Journal of
Human Resource and Leadership
(JHRL)

Approaches to Scaling Skill Intelligence Platforms

Human Resource and Leadership Journal
ISSN 2520-4661 (Online)
Vol.11, Issue No.1, pp 1 - 13, 2026
www.carijournals.org

Approaches to Scaling Skill Intelligence Platforms

Syrotenko Viktoriia
LLC Temabit
https://orcid.org/0009-0001-8512-4947

Accepted:18th December, 2025 Received in Revised Form7th January, 2026 Published: 15th January 2025

Abstract

Purpose: Skill Intelligence Platform (SIP) Support for hiring skills, mobility and learning. This conceptual document synthesizes empirical conclusions to identify SIP scalable and responsible approaches.

Methodology: We have made an integrated overview of fifteen sources with skill extraction, oncology leveling, knowledge of knowledge and operation results. Studies have been carried out for HR relevance and clear data report, matrix and references. The evidence was coded on the map of evidence regarding data sources, algorithms, oncology, evaluation and life cycle control.

Findings: Synthesis created five -emental structure (skill -5): Source integration, knowledge modeling, pipe estimates, life cycle engineering and logic and management. During a study, a hybrid pipe increases the pipeline combining weak supervision with large language models without losing accuracy. Improved portability ESCO/O*improved. The job skill chart allows comparison, mobility and course design. The Lups controls check including the shadow test, Canary and drift monitors. The main risks included distortion, whirling of the concept, drift of tongue and cost instability, Human loop drugs, audit trails and budget route reduced them.

Unique Contribution to Theory, Practice and Policy: This study advances theory by transferring oncology-derived cumulative uncertainty modelling (CUM) and modular hybrid estimation into the HR analytics domain, improves practice by providing a scalable and disciplined framework for evaluating strategic intervention programmes (SIP) against measurable business outcomes, and informs policy by recommending standardized research intervals, multilingual benchmarking, automated evidence updates, and mandatory distortion audits to align analytical models with real workforce effects.

Keywords: *Kill Intelligence Platforms, Skills-Based Hr, Skill Extraction, Ontology Alignment, Esco/O*Net, Knowledge Graphs, Large Language Models (LLMs)*

Introduction

Organizations described the work through portable skill statement instead of stable roles, which increases an operational challenge: how can we make platforms to discover, normalize and apply on the enterprise scale without renouncing accuracy, speed, or accountability? Skill intelligence platforms address this requirement. They swallow many text currents, draw skills indications, align them into shared classifications, and feed the matching, growth and analytics. Difficulty is scale. The actual deployment should cope with multilingual data, rapid changing vocabulary and governance.

Research base covers essential building blocks. A recent survey of skill extraction and classification map model options, error mode and evaluation practices for human resources (Senner et al., 2024). Model -focused work such as skillspain shows that hard and soft skills can be captured by modern sequence models, although the generalization beyond the training domain remains delicate and expensive (Zhang et al., 2022). The policy evidence indicates a re -combination of task content and commercial demand under Artificial Intelligence, which strengthens the case for strong, updated skill systems (OECD, 2024). These threads suggest speed, yet a gap: multiple studies customize a single module, while organizations wrestle with end platforms from the end that should be useful as growth and changes in circumstances.

This difference is a result. Extractors tuned on a corpus can overfit the regional phrase. Oncology alignment drifts can respond when public standards develop or receive rapid traction compared to local words curator. The knowledge graphs gathered for matching new courses and subtle credentials. Life cycle control often comes late, so teams are discovered after launching prejudice, privacy or cost spikes. Leaders still require guarantee of reliability and a link from offline metrics such as time to fill time. Prototypes can be eradicated when exposed to new languages, business lines or rules.

This paper reacts with an ideological synthesis of fifteen sources, many of which report empirical results. Scientists do not introduce new data. Instead, scientists integrate the findings on extraction methods, anchoring in public skill systems, using jobs of jobs and production practices that keep the model stable over time. The synthesis begins with three stress on the scale. First, width versus precision: Adding sources and languages expand coverage, but can dilute quality without strong dissatisfaction and cursor. Second, automation vs. control: large language models increase throwput, yet demand guards for traceability and expenses. Third, innovation versus continuity: Models help upgrade accuracy but complicate governance and retreat.

