, to support recognition and formal credentialing.
- **Methodology:** A DACUM-led occupational analysis conducted across 5 workshops and 43 participants with three key partners/organizations.
- **AISF Boundaries:** Defines the AISF's scope, target workforce, and distinction from related AI/ML engineer roles.

Page 12-13

Demystifying the AI Data Value Chain

Maps the AI-data value chain from data generation to model maintenance, highlighting worker roles across BPO, gig and social enterprise models.

Page 14-15

Understanding the AISF – Blueprint and Parameters

Framework Architecture: Introduces the five foundational parameters: Knowledge, Technical Skills, Transversal and Behavioural Skills, Complexity, and Accountability.

Framework Snapshot: Maps the five personas from Data Associate to Data Specialist, including their distinctive human capabilities, accountability levels, and corresponding NSQF/NCrF bands.

Page 16-30

Detailed Personas, Frictions, and Utility

Competencies and Frictions: Moving from persona to operational realities, this section details the competencies associated with each persona and the common frictions encountered across roles.

Demonstrates how each persona can inform actions by government, industry, and civil society.

Page 31 onwards

Conclusion

Key takeaways and resources for applying the framework across use cases such as workforce planning, policy, skilling, and credentialing by different stakeholder groups.

About

THE/NUDGE

www.thenudge.org

The/Nudge is an action institute building resilient livelihoods to alleviate poverty. We work with social entrepreneurs, women, farmers, tribals and youth on rural development, agriculture, skilling and economic inclusion, along with 15 central and state government partners. Set up with support from 90+ eminent philanthropists, 40+ corporates and 15+ foundations, The/Nudge is contributing towards a "**poverty-free India, within our lifetime**".



www.gxdhub.org | www.ifmrlead.org

The Gender x Digital hub (GxD hub) is a cross-disciplinary platform dedicated to **advancing meaningful digital connectivity among women and girls in India**. GxD hub is an initiative of LEAD at Krea University (IFMR). LEAD at Krea University is an action-oriented research centre housed at the Institute for Financial Management and Research (IFMR), a not-for-profit society which is also the Sponsoring Body of Krea University.

Acknowledgements

This work would not have been possible without the partnership and support of several colleagues and organisations. We are grateful to our colleagues at The/Nudge for their continued support and partnership on this work, and for developing the first iteration of the framework (CDSF) – Anurag Vaishnav, Mohit Chelani, Priyanshi Toshniwal, Chetana Koulagi and Meera Vijay. We also thank Kanishka Chatterjee and Anuradha Sharma at The/Nudge, and Sharon Buteau at LEAD at Krea University and Yashita Jhurani for their thought partnership in shaping this work.

We are also deeply grateful to the management and teams at the organisations who made this research possible. At NextWealth, we thank Madhukar Yarra, Sugathan R, and Laxmi Gunupudi for their close engagement with the framework, their guidance throughout, and their willingness to share their processes, facilitate interviews with their associates, and host our field visits. At Myna Mahila, we thank Kirti Shukla, Dheeraj Verma and the entire team for sharing the organization's journey, platform processes, quality assurance and community-engagement processes, and for facilitating interviews with gig workers. At Karya, we thank Manu Chopra, Safiya Husain, and their team, for their valuable inputs on the framework. Finally, and most importantly, we thank the gig and BPO workers who generously gave their time to share their experiences, knowledge, and insights with us. This report exists because of their willingness to be candid about their working lives, and we hope it does justice to what they shared.

Technical Leads: Sabeena Mathayas, Mahima Taneja

Research Team: Deepti, Vaidehi Sahasrabhojane

Editorial Support: Diksha Singh

Report Production & Design: Keerthana Ramaswamy, Sakthivel Arumugam

© LEAD at Krea University (IFMR), 2026.

Illustrations in this report were developed using AI-assisted tools, with human input and refinement.

Introduction: Defining India's Human Infrastructure for AI

Through initiatives like the India AI Mission and the Centers of Excellence for AI, India has committed to an investment of INR 10,300 crore over five years, projecting that AI could add as much as USD 1.7 trillion to India's economy by 2035 (PIB Research Unit 2025). This optimism parallels the global language of inevitability, efficiency, automation, and algorithms, and rests on significant digital labour of AI-data workers i.e., the human layer that generates, annotates, manipulates, validates data and maintains AI systems. The opportunity this opens is real and growing — India's data annotation and labelling workforce, for example, is projected to reach 1 million workers as the sector's market value crosses USD 7 billion by 2030 (NASSCOM 2021), with the country's linguistic diversity driving additional demand for language data generation and processing.



**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

Data work for AI models and agents can be a dignified employment opportunity at scale, especially when paired with targeted, competency-based, task-specific training, and organizational scaffolds. Men and women from rural areas and Tier 2/3 towns across India are already powering AI-data systems, aggregated through a range of organizations including AI-specific data BPOs (NextWealth, iMerit), generalist BPOs adapting to AI demand, open online-labour platforms (MTurk, Appen), managed AI-data platforms (Shaip, Playment), in-house data teams at large technology firms, and social enterprises and non-profits (Anudip Foundation, Karya, Myna Mahila) that recruit and skill workers (Sarangi et al 2025). Only a handful of these organizations carry a social purpose beyond throughput by providing the training, community support, and grievance redress that convert this work into a stable livelihood, especially for women, rather than a transaction.

At the national level, the systems for recognizing, training, and progressing this workforce have not caught up either. This workforce remains statistically invisible and absent from formal skilling curricula and competency frameworks. Left unaddressed, this risks leaving the most vulnerable humans-in-the-loop — especially women, who are concentrated in upstream data preparation and downstream service functions of the AI-data value chain (UNESCO 2026) — stranded at the periphery of a progressively digital economy.

What is the AISF?

The **AI-Data Skills Framework (AISF)** is a competency framework that identifies, names, and standardizes the skills of the human workforce that trains, labels, annotates, reviews, and maintains the datasets powering artificial intelligence (AI) systems. It is a recognition of both the potential and the risks opened by the AI-data economy, intended to act as a bridge governance system that recognizes, advocates for, and accounts for an emergent occupation — a digital trade powered by human curation.

The Framework identifies and recognizes the **skills and competencies of a workforce** that currently sits outside conventional job classifications and mainstream skilling systems

Skill Identification



Maps those skills and competencies (across five personas) to India's **National Skills Qualification Framework (NSQF)** and **National Credit Framework (NCrF)**, creating a base for industry-aligned curricula and ensuring that the expertise acquired through AI-data work is portable to other learning and earning opportunities, rather than remaining invisible

Skill Mapping



Provides employers, skilling providers, and workers a shared vocabulary for **recruitment, training, assessment, credentialing, retention, and career mobility**, by creating clear categories of skills that are transferable across domains in the AI-data value chain

Shared Vocabulary



Methodology: How we arrived here

This framework is the outcome of an intensive consultation with AI-data enterprises and their workers in India, conducted through DACUM (Developing A Curriculum) workshops, Key Informant Interviews (KIIs), and In-Depth Interviews (IDIs) with 43 voluntary participants (managers, data associates, team leads, gig workers, community program managers etc) across three organisations:



NextWealth Entrepreneurs Pvt. Ltd. - A structured AI-data BPO/KPO that employs full-time and contractual workers in tier 2 and tier 3 towns.



