WORKING PAPER

India Iceberg Index: Measuring AI Occupational Exposure Across 744 Districts

Somesh Mohapatra[1]

[1]Independent.

Contributing authors: someshm@alum.mit.edu

Abstract

Generative artificial intelligence represents a general-purpose technology targeting cognitive capabilities rather than routine manual labor, yet discourse on AI's workforce impact in India remains concentrated on the information technology sector. We introduce the India Iceberg Index, adapting the Project Iceberg methodology to quantify AI occupational exposure across 744 districts and 36 states and union territories using Periodic Labour Force Survey (PLFS) 2023–24 microdata covering 415,549 individuals. Our analysis reveals that approximately 41.5% of India's wage-weighted employment faces technical exposure to current-generation AI systems. The national mean Iceberg Index is 37.74 (SD = 9.38) at the district level, with values ranging from 13.41 to 70.56. Critically, visible technology-sector exposure (Surface Index) constitutes only 0.96% of total employment, while hidden cognitive exposure in administrative, financial, and clerical occupations—the "submerged mass"—extends to districts conventionally considered outside AI's reach. Districts exhibiting maximum "surprise" include The Dangs (Gujarat), Central Delhi, and Chitrakoot (Uttar Pradesh), where Iceberg Index values exceed 60 despite near-zero technology-sector presence. Rural areas show higher mean exposure (49.52) than urban areas (31.46), driven by formal administrative employment in government and banking. Industry concentration analysis using the Herfindahl-Hirschman Index reveals that 55.6% of districts exhibit distributed exposure patterns requiring multi-sector policy coordination. These findings challenge the prevailing "tech problem" framing and suggest that India's service-sector development model faces structural vulnerability as AI capabilities expand into entry-level cognitive work.

Keywords: Artificial Intelligence, Labor Markets, Occupational Exposure, India, Automation, Service Sector, Digital Economy

1 Introduction

Artificial intelligence is transforming global labor markets through mechanisms fundamentally different from prior waves of automation. Where industrial robots and computerization targeted routine manual and computational tasks, large language models (LLMs) demonstrate capability across cognitive functions: information synthesis, text generation, scheduling, coordination, and analysis [1]. This technological shift carries particular implications for economies that have built development strategies around service-sector expansion.

India represents a critical case for understanding these dynamics. The country's post-liberalization growth model has relied substantially on service-sector employment as the primary channel for upward mobility [3]. The information technology and business process outsourcing (IT-BPO) industry, employing approximately 5.4 million workers directly and generating $245 billion in revenue, has anchored this strategy [4]. Entry-level positions in call centers, data processing, and software development have served as "first rungs" on the formal employment ladder for graduates from India's expanding higher education system.

The prevailing discourse on AI's labor market impact in India frames it as a sectoral challenge confined to technology workers. Headlines focus on potential disruption to IT services, coding jobs, and data science roles. This framing treats AI as a vertical shock to specific industries rather than a horizontal transformation of cognitive work across the economy.

We argue this interpretation fundamentally mischaracterizes the nature of LLM-based AI systems. Unlike previous automation technologies, current AI capabilities target the core cognitive tasks that define white-collar employment: processing text, synthesizing information, generating reports, and coordinating workflows. These tasks are not exclusive to technology workers; they constitute the operational foundation of government administration, banking, education, and professional services throughout India.

To quantify this broader exposure, we adapt the Project Iceberg methodology [2], which measures the wage-weighted share of occupational tasks that AI systems can technically perform. The framework distinguishes between the "surface"—visible technology-sector disruption concentrated in software development and data science—and the "iceberg"—the larger submerged mass of cognitive exposure extending through administrative, financial, and clerical occupations nationwide.

Our analysis processes PLFS 2023–24 microdata covering 415,549 individuals across 744 districts, mapping 3,445 National Classification of Occupations (NCO-2015) codes to AI Occupational Exposure (AIOE) scores derived from the Felten et al. framework [1]. We calculate wage-weighted exposure indices at district, state, and sectoral levels, decomposing by urban-rural geography and industry concentration.

The results reveal that approximately 41.5% of India's wage-weighted employment faces technical exposure to current AI systems—a figure substantially higher than visible technology-sector adoption would suggest. The technology sector accounts for only 0.96% of total employment, yet cognitive exposure extends to districts with minimal

technology presence. This pattern indicates structural vulnerability in India's service-sector development model that transcends the boundaries of IT parks and startup ecosystems.

This paper makes three contributions. First, we provide the first comprehensive district-level mapping of AI occupational exposure for India, enabling geographically targeted policy analysis. Second, we identify substantial "hidden" exposure in administrative and clerical occupations that may not be recognized in technology-focused workforce planning. Third, we characterize the industry concentration structure of exposure, distinguishing districts requiring sector-specific intervention from those requiring broad-based coordination.

2 Introduction

Artificial intelligence is transforming global labor markets through mechanisms fundamentally different from prior waves of automation. Where industrial robots and computerization targeted routine manual and computational tasks, large language models (LLMs) demonstrate capability across cognitive functions: information synthesis, text generation, scheduling, coordination, and analysis [1]. This technological shift carries particular implications for economies that have built development strategies around service-sector expansion.

India represents a critical case for understanding these dynamics. The country's post-liberalization growth model has relied substantially on service-sector employment as the primary channel for upward mobility [3]. The information technology and business process outsourcing (IT-BPO) industry, employing approximately 5.4 million workers directly and generating $245 billion in revenue, has anchored this strategy [4]. Entry-level positions in call centers, data processing, and software development have served as "first rungs" on the formal employment ladder for graduates from India's expanding higher education system.

The prevailing discourse on AI's labor market impact in India frames it as a sectoral challenge confined to technology workers. Headlines focus on potential disruption to IT services, coding jobs, and data science roles. This framing treats AI as a vertical shock to specific industries rather than a horizontal transformation of cognitive work across the economy.

We argue this interpretation fundamentally mischaracterizes the nature of LLM-based AI systems. Unlike previous automation technologies, current AI capabilities target the core cognitive tasks that define white-collar employment: processing text, synthesizing information, generating reports, and coordinating workflows. These tasks are not exclusive to technology workers; they constitute the operational foundation of government administration, banking, education, and professional services throughout India.

To quantify this broader exposure, we adapt the Project Iceberg methodology [2], which measures the wage-weighted share of occupational tasks that AI systems can technically perform. The framework distinguishes between the "surface"—visible

technology-sector disruption concentrated in software development and data science—and the "iceberg"—the larger submerged mass of cognitive exposure extending through administrative, financial, and clerical occupations nationwide.

Our analysis processes PLFS 2023–24 microdata covering 415,549 individuals across 744 districts, mapping 3,445 National Classification of Occupations (NCO-2015) codes to AI Occupational Exposure (AIOE) scores derived from the Felten et al. framework [1]. We calculate wage-weighted exposure indices at district, state, and sectoral levels, decomposing by urban-rural geography and industry concentration.

The results reveal that approximately 41.5% of India's wage-weighted employment faces technical exposure to current AI systems—a figure substantially higher than visible technology-sector adoption would suggest. The technology sector accounts for only 0.96% of total employment, yet cognitive exposure extends to districts with minimal technology presence. This pattern indicates structural vulnerability in India's service-sector development model that transcends the boundaries of IT parks and startup ecosystems.