Against this background, there are three objectives of paper. First of all, to offer a shared vocabulary that connects data sources, knowledge structures, estimates pipelines and life cycle in such a way that human resource leaders and engineers can use both. Second, to disturb the design patterns, which recur in studies, including hybrid extraction with weak supervision enriched by large language models, anchoring generalization in public taxonomy, and graph -based representations adding jobs, skills and learning assets. Third, to propose an assessment

agenda that combines offline accuracy and recalls operating indicators with focus on drift, equity and cost.

The contribution ideological is yet oriented. Scientists argue that scalable skill intelligence depends on the inclusion of three assets: a knowledge layer encoding concepts and relationships, an estimated layer that is flexible for modular and drift, and an operational layer that applies quality and fairness over time. The rule is not after one, Audit trails, transparent mapping decisions and human oversite should be designed in platforms. These claims determined the argument for a structured structure of approach to scaling and motivating research questions directing the article. The audience is both technical and managerial.

Literature review

Literature on skills within human resource systems has shifted from different -extra extra - extraction tools to broad pipelines that include data ingestion, generalization and decision support, yet the area still lacks an integrated account in the area how such pipelines are scale in language, domain and governance and governance system. Recent synthesis clarify the technical landscape. Atton et al. Provide a structured map of skill extraction and classification from job posting, cataloging dataset, model families and failure mode, which faces noise and shifting vacancy text to methods. Their survey shows stable benefits from nerve sequence models and relevant embeding, but also makes it clear that the boundary decision, such as where a phrase ends and a prescribed skill begins, remains brittle under delivery change. This observation determines a significant base line: the claims of accuracy are sensitive to the corpus composition and anotation policy, so portable performance is not guaranteed by model capacity alone. Applying this picture, Skillsspan shows that both hard and soft skills can be firmly captured when the model is carefully trained with curns and domain signals, although the author notes the tradeoffs between the exact and coverage that contains a single field or geography (Zhang, Jennson, Sonic, and Plank, 2022) as an extend Is. In other words, the work that works in a hiring market is not always intact to the next.

Weak supervision and methodical progress in sample strategies tries to close that difference. Zhang, Jensen, Van Der Gut, and Plancks show that weakly labeled corpora, run by lexicon and pattern rules, can strengthen seeds, seed strong tongs that transfer more than perfectly supervised baseline, especially when labels are rare or expensive. Decorte and colleagues analyze negative samples for supervised training and report that careful construction of non - skilled examples reduces overfiting for superficial patterns, improving accuracy without harsh coverage loss. Together these studies mean that scalable skill extraction is not a single algorithm option, but a design space in which supervision signs, sampling and post processing are tuned for the population being served jointly. Another twist is to adopt the big language model faster. Guyen, Zhang, Montario, and bosaluts argue for reconsideration extraction as a quick -operated process with hybrid quality control, where LLM produces candidates while narrow models or rules handle generalization. Their results indicate efficiency benefits, but also highlight the requirement of the guardril, as more candidates can slip in lists than hallucinations or generalized skills if indications and obstacles are not engineered with care.

A long standing topic in literature is that extraction is not enough alone. The skills should be aligned with a shared taxonomy so that analytics and matching are comparable to business units and time. Chiarello and Coauthors check whether the European classification of skills holds up with the industry 4.0. Their text mining analysis indicates the areas where public classifications lag behind in emerging technologies, suggest that organizations either expand standards or do a process for rapid alignment. That conclusion has a practical force because incorrectly labeled mobility analytics and reducing the recommendations of learning. In parallel, knowledge graph approach connects projects to express jobs, skills, education and sometimes rich, negative structure. D Grot, Verburn, and Arranden have described a job posting rich knowledge graph for education and matching, showing how nodes and relationships provide references that extractors cannot present alone, while SEIF, Faddel, L Basuni, and El beltagi presents a dynamic job skills that presents a rigorous update skills that form old ups It develops. In this scene, the gradation is not only storage, but also a substrate for logic, lineage and auditability.