Myna Mahila Foundation - A non-profit that sources gig workers door-to-door, building on relationships from other community welfare work, largely in low-income urban areas, with expanding presence in rural geographies



Karya - A social enterprise which sources gig workers from rural areas through community-based organizations and self-help groups onboarded onto its proprietary platform.

This is not a representative sample as participation depended on organizations' openness to share information, scheduling, and the constant paucity of time and resources. Against this context, the AISF should be read as a dynamic tool: one that will benefit from wider consultation and continue to evolve with the realities of industry and workers.

The research was anchored by the Gender x Digital (GxD) hub and approved by IFMR's ethics committee.

Defining the Boundaries: Who is the framework about?

The AISF represents AI-data workers who collaborate closely with data scientists, machine learning (ML) engineers, and other AI specialists, but occupy a distinct professional lane. The hand-offs between these roles are iterative and, at times, fluid, which makes strict boundaries hard to draw; that's more a reflection of the opacity of AI as we've come to know it since 2023 than any imprecision in the framework. Behind every model response is data, and data is created by people. If the AISF proves one thing, it is how many people it takes to power AI.

To understand the AISF, it is important to understand the difference between ML engineers, data scientists, AI specialists and the like, and AI-data workers. Both are essential to building AI, but they solve two different sides of the same equation.



AI, ML, data specialists

Machine Learning Engineers, Data Scientists, and other AI specialists and innovators focus on the engine: writing code, designing algorithms, building computational infrastructure, and optimizing models for raw capability, speed, and scale.



AI-Data Workers

The facilitators focus on the steering: curating datasets, annotating nuance, validating outputs, and injecting domain-specific human context so the AI stays accurate, safe, and reliable. They supply the foundational reality and ethical logic the architecture processes, relying on human cognitive and contextual judgment to translate reality into actionable logic.¹

Put simply, ML engineers, data scientists, and AI specialists build how AI computes; AI-data workers govern what AI actually understands. The two must collaborate to bridge human nuance and model logic.

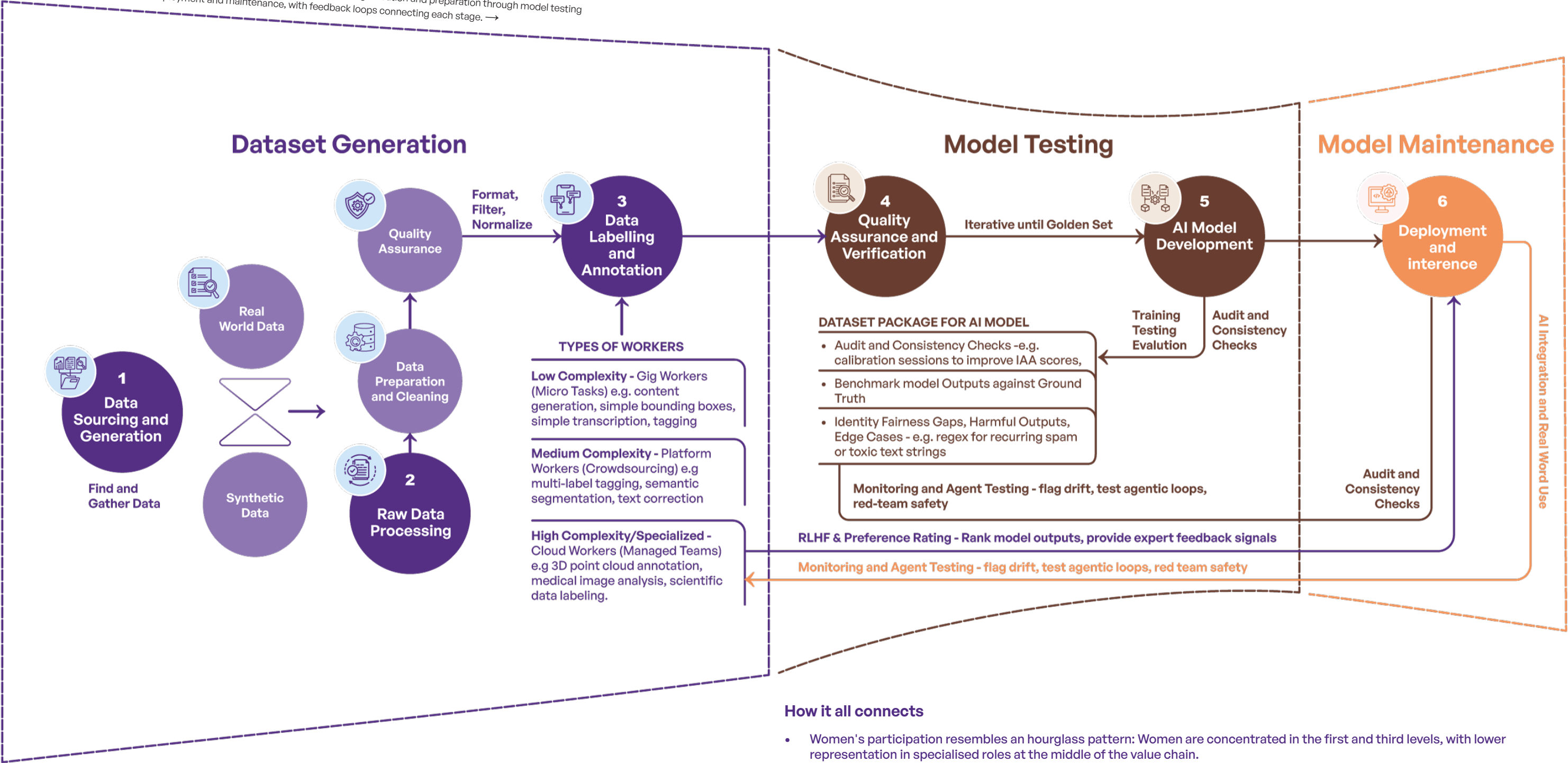
This distinction is what gives AI-data workers formal visibility in the labour ecosystem: high-level technical engineering depends as much on the human judgment, domain expertise, and data validation this workforce provides as it does on code. The AISF focuses on the AI-data workers who are also called AI facilitators (Ferrone et al 2026), as their knowledge, skills, and attributes are distinct enough from engineers' and people managers' to warrant a professional identity of their own. This is the necessary first step before that identity can be built into India's broader economic and educational systems.

¹This distinction is in line with the three distinct categories of AI labour stack posited by Ferrone et al (2026). 1.) innovators, who develop the underlying AI infrastructure and models; 2.) facilitators, the technical workers who deploy, integrate, and enable the use of AI; and 3.) users, whose work is altered by applying AI across the economy. AI-data workers as facilitators provide the human oversight needed when AI systems are developed by innovators and deployed to 'users'.

Demystifying the AI-Data value chain

Read from left to right, the value chain traces how data moves from generation and preparation through model testing and development to deployment and maintenance, with feedback loops connecting each stage. →

While AI is framed as autonomous, supervised machine learning requires massive volumes of high-quality, human-annotated training data to function, depending on "human-in-the-loop" (HITL) techniques like Supervised Fine-Tuning (SFT) and Reinforcement Learning with Human Feedback (RLHF). Further, automated systems return high-confidence predictions immediately, but route low-confidence predictions to human review, creating a permanent structural demand for human labour. The AI-data workforce is thus not a temporary bridge to automation but a persistent requirement that turns low-confidence data into high-confidence intelligence.



