An AI-Driven Lens on The Demand side of the Egyptian Labor Market (2021-To Date)

Part II: An Agentic-AI System for ISCO-08 Occupational Classification

WP No. 242

July 2025

This working paper is the second in a series of papers by the Egyptian Center for Economic Studies (ECES), explaining the methodology for assessing skill demand in Egypt's labor market using AI. It is authored by Abdallah El-Lawah, External AI Consultant; Ahmed Habashy, AI Engineer; and Youssef Nasr, Research Analyst. The core project team also includes Ahmed Dawoud, Head of the Data Analytics Unit at ECES; Sondos Samir, Research Analyst; and Aya Saleh, Research Analyst.

Abstract

Accurate and scalable occupational classification is a foundational challenge for modern labor market analysis. This paper documents the methodological evolution of job classification within the ECES Labor Demand project, charting a deliberate progression from manual methods to a sophisticated AI system. The culmination of this work is JobIt-CLF, a novel system built on an agentic architecture that combines a Large Language Model (LLM) with a specialized SQL Agent for granular, 4-digit ISCO-08 coding. By grounding its reasoning in a comprehensive database of official ILO definitions, the system mimics expert human logic through a transparent, hierarchical process. Rigorously validated against both large-scale automated audits and expert human review, JobIt-CLF achieves 94–97% accuracy. By detailing this multi-stage evolution, we present a robust, replicable framework for aligning large-scale labor market data with international standards, ensuring more impactful economic analysis.

ملخص

يُعد تصنيف الوظائف بدقة وعلى نطاق واسع، تحديًا أساسيًا في التحليلات الحديثة لسوق العمل. توثق هذه الورقة التطور المنهجي الذي اتبعه مشروع "الطلب على الوظائف" بالمركز المصري للدراسات الاقتصادية، حيث ترصد التحول المدروس من الأساليب اليدوية إلى منظومة ذكاء اصطناعي متطورة. وقد أثمر هذا الجهد عن تطوير "JobIt-CLF"، وهو نظام مبتكر قائم على بنية وكيل ذكي يدمج بين قدرات النماذج اللغوية الكبيرة (LLM) ووكيل SQL متخصص، بهدف إجراء تصنيف مهني دقيق يصل إلى المستوى الرابع من التصنيف الدولي الموحد للمهن (ISCO-08). ومن خلال إرساء استنتاجاته على قاعدة بيانات شاملة من التعريفات الرسمية الصادرة عن منظمة العمل الدولية، يحاكي النظام منطق الخبراء البشريين عبر آلية هرمية شفافة. وبعد إخضاعه لعمليات تدقيق صارمة، شملت مراجعات آلية واسعة النطاق وتقييمًا من قبل خبراء بشريين، أثبت النظام دقته العالية التي تتراوح بين 94% و97%. عبر استعراض هذا التطور متعدد المراحل، فإننا نقدم إطار عمل راسخًا وقابلًا للتطبيق، يهدف إلى مواءمة بيانات سوق العمل الضخمة مع المعايير الدولية، بما يضمن الخروج بتحليلات اقتصادية أكثر تأثيرًا وعمقًا.

Disclaimer

The findings and analysis presented in this study are based exclusively on online job postings sourced from trusted and credible platforms. While this approach does not capture the entirety of labor market demand in Egypt, it offers a valuable and timely perspective into employer needs. Online data provides real-time, continuously updated insights that reflect evolving market trends and recruitment practices. As digital platforms increasingly become the primary channel for job advertising, this method represents a forward-looking approach to understanding the demand side of the labor market with greater relevance, accuracy, and immediacy.

Contents

List of Figures

1. Introduction and Rationale

The fidelity of job classification is the bedrock of reliable labor market analysis. For any large-scale analysis, an error at this foundational stage can corrupt an entire dataset, rendering subsequent policy recommendations unreliable. Recognizing this high-stakes challenge from the outset of the ”Demand for Skills” project, our team made a core commitment to best practices. Early feedback from stakeholders confirmed that adherence to the International Labour Organization’s (ILO) framework was not merely advisable, but essential for the project’s credibility. This imperative became the catalyst for a dedicated, multi-year developmental effort.