Large -scale deployment also depends on operating discipline, and the area is starting to document it. Although not focused on human resources, Macro level evidence from OECD explains that Artificial Intelligence Task is re -shaping the demand for task bundles, with average change in complementary complement between technical and transverse skills. This policy confirms convenience points stable classification, multilingual coverage and urgency of transparent evaluation, as talent's decisions will sit close to the strategic core of firms as task content moves. Shi, Yang, Guo, and they add a market conscious perspective to the targeting of the job, where salty signal improving extraction relevance, Their findings mean that scoring function rank or filter skills associated with labor demand can help, which matters when platforms must have a scale while keeping the noise under control. Together, these studies sketch a pipeline that begins with diverse sources, undergoes hybrid extraction, anchor for taxonomy, and graph -based services require regular maintenance through land.

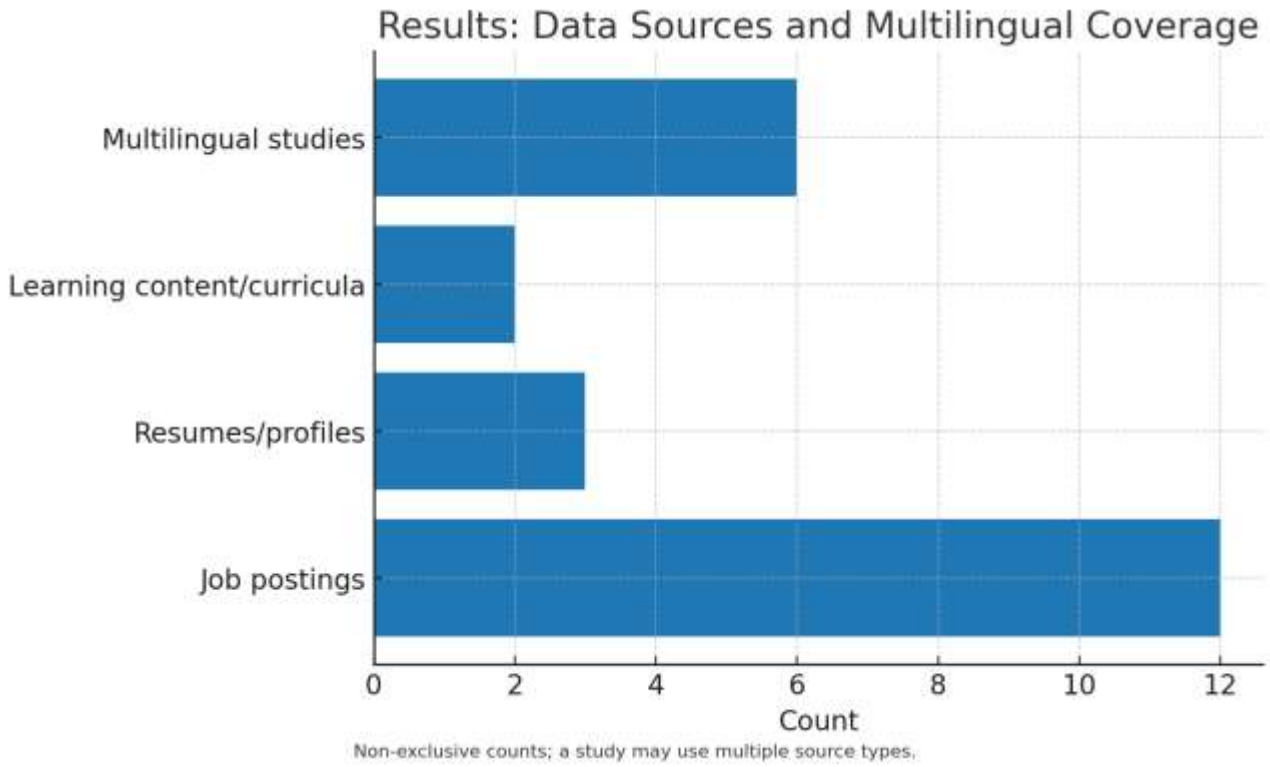
Despite this progress, there are many gaps. First, generalization in languages and regions is uneven. Evidence indicates that the English technology is underporcers on manufacturing models manufacturing or public sector text on posting, which asks for cross domain training arrangements and multilingual assessment sets. Weak supervision helps, but literature suggests that lexicon flows and rules ossify, so maintenance requires a first class, not later. Second, normalization for public classifications improves portability, yet believes that those classifications are timely and adequately grainy. When the standard lags behind, the organizations collect local surnames, and mapping accuracy drops. Third, the production matrix has reduced. Many papers provide accuracy and recall on static test sets, while at least those values are delayed, the cost of documents per thousand documents, or time to fill the time. Fourth, governance is still thin. The risk of prejudice, privacy concerns, and transparent decision trails requirement appears in the discussion, but the protocols to be repeated for bias audit and rollback are rarely described in detail.