How it all connects

- Women’s participation resembles an hourglass pattern: Women are concentrated in the first and third levels, with lower representation in specialised roles at the middle of the value chain.
- Progression is not necessarily linear: Roles draw on different combinations of technical skills, contextual knowledge, judgement and accountability, creating multiple pathways across the value chain.
- The value chain is iterative with multiple feedback loops: Testing, auditing and maintenance feed back into earlier stages, keeping human judgement and quality assurance integral throughout the AI lifecycle.



This builds on two bodies of work. The first is the GSMA (2018) which maps this hidden production process as a five-stage Global Data Value Chain: data acquisition, storage and validation, data processing, data analysis, and monetization, naming the human role at each stage (Capturers, Collectors, Processors, Analysts, Providers). The second is the Competition Commission of India's (2025) framing of the "AI Stack in India," which divides the ecosystem into upstream layers, where data and foundational models are prepared, and downstream layers, where AI is adapted and deployed for real-world use.

Within the upstream layers, the CCI identifies the Data Layer (gathering raw inputs, managing data preparation, and enforcing data governance) as the foundational bedrock of the entire AI ecosystem. In both framings, this is not a simple linear pipeline: continuous feedback loops (upstream data trains models, which are adapted downstream and generate fresh data that flows back upstream) run through vertical support layers such as Computing Infrastructure and Quality Management, Governance and Orchestration, which span the entire structure. Our three-stage grouping reflects this same cyclical logic, and it is the basis on which we developed the AISF's five personas.



The volume of work at each stage still needs validating through an industry-led census or survey of employment. Our current analysis finds AI-data workers largely map onto Stage 1 and Stage 3.²

The AISF acts as the operational interface between AI infrastructure and human-directed refinement by translating the chain's functional stages into clearly defined human capabilities. The framework anchors five professional personas — the Data Associate, Data Enricher, Data Auditor, Data Analyst, and Data Specialist — each representing a localized "Human Advantage" that maps the cognitive, technical, and behavioural parameters required to handle discriminative ambiguity, resolve socio-cultural grey areas, and deconstruct expert intuition into agentic reasoning chains. By structuring these parameters against national standards like the NSQF and NCrf, the AISF transforms what was previously obscured as commoditized "ghost work" (Gray and Suri, 2019) into a standardized, rights-bearing digital trade.

²Women's participation in AI as a whole also follows the same dual pattern with participation squeezing at stage 2. This finding is consistent with UNESCO's recent research, which finds that in South Asia, women in AI are concentrated at the data-preparation end and the downstream service end of the chain, with the least access to the model-building work in between (UNESCO 2026).

The Blueprint: The AI-Data Skills Framework (AISF)

Persona	Value Chain Phase	Employment	Human Advantage	Accountability	NSQF/ NCrF
Data Associate 	Dataset Generation	Gig/Platform Workers	Collecting or generating real-world data and establishing "Ground Truth."	Responsible for immediate output volume following linear, zero-ambiguity SOPs.	Level 2.5–3.0
Data Enricher 	Dataset Generation → Model Training	Gig/ Platform &/ or Salaried/ Contracted	Mapping chaotic reality into machine-readable taxonomies with precision.	Responsible for batch accuracy and navigating discriminative ambiguity in defined menus.	Level 4.0
Data Auditor 	Dataset Generation → Model Training	Salaried or Contracted	Resolving "gray areas" using empathy, sarcasm detection, cultural nuance.	Power user capable of troubleshooting "noisy" data and interpreting complex policy.	Level 4.5–5.0
Data Analyst 	Model Training → Model Maintenance	Salaried or Contracted	Identifying systemic logic failures or "drift."	Bridging raw work with final production models through statistical validation and diagnostics.	Level 5.5–6.0
Data Specialist 	Model Maintenance → Dataset Refinement/ Development → Model Training	Salaried or Contracted	Deconstructing professional wisdom into modular reasoning chains for autonomous agents.	Translating high-level goals into actionable constraints for specialized professional use cases.	Level 6.5–7.0

The AISF provides a comprehensive, five-tier blueprint designed to standardize and formalize the AI-data labor ecosystem. Moving beyond both precarious "click-work" stigmas and traditional corporate hierarchies, the AISF meticulously maps the AI data value chain across five distinct professional archetypes: the Data Associate, Data Enricher, Data Auditor, Data Analyst, and Data Specialist. Central to this framework is the structural definition of the "Human Advantage" — the unique cognitive moat that values human labour against the rise of synthetic data generation.

Navigating the AISF Parameters

The AISF offers a shared vocabulary for roles, skills, and talent management.

The key parameters of the AISF are:

- **Knowledge** — captures the depth of theoretical, conceptual, and/or domain understanding ("knowing") required to perform duties efficiently, scaling from basic digital literacy to deep first-principles mastery. Each role may have specific domain or discipline requirements based on the data modality or purpose. Consequently, the "Domain Knowledge and Experience" is intentionally open-ended because project requirements determine the specificity. However, the *minimal* formal educational requirements are detailed, and the expectations of functional literacy that accompany it. The AISF represents a digitally connected population but found that employable digital skills tend to be presumed and undefined. This section details the *minimal* digital and critical thinking skills required at each level to complete the tasks that follow.
- **Technical Skills** — captures the precise tasks and activities to perform a role. Here it measures the workers manual, digital or cognitive dexterity in using tools, platforms, languages, and processes to achieve a result. In the AISF, the technical skills differ across data modalities and at each workstream in the AI-Data Value Chain. This is a representative, overarching list of tasks and their performance metrics and is not meant to represent the specificity of any singular job role.
- **Transversal and Behavioural Skills** — capture the transversal professional skills, interpersonal abilities, and behaviors and attitudes required to be successful — communication, ethics, self-management, coachability, curiosity and creativity. The AISF specifically optimizes for data management, project management, and platform navigation to manage work assignments, quality metrics and security.

Two further parameters, not specifically detailed at each level but critical to establishing the precise NSQF/NCrF level, are Accountability (scope of impact and burden of consequences) and Complexity (ambiguity, cognitive load, and autonomy).

On Levels and Progression: The AISF presents five stages or levels of AI-Data activity. However, the levels and personas are not strictly progressive — they mirror the AI-data value chain and represent iterative, layered data work flowing through multiple feedback loops with increasing complexity and accountability. Each level and persona therefore maps to an NSQF-NCrF band rather than a single level; the complexity and accountability required to manage data through any given feedback loop determines the final level. DACUM workshops revealed that progression in AI-Data work is primarily through increasing specialization in a discipline (for e.g. data sciences, finance, healthcare) or through increasing people-management skills (for e.g. Associate → Quality Assurance → Assistant Team Leader → Trainer/Peer Coach → Team Lead → Process Lead → Center Head). Progression was not linear nor always visible, especially for purely micro- and gig-work; the AISF should be navigated through the personas inhabiting each stage of the value chain, not as a guaranteed ladder.

On Personas: The use of personas offers distinct structural advantages. In the micro-work economy, roles are frequently commoditized into isolated "tasks," stripping workers of professional identity. Personas act as a cognitive bridge, simplifying competency matrices into recognizable human identities: calling a worker "The Data Associate" instead of a "click-worker" provides psychological legitimacy, fostering self-efficacy and confidence, particularly for women and other underrepresented groups, and creates a clear narrative for growth. Operationally, personas enable "Empathy-Based Design" during functional job analysis and curriculum development, and ensure that resulting roles meet the strict, task-aligned taxonomy required by bodies like the NCVET for formal compliance and credentialing.