Meeting this challenge was not a single action but a deliberate, multi-stage evolution. This paper documents that journey—from initial, coarse methods and intensive manual reviews to the sophisticated AI engine we deploy today. Our objective is to provide a transparent, step-by-step account of the methodology that now powers our high-accuracy classification system, detailing the rationale and key learnings that shaped its final, validated form.

2. Evolution of the Job Classification Methodology

The path to our current classification engine was a pragmatic one, driven by a cycle of identifying a core limitation and engineering its solution. Each methodological phase served as a direct response to the shortcomings of its predecessor, creating a clear, evidence-based case for the next technological step. This iterative process ensured that our final system was not merely designed in theory but forged in practice. This section examines that progression, detailing how each of the following four stages not only improved upon the last but also revealed the necessity for the next:

- Stage 1: Source-Based Classification
- Stage 2: Manual Classification
- Stage 3: Basic LLM-Based Classification with Expert Validation
- Stage 4: Advanced Agentic ISCO Classification (JobIt-CLF)

2.1. Stage 1: Source-Based Classification

In its foundational phase, the project required a swift, scalable method to impose initial order on a vast and unstructured dataset. The solution was a source-based classification strategy, which operated on a simple assumption: a job’s category could be inferred from its platform of origin. Job postings from websites like Forasna, known for manual and service-oriented roles (e.g., plumbers, drivers, chefs), were systematically tagged as *blue-collar*. Conversely, those from professional networks like LinkedIn were categorized as *white-collar*.

While this source-based logic provided a rapid first-pass classification, its inherent flaw quickly became apparent. Internal validation revealed an accuracy ceiling of approximately 70%. The primary issue was platform bleed: a significant volume of professional roles appeared on traditionally blue-collar sites, and vice versa. This 30% misclassification rate represented an unacceptable margin of error, making it clear that a more granular, content-aware approach was not just an improvement, but a necessity.

2.2. Stage 2: Manual Classification

The clear limitations of source-based categorization prompted a shift to a human-in-the-loop model. The strategy was to leverage the initial automated classification as a first pass, but to then deploy ECES economists to manually audit and correct the inevitable misclassifications. Their task was to meticulously examine the data and re-categorize any job posting where the title was incongruous with its source platform's general theme.

This two-step process significantly enhanced the accuracy of the dataset and marked a crucial move toward more granular classification. Yet, the victory for precision revealed a critical operational flaw. The sheer volume of data made the manual correction process a severe constraint on the project's speed and agility. It became evident that while expert judgment was essential, relying on it for high-volume correction was unsustainable. The project needed a tool that could absorb this expert logic and apply it automatically, leading directly to the exploration of AI-driven methods.

2.3. Stage 3: Basic LLM-Based Classification with Expert Validation

The unscalable nature of manual review created a clear imperative: to find a way to automate the expert logic developed in the previous phase. This led to the project's first strategic step into Artificial Intelligence. Launched in the first half of 2024, this pilot phase had a deliberately focused objective: to test whether a Large Language Model (LLM) could be trained to reliably perform the broad "white-collar" versus "blue-collar" classification, with human experts serving as the ultimate judge of accuracy.

The core challenge was to translate the nuanced intuition of a human expert into a structured, machine-executable prompt. To ensure the model's logic was grounded in established labor market theory, the classification criteria were developed through extensive desk research of International Labour Organization (ILO) guidelines and subsequently validated with industry stakeholders. This process yielded seven distinct criteria against which the LLM was mandated to evaluate each job title:

1. **Work Environment:** Does the role primarily take place in an office setting (+1), or in a non-office environment like a factory, construction site, or on the road (−1)?
2. **Physical Labor:** Is the job's core function cognitive and analytical with minimal physical exertion (+1), or does it involve significant hands-on, physical tasks (−1)?
3. **Education Requirements:** Does the role typically require a bachelor's degree or higher (+1), or is it more commonly associated with vocational training or a high school diploma (−1)?
4. **Primary Skills:** Does the job rely on analytical, strategic, or creative abilities (+1), or does it depend on technical, mechanical, or manual skills (−1)?
5. **Typical Attire:** Is the expected dress code business or business casual (+1), or does the role require a uniform or protective gear (−1)?
6. **Industry Association:** Is the job in a sector like finance, law, or technology (+1), or is it in manufacturing, construction, or agriculture (−1)?
7. **Decision-Making Level:** Does the position involve a high degree of autonomy and strategic input (+1), or does it primarily involve following established procedures under direct supervision (−1)?