This paper makes three contributions. First, we provide the first comprehensive district-level mapping of AI occupational exposure for India, enabling geographically targeted policy analysis. Second, we identify substantial "hidden" exposure in administrative and clerical occupations that may not be recognized in technology-focused workforce planning. Third, we characterize the industry concentration structure of exposure, distinguishing districts requiring sector-specific intervention from those requiring broad-based coordination.

3 Methods

This study adapts the Iceberg Index methodology [2] to measure district-level AI workforce exposure in India. The analysis integrates four data sources through a multi-stage pipeline: (1) cross-national occupation mapping from India's NCO to the U.S. O*NET system, (2) AI automatability scoring using the Language Modeling AIOE framework [1], (3) workforce exposure calculation from PLFS microdata, and (4) validation against Census socioeconomic indicators. Figure 1 illustrates the complete data flow.

3.1 Data Sources

The analysis draws on five primary data sources, each selected for its authoritative coverage and methodological suitability:

Periodic Labour Force Survey (PLFS) 2023–24

The PLFS [5] conducted by India's National Sample Survey Office provides the primary workforce data. The survey covers 415,549 individuals across 101,957 households, representing India's entire civilian non-institutional population. We use two files: the person-level file (cperv1.csv) containing occupation codes, employment status, earnings, demographics, and geographic identifiers; and the household-level file (chhv1.csv) containing monthly consumer expenditure for wage imputation. The PLFS was selected because it is India's official quarterly labour force survey, providing

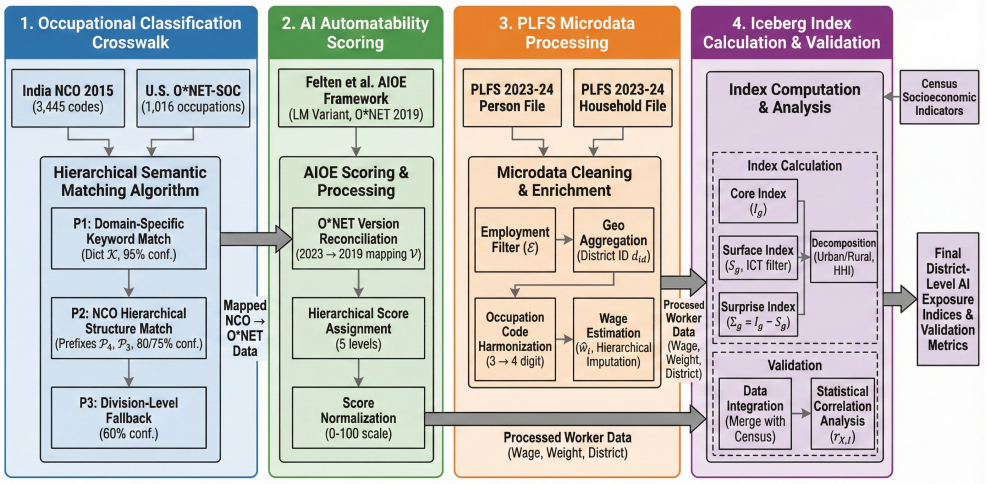

Fig. 1 Methodological pipeline for computing the India Iceberg Index. The framework integrates four components: (1) Occupational Classification Crosswalk mapping 3,445 India NCO-2015 codes to 1,016 U.S. O*NET-SOC occupations through hierarchical semantic matching; (2) AI Automatability Scoring using the Felten et al. AIOE framework with O*NET version reconciliation and hierarchical score assignment; (3) PLFS Microdata Processing including employment filtering, geographic aggregation, occupation code harmonization, and wage estimation; and (4) Iceberg Index Calculation & Validation computing core, surface, and surprise indices with decomposition by urban-rural sector and industry concentration (HHI), validated against Census socioeconomic indicators.

nationally representative data with sufficient geographic granularity for district-level analysis. The survey's stratified multi-stage sampling design and comprehensive occupational coding make it the authoritative source for contemporary Indian workforce composition.

National Classification of Occupations (NCO) 2015

India's NCO-2015 [6] provides the occupational classification framework, comprising 3,445 occupation codes structured hierarchically following the International Standard Classification of Occupations (ISCO-08). The NCO was developed by India's Ministry of Labour and Employment specifically to classify the Indian workforce, incorporating India-specific occupations (e.g., Ayurveda practitioners, tabla makers) while maintaining international comparability. We use NCO-2015 as the bridge between PLFS occupation codes and international skill taxonomies.

*O*NET Occupational Information Network*

The U.S. Bureau of Labor Statistics O*NET-SOC system [7] provides detailed skill, ability, and task information for 1,016 occupations. O*NET was selected as the target for crosswalk mapping because the Felten et al. AIOE scores are calibrated to O*NET occupation codes, and O*NET provides the richest publicly available occupational skill taxonomy. The O*NET database includes both quantitative ratings (importance and level scores on 1–5 and 0–7 scales) and qualitative descriptions for 52 human abilities.

AI Occupational Exposure (AIOE) Scores

We adopt the Language Modeling variant of AIOE scores from Felten et al. [1], calibrated to GPT-3/ChatGPT-era capabilities. This framework was selected for three reasons: (1) it provides validated, occupation-level exposure scores rather than task-level capability assessments; (2) the Language Modeling variant specifically captures LLM capabilities most relevant to cognitive automation; and (3) the scores map directly to O*NET occupation codes, enabling systematic crosswalk from Indian occupational data. The AIOE framework maps 10 AI application areas to 52 human abilities through a crowd-sourced relatedness matrix, producing scores ranging from -1.854 (least exposed) to $+1.926$ (most exposed) across 774 O*NET occupations.

Census of India 2011

District-level socioeconomic indicators from Census 2011 [8] (with 2024 projections where available) provide validation benchmarks. Census data offers the most comprehensive district-level coverage of literacy rates, urbanization, internet access, and demographic composition. The 10+ year gap between Census 2011 and PLFS 2023–24 introduces some temporal mismatch; we account for this by using projected indicators where available and interpreting validation results as structural correlations rather than precise calibration.

3.2 Occupational Classification Crosswalk

Traditional fuzzy text matching produces semantically incorrect mappings (e.g., matching "mycologist" to "dentist" based on suffix similarity rather than to "biologist" based on occupational domain). We therefore implement a hierarchical semantic matching algorithm with three priority levels:

Priority 1: Domain-Specific Keyword Matching

We construct a dictionary $\mathcal{K}$ of approximately 200 domain-specific keywords mapped to O*NET codes. Keywords span legal professions, biological sciences, engineering specializations, healthcare, and India-specific occupations. For a given NCO title t, keywords are matched in descending order of string length to ensure specific terms take precedence:

$$\text{match}_1(t) = \arg\max_{k \in \mathcal{K}} \{|k| : k \subseteq t_{\text{lower}}\} \tag{1}$$

where t_{lower} denotes the lowercase transformation of t and $|k|$ is keyword length. Matches at this level receive confidence score $s = 95$.