PERSONA 1

Data Associate

Integration phase:

Dataset Generation

Dominant employment arrangement:

Gig/Platform Workers

Human Advantage:

Collecting or generating real-world data and establishing "Ground Truth"

Accountability:

Responsible for immediate output volume following linear, zero-ambiguity SOPs

NSQF/NCrF level:

Level 2.5 – 3.0 (Equivalent to Grade 10–12 /Early Vocational Certificate)



The Operational Reality

This category represents the high-volume, highly decentralized baseline of the AI-data engine. Workers are typically algorithmically managed or platform-based with operational digital literacy. They are responsible for immediate output and follow strict, linear SOPs. Workers often log in via shared mobile devices in noisy, informal environments, completing rapid, non-ambiguous tasks like reading a sentence aloud, recording a local gesture, or tagging a basic image to establish 'Ground Truth'. Fundamentally, they produce the raw data that later becomes part of large datasets. However, the rise of synthetic data generation has made this function more intermittent, specialized, and increasingly precarious.

This category is best understood as an evolution of the erstwhile "Data Entry Operator" role.³ Unlike the organized Data Entry Operator, work arrangements can vary from micro- to gig-work, with contracts and worker protection (if any) handled by task-aggregator platforms or community-based organizations and social enterprises, which can also act as a form of community mobilization, training, and support for women in this category.

³ <https://www.nqr.gov.in/sites/default/files/MC-SSCQ2212-%20DDEO%20.pdf>

Knowledge

Domain Knowledge or Experience

- Dependent on project brief and requirements
- Skills are acquired through intensive project training and community support where available
- Completion of at least a high school (10 or 10+2) education is preferred. Prior work experience is not necessary but preferred

Foundational Critical Thinking

- Identify gaps in knowledge or domain
- Distinguish amongst multiple sources of information
- Process and interpret briefs and received information
- Compare data against visual or textual examples following strict, zero-ambiguity guidelines

Basic and Operational Digital Skills

- Device familiarity and managing access; at least a smartphone and internet connectivity
- Interact with digital technologies like sending/receiving text messages, voice and video calls and messages, etc.
- Copy or move files (such as documents, data, images, video) between folders, devices (via email, texting/messaging, USB, etc.) or on the cloud
- Use word processing and spreadsheet software

Technical Skills

Data Modalities – Audio, Video, and Egocentric, Image, Text, and LIDAR Audio data

- Capture high-fidelity audio using external microphones in a noise-free environment
- Maintain an audio rejection rate of less than 5% due to background noise or mispronunciation
- Listen to peer audio recordings to identify speech cuts, background disturbances, or mispronunciations

Visual & Egocentric data

- Operate mounted cameras to capture stable, centred egocentric video feeds
- Convert file formats (e.g., .png to .jpg) using provided software tools
- Utilize mobile data apps to geo-tag and time-stamp field-sourced images
- Synchronize multiple data inputs (e.g., audio and video)

Text Data

- Transcribe conversational speech audio files containing dense regional dialects into standardized text using local script keyboards
- Transcribe spoken audio using verbatim rules (including stutters and filler words like “um”)
- Operate transcription pedals to control audio playback speed
- Execute keyboard macros to insert recurring linguistic tags (e.g., [noise], [silence], [overlap])

Quality Management

- Adhere to accuracy thresholds
- Manage task-completion speed to balance the price-per-task with high quality
- Verify source authenticity by checking timestamps and location metadata against project rules
- Flag incomplete records before submitting the batch to the next stage of processing

Transversal and Behavioural Skills

Data management

- Perform basic file naming according to the project-specific naming convention
- Execute batch uploads using the designated platform or secure folder structure and formatting rules
- Format text responses using prescribed formats and syntax
- Maintain data confidentiality by following strict privacy protocols

Task and Platform Management

- Navigate task queues to find high-priority assignments on the dashboard
- Execute standardized workflows to meet the platform’s specific “golden set” standards
- Apply version controls to ensure the most recent guidelines are followed
- Respond to automated feedback by adjusting outputs

Behavioural Skills

- Recognize emotional triggers when interacting with virtual assistants, chatbots, or digital platforms
- Balance online interactions with offline routines to protect basic physical, mental, and social health
- Employ appropriate digital etiquette (e.g., tone and emoji choices) to reflect a respectful online disposition during human-to-human communications
- Demonstrate coachability; i.e., adaptability, a willingness to work, and openness to adjust work habits after receiving automated feedback, training, or other calibration coaching
- Sustain visual and cognitive focus during long shifts coachability; i.e., adaptability, a willingness to work, and openness to adjust work habits after receiving automated feedback, training, or other calibration coaching

Frictions

Piece-rate compensation actively incentivizes speed over accuracy, structurally compromising the very dataset workers are hired to build.

Ad-hoc, piecemeal nature of the work drives high turnover, more acute for women in this workforce. The lack of work stability and income security undercuts the value of this category as a gateway to seek further employment, skill development, or community visibility.

Utility for Stakeholders

Government:

Located at NSQF Level 2.5–3.0, this persona is the crucial entry point for the digitally marginalized. The primary policy intervention is rapid deployment of Recognition of Prior Learning (RPL) programs, transforming invisible, informal piece-rate workers into statistically recognized economic participants.

Industry and Civil Society:

- Community or Resourcing teams may utilize this persona to structure large-scale, decentralized sourcing initiatives, developing screening evaluations and training tasks within the worker management platform.
- Civil society actors can structure training, peer support groups, and skill-equivalence programs that recognize and value the skills listed under this persona, focusing on foundational digital skills and practices (secure logins, device troubleshooting, digital etiquette) instead of complex AI theory
- **Compensation:** Aggregating agencies can shift from pure piece-rate volume to "quality-gated batch bonuses" that pass automated quality checks without rejection.

PERSONA 2

Data Enricher

Integration phase:

Dataset Generation → Model Training

Employment arrangement:

Gig/Platform Workers &/or Salaried or Contracted Workers

Human Advantage:

Mapping chaotic reality into machine-readable taxonomies with precision

Accountability:

Responsible for batch accuracy and navigating discriminative ambiguity in defined menus

NSQF/NCrF level:

Level 4.0 (Equivalent to Higher Secondary Vocational/Undergraduate Entry)



The Operational Reality

This is the engine room of Supervised Fine-Tuning (SFT). Workers here provide syntactic coherence by executing tasks that require intense structural dexterity, such as drawing pixel-perfect bounding boxes around LiDAR data, navigating complex, multi-layered taxonomy menus to categorize text, and maintaining syntactic coherence.

A worker here demonstrates digital fluency, showcases domain &/or procedural knowledge, and can navigate platform settings, UI, and basic troubleshooting. They are responsible for the accuracy of their own batch/work and are comfortable with discriminative ambiguity, i.e., they choose accurate rule(s) from defined menus and options. They are comfortable implementing pre-determined taxonomies and categorizations. With greater experience or domain knowledge, these workers may even contribute to refining taxonomies or categories or to evaluating an automation solution or an AI model's veracity for a purpose. They do not design data pipelines or write code. Their job is solely to construe the raw, multi-modal human inputs that feed the pipeline.

Commonly associated with the "Data Annotator" role, this category encompasses the spectrum of data labelling and annotation roles in AI training.⁴ Worker arrangements ran from micro-work to full-time employment, with greater value placed on workers with domain expertise navigating high-complexity, high-ambiguity datasets. It was observed during the field research that the degree of complexity and accountability demanded by the data determined the fully remote or office-based nature of the function.