To ensure a definitive judgment and avoid classifying ambiguous roles, we implemented a quantitative scoring system with a high threshold for classification. A job title's final score was the sum of its points across the seven criteria, with the following rules:

- A total score **greater than** $+4$ resulted in a **"White-collar"** classification.
- A total score **less than** -4 resulted in a **"Blue-collar"** classification.

This strict threshold intentionally created a wide "zone of ambiguity" for any job scoring between -4 and $+4$. A job had to show a very strong consensus among the criteria—with at least five more points in one direction than the other—before a firm classification was assigned. As before, the LLM had to provide a step-by-step rationale for its scoring to ensure its reasoning was fully auditable.

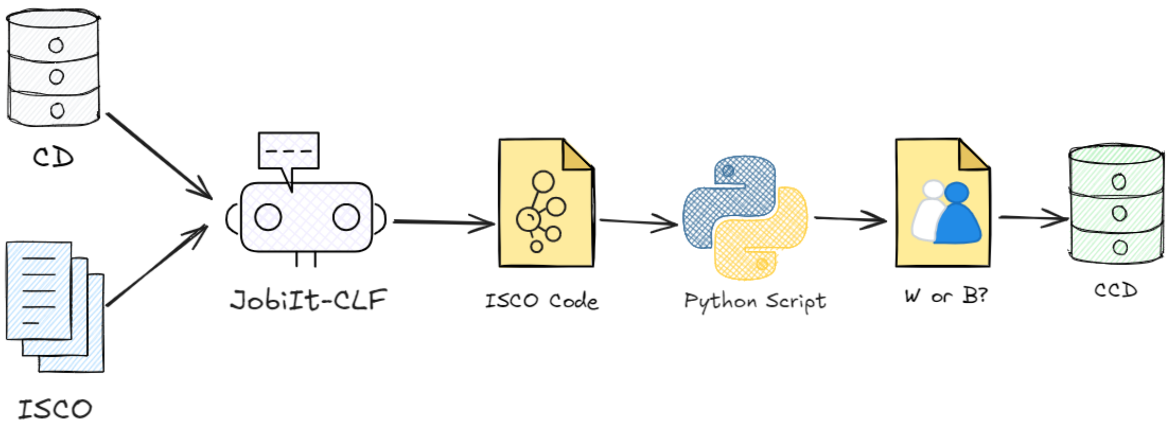
To rigorously measure the pilot's effectiveness, a validation procedure was conducted by ECES's trained economists. Three random samples, each containing 100 job postings, were drawn from the dataset. Our economists manually classified each job in the samples, and their classifications were used as the ground truth. The LLM's automated results were then compared against this benchmark.

This process yielded an average accuracy rate of **90%**. While not perfect, this was a highly promising result. It demonstrated that an AI-driven approach could achieve great accuracy, successfully automating the vast majority of the classification workload. This strong validation provided the quantitative justification and institutional confidence needed to proceed to the final phase: developing a more sophisticated system for granular, 4-digit international standard classification.

2.4. *Stage 4: Advanced Agentic ISCO Classification (JobIt-CLF)*

The success of the AI pilot was a critical green light, but it was only the beginning. The ultimate goal was far more ambitious than a simple two-way sort. We needed a system capable of assigning a precise, 4-digit International Standard Classification of Occupations (ISCO-08) code to every job posting—a task of immense complexity. This required moving beyond a simple pilot to engineer an industrial-strength, specialized AI system. This final phase marks the creation and deployment of that system: **JobIt-CLF (JobIt Classifier)**, whose workflow is illustrated in Figure 2.1..

Figure 2.1.: Job Classification Workflow



Source: Author's own illustration.