Priority 2: NCO Hierarchical Code Structure

When no keyword match exists, the algorithm exploits NCO's hierarchical structure. The first four digits of an NCO code identify the unit group (occupational family), while the first three digits identify the minor group. We maintain lookup tables $\mathcal{P}_4$ and $\mathcal{P}_3$ mapping these prefixes to appropriate O*NET codes:

$$\text{match}_2(c) = \begin{cases} \mathcal{P}_4[c_{1:4}] & \text{if } c_{1:4} \in \mathcal{P}_4, \quad s = 80 \\ \mathcal{P}_3[c_{1:3}] & \text{if } c_{1:3} \in \mathcal{P}_3, \quad s = 75 \end{cases} \tag{2}$$

where $c_{1:n}$ denotes the first n digits of NCO code c.

Priority 3: Division-Level Fallback

For unmatched occupations, the algorithm assigns default O*NET codes based on the NCO major division (first digit), ensuring complete coverage. These mappings receive confidence score $s = 60$.

The final crosswalk achieves 100% coverage across 3,445 NCO occupations with the following distribution: semantic keyword matches 29.5%, NCO prefix matches 70.2%, and division-level fallbacks 0.3% (see Table 4 and Appendix A).

3.3 AI Automatability Scoring

For each O*NET occupation o with ability importance vector $\mathbf{a}_o \in \mathbb{R}^{52}$ and prevalence vector $\mathbf{p}_o \in \mathbb{R}^{52}$, the Language Modeling AIOE score is:

$$\text{AIOE}_{\text{LM}}(o) = \sum_{j=1}^{52} a_{o,j} \cdot p_{o,j} \cdot \left(\sum_{i=1}^{10} w_i \cdot R_{i,j} \right) \tag{3}$$

where w_i represents the weight assigned to AI application i for language modeling capabilities and $R \in \mathbb{R}^{10\times 52}$ is the crowd-sourced relatedness matrix.

3.3.1 O*NET Version Reconciliation

The Felten scores use O*NET 2019 codes, while our NCO crosswalk targets O*NET 2023 codes. We implement a version mapping table $\mathcal{V}$ for 97 restructured codes. Key restructurings include computer systems analysts (15-1211.00 → 15-1121.00) and software developers (15-1252.00 → 15-1132.00).

3.3.2 Hierarchical Score Assignment

For each NCO occupation mapped to O*NET code c, automatability scores are assigned through a five-level hierarchy: (1) Direct match (87.6%), (2) Version mapping (5.0%), (3) Base SOC match (2.6%), (4) Prefix-5 average (4.8%), and (5) Major group mean (0%).

3.3.3 Score Normalization

Raw AIOE scores are normalized to a 0–100 scale:

$$\text{AI}_{\text{auto}}(o) = \frac{\text{AIOE}_{\text{LM}}(o) - \text{AIOE}_{\min}}{\text{AIOE}_{\max} - \text{AIOE}_{\min}} \times 100 \tag{4}$$

where $\text{AIOE}_{\min} = -1.854$ and $\text{AIOE}_{\max} = 1.926$.

3.4 PLFS Microdata Processing

Workers are classified as employed based on Principal Activity Status codes $\mathcal{E} = \{11, 12, 21, 31, 41, 51\}$ representing self-employed, unpaid family helpers, regular wage workers, and casual laborers.

3.4.1 Wage Estimation

India's large informal sector requires hierarchical wage imputation. For worker i, estimated monthly wage $\hat{w}_i$ follows priority order:

$$\hat{w}_i = \begin{cases} \text{CWS}_{\text{salaried},i} \times 4.33 & \text{if available} \\ \text{CWS}_{\text{self-emp},i} \times 4.33 & \text{elif available} \\ \sum_{d=1}^{7} \sum_{a=1}^{2} w_{i,d,a} \times 4.33 & \text{elif daily wages available} \\ \text{MCE}_h / \text{HH\_Size}_h & \text{otherwise} \end{cases} \tag{5}$$

where CWS denotes Current Weekly Status earnings and MCE_h is monthly consumer expenditure for household h.

3.5 Iceberg Index Calculation

Following the Project Iceberg methodology [2], the Iceberg Index for geographic unit g measures the wage-weighted proportion of occupational value where AI systems demonstrate technical capability:

$$I_g = \frac{\sum_{i \in g} \hat{w}_i \cdot m_i \cdot A_i}{\sum_{i \in g} \hat{w}_i \cdot m_i} \times 100 \tag{6}$$

where $\hat{w}_i$ is estimated monthly wage, m_i is the survey multiplier, and $A_i \in [0, 1]$ is the normalized automatability score.

The Surface Index captures visible technology-sector exposure:

$$S_g = \frac{\sum_{i \in g : o_i \in \mathcal{T}} \hat{w}_i \cdot m_i \cdot A_i}{\sum_{i \in g} \hat{w}_i \cdot m_i} \times 100 \tag{7}$$

where $\mathcal{T} = \{251, 252, 351, 352\}$ represents ICT occupation divisions.

The Surprise Index quantifies hidden white-collar exposure: $\Sigma_g = I_g - S_g$.

3.5.1 Industry Concentration

We compute the Herfindahl-Hirschman Index (HHI) [9] of exposure shares across NIC industry codes (see Appendix A for methodology):

$$\text{HHI}_g = \sum_j \left(\frac{E_{g,j}}{\sum_j E_{g,j}} \right)^2 \times 10000 \tag{8}$$

3.6 Census Validation

District-level Iceberg Index values are merged with Census 2011 [8] indicators using normalized district names. The merge achieves 463 matched districts (67% of 694 total). We examine literacy rate, internet access, urbanization rate, and youth population share.

3.7 Implementation

All analyses are implemented in Python 3.10. The complete pipeline executes in approximately 15 minutes. Code and datasets are available at https://github.com/pikulsomesh/india-iceberg-index.

4 Results

4.1 National Exposure Profile

Analysis of the PLFS 2023–24 microdata yields AI occupational exposure estimates for 744 districts across 36 states and union territories. The wage-weighted national Iceberg Index stands at 41.5%, indicating that approximately two-fifths of India's formal labor market wage value involves tasks where current AI systems demonstrate technical capability.

At the district level, the mean Iceberg Index is 37.74 (SD = 9.38), ranging from a minimum of 13.41 to a maximum of 70.56 (Table 1). The distribution exhibits moderate positive skew, with the median (37.91) falling slightly above the mean.

Table 1 Summary Statistics: District-Level Iceberg Index (N = 744)

Statistic	Mean	SD	Min	Median	Max
Iceberg Index	37.74	9.38	13.41	37.91	70.56
Surface Index	22.63	28.53	0.00	0.00	68.05
Surprise Index	15.10	28.74	−36.52	32.44	70.56
Industry HHI	1645	956	372	1380	9573

The Surface Index, measuring exposure within explicitly technology-sector occupations, averages 22.63 across districts. However, only 267 of 744 districts (35.9%) show any technology-sector employment at all. Nationally, technology-sector employment constitutes merely 0.96% of total weighted employment.

The Surprise Index—the difference between Iceberg and Surface indices—quantifies hidden cognitive exposure beyond visible technology adoption. The national mean Surprise Index of 15.10 indicates that the typical district's cognitive automation potential exceeds its technology-sector exposure by approximately 15 percentage points.