⁴ <https://nqr.gov.in/qualifications/13281>

Knowledge

Domain Knowledge or Experience

- Dependent on project requirements, but at this stage, any domain knowledge or skill must be demonstrable with functional fluency
- Skills in any labeling or annotation domain are typically acquired on the job through training, peer support, and practice assessments
- A graduate education (12 + UG) is preferred. Prior work experience is not necessary but preferred

Applied Critical Thinking

- Capacity to navigate "discriminative ambiguity" by differentiating overlapping categories and applying hierarchical logic from project taxonomies
- Reflect and organize information such as data, evidence, statements, concepts, questions, opinions, and other forms of representation to create sense and meaning
- Distinguish fact from information from opinion from hallucination, by identifying and evaluating factors such as currency, reliability, relevance, authorship, completeness, or veracity
- Apply fundamental logical reasoning, i.e. demonstrating the ability to reason through sets of propositions, rules, conditions, statements, and premises to arrive at a valid conclusion

Engaged Digital Skills

- Comfort with using collaborative office suites and/or platforms
- Create and manage files and folders on a desktop, laptop, tablet, or mobile phone. Moving files between folders, devices, or cloud-based platforms
- Search, locate, and verify information on the internet across devices
- Use advanced features of word processing or spreadsheet software to organize, analyze, structure, or modify data

Technical Skills

Data Modalities – Audio, Video, and Egocentric, Image, Text, and LiDAR

Labelling

- Assign intent and emotional sentiment labels to vernacular text strings based on localized cultural contexts and idioms
- Draw 2D bounding boxes around prominent, non-occluded target objects (e.g., vehicles, pedestrians) according to baseline labelling guides
- Assign pre-defined tags using a binary (Positive/Negative) selection menu

Annotation

- Outline cell boundaries (polygons, polylines) in images using specialized segmentation tool hotkeys.
- Maintain a boundary precision tolerance within a 2-pixel variance.
- Segment occluded or partially hidden objects in complex images by applying specific "visibility and inference" rules
- Outline irregular object boundaries (polygons or polylines) for overlapping or partially blocked targets according to complex pixel-edge rules
- Achieve a pixel-level classification accuracy of 95% or higher
- Select the correct fine-grained attribute tag from hierarchical taxonomies based on context

Quality Management

- Achieve an Intersection-over-Union (IoU) overlap score of 90% or higher across all assigned baseline image frames
- Review pre-labelled image assets to confirm the presence or absence of a specified attribute tag using a binary (Yes/No) toggle
- Match the verification consensus with a 99% accuracy rate compared to the platform's embedded control data
- Audit rejected or low-confidence annotation batches submitted by (junior) workers to diagnose recurring edge-case errors
- Adhere to complexity, accuracy, and precision thresholds as defined by the organization or platform with respect to golden datasets
- Flag incomplete records before submitting the batch to the next stage of processing

Transversal and Behavioural Skills

Data management

- Perform basic file naming according to the project-specific naming convention
- Execute batch uploads using the designated platform or secure folder structure and formatting rules
- Foundational proficiency in querying databases like spreadsheets in MS Excel
- Implement processes and tools to ensure accuracy, completeness, and consistency in data quality
- Help identify and report issues and discrepancies
- Monitor and maintain data quality through regular reviews and validation checks
- Maintain data confidentiality by following strict privacy protocols

Task and Platform Management

- Navigate task queues on the dashboard
- Execute standardized workflows to meet the platform's specific "golden set" standards
- Apply version controls to ensure the most recent guidelines are followed
- Protect personal privacy by safely managing online accounts and limiting passive data footprint tracking
- Respond to automated feedback by adjusting outputs.

Behavioural Skills

- Recognize emotional triggers when interacting with virtual assistants, chatbots, or digital platforms
- Work closely with cross-functional teams, including IT, data scientists, and business leaders, to achieve organizational goals
- Balance online interactions with offline routines to protect basic physical, mental, and social health
- Employ appropriate digital etiquette (e.g., tone and emoji choices) to reflect a respectful online disposition during human-to-human communications

Frictions

Cognitive Attrition:

This level requires sustained focus but is often managed with high-turnover, low-touch methods, and when edge cases arise outside the prescribed taxonomy, workers are forced to guess, leading to downstream data drift.

Utility for Stakeholders

Government:

Mapping to NSQF Level 4.0, this category provides the broadest foundation for formal training and the largest possible set of micro-credentials, trackable by region, enterprise, language, and labour configuration. Skilling bodies (SSC-NASSCOM, NSDC) must align curricula directly to these technical skills.

Industry and Civil Society:

- This is the workhorse layer organizing unstructured reality into machine-readable formats, covering the greatest diversity of processes. Industry can use AISF to recruit, and to standardize training on data modalities and technical skills.
- **Implement Ambiguity Escalation Protocols:** Industry can build explicit workflows for edge case management, with workers trained and incentivized to flag anomalies rather than force a wrong answer.
- **Establish Floor Wages for Complexity:** Transition to tiered pricing based on cognitive load, providing wage stability for reliable supplementary income.
- **Program Design:** implement ergonomic interventions against visual and physical fatigue. Training program designs must also focus on transversal skills and deepening human collaboration for long term career management.

PERSONA 3

Data Auditor

Integration phase:

Dataset Generation → Model Training

Employment arrangement:

Salaried or Contracted Workers

Human Advantage:

Resolving "gray areas" using empathy, sarcasm detection, and cultural nuance

Accountability:

Power user capable of troubleshooting "noisy" data and interpreting complex policy

NSQF/NCrF level:

Level 4.5 – 5.0 (Equivalent to Advanced Diploma/ B.Voc / Undergraduate Degree)



The Operational Reality

The Data Auditor functions as the cultural translator and quality referee. They review the outputs of Level 2 workers or audit the safety of large language model responses, resolving complex "grey areas" where binary rules fail, and human empathy or cultural nuance is required. The worker is typically a power user with expertise who showcases a thorough knowledge of ethics, shortcuts, formulas, and debugging for advanced annotation, content moderation, and model evaluation. As an evaluative thinker and interpretive communicator, they may be responsible for small teams or specific project modules. They sometimes demonstrate leadership through peer mentorship and training, capable of troubleshooting and cutting through "noisy" data. They do not program safety algorithms or filters. They provide the highly contextual human judgment that trains the safety algorithms to recognize toxicity or bias.

This is a harder job role to pin down in the industry since there is a small degree of progression from Level 2 to Level 3. Job titles vary widely — "AI Data Quality Manager," "Content Moderator," "AI Quality Analyst," "Trust and Safety Specialist" — and are in no way exhaustive. Field research indicated that workers who demonstrated greater ethical and critical thinking through attention to detail, complexity, and contextual cues did "progress" into a data quality manager or a content moderator and evaluator of AI outputs. Work arrangements in India tend to be full-time; globally, contractual or gig-work arrangements are more common.

