8th December 2025

Transforming Zambia's Labour Force Survey Using Artificial Intelligence

Ty Rossow[2], Tobias Edison[1], Rory Hardie[1], Harry Mayne[1, 3]

1 – International Growth Centre (IGC)
2 – International Monetary Fund, ex-IGC (paper produced solely at IGC)
3 – University of Oxford

IGC

Our Paper (in One Slide)

- Human assignment of labor codes is **time-consuming** and **unreliable**
- LLMs improve **performance** at **low cost**
- This can hasten **statistical production** and improve our understanding of the labor force **composition**
- **Scalability** is possible and desirable

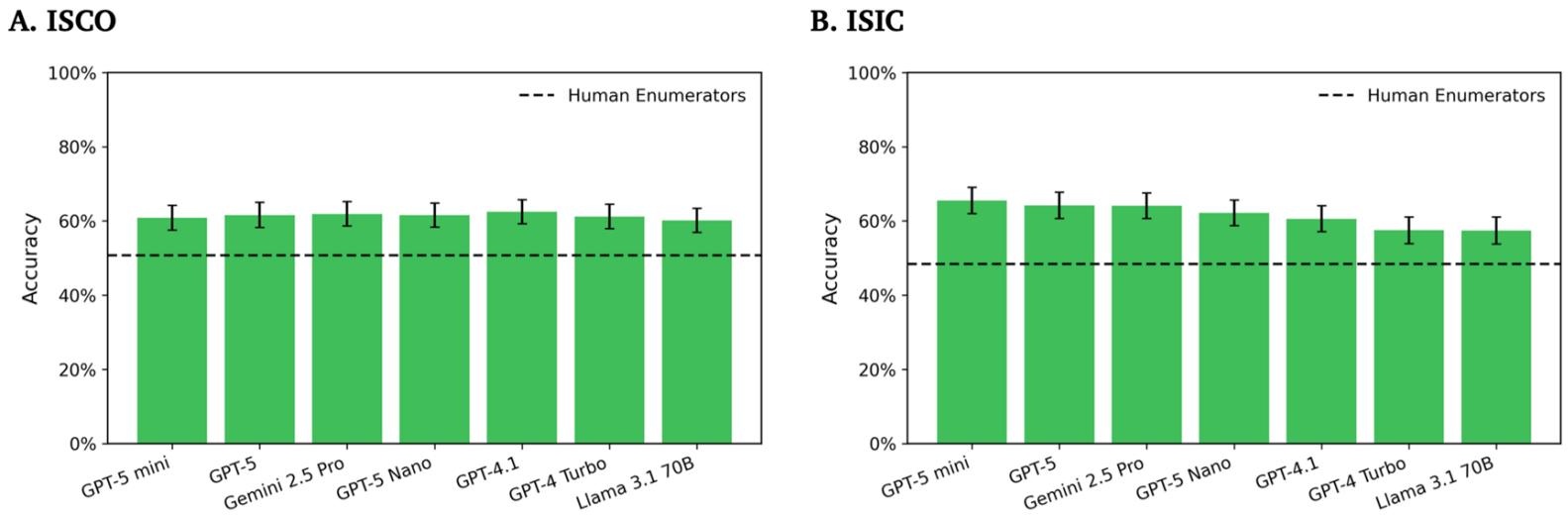
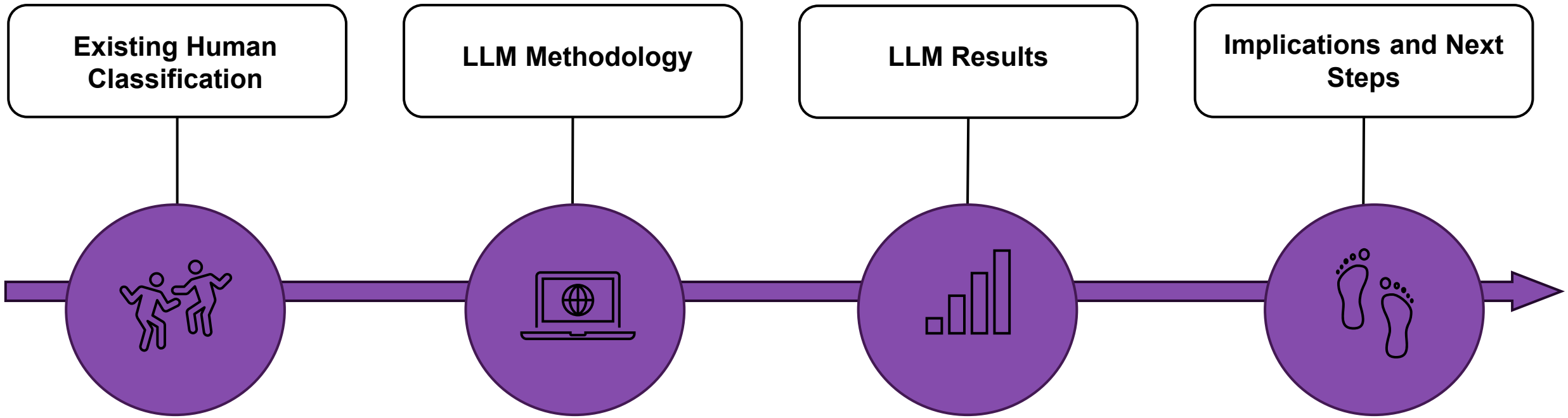