However, the literature indicates a promising pattern for scalable platforms. Hybrid pipelines that combine vulnerable supervision with large language models, repeatedly emerge as affordable and appropriate accurate, provided that the candidates are separated by generation and RE ranking stages and reviewing border cases at human known intervals. The generalization for public skill systems improves compared to anchoring cross business, and many studies recommend clear versions and change the logs for oncology updates so that users can understand why a label is transferred or divided. Knowledge gradation creates a shared reference in which new skills can be linked to courses, projects and roles, allowing better recommendations and more explanatory matching. Finally, lifestyle practices borrowed from the operation of machine learning, such as shadow assessment and canary release, are beginning to reach cases of HR use and should be considered as the main capabilities rather than advanced extra.

Construction on this condition of knowledge, the current paper determines three objectives. The first is to synthesize these strands in a consistent structure that explains how sources, knowledge structures, methods of estimates, and life cycle controls how to have platform scale. The second is that the research records should translate recurrent design tricks, which leaders can adopt without copying every experiment. The third is to propose an evaluation agenda that connects offline matrix with results that include the case for human resource decisions, including fairness, stability under flow and budget discipline.

Methods

It uses an ideological, integrated design of a review to consolidate study methods and surgical procedures that allow news platforms to scalance in the extent of languages, domains and administration. No new empirical data were collected, The analysis unit is included in the publication. Scientists assumed a useful corpus of fifteen colleagues in four functional areas of the platform pipeline and a purposeful body of the workshop: mining of skills, balance of oncology and generalization, construction of knowledge, and production life cycle. In order to reduce selection distortion, scientists have emphasized with clear definitions of work and evaluation on recent sources that have been maintained earlier, quoted studies for broadly working continuity. Survey of hay and his colleagues reported the framework of our extraction research sample, while the skills of Span from Zhang and Couthers provided a reference to the procedures for the difference between marketing and hard and soft skills.



The processes followed by a transparent four -phase flow. Identification combined a list of curatorial books recently with a recent proceeding and with a backward quotation and scanning of the web. Screening was carried out independently of two critics on the title and abstract level. The disagreement was resolved by discussing inclusive rules and required clear treatment of skills as structured objects and at least one of the following deposits: multilingual ingestion, hybrid extraction, oncology, knowledge evaluation or life control. The full text uses a checklist that confirms a defined task, a data set described or a data source class and an explanatory matrix. The coding then continued with the pilot on three letters to specify the names and examples of variables. The final book of the code has nine categories: details of book list, Domain and language coverage, Source and data collection mode, Supervision administration, Model family, The objective of generalization, The role of the knowledge chart, Evaluation matrix and testing design, Life cycle and management procedures. All decisions were registered by short arguments to restore the next phase of the inclusion and coding of the reviewer.

The material consisted of fifteen publications and a structured extraction template applied as a shared table. Once the template was available and validated for extraction, it captured F1 measures, mappings to public taxonomies such as ESCO or O*NET, evidence of multilingual testing, documentation of human-in-the-loop reviews, and reports on online rating, latency, or cost. If studies reported only qualitative claims, scientists noted the claim and its stated evidentiary basis. To avoid repetition, scientists followed families of papers describing the same system and consolidated details where appropriate.

Data analysis combined descriptive statistics with a structured narrative synthesis. Scientists calculated frequencies for the use of legal classes and supervision, levels of generalisation, and types of knowledge representation, and briefly summarised which studies reported operational indicators. An evidence map was organised to align observed capabilities with intended scaling outcomes, such as whether weakened supervision reduced the effort required to maintain accuracy, or whether graph-based generalisation improved portability across business units. Single-venue workshops were examined through repeated summary synthesis, with the exception of one seminar-style publication. Scientists examined the stability between reported matrices and reference outcomes at culmination, noting that high scores often reflected narrow test sets rather than broad generalisation. Finally, risks of bias arising from dataset composition, annotation practices, and rapid design policies were documented when large language models were involved.