4.2 Geographic Distribution

State and union territory-level aggregation reveals substantial regional variation in AI occupational exposure (Table 2). The highest Iceberg Index values appear in Chandigarh (56.43), Puducherry (52.42), Sikkim (50.44), Goa (49.97), and Kerala (49.01). Complete state-level statistics are provided in Appendix D.

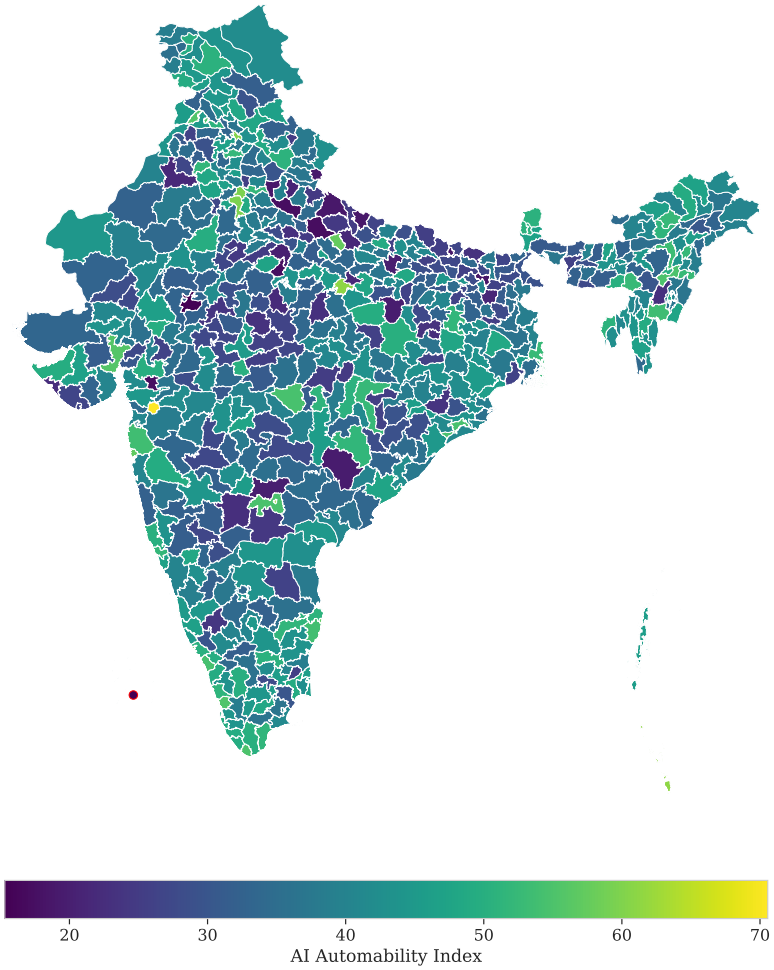

Fig. 2 Geographic distribution of AI Automability Index across India's 744 districts. The choropleth map displays district-level Iceberg Index values ranging from approximately 20 (dark purple, lowest exposure) to 70 (yellow, highest exposure).

Table 2 State and Union Territory-Level Iceberg Index: Top and Bottom Five

State	Iceberg	Surface	Surprise	Districts
Highest Exposure				
Chandigarh	56.43	63.47	−7.04	1
Puducherry	52.42	64.05	−11.63	4
Sikkim	50.44	66.19	−15.75	2
Goa	49.97	68.05	−18.08	2
Kerala	49.01	58.42	−9.41	14
Lowest Exposure				
Bihar	29.91	14.57	15.34	38
Madhya Pradesh	33.83	17.09	16.74	50
Uttar Pradesh	34.35	16.31	18.04	71
Jharkhand	36.75	17.62	19.13	24
Assam	37.35	9.37	27.98	33

The lowest state-level exposure appears in Bihar (29.91), Madhya Pradesh (33.83), and Uttar Pradesh (34.35). Notably, all five lowest-exposure states show positive Surprise Index values, indicating that their cognitive exposure substantially exceeds visible technology adoption.

Regional aggregation reveals: South India (42.06) exhibits the highest exposure, followed by West India (39.29), North India (37.28), and East India (36.66).

4.3 Districts with Maximum Hidden Exposure

The Surprise Index identifies districts where AI occupational exposure extends far beyond visible technology-sector presence. Table 3 presents the ten districts with highest Surprise Index values. For detailed case studies of these districts, see Appendix C.

Table 3 Districts with Highest Surprise Index (Hidden Exposure)

State	District	Iceberg	Surface	Surprise
Gujarat	The Dangs	70.56	0.00	70.56
Delhi	Central	67.72	0.00	67.72
Uttar Pradesh	Chitrakoot	61.19	0.00	61.19
Delhi	New Delhi	59.75	0.00	59.75
Nagaland	Kohima	54.82	0.00	54.82
Jammu & Kashmir	Ramban	54.43	0.00	54.43
Manipur	Churachandpur	53.45	0.00	53.45
Assam	Golaghat	53.14	0.00	53.14
Assam	Hojai	52.62	0.00	52.62
Delhi	South West	52.54	0.00	52.54

The Dangs district in Gujarat, a predominantly tribal area with limited industrial development, registers the highest Iceberg Index (70.56) in the entire dataset.

This counterintuitive finding reflects concentration of formal employment in government offices, cooperative banks, and educational institutions where clerical and administrative tasks dominate.

4.4 Urban-Rural Decomposition

Contrary to the assumption that AI exposure concentrates in urban technology hubs, our analysis reveals higher mean exposure in rural areas. Among 633 districts with both urban and rural employment data, the mean rural Iceberg Index (49.52) substantially exceeds the urban mean (31.46)—a difference of 18.06 percentage points.

This counterintuitive finding reflects the composition of formal employment in rural India: government offices, public banking, primary education, and agricultural extension services concentrate in rural formal employment, all involving substantial text processing, record-keeping, and coordination tasks with high AI exposure scores.

4.5 Industry Concentration Analysis

The Herfindahl-Hirschman Index (HHI) of industry exposure shares reveals whether districts face concentrated or distributed automation risk. Of 744 districts, 414 (55.6%) exhibit distributed exposure patterns, 252 (33.9%) show moderate concentration, and 78 (10.5%) display concentrated exposure.

Districts with distributed exposure require coordinated multi-sector intervention. Districts with concentrated exposure—often government-dominated northeastern districts or manufacturing-intensive industrial zones—may benefit from targeted sector-specific programs.

4.6 Socioeconomic Correlates

Census validation reveals moderate correlations between the Iceberg Index and socioeconomic development indicators. Across 463 matched districts, the Iceberg Index correlates positively with literacy rate ($r = 0.50$), internet access ($r = 0.49$), and urbanization ($r = 0.47$). A weak negative correlation appears with youth population share ($r = -0.23$).