Figure 1: LLMs outperform human enumerators on labor force classification task.

Road Map


Existing Human Classification
LLM Methodology
LLM Results
Implications and Next Steps

Labour Force Survey (LFS) Context

- **Household-based sample survey** administered by the Zambia Statistics Agency (ZamStats) in partnership with the Ministry of Labour and Social Security (MLSS)
- **10,400** households interviewed by enumerators annually
- Face-to-face interviews quarterly

Source: Coauthor Rory Hardie with LFS Enumerators

Labour Force Survey (LFS) Context

- Core component: '*Characteristics of the Main Job*' module to identify occupation and industry
  - **Occupation**: In his/her main job/business, what kind of work does (NAME) usually do? (Write occupation title, if any, and main duties and tasks)
  - **Industry**: In (NAME) workplace what kind of business/activity is mainly carried out? (Write name of establishment, if any, and main activity, goods, or services)

- Enumerators use to assign four-digit International Standard Classification of Occupations (ISCO) and the International Standard Industrial Classification (ISIC) codes

- ISCO describes **job**, while ISIC describes **industry**

Example Classification

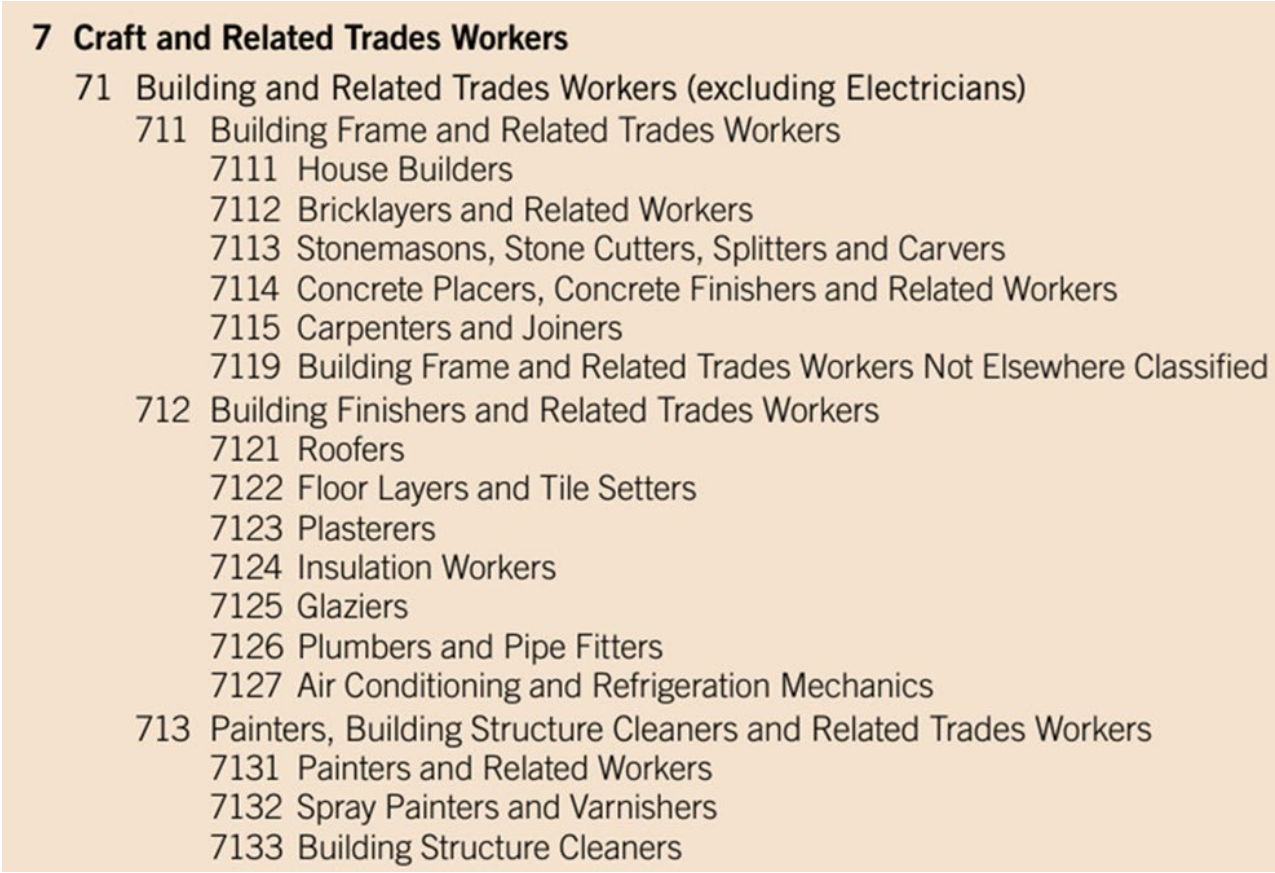
- Recorded observations to main work and business activity
- `D1_TITLE`: Subsistence Crop Farmer
- `D1_DESC`: Subsistence Crop Farmer Growing of Maize
- `D2_MAIN_ACTIVITIES`: Subsistence Crop Farmer Growing Crops for Sale