Knowledge

Domain Knowledge or Experience

- Dependent on project requirements, but at this stage, any domain knowledge or skill must be demonstrable with high proficiency
- Skills in advanced annotation, content moderation and verification, and quality control are typically acquired on the job through training, peer support, work experience, coaching, and practice assessments
- A graduate education (12 + UG) is preferred. Prior work experience is preferred due to handling of sensitive PII data

Analytical Critical Thinking

- Synthesize multi-modal context (e.g., combining text and images) and evaluate cultural idioms or nuances to resolve "grey area" policy disputes
- Apply logical reasoning, i.e. demonstrating the ability to reason through sets of propositions, rules, conditions, statements, and premises to arrive at a valid conclusion. This includes the ability to identify fallacies and technical flaws in various representations of reasoning
- Identify and evaluate implicit assumptions, biases, values, and beliefs operating within one's own or another's reasoning
- Identify criteria for decision making, including an analysis of the current situation, and the ability to recognize what would constitute an ideal outcome
- Distinguish fact from information from opinion from hallucination, by identifying and evaluating factors such as currency, reliability, relevance, authorship, completeness, or veracity

Managed Digital Skills

- Comfort with using collaborative office suites and/or platforms
- Create and manage files and folders on a desktop, laptop, tablet, or mobile phone. Move files between folders, devices, or cloud-based platforms
- Search, locate and verify information on the internet across devices
- Use advanced features of word processing or spreadsheet software to organize, analyze, structure, or modify data
- Navigate the internet and social platforms to verify information, products, services, reviews, and prices
- Ability to identify and use online learning resources
- Comfort with safely and ethically utilizing one's own and others' Personally Identifiable Information (PII) to navigate verifications, online banking, and other civic services

Technical Skills

Data Modalities – Audio, Video, and Egocentric, Image, Text, and LIDAR Audio data

- Label named entities and coreference chains within multi-turn conversational transcripts according to specialized domain schemas
- Map semantic dependency relations between nested clauses in low-resource vernacular text using hierarchical graph tools
- Evaluate multi-turn AI agent or chatbot responses for semantic drift, factual hallucination, and instructional compliance using hierarchical evaluation rubrics
- Achieve an inter-annotator agreement rate (Cohen's Kappa) of 0.85 or greater across all assigned text corpora
- Maintain an F1-score of 0.92 or higher on cross-validated gold standard test batches

Content Moderation

- Rank competing LLM response outputs for safety, factuality, and instructional compliance using multi-dimensional rubric scales
- Analyze borderline policy-violating content uploads by applying deep contextual evaluation to resolve complex grey areas not explicitly covered in the baseline safety manual
- Audit completed peer moderation logs using targeted spot-checks to uncover systemic reviewer bias or regional pattern misunderstandings within localized queues

Quality Control

- Verify completed image annotation batches against primary gold standard control samples before final client submission
- Flag faulty or misaligned boundaries with a matching consensus rate of 98% relative to the reviewer matrix.
- Maintain an inter-moderator alignment consensus rate of 90% or higher when processing high-ambiguity socio-cultural safety escalations
- Translate client policy shifts or model trend data into simplified edge-case guidelines and explanatory rubrics for junior cohorts to review

Transversal and Behavioural Skills

Data management

- Execute batch uploads using the designated platform or secure folder structure and formatting rules
- Functional proficiency in querying databases like spreadsheets in MS Excel
- Implement processes and tools to ensure accuracy, completeness, and consistency in data quality
- Maintain data confidentiality by following strict privacy protocols

Task and Platform Management

- Allocate incoming data annotation micro-tasks across decentralized worker queues using platform management dashboards to maintain optimal delivery timelines
- Diagnose annotation tool interface glitches and operational exceptions by reviewing active user error logs before escalating technical bugs to the engineering team
- Meet 100% of weekly client service level agreements (SLAs) by dynamically shifting task loads across active platform delivery nodes. Execute standardized workflows to meet the platform's specific "golden set" standards.
- Resolve at least 90% of localized platform user interface exceptions without requiring backend intervention
- Investigate abnormal machine learning behaviours or edge cases to discover the underlying root cause of structural dataset failure

Behavioural Skills

- Practice active empathy by taking others' diverse viewpoints, cultural backgrounds, and preferences into account across collaborative digital groups
- Provide constructive performance feedback and tactical guidance to junior worker cohorts to improve collective batch quality thresholds
- Resolve inter-team data classification disputes and calibration differences using objective metric scorecards to minimize downstream processing delays
- Initiate collaborative workflows that constructively combine human talents with automated functionalities
- Work closely with cross-functional teams, including IT, data scientists, and business leaders, to achieve organizational goals
- Balance online interactions with offline routines to protect basic physical, mental, and social health
- Prioritize and model respectful digital behaviours that support inclusion and protect the digital reputations of all team members

Frictions

Subjectivity vs. Uniformity:

Auditors are hired for local cultural nuance but strictly evaluated against a rigid "Golden Set" developed by engineers in different socio-cultural contexts; if localized judgment deviates from the global baseline, they are penalized for being "wrong."

Utility for Stakeholders

Government:

At NSQF Level 4.5 – 5.0 (requiring an undergraduate degree or equivalent), regulators should use this persona to develop advanced credentialing pathways critical for national AI sovereignty, ensuring global models comply with local cultural and legal standards.

Industry:

This persona is vital for corporate risk mitigation — the frontline defence against AI hallucinations, toxicity, and systemic bias. This layer is most likely to progress along the people-management track of an enterprise, and should be nurtured as such.

- **Mandate Psychological Scaffolding:** Institutionalize mental health debriefing protocols and limit continuous exposure hours for safety and content moderation workflows.
- **Transition to Formal Salaries:** Move these workers entirely away from piece-rates to stable, contracted salaries to achieve wage parity commensurate with recognized domain expertise.

PERSONA 4

Data Analyst

Integration phase:

Dataset Generation → Model Training

Employment arrangement:

Salaried or Contracted Workers

Human Advantage:

Identifying systemic logic failures or "drift" where statistics look sound, but reality is flawed

Accountability:

Bridging raw work with final production models through statistical validation and diagnostics

NSQF/NCrF level:

Level 5.5 – 6.0 (Equivalent to Professional Bachelor's Degree / Post-Graduate Diploma)



The Operational Reality

These are the system mechanics. They manage the statistical health of the pipeline, analyze error distributions, translate raw data metrics into performance dashboards, and serve as the operational bridge between the grassroots annotators and the machine learning engineers. Workers are practiced data experts who conduct statistical validation, automated auditing, and performance diagnostics. They understand the up- & down-stream dependencies and systems, and bridge the raw work of the annotators with the final production model. They may be responsible for entire workflows or functional health and adapt principles/definitions to new/unfamiliar situations. They do not design the foundational neural network architecture. They measure, diagnose, and calibrate its accuracy based strictly on human data inputs.

Best mapped to "Data Analyst"⁵ or "AI-Data Quality Analyst"⁶ in existing qualification lists. These workers are explicitly not data scientists. Of all AISF categories, the Data Analyst has the most established credentials and progression pathways; workers at this stage typically held at least an undergraduate engineering or computer science degree. There is little to no evidence of progression from Levels 1, 2, or 3 into this level. Work arrangements in this category tend to be full time employment contracts.

Note in box: Data analysts under AISF are not the same as Data Scientists. Data Scientists write the code to design and build the foundational neural network architecture, whereas the Data Analysts in the AISF validate, diagnose, and ensure the quality of the human-curated datasets that train those systems.