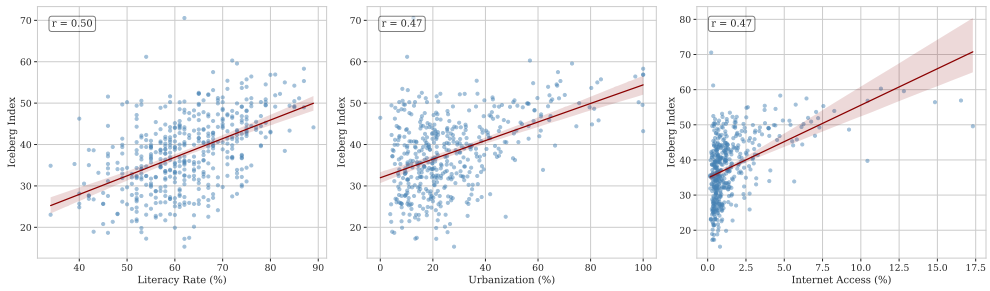

Fig. 3 Correlation between Iceberg Index and Census 2011 socioeconomic indicators across Indian districts. $N = 463$ matched districts.

4.7 Validation with Real-World Patterns

Our methodology produces exposure estimates consistent with observed AI adoption patterns. Districts containing major IT hubs (Hyderabad, Bangalore Urban, Pune, Gurgaon) rank among the highest in Surface Index, confirming alignment between our framework and actual technology-sector presence.

The wage estimation procedure successfully imputes values for 94.7% of employed individuals in the PLFS sample. Table 4 and Table 5 present detailed match quality statistics.

Table 4 NCO to O*NET Crosswalk Match Quality

Match Type	Count	Percentage	Confidence Score
Semantic keyword	1,016	29.5%	95
NCO prefix (4-digit)	2,417	70.2%	80
Division fallback	12	0.3%	60
Total	3,445	100.0%	—

Table 5 AIOE Score to O*NET Code Match Types

Strategy	Count	Percentage
Direct match	3,019	87.6%
Version mapping (2023→2019)	172	5.0%
Base SOC match	90	2.6%
Prefix-5 average	164	4.8%
Total	3,445	100.0%

Table 6 AI Automatability Score Distribution

Category	Score Range	Count	Percentage
Low	0–25	1,349	39.1%
Medium-Low	25–50	982	28.5%
Medium-High	50–75	666	19.3%
High	75–100	448	13.0%
Mean (SD)			40.6 (25.0)

5 Discussion

Our analysis reveals that AI occupational exposure in India extends substantially beyond the technology sector that dominates public discourse. The 41.5% wage-weighted national exposure, combined with technology-sector employment of less than 1%, indicates that the "submerged mass" of the iceberg—cognitive exposure in administrative, financial, and clerical occupations—constitutes the dominant share of automation potential.

The geographic distribution of exposure challenges conventional assumptions about AI's reach. Districts like The Dangs, Chitrakoot, and Ramban—conventionally associated with agriculture, pilgrimage, and rural heritage—register among the highest Iceberg Index values nationally (see Appendix C for detailed case studies). Their exposure derives not from technology-sector employment but from formal administrative work in government offices, banking institutions, and educational establishments. The cognitive tasks performed by clerks, data entry operators, and administrative assistants in these settings exhibit high alignment with current AI capabilities.

The counterintuitive urban-rural pattern merits particular attention. Rural formal employment in India concentrates in occupations with high AI exposure: government clerks, banking correspondents, primary school teachers, and agricultural extension workers. Urban areas, by contrast, contain both high-exposure (financial services) and low-exposure (retail, hospitality) occupations. This pattern suggests that AI's impact on rural India may arrive through formal administrative channels rather than agricultural automation.

The positive Surprise Index in lower-income states (Bihar, Uttar Pradesh, Madhya Pradesh, Jharkhand, Assam) indicates a structural mismatch between preparation and exposure. These states show minimal technology-sector presence yet substantial cognitive exposure in government and public sector employment. Workforce planning focused on visible technology disruption may substantially underestimate transformation potential in these regions. Appendix B provides detailed state-level case studies examining these patterns.

Our findings carry implications for India's service-sector development model. The entry-level cognitive jobs that have served as mobility pathways—data entry clerks, junior accountants, administrative assistants—represent precisely the tasks where AI demonstrates strongest capability. If these positions contract without corresponding growth in complementary roles, the "first rung" of the formal employment ladder may weaken.

Several limitations bound our analysis. The PLFS occupation codes aggregate to 3-digit NCO divisions, limiting within-division variation capture. The wage imputation procedure, while necessary given India's informal employment structure, introduces measurement error. The AIOE scores derive from US occupational characteristics and may not fully reflect Indian workplace contexts. Finally, the index measures technical capability overlap rather than adoption likelihood or displacement outcomes.

6 Conclusion

The India Iceberg Index provides the first comprehensive district-level quantification of AI occupational exposure for India's labor market. Our analysis of 744 districts reveals that approximately 41.5% of wage-weighted employment faces technical exposure to current AI systems, with the "hidden" cognitive exposure in administrative and clerical occupations substantially exceeding visible technology-sector disruption.

The findings suggest that framing AI as a "tech problem" fundamentally mischaracterizes its scope. Districts without technology-sector presence register among the highest exposure values, driven by formal administrative employment in government, banking, and education. Rural areas show higher mean exposure than urban areas, reflecting the concentration of formal cognitive work in non-metropolitan settings.

Policy implications flow from the industry concentration analysis. The majority of districts (55.6%) exhibit distributed exposure patterns requiring coordinated multi-sector intervention rather than targeted industry programs. States with positive Surprise Index values face particular urgency in recognizing cognitive exposure that current workforce planning may overlook.

Future work should extend this analysis temporally to track exposure evolution, incorporate firm-level adoption data to connect capability to implementation, and examine occupation-specific reskilling pathways. The India Iceberg Index provides a baseline for such investigations and a framework for evidence-based workforce policy in the AI era.

Declarations

- **Funding**: Not applicable.
- **Conflict of interest**: The author declares no competing interests.
- **Ethics approval**: Not applicable (secondary analysis of publicly available survey data).
- **Consent for publication**: Not applicable.
- **Author contribution**: S.M. conceived the study, developed the methodology, conducted the analysis, and wrote the manuscript.

Data Availability. The complete dataset, analysis code, and interactive visualization are publicly available at https://github.com/pikulsomesh/india-iceberg-index. The Streamlit application provides district-level exploration at https://india-iceberg-index.streamlit.app/.

Appendix A Data Sources and Technical Details

This appendix provides detailed technical documentation on data sources, crosswalk methodology, and scoring procedures.

A.1 Data Source Specifications

Table A1 summarizes the primary data sources, their provenance, and role in the analysis.

Table A1 Primary Data Sources

Data Source	Provider	Role in Analysis
PLFS 2023–24	NSSO, Government of India	Individual-level employment, occupation codes, earnings, geographic identifiers
NCO-2015	Ministry of Labour and Employment	3,445 occupation codes with hierarchical structure following ISCO-08
O*NET 28.0	U.S. Bureau of Labor Statistics	Skill and ability taxonomies for 1,016 occupations; target for crosswalk
AIOE Scores	Felten et al. (2023)	Language Modeling exposure scores for 774 O*NET occupations
Census 2011	Office of Registrar General	District-level socioeconomic indicators for validation

A.2 Industry Concentration Methodology

The Herfindahl-Hirschman Index (HHI) quantifies whether AI exposure concentrates in few industries or distributes broadly. Originally developed for market concentration analysis, HHI has been validated in labor economics for measuring employer dominance and sectoral risk [9].