Number of Digits	ISCO	ISIC
One	Major Group **6**310 – Skilled Agricultural, Forestry, and Fishery Workers	N/A
Two	Sub-major Group **63**10 – Subsistence Farmers, Fishers, Hunters, and Gatherers	Division **01**11 – Crop and Animal Production, Hunting, and Related Service Activities
Three	Minor Group **631**0 – Subsistence Crop Farmers	Group **011**1 – Growing of Non-Perennial Crops
Four	Unit Group **6310** – Subsistence Crop Farmers	Class **0111** – Growing of Cereals (Except Rice), Leguminous Crops, and Oil Seeds

Table 1: Classification example.

Classification Challenges

- Existing challenges:
- **Complex codes**: 436 ISCO and 419 ISIC, codebooks > 300 pages.
- **Description detail**: time constraints reduce description detail.
- **Spelling errors**: misspelled keywords prevent correct retrieval.



7 Craft and Related Trades Workers

71 Building and Related Trades Workers (excluding Electricians)

711 Building Frame and Related Trades Workers

7111 House Builders
7112 Bricklayers and Related Workers
7113 Stonemasons, Stone Cutters, Splitters and Carvers
7114 Concrete Placers, Concrete Finishers and Related Workers
7115 Carpenters and Joiners
7119 Building Frame and Related Trades Workers Not Elsewhere Classified

712 Building Finishers and Related Trades Workers

7121 Roofers
7122 Floor Layers and Tile Setters
7123 Plasterers
7124 Insulation Workers
7125 Glaziers
7126 Plumbers and Pipe Fitters
7127 Air Conditioning and Refrigeration Mechanics

713 Painters, Building Structure Cleaners and Related Trades Workers

7131 Painters and Related Workers
7132 Spray Painters and Varnishers
7133 Building Structure Cleaners

Figure 2: Granularity of classification categories.

LLM Task and Dataset

- **Task**: LLM interprets free-text descriptions and return ISCO or ISIC classification (sequential), return JSON containing code
- **Dataset**: 1,000 observations labelled by Zambia Statistics Agency LFS team
- **Ground truth codes** independent of enumerator classifications

Figure 3: Comparison groups.