To build an AI that could master this task, we first had to teach it the language of international occupational standards. The ISCO-08 framework, developed by the ILO, provides a powerful and hierarchical system for organizing the world of work. Each job is assigned a 4-digit code that places it within a nested structure of increasing specificity:

- **1-digit Major Group:** The broadest category (e.g., 2 for "Professionals").
- **2-digit Sub-Major Group:** A more specific field within the major group (e.g., 21 for "Science and Engineering Professionals").

- **3-digit Minor Group:** A further specialization (e.g., `213` for "Life Science Professionals").
- **4-digit Unit Group:** The most detailed role, one of 436 unique occupations (e.g., `2132` for "Farming, Forestry and Fisheries Advisers").

Mastering this hierarchy is essential for the deep, meaningful labor market analysis that is the project's entire purpose.

2.4.1 The Architecture of JobIt-CLF

At the heart of this phase lies the **JobIt-CLF** engine. To build it, we did not start from scratch. We armed the system with a comprehensive knowledge base—a database containing the ILO's own detailed definitions for every ISCO code. This database includes detailed descriptions of tasks, required skills and qualifications, and even official example job titles for all 436 occupations.

With this knowledge base in place, JobIt-CLF processes each job posting from ECES's clean dataset, focusing on its title, description, and listed skills. Instead of making a single, high-stakes guess at the 4-digit code, the engine follows a logical, top-down process of elimination that mimics expert human reasoning:

1. **Step 1: Choose the Major Group.** The LLM first analyzes the job posting to make the broadest possible decision: "Is this a 'Manager', a 'Professional', a 'Technician', or a 'Clerical Support Worker'?" It selects the most appropriate 1-digit code and, crucially, provides a written justification for its choice.
2. **Step 2: Navigate the Hierarchy.** With the Major Group selected, the engine drills down. If it chose "Professionals" (Code 2), its next question is, "Is this a 'Science and Engineering Professional' (21), a 'Health Professional' (22), or perhaps a 'Business and Administration Professional' (24)?"
3. **Step 3: Executing a Tool-Use Cycle with the SQL Agent.** To move from a broad category to a precise code, the system must consult the ISCO database. It does this by initiating a sophisticated tool-use cycle. The LLM, acting as the project manager, recognizes when it has reached the limits of its own general knowledge and needs a specialist. At this moment, it activates its most important tool: the ***SQL Agent***.

   This process of intelligent delegation is best understood with an example. Imagine the system is analyzing a job posting for a Marketing Manager." The LLM might quickly determine it belongs to Major Group '1' (Managers) and Sub-Major Group '12' (Administrative and Commercial Managers). However, to select the correct 3-digit Minor Group, it need to know the precise difference between, for example, ‘122‘ (Sales, Marketing and Development Managers) and ‘123‘ (Services Managers). Here is how the cycle unfolds:

(a) **The Question:** The LLM formulates a clear, natural-language question for its specialist, such as: I need the official ILO definitions, typical tasks, and required skills for ISCO minor groups 122 and 123.”

(b) **The Translation:** This question is passed to the SQL Agent. The agent’s sole purpose is to act as an expert translator. It takes the English question and converts it into a precise command that our database understands—a structured SQL query. It does not reason about the job posting itself; it simply translates the request into SQL code.

(c) **The Retrieval:** The agent executes this SQL query on our dedicated knowledge base, which contains all the official ISCO-08 information. It pulls the exact, structured data requested—the official definitions for codes ‘122‘ and ‘123‘—and returns it to the LLM.

(d) **The Synthesis and Selection:** Finally, the LLM synthesizes the retrieved information. It critically evaluates the official ISCO definitions, tasks, and skills, comparing them directly against the content of the job posting. This evidence-based process allows it to select the code that best reflects the job’s core functions and responsibilities.

4. **Step 4: Arrive at the Final Code.** This methodical process of “query, compare, and justify” is repeated for the 3-digit and finally the 4-digit level, ensuring the final classification is not a guess, but the result of a transparent and auditable deductive process.

2.4.2 Validation: A Two-Pronged Approach to Proving Accuracy

A system this critical cannot simply be claimed to be accurate; its reliability must be proven. To achieve this, we designed a rigorous, two-pronged validation strategy.