For each district g, we calculate the exposure contribution from each 2-digit NIC industry code j:

$$E_{g,j} = \sum_{i \in g : \text{NIC}_i = j} \hat{w}_i \cdot m_i \cdot A_i$$

The industry share of exposure is $s_{g,j} = E_{g,j} / \sum_j E_{g,j}$, and the district HHI is $\text{HHI}_g = \sum_j s_{g,j}^2 \times 10000$.

We classify districts as: Distributed (HHI $<$ 1500), Moderate (1500 $\leq$ HHI $<$ 2500), or Concentrated (HHI $\geq$ 2500).

Appendix B State-Level Case Studies

This appendix presents detailed case studies of states exhibiting extreme patterns across the Iceberg, Surface, and Surprise indices.

B.1 Uniform High Exposure: Nagaland

Nagaland presents a striking example of uniform high exposure with complete absence of technology-sector employment. Across 11 districts, the mean Iceberg Index is 44.62 (SD = 7.03), while the Surface Index is uniformly zero—no district records any technology-sector employment.

Exposure Profile

The state's Iceberg Index ranges from 34.05 (Mon district) to 54.82 (Kohima). The exposure distribution is: distributed (5 districts), moderate (3 districts), and concentrated (3 districts).

Driving Factors

Nagaland's workforce structure is dominated by government employment. According to the Nagaland State Government's Personnel Information Management System, approximately 1.23 lakh employees were on the state government payroll as of March 2021 [12]. The 2025 Survey on Employment, Unemployment, Skill and Migration conducted by the Directorate of Economics and Statistics, Government of Nagaland, found that salaried employees accounted for 16% in the public sector and 13% in the private sector, with 27% of migrants citing public sector employment as their reason for migration [13]. District headquarters concentrate employment in occupations like clerks (NCO 411-413), data entry operators (NCO 413), and administrative assistants (NCO 334)—all with high AI exposure scores.

Policy Implications

With zero technology-sector employment, conventional "tech disruption" frameworks provide no guidance for Nagaland. Policy responses should focus on: (1) augmenting administrative capacity through AI tools rather than workforce reduction; (2) retraining programs for clerical workers toward supervision and citizen-interface roles; (3) digital infrastructure investment to enable human-AI collaboration in service delivery.

B.2 Technology-Dominated Exposure: Andhra Pradesh

Andhra Pradesh exemplifies technology-dominated exposure patterns, where visible Surface Index exceeds hidden cognitive exposure. The state's mean Surface Index (60.65) is the highest nationally among states with 10+ districts, while the mean Surprise Index (-20.01) is the most negative.

Exposure Profile

Across 13 districts, the mean Iceberg Index is 40.64 (SD = 4.22)—relatively uniform. However, the Surface Index ranges from 0.00 (Vizianagaram) to 68.05 (West

Godavari), reflecting extreme heterogeneity in technology-sector presence. The state's primary IT hub is Visakhapatnam, which hosts the Fintech Valley initiative launched in 2016, IT Special Economic Zone, and multiple software development centers [14, 15].

District-Level Patterns

The contrast between adjacent districts is instructive:

- **Visakhapatnam**: Iceberg = 47.64, Surface = 66.97, Surprise = −19.33. The state's primary IT hub with Fintech Valley and IT SEZ creates substantial visible technology employment. The Visakhapatnam Special Economic Zone operates under the Ministry of Commerce and Industry, Government of India [14]. The district anchors Andhra Pradesh's technology economy post-bifurcation from Telangana (which retained Hyderabad).
- **Vizianagaram**: Iceberg = 42.15, Surface = 0.00, Surprise = 42.15. An adjacent district classified among India's 250 most backward districts under the Backward Regions Grant Fund Programme by the Ministry of Panchayati Raj in 2006 [16]. Despite zero technology-sector employment, the district shows substantial hidden exposure driven by government administrative employment and tertiary sector services. This case demonstrates that geographic proximity to IT hubs does not translate to technology-sector employment spillover, while administrative and service sector formalization creates AI exposure through entirely different occupational channels.

Policy Implications

Andhra Pradesh requires bifurcated policy: technology-sector reskilling in Visakhapatnam and other hub districts, combined with administrative-sector adaptation in districts like Vizianagaram where formal employment concentrates in government services. The Vizianagaram pattern—high hidden exposure in a backward district adjacent to an IT hub—suggests that regional development spillovers may be more limited than commonly assumed.

B.3 Uniform Low Exposure: Bihar

Bihar exhibits the lowest mean Iceberg Index (30.15) among major states, with moderate standard deviation (SD = 6.06) indicating relatively uniform low exposure across its 38 districts.

Exposure Profile

The Iceberg Index ranges from 18.67 (Khagaria) to 44.19 (Patna). The Surface Index averages 10.40, with only 9 of 38 districts showing any technology-sector employment.

Driving Factors

Bihar's low exposure reflects its workforce structure: large agricultural employment share, limited formal-sector penetration, and constrained service-sector development. The informal economy dominates employment and receives low AI exposure scores.

District Extremes

- **Patna**: Iceberg = 44.19, the state's only district approaching national mean.
- **Khagaria**: Iceberg = 18.67, among the lowest nationally, with minimal formal-sector employment.

Policy Implications

Bihar's low exposure might appear advantageous, but it reflects underdevelopment rather than resilience. Early planning could enable AI-augmented pathways for new formal-sector entrants.

B.4 Maximum Hidden Exposure: Northeastern States

The northeastern states collectively exhibit the highest Surprise Index values nationally, indicating that AI exposure is almost entirely "hidden" in administrative occupations with negligible technology-sector presence.

Aggregate Profile

- **Nagaland**: Mean Surprise = 44.62, Surface = 0.00
- **Mizoram**: Mean Surprise = 37.73, Surface = 5.74
- **Arunachal Pradesh**: Mean Surprise = 36.84, Surface = 5.74
- **Manipur**: Mean Surprise = 15.47, Surface = 24.66

Common Structural Features

These states share: (1) Government employment dominance; (2) Limited private technology sector; (3) Banking sector formalization through financial inclusion programs; (4) Large public education workforces.

Policy Implications

The northeastern states face AI exposure through government employment rather than private technology sectors. Policy coordination between state governments and central agencies becomes critical.

B.5 Heterogeneous Exposure: Delhi

Delhi presents an unusual case of extreme within-state heterogeneity, with the highest standard deviation in Surprise Index (35.59) among states with 5+ districts.

Exposure Profile

Across 9 districts, the mean Iceberg Index is 51.44, but values range from 43.00 (North East) to 67.72 (Central). The Surprise Index spans 92 percentage points (-24.48 to 67.72).

District Extremes

- **Central Delhi**: Iceberg = 67.72, Surface = 0.00. Houses government ministries with clerical and administrative employment.

- **South West Delhi**: Iceberg = 43.57, Surface = 68.05. Contains Dwarka's IT corridor creating substantial technology-sector employment.

Policy Implications

Delhi's heterogeneity requires district-specific responses within a unified city framework.

Appendix C District-Level Case Studies

This appendix examines individual districts that represent extreme or unusual patterns in AI occupational exposure.

C.1 Maximum Deviation Districts

Table C2 presents districts with the largest positive and negative deviations from their respective state means.