LLM Methodology

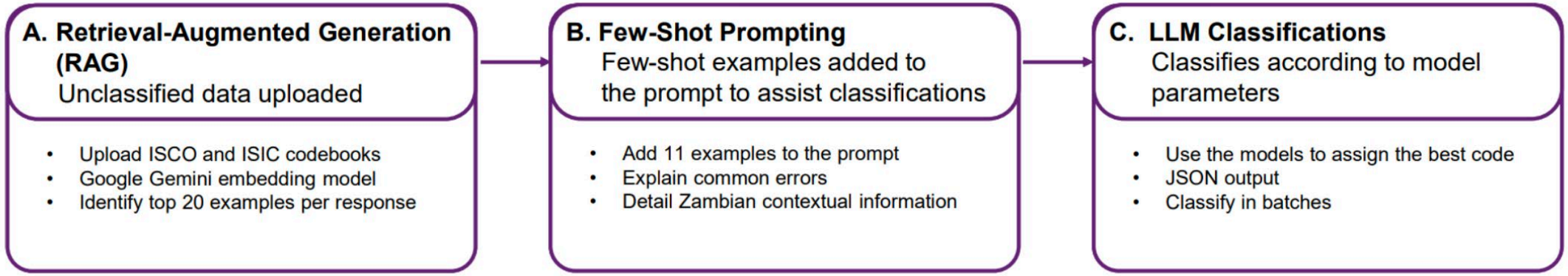


Figure 4: Methodology.

- RAG uses semantic similarity to reduce **cost** and **input** load
- Few-shot prompting adds **Zambian context**
- **LLM** classifies based on main prompt and examples (seven leading LLMs)

Few-Shot Prompting and Zambia Knowledge

CASE C.2.1. Assumed economic sector

D1_TITLE:	Selling chicken for sale
D1_DESC:	Selling chicken
D2_MAIN_ACTIVITIES:	Selling chicken

CASE C.2.2. Developing country knowledge required

D1_TITLE:	Business lady (retailer)
D1_DESC:	Buying and selling of fried cassava
D2_MAIN_ACTIVITIES:	Buying and selling of fried cassava

CASE C.2.3. Zambia knowledge required

D1_TITLE:	Kachasu brewer
D1_DESC:	Selling kachasu at whole sale price
D2_MAIN_ACTIVITIES:	Selling kachasu beer at whole price

Scoring and Assumptions

- **Exact** and **partial** match
- Assumption: ISCO and ISIC codes are correctly assigned (**ground truth**)
- Assumption: sample is **representative** (internal and external validity)
- Assumption: ISCO and ISIC codes capture Zambia's occupational and industry diversity (**informal sector**)
- Assumptions determine **benchmark validity** (Alaa et al., 2025): does the test capture the real-world phenomenon