⁵ <https://nqr.gov.in/qualification/file/QF-Junior%20Data%20Analyst.pdf>

⁶ https://www.nqr.gov.in/qualification/file/SSC%20Q8101_AI%20Data%20Quality%20Analyst_V3.pdf

Knowledge

Domain Knowledge or Experience

- Defined by the data needs of the project requirements. At this stage, domain knowledge or skills in data analytics and management must be demonstrable with high proficiency
- Data analytics and management skills may be acquired on the job based on project or domain requirements
- A graduate education (12 + UG) is preferred. Prior work experience and comfort in business analytics are preferred

Evaluative Critical Thinking

- Reflect and organize information such as data, evidence, statements, concepts, questions, opinions, and other forms of representation to create sense and meaning
- Categorize and synthesize facts from information, from opinion and hallucination, by identifying and evaluating factors such as currency, reliability, relevance, authorship, completeness, or veracity
- Apply advanced logical reasoning, i.e., demonstrating the ability to reason through sets of propositions, rules, conditions, statements, and premises to arrive at a valid conclusion
- Clearly articulate the parts of a problem, identify procedurally compliant solutions and consequences

Productive Digital Skills

- Advanced use of collaborative office suites and/or platforms
- Advanced ability to download and customize or change settings in software, apps, or devices
- Use advanced features of spreadsheet software (functions, formulas, macros, and other developer functions) to organize, analyze, structure, and modify data
- May have some knowledge of coding languages and programming scripts to search, align, and structure data

Technical Skills

Data Modalities – Audio, Video, and Egocentric, Image, Text, and LiDAR

- Assist in data preparation and analysis activities under supervision. Process and validate data to support analytics and generate standard reports and insights using established tools and methods
- Analyze error distributions to identify systemic pattern failures in the AI model or project guidelines
- Define quality protocols to validate the accuracy and safety of incoming dataset submissions
- Perform random batch audits to validate specific metrics, like word error rates and signal-to-noise ratios
- Synchronize raw data inputs (such as audio timestamps and speaker IDs) to normalize complex datasets
- Format finalized data sets into structured JSON or CSV files for official client delivery
- Evaluate key performance metrics (such as Accuracy, Precision, Recall, and F1 scores) to monitor overall project health

Grassroots Calibration and Stakeholder Communication

- Communicate observed data patterns to clients to refine the definitions of Standard Operating Procedures (SOPs) and project guidelines
- Translate technical client feedback into actionable coaching for the grassroots labelling and annotation workforce
- Manage continuous feedback loops between the technical production team and the data generation workforce

Team Management

- Manage overall team throughput to ensure high productivity and service level agreements at the project or center level
- Coordinate worker mobilization and training sessions to fulfill specific demographic or project data requirements

Transversal and Behavioural Skills

Data management

- Advanced proficiency in querying databases like spreadsheets in MS Excel
- Implement processes and tools to ensure accuracy, completeness, and consistency in data quality
- Maintain data confidentiality by following strict privacy protocols

Task and Platform Management

- Execute standardized workflows to meet the platform's specific "golden set" standards
- Apply version controls to ensure the most recent guidelines are followed
- Protect organizational and personal privacy by safely managing online accounts and limiting passive data footprint tracking

Behavioural Skills

- Work closely with cross-functional teams, including IT, data scientists, and business leaders, to achieve organizational goals
- Secure formal stakeholder alignment and resource sign-off on major pipeline changes
- Utilize persuasive storytelling techniques to frame complex pipeline blockages into clear, narrative-driven optimization solutions for non-technical leadership
- Negotiate project delivery parameters and technical values with internal stakeholders to protect the mental load and retention thresholds of grassroots worker cohorts
- Balance online interactions with offline routines to protect basic physical, mental, and social health
- Employ appropriate digital etiquette (e.g., tone and emoji choices) to reflect a respectful online disposition during human-to-human communications

Frictions

Frictions & Contradictions: The Middle-Management Squeeze

Middle managers are responsible for fixing dataset logic failures and throughput bottlenecks, but rarely have the administrative authority to alter the underlying engineering architecture—forcing them to solve systemic problems through operational workarounds.

Utility for Stakeholders

Government:

Mapping to NSQF Level 5.5–6.0, aligned with professional degrees and post-graduate diplomas, interventions should integrate these competencies into higher education (data science, business analytics) to build the "middle-management" layer required to scale the national AI ecosystem.

Industry:

Enterprises use this persona to bridge raw data generation and the final production dataset, continuously calibrating guidelines used by the grassroots workforce. Industry can:

- **Align Compensation with Systemic Health:** tie salaries to pipeline health metrics — SLA adherence, IoU improvements, aggregate error reduction.
- **Train for Strategic Influence:** pivot professional development toward abductive reasoning and persuasive storytelling for non-technical leadership.
- **Program Design:** build capacity to synthesize conflicting data and communicate root-cause problems persuasively.

PERSONA 5

Data Specialist

Integration phase:

Model Maintenance → Dataset Refinement/Development → Model Training

Employment arrangement:

Salaried or Contracted Workers

Human Advantage:

Deconstructing professional wisdom into modular reasoning chains for autonomous agents

Accountability:

Translating high-level goals into actionable constraints for specialized professional use cases

NSQF/NCrF level:

Level 6.5 – 7.0 (Equivalent to Master's Degree / Expert Professional)



The Operational Reality

There are two types of users here — (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 (like healthcare, education, or finance).⁷ The first are ‘business intelligence analysts’ and overall business leaders in the AI ecosystem, steering client relationships, innovation, and model development. The latter are elite, highly credentialed professionals (e.g., doctors, lawyers, financial analysts) operating as the Human-Pretend Agent (Fragiadakis, et al. 2024; IEEE 2025). They inject ground-truth professional wisdom directly into the model through Reinforcement Learning from Human Feedback (RLHF) and complex prompt architecture. Both deploy strategic professional insights to define and refine AI models and applications pre- and post- deployment. Neither write the backend codebase.

⁷https://nqr.gov.in/sites/default/files/QP_SSC_Q8102_v2.0_AI%20Business%20Intelligence%20Analyst.pdf

Knowledge

Domain Knowledge or Experience

- Expertise at this level is intersectional and experiential. This requires a demonstrated expert-fluency in a domain of expertise
- A post graduate education and prior work experience is mandatory

Generative Critical Thinking

- Advanced evaluation of information such as data, evidence, statements, concepts, questions, opinions, and other forms of representation to create sense and meaning
- Synthesize information from diverse and sometimes contradictory sources to form fresh insights and frameworks
- Apply systems thinking by understanding how individual components and subsystems interact with and impact broader systems
- Reflect on one's own thought processes and actively recognize cognitive and socio-economic biases

Strategic Digital Skills

- Workflow automation and deployment of strategic digital tools and platforms, i.e., integrating generative AI and machine learning tools to optimize operations, enhance productivity, and redesign daily processes
- Data-driven decision-making; i.e., the ability to interpret analytics, identify market trends, and measure Key Performance Indicators (KPIs) to guide business strategy
- Lead teams through technological transitions, overcome resistance to change, and build capacity across the organization
- Set privacy strategy and cybersecurity guardrails as a risk-management strategy to protect corporate data and customer trust

Technical Skills

Agentic Auditing & Ethical Governance

- Architect reasoning chains and complex prompt hierarchies to deconstruct expert workflows
- Audit Chain of Thought (CoT) logs to evaluate the decisions, outputs, and intents of independent AI agents
- Evaluate AI-generated creative hallucinations to distinguish between innovative leaps and factual errors.
- Assess model responses to ensure factual accuracy, contextual relevance, tonality, and user safety
- Advocate domain-specific ethics to prevent the AI from generating "correct but harmful" advice

Define ethical boundaries and reasoning chains to comply with legal frameworks (such as GDPR) and ensure transparency