Table C2 Districts with Maximum Deviation from State Mean Iceberg Index

State	District	Iceberg	State Mean	Deviation
Highest Positive Deviation				
Gujarat	The Dangs	70.56	37.03	+33.53
Telangana	Hyderabad	62.07	33.97	+28.10
Uttar Pradesh	Chitrakoot	61.19	33.74	+27.45
Uttar Pradesh	Lucknow	57.49	33.74	+23.75
Telangana	Rangareddy	54.77	33.97	+20.80
Highest Negative Deviation				
Telangana	Mulugu	13.72	33.97	−20.25
Chhattisgarh	Kondagaon	15.38	35.48	−20.10
Gujarat	Narmada	17.46	37.03	−19.57
Jammu & Kashmir	Reasi	22.02	40.63	−18.61
Karnataka	Yadgir	17.24	35.49	−18.25

C.2 Case Study: The Dangs, Gujarat

The Dangs district records India's highest Iceberg Index (70.56) despite being classified as a Scheduled Tribe-dominated, primarily forested, economically backward district.

District Profile

According to the District Collectorate, Government of Gujarat, The Dangs has a population of approximately 228,291 (Census 2011), with about 94% belonging to Scheduled Tribes—the highest concentration in Gujarat [10]. The district's demography portal reports a population density of 129 inhabitants per square kilometre with forest cover exceeding 75% of land area [11]. The district has no urban centers above 10,000 population and no industrial zones.

Employment Composition

The PLFS data reveals highly concentrated formal employment:

- **Public administration** (NIC 84): 97.8% of formal employment exposure
- Top occupations: Clerks, data entry operators, government administrative officers
- Zero technology-sector employment (Surface Index = 0.00)

Explanation

The Dangs' extreme Iceberg Index reflects the *composition* rather than *volume* of formal employment. As a tribal-dominated district, The Dangs receives substantial central and state government attention through tribal welfare departments, Integrated Tribal Development Agency offices, forest department headquarters, banking correspondents under financial inclusion mandates, and primary schools with teacher and administrative staff.

When formal employment is almost entirely government administrative work, the Iceberg Index mechanically approaches the AI exposure score of clerical occupations (which exceeds 70).

Policy Implications

The Dangs' case demonstrates that high Iceberg Index values do not necessarily indicate large absolute numbers of exposed workers. The district's total formal employment is small. However, AI adoption in government services could affect a large share of the district's formal employment base.

C.3 Case Study: Hyderabad, Telangana

Hyderabad district shows the second-highest positive deviation from state mean (+28.10), but for reasons entirely different from The Dangs.

District Profile

Hyderabad is India's fourth-largest metropolitan economy, with substantial presence in IT services, pharmaceuticals, financial services, and government administration. According to the Department of Information Technology, Electronics and Communications (ITE&C), Government of Telangana, IT/ITeS exports from Telangana reached Rs 2,41,275 crore in FY 2022-23, with 9,05,715 employees in the IT/ITES sector working in more than 1,500 companies [17]. The HITEC City development, established in 1998, hosts major global technology companies including Microsoft, Amazon, and Google [18].

Exposure Profile

- Iceberg Index: 62.07 (2nd highest nationally)
- Surface Index: 68.05 (near-maximum)
- Surprise Index: −5.98 (tech-dominated)

Explanation

Unlike The Dangs, Hyderabad's high exposure reflects genuine technology-sector concentration. The HITEC City and Gachibowli IT corridors employ over 600,000 workers in software development, data analytics, and business process outsourcing.

The slightly negative Surprise Index indicates that technology-sector employment *exceeds* what would be predicted from occupational composition alone.

Contrast with State

Telangana's state mean Iceberg Index (33.97) reflects the averaging of Hyderabad's high exposure with rural districts like Mulugu (13.72). This 48-point within-state range illustrates challenges of state-level policy when district conditions vary dramatically.

C.4 Case Study: Mulugu, Telangana

Mulugu district shows the lowest Iceberg Index nationally (13.72) and the largest negative deviation from state mean (−20.25).

District Profile

Mulugu is a newly created (2019) predominantly tribal district in northern Telangana, dominated by the Eturnagaram Wildlife Sanctuary with limited non-agricultural economic activity.

Exposure Profile

- Iceberg Index: 13.72 (lowest nationally)
- Surface Index: 0.00
- Industry HHI: 6646 (highly concentrated)
- Top industry: Agriculture (NIC 11), 79.8% of exposure

Explanation

Mulugu's near-absence of AI exposure reflects an economy dominated by primary-sector activities with minimal formal service-sector development.

Policy Implications

Mulugu represents districts where economic structure precludes significant automation impact in the current AI capability regime. Current policy attention should focus on development fundamentals rather than AI workforce adaptation.

C.5 Urban-Rural Inversions

Several districts exhibit counterintuitive patterns where rural exposure exceeds urban exposure substantially.

Srinagar, Jammu & Kashmir

- Urban Iceberg: 21.22
- Rural Iceberg: 51.04
- Urban-Rural Gap: −29.82

Srinagar's inversion reflects the security situation's impact on economic structure. Urban areas contain diverse informal-sector employment with low AI exposure. Rural areas contain formal government offices relocated from urban security zones, creating concentrated administrative employment.

C.6 Technology Islands in Agricultural States

Several districts show high Surface Index values despite being located in states with minimal overall technology-sector presence.

Lucknow, Uttar Pradesh

- Iceberg Index: 57.49
- Surface Index: 60.40
- State Mean Iceberg: 33.74
- Deviation: +23.75

Lucknow's IT parks create technology-sector employment contrasting sharply with surrounding agricultural districts.

Patna, Bihar

- Iceberg Index: 44.19
- Surface Index: 65.67
- State Mean Iceberg: 30.15
- Deviation: +14.04

Patna represents Bihar's only substantial formal-sector employment concentration.

C.7 Policy Targeting Implications

These case studies illustrate several principles for policy design:

1. **State-level policies are insufficient**: The 50+ percentage point within-state ranges indicate that state-level workforce programs will systematically misallocate resources. District-level targeting is essential.
2. **High exposure can reflect underdevelopment**: The Dangs' maximum Iceberg Index reflects concentrated government employment in an underdeveloped economy, not technology-sector strength.
3. **Technology islands require regional coordination**: Districts like Lucknow and Patna concentrate state-level technology employment, creating regional disparities that may widen as AI adoption accelerates.
4. **Rural exposure pathways differ**: Rural AI exposure primarily flows through government services, banking correspondents, and education—all sectors subject to central policy decisions.

Appendix D State-Level Aggregation Statistics

This appendix presents comprehensive state-level aggregation statistics for all three exposure indices. Complete tables are provided for the Iceberg Index, Surface Index, and Surprise Index across all states and union territories.