Results: Partial Scoring

LLMs outperform human enumerators at every level of code granularity

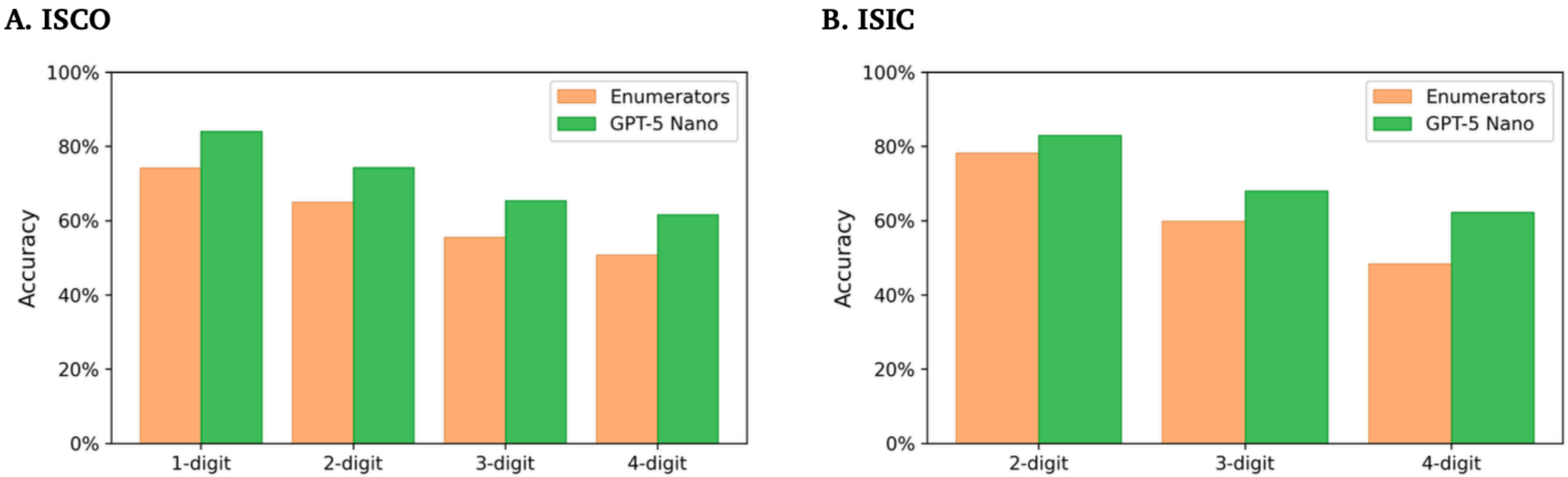


Figure 5: Comparison of GPT-5 Nano, the most cost-effective model, against enumerators across all digits. All results are statistically significant at the 99.9 percent confidence level. Error bars are omitted as confidence intervals are only informative on the pairwise difference between the LLMs and enumerators, rather than on the individual proportions.

LLM Results (ISCO)

Cumulative Digits	GPT–4.1	GPT–5	GPT–5 mini	GPT–5 Nano	Gemini 2.5 Pro
One	0.1140*** (0.0132)	0.1070*** (0.0138)	0.0940*** (0.0135)	0.0990*** (0.0131)	0.1010*** (0.0133)
Two	0.0970*** (0.0165)	0.1000*** (0.0163)	0.0850*** (0.0162)	0.0930*** (0.0154)	0.0890*** (0.0160)
Three	0.1080*** (0.0167)	0.1030*** (0.0170)	0.1000*** (0.0168)	0.0990*** (0.0163)	0.1060*** (0.0165)
Four	0.1180*** (0.0170)	0.1090*** (0.0172)	0.1020*** (0.0169)	0.1090*** (0.0165)	0.1120*** (0.0166)

Table 2: Accuracy Gain Over Enumerators (ISCO). Each cell shows the LLM's accuracy gain over enumerators for the ISCO classification task, with p-values computed using McNemar's test and Wald standard errors in parentheses. Performance results for Llama 3.1 70B and GPT-4 Turbo are included in Appendix B due to spacing constraints.

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

LLM Results (ISIC)

Cumulative Digits	GPT–4.1	GPT–5	GPT–5 mini	GPT–5 Nano	Gemini 2.5 Pro
Two	0.0400** (0.0137)	0.0680*** (0.0130)	0.0660*** (0.0130)	0.0480*** (0.0129)	0.0590*** (0.0133)
Three	0.0510** (0.0166)	0.1040*** (0.0162)	0.1070*** (0.0159)	0.0810*** (0.0158)	0.0960*** (0.0163)
Four	0.1220*** (0.0182)	0.1580*** (0.0179)	0.1710*** (0.0176)	0.1380*** (0.0176)	0.1570*** (0.0178)

Table 3: Accuracy Gain Over Enumerators (ISIC). Each cell shows the LLM's accuracy gain over enumerators for the ISIC classification task, with p-values computed using McNemar's test and Wald standard errors in parentheses. Performance results for Llama 3.1 70B and GPT-4 Turbo, which were weaker on ISIC digits, are included in Appendix B due to spacing constraints.
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Cost Implications