Human in the Loop (HITL) & System Architecture

- Design Standard Operating Procedures (SOPs) to establish overarching data structures, guidelines, and quality standards
- Monitor key success metrics to lead the continuous optimization of the entire AI project
- Deconstruct expert-level workflows into modular steps that an AI agent can emulate
- Architect complex prompt hierarchies to test multi-step reasoning capabilities in the AI agent
- Translate high-level project goals into actionable constraints for the AI training pipeline

Domain Veracity & Golden Data Management

- Synthesize contradictory domain data to establish a "Unified Truth" for the model to follow
- Evaluate highly complex edge cases to advise on cultural relevance, local context, and domain accuracy
- Inject high-quality "golden samples" to fine-tune linguistic and computational models
- Engineer system-level personas to ensure the agent maintains a consistent, contextually appropriate professional voice (e.g., an SRH medical bot)
- Execute Reinforcement Learning from Human Feedback (RLHF) preference rankings to reward the model based on expert-level domain criteria

Process Evaluation & Capacity Building

- Conduct process evaluations to audit overall AI model performance and outputs
- Mentor Quality Assurance (QA) teams to explain the "why" behind complex, expert-level decisions
- Standardize evaluation rubrics to provide a clear "compass of quality" for annotators and coordinators
- Collaborate with machine learning engineers to bridge the gap between human nuance and model logic

Transversal and Behavioural Skills

Data management

- Perform basic file naming according to the project-specific naming convention.
- Execute batch uploads using the designated platform or secure folder structure and formatting rules
- Format text responses using prescribed formats and syntax
- Maintain data confidentiality by following strict privacy protocols

Task and Platform Management

- Navigate task queues to find high-priority assignments on the dashboard
- Execute standardized workflows to meet the platform's specific "golden set" standards
- Apply version controls to ensure the most recent guidelines are followed
- Respond to automated feedback by adjusting outputs

Behavioural Skills

- Recognize emotional triggers when interacting with virtual assistants, chatbots, or digital platforms
- Balance online interactions with offline routines to protect basic physical, mental, and social health
- Employ appropriate digital etiquette (e.g., tone and emoji choices) to reflect a respectful online disposition during human-to-human communications
- Demonstrate coachability; i.e., adaptability, a willingness to work, and openness to adjust work habits after receiving automated feedback, training, or other calibration coaching
- Sustain visual and cognitive focus during long shifts coachability; i.e., adaptability, a willingness to work, and openness to adjust work habits after receiving automated feedback, training, or other calibration coaching

Frictions

Deconstructing Intuition:

Experts are accustomed to associative, intuitive problem-solving based on tacit knowledge, but AI training requires breaking these leaps down into rigid, sequential Chain of Thought logs — forcing a fluid human expert to think like a linear algorithm often causes immense frustration.

Utility for Stakeholders

Government:

At NSQF Level 6.5–7.0, policymakers must engage this tier to draft national standards, audit public-sector AI deployments in sensitive welfare sectors, and ensure AI systems in critical areas (healthcare, education, finance) adhere strictly to legal frameworks like the DPDP Act.

Industry:

Enterprises depend on this persona to deconstruct complex professional wisdom into modular reasoning chains that autonomous agents can emulate — the architects of the "Unified Truth" who can strategize and lead further AI solution development. Industry can:

- **Implement Competitive Compensation Models to Fuel Innovation:** expert-level consultancy retainers or equity-based compensation matching or exceeding standard engineering roles.
- **Program Design:** focus training entirely on "Prompt Engineering" and cognitive deconstruction — teaching experts to break their own expertise into machine-readable steps.

From Framework to Action: Building the Skills Infrastructure for India's AI Economy

India accounts for 24 per cent of the global online labour market and AI talent pool, while AI is projected to contribute substantially to the economy over the coming decade. As AI-related services and employment opportunities expand, there is an opportunity to build the skills infrastructure needed to support the scale, quality and mobility of this emerging workforce.

The AI-Data Skills Framework (AISF) provides a practical starting point. By defining roles across the AI-data value chain and mapping measurable competencies across knowledge, technical skills, transversal and behavioural skills, complexity and accountability, it creates a common language for understanding how this work is evolving and the capabilities it requires.

The value of the framework lies in its application. Upstream, it can strengthen labour-market intelligence by helping industry, skilling institutions and policymakers better understand emerging roles, skill requirements and gaps. Downstream, it can inform targeted skilling, recognition of existing capabilities, formal credentialing and clearer pathways across roles. Applied across the ecosystem, these building blocks can support a more responsive skills system and contribute to better-quality employment as India's AI-data economy matures.⁸

AISF's tripartite utility

1

For government regulators and standard-setting bodies, it provides rigorous, task-aligned taxonomy necessary for formal credentialing and national wage mapping.

2

For industry and the private sector, it functions as an agile operational tool to optimize workflows, predictably map data quality, and reduce attrition through clear progression pathways.

3

For civil society, the AISF serves as an instrument of program design and community empowerment, providing a shared, dignified vocabulary that transforms informal gig work into recognized, scalable digital trades.

To convert this structural architecture into tangible, real-world value, the framework aims to be systematically deployable across three critical pillars: Policy and Governmental Institutions, Private Industry, and Civil Society. The specific Toolkits for these pillars will be available shortly on our website.

⁸ Prosus, BCG and MeitY, 2026

References

- Anwar, Mohammad Amir. 2024. "Value Chains of AI: Data Training Firms, Platforms, and Workers." *The Future of Labour* (Routledge) 198-213.
- Competition Commission of India . 2025. Artificial Intelligence and Competition. Market Study, New Delhi: CCI.
- Fragiadakis, George, Christos Diou, George Kousiouris, and Mara Nikolaidou. 2024. "Evaluating Human-AI Collaboration: A Review and Methodological Framework." *arXiv.org*, July .
- Groupe Special Mobile Association (GSMA). 2018. Data Value Chain - Structure and Activities. Technical Report, London: GSMA.
- IEEE. 2025. P7015 - Draft Standard for Data and Artificial Intelligence (AI) Literacy, Skills, And Readiness. Standard, New York: The Institute of Electrical and Electronics Engineers, Inc.
- —. 2023. "NATIONAL SKILLS QUALIFICATION FRAMEWORK-2023." Page 1 of 40 F.No-22001/01/2023/NCVET. New Delhi: Government of India.
- NASSCOM. 2021. Data Annotation: Billion Dollar Potential Driving the AI Revolution. NASSCOM. <https://community.nasscom.in/communities/ai/data-annotation-billion-dollar-potential-driving-ai-revolution>
- 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. Accessed January 5, 2026. <https://www.pib.gov.in/PressNoteDetails.aspx?NotelId=156786&ModuleId=3®=3&lang=2#:~:text=AI%20Ecosystem%20in%20India%20at,than%20500%20focused%20on%20AI>.
- 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>.
- Sahasrabojanee, V. & Taneja, M. 2026. When Flexibility Becomes a Constraint: The Paradox of Women's Participation in AI-based Gig Microwork. GxD hub, LEAD at Krea University, IFMR. Accessed 19 August 2026 from <https://gxdhub.org/wp-content/uploads/2026/07/Womens-Participation-in-AI-Gig-Microwork-Evidence-Brief.pdf>
- UNESCO. 2026. Outlook study on artificial intelligence and gender in South Asia. UNESCO Regional Office for South Asia, New Delhi. Published 8 July 2026. Available at: <https://www.unesco.org/en/articles/outlook-study-artificial-intelligence-and-gender-south-asia>



www.gxdhub.org

THE/NUDGE

www.thenudge.org