- **Machine-to-Machine Audit:** First, we conducted a massive-scale audit where a more advanced, “expert” LLM was used to review the work of JobIt-CLF. This involved processing **100 random samples, each containing 100 unique job postings** (a total of 10,000 classifications). This provided a large-scale statistical measure of the system’s consistency and accuracy.

- **The Human Gold Standard:** Second, we subjected the system to the ultimate test: human expert review. A senior ECES economist with deep expertise in occupational classification independently reviewed **five additional random samples of 100 job postings each**.

Across both validation methods, the results were consistent and exceptional. JobIt-CLF achieved an **accuracy rate between 94% and 97%**, confirming its high reliability for large-scale, automated classification and far exceeding the performance of any previous method.

2.4.3 The Final Output: The Classified Clean Dataset (CCD)

The culmination of this entire four-phase journey is the **Classified Clean Dataset (CCD)**. This is the project's master data asset. It contains tens of thousands of clean, structured job postings, where each posting has been enriched with a validated, 4-digit ISCO-08 code. This high-quality, standardized dataset serves as the single source of truth for all subsequent analysis, visualizations, and research reports generated by the project.

3. Discussion and Conclusion

The journey detailed in this paper—from coarse, source-based assumptions to a sophisticated, AI-driven classification engine—underscores a fundamental principle in modern economic analysis: the quality of insight is directly constrained by the quality of the underlying data. Our initial challenge was not merely to process a large volume of job postings, but to enrich them with a standardized, verifiable meaning that would permit reliable analysis of Egypt's labor market. This paper has chronicled the multi-stage effort to meet that challenge, culminating in the development of JobIt-CLF.

The evolution from manual classification to an AI-powered system was not a simple replacement of human effort with machine automation. Rather, it was a process of codifying and scaling human expertise. The rigorous logic and nuanced criteria developed during the manual and basic LLM stages became the intellectual blueprint for JobIt-CLF. The system's true innovation lies not in its use of a Large Language Model, but in its agentic architecture. By delegating specialized data retrieval to an SQL Agent, the system grounds its reasoning in the authoritative ISCO-08 knowledge base. This tool-use cycle transforms the LLM from a probabilistic text generator into a deductive reasoning engine, whose step-by-step logic is both transparent and auditable.

The result of this work, the Classified Clean Dataset (CCD), is more than a final output; it is the foundational asset for the next phase of our research. With its high-accuracy, 4-digit ISCO-08 codes, the CCD transforms a noisy stream of raw data into a structured resource for high-stakes analysis. It enables ECES to track granular shifts in skill demand, identify emerging and declining occupations, and provide policymakers with a robust evidence base for education and labor market interventions.

Ultimately, JobIt-CLF serves as a validated blueprint for other research institutions and government bodies facing the critical task of making sense of large-scale labor market data. It demonstrates a powerful paradigm where AI does not replace domain expertise but amplifies it, enabling economic insights at a scale and depth that were previously unattainable.

Bibliography

Achananuparp, P., & E.-P. Lim 2025. "A Multi-Stage Framework with Taxonomy-Guided Reasoning for Occupation Classification Using Large Language Models". *arXiv preprint arXiv:2503.12989*.

International Labour Office. 2012. "*International Standard Classification of Occupations: ISCO-08*". Geneva: ILO.

Pourreza, M., & D. Rafiei 2023. "DIN-SQL: Decomposed In-Context Learning of Text-to-SQL with Self-Correction". *NeurIPS 2023*.

Van der Toren, R. G. D., R. Schram, R. Spijkerman, A. J. van der Beek, & T. Meijster 2023. "A Comparative Analysis of Three Automated Job Coding Tools Based on Free-Text Job Descriptions: AUTONOC, CASCOT, and LabourR". *Annals of Work Exposures and Health, 67*(5), 663–672.

Wang, B., C. Ren, J. Yang, X. Liang, J. Bai, L. Chai, D. Yin, X. Sun, & Z. Li 2025. "MAC-SQL: A Multi-Agent Collaborative Framework for Text-to-SQL". *Proceedings of the 31st International Conference on Computational Linguistics*, 540–557.