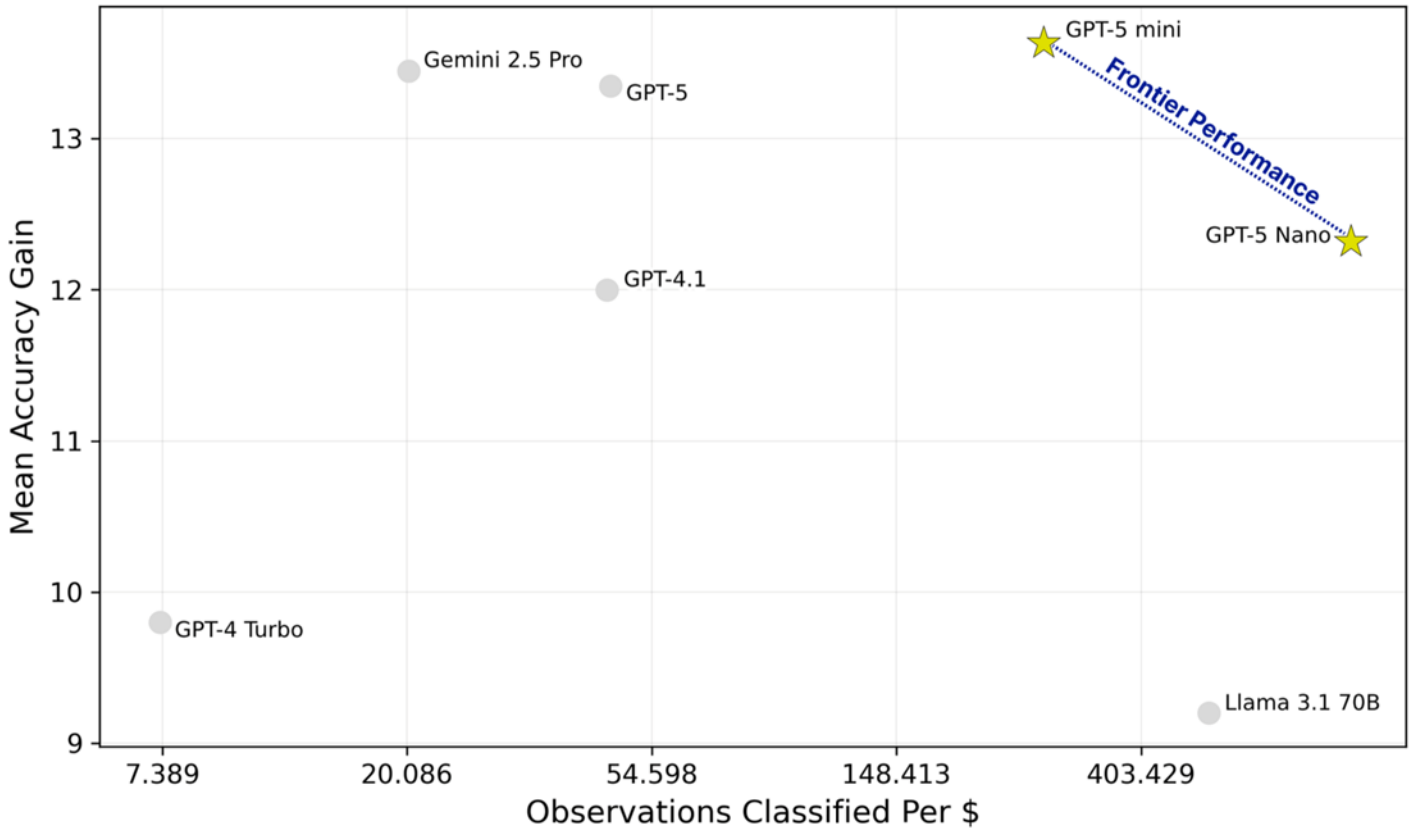


Figure 6: Cost against performance comparison (log scale).

Provider	Model	Total Cost ($)
OpenAI	GPT-4 Turbo	137.12
Google	Gemini 2.5 Pro	49.60
OpenAI	GPT-4.1	21.90
OpenAI	GPT-5	21.69
OpenAI	GPT-5 mini	3.60
Nebius AI Studio	Llama 3.1 70B	1.88
OpenAI	GPT-5 Nano	1.07

Table 4: LLM costs per 1,000 classifications

- Cheaper models exhibit **frontier performance**
- With GPT-5 Nano, LFS for **under $10 annually** given typical labor participation

Statistical Implications

- **Time savings** for national statistics agencies (43 working days if 1 minute per code)
- Speed up statistical production
- Improved **sectoral** understanding