D.1 Iceberg Index by State

Table D3: State-Level Iceberg Index Statistics

State/UT	n	Mean	SD	Min	Max
Andaman & Nicobar	2	53.54	10.80	45.90	61.18
Andhra Pradesh	23	37.93	9.49	20.06	62.07
Arunachal Pradesh	15	42.66	5.31	34.33	52.26
Assam	23	37.26	6.86	26.06	53.14
Bihar	37	30.15	6.17	18.67	44.19
Chandigarh	1	56.43	—	56.43	56.43
Chhattisgarh	16	38.07	10.40	19.53	52.27
Delhi	1	59.75	—	59.75	59.75
Goa	2	49.38	3.55	46.87	51.89
Gujarat	25	38.45	11.29	17.46	70.56
Haryana	19	43.67	9.34	27.03	60.29
Himachal Pradesh	12	42.08	6.09	29.72	48.12
Jammu & Kashmir	14	43.16	5.40	34.66	50.56
Jharkhand	22	35.77	8.19	22.35	51.58
Karnataka	27	36.59	7.19	22.76	48.33
Kerala	14	48.01	4.57	38.46	55.33
Madhya Pradesh	48	33.72	8.25	15.32	50.39
Maharashtra	34	40.24	7.87	27.28	57.08
Manipur	9	40.13	9.04	22.99	53.45
Meghalaya	7	38.86	9.66	27.58	52.62
Mizoram	8	43.47	4.34	34.03	49.00
Nagaland	8	45.98	7.50	34.05	54.82
Odisha	30	38.26	7.19	23.64	54.48
Puducherry	4	50.33	7.15	43.22	58.31
Punjab	17	38.29	8.33	25.90	55.41
Rajasthan	32	35.74	7.13	21.69	49.82
Sikkim	4	49.79	1.74	47.18	50.66
Tamil Nadu	30	43.78	7.91	25.65	56.91
Tripura	4	44.30	4.52	40.71	50.91
Uttar Pradesh	70	33.89	9.72	17.26	61.19
Uttarakhand	13	37.36	8.68	19.05	50.72
West Bengal	19	39.09	8.02	26.29	56.84

Continued on next page

Table D3: State-Level Iceberg Index Statistics (continued)

State/UT	n	Mean	SD	Min	Max
All India	**594**	**38.21**	**9.13**	**15.32**	**70.56**

Notes: Statistics computed from district-level index values. SD = standard deviation; n = number of districts. States with single districts show no SD. The "All India" row aggregates across all districts with complete data. Index values are expressed on a 0–100 scale. Some smaller union territories are omitted for brevity.

References

[1] Felten, E.W., Raj, M., Seamans, R.: How will language modelers like ChatGPT affect occupations and industries? arXiv preprint arXiv:2303.01157 (2023)

[2] Chopra, A., Bhattacharya, S., Salvador, D., et al.: The Iceberg Index: Measuring skills-centered exposure in the AI economy. arXiv preprint arXiv:2510.25137v2 (2025)

[3] Rodrik, D.: Premature deindustrialization. Journal of Economic Growth **21**(1), 1–33 (2016)

[4] NASSCOM: Technology Sector in India 2024: Strategic Review. National Association of Software and Service Companies, New Delhi (2024)

[5] National Sample Survey Office: Periodic Labour Force Survey (PLFS) Annual Report (July 2023–June 2024). Ministry of Statistics and Programme Implementation, Government of India (2024)

[6] Ministry of Labour and Employment: National Classification of Occupations (NCO-2015). Government of India (2015)

[7] National Center for O*NET Development: O*NET OnLine. U.S. Department of Labor, Employment and Training Administration (2024). https://www.onetonline.org/

[8] Office of the Registrar General & Census Commissioner: Census of India 2011. Ministry of Home Affairs, Government of India (2011)

[9] Azar, J.A., Marinescu, I., Steinbaum, M.I., Taska, B.: Concentration in US labor markets: Evidence from online vacancy data. Labour Economics **66**, 101886 (2020)

[10] District Collectorate, The Dangs: About Dang. Government of Gujarat (2024). https://dangs.gujarat.gov.in/about-dang. Accessed 15 December 2025

[11] District Administration, The Dangs: Demography. Government of Gujarat, National Informatics Centre (2024). https://dangs.nic.in/demography/. Accessed 15 December 2025

[12] Department of Personnel and Administrative Reforms: Personal Information Management System (PIMS). Government of Nagaland (2021). https://nglemployeedirectory.in/. Accessed 15 December 2025

[13] Directorate of Economics and Statistics: Survey Report on Employment, Unemployment, Skill and Migration in Nagaland, 2025. Government of Nagaland (2025). https://ipr.nagaland.gov.in/. Accessed 15 December 2025

[14] Visakhapatnam Special Economic Zone: IT SEZs in AP. Ministry of Commerce and Industry, Government of India (2024). https://vsez.gov.in/it-sezs-in-ap/. Accessed 15 December 2025

[15] Government of Andhra Pradesh: Fintech Valley Vizag Initiative. Department of Information Technology, Electronics and Communications (2024). https://apit.ap.gov.in/. Accessed 15 December 2025

[16] Press Information Bureau: Districts under Backward Regions Grant Fund. Ministry of Panchayati Raj, Government of India (2012). https://www.pib.gov.in/newsite/PrintRelease.aspx?relid=91066&reg=3&lang=2. Accessed 15 December 2025

[17] Department of Information Technology, Electronics and Communications: IT/ITES Sector. Government of Telangana (2023). https://it.telangana.gov.in/. Accessed 15 December 2025

[18] Invest Telangana: IT/ITES Sector Overview. Government of Telangana (2024). https://it.telangana.gov.in/why-invest/. Accessed 15 December 2025

About the Author

Dr. Somesh Mohapatra is a global technology leader in artificial intelligence with experience spanning strategy, product, and data science across India, Japan, Singapore, and the United States. He has led AI-driven digital transformation across large, complex organizations, with a track record of building and scaling data science, analytics, and product teams to solve high-impact problems in industrial and socio-technical systems. His leadership work centers on deploying advanced machine learning, generative AI, and decision intelligence tools to shape the future of work, infrastructure, and industrial operations at scale.

In his industry career, Dr. Mohapatra has led data science and product organizations driving digital and AI transformation across Fortune 100 companies and public sector institutions. At Caterpillar Inc., he currently leads global data science and product management teams deploying AI for digital transformation in the manufacturing sector. He has shaped and implemented enterprise-wide digital and AI strategy for supply resiliency and electrification. Previously, at Google, he co-developed software platforms that modernized model building, hyperparameter tuning, and attribution workflows, significantly accelerating development cycles for advanced machine learning systems. His leadership portfolio also includes technical responsibilities at the Unique Identification Authority of India (UIDAI), where he contributed to data, privacy, and business frameworks for one of the world's largest digital identity programs.

On the academic front, Dr. Mohapatra holds a PhD in Artificial Intelligence from the Massachusetts Institute of Technology and an MBA from the MIT Sloan School of Management, where he was a Leaders for Global Operations Fellow, and he is a Gold Medalist from the Indian Institute of Technology Roorkee. His research record includes 30+ publications—spanning peer-reviewed papers, preprints, and book chapters—with 500+ citations across leading AI and machine learning venues such as the Conference on Neural Information Processing Systems (NeurIPS), the International Conference on Learning Representations (ICLR), and the International Conference on Machine Learning (ICML) - and multiple journals in the famed Nature portfolio. His scholarship bridges foundational advances in machine learning with real-world deployment in complex industrial and societal contexts, reinforcing his profile as a versatile, general-purpose leader in the global artificial intelligence ecosystem.