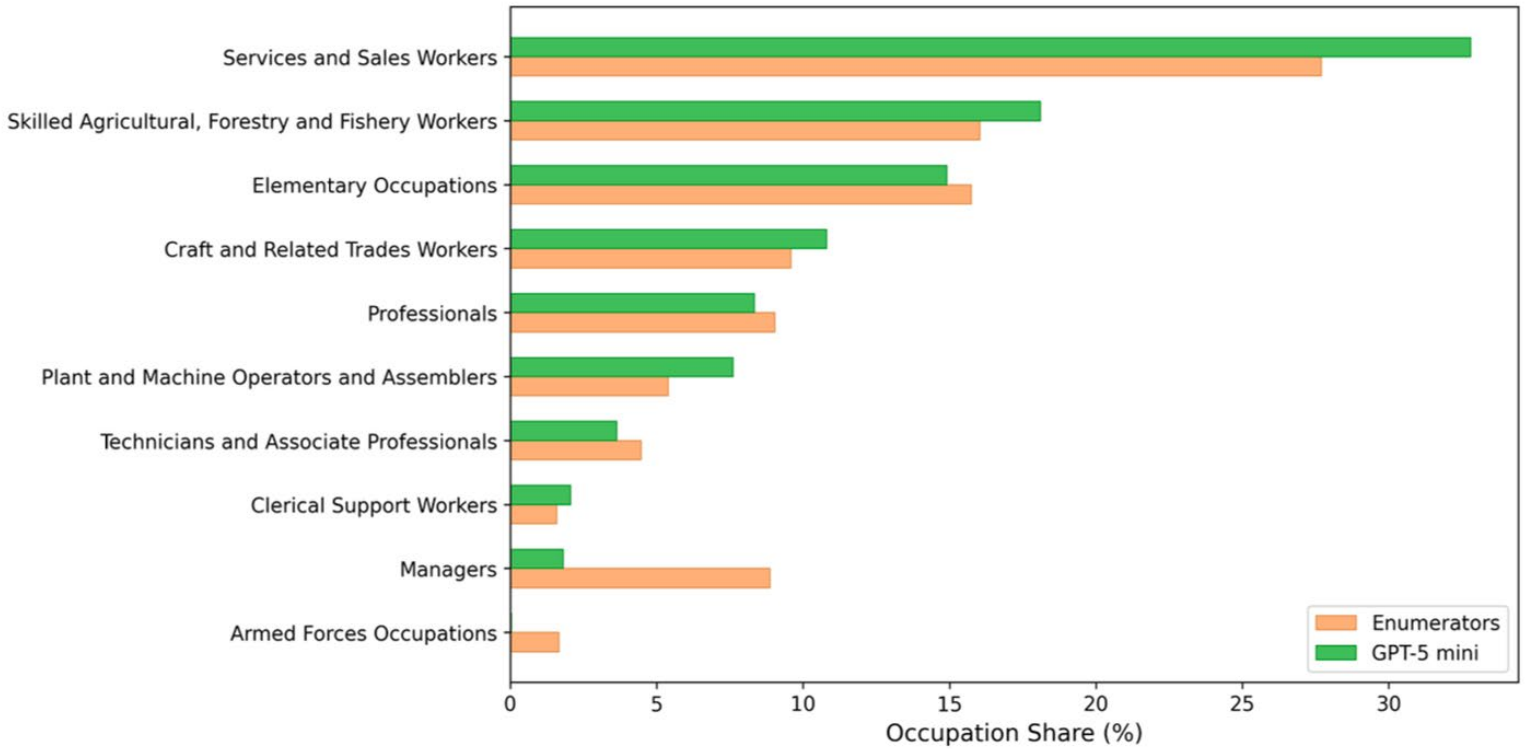


Figure 7: Occupation shares in 2022 LFS. All differences are statistically significant for $p < 0.05$. Confidence intervals are omitted as they are based on the pairwise difference between GPT-5 mini and enumerators rather than the individual proportions.

Scalability

- Graphical User Interface (GUI) for **no code** implementation
- Scalability in Zambia: **Census** data (4 million households, LLM cost $4,250)
- Scalability elsewhere: **national labor surveys** with local context

Figure 8: GUI developed by IGC for Zambia Statistics Agency implementation.

Limitations

- Limited **scope**: Zambian LFS (generalizability)
- Sample size: too small for statistically significant differences **between LLMs**
- Ground truth data: possible errors and lack of **inter-annotator agreement** to measure reliability
- Developing gold standard with **multiple human annotators** and **larger sample**

IGC

Conclusion

- LLM-based methods significantly outperform human enumerators (up to **17.1 percentage points** on exact match)
- Cost-effectiveness as low as $1.07 per 1,000 records and **$10 annually**
- Hastens **statistical production** and improves understanding of **labor force composition**
- Scalability to **Census**, other countries with GUI and local context

IGC