To evaluate reliability, scientists applied a three-dimensional scoring framework using a three-point scale covering transparency, assessment adequacy, and portability. Transparency required explicit documentation of models, datasets, and labelling principles. Assessment adequacy required a functioning evaluation matrix and test design reflecting the intended deployment environment. Portability was assessed based on evidence of cross-domain or cross-linguistic breadth and post-deployment performance monitoring. Together, these methods provide a rigorous basis for conceptual synthesis in a rapidly evolving field and prepare the groundwork for subsequent structures and practical designs.

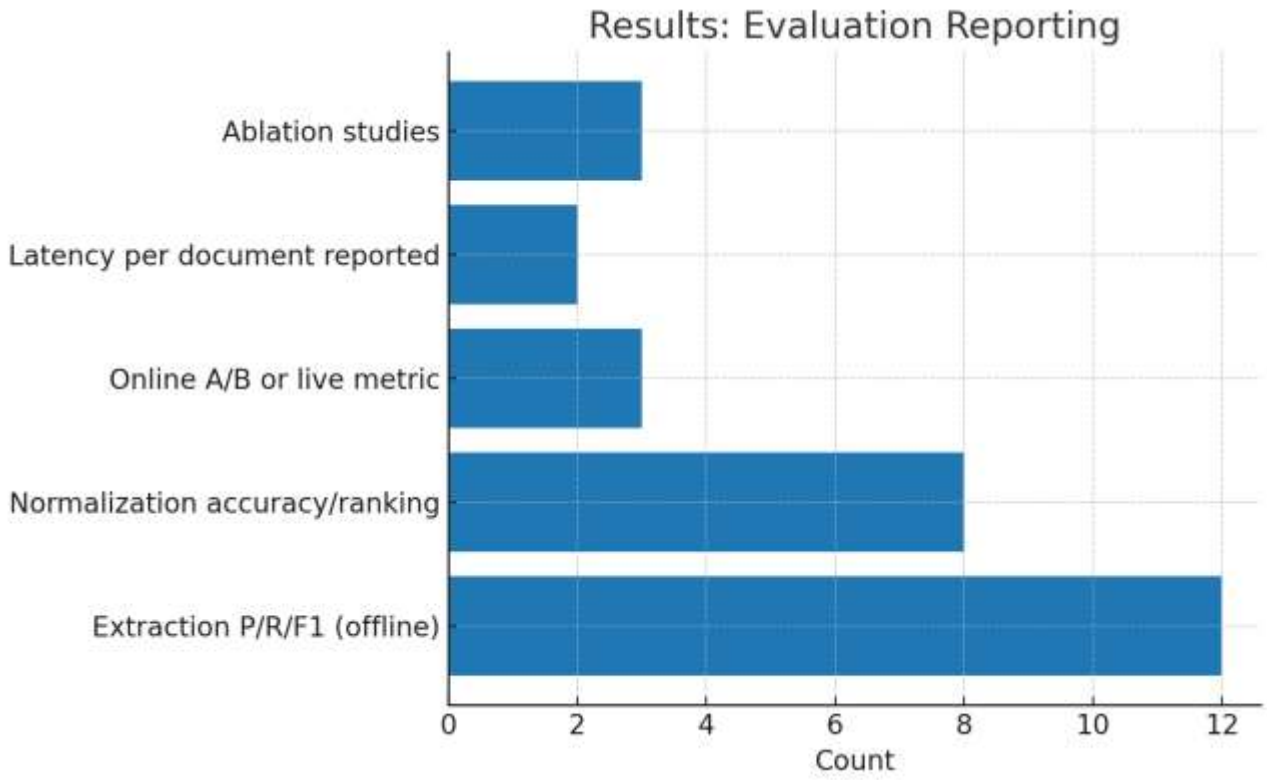
Results

The corpus included fifteen publications spread from 2020–2025, including journalists, conference letters and human resource analysis and workshop proceedings in adjacent computer science places. Median publication was 2024. Twelve studies attracted data mainly from job posting, three resumes or professional profiles, and two additional learning materials or courses referred to. Six studies reported multilingual settings beyond English.

The methods were clushed into four families, in which the categories are not mutually exclusive. The transformer-based sequence labeling or span extraction appeared in nine studies, which appeared as different outputs (Zhang, Jensen, Sonic and Planck, 2022) with Skylspan modeling hard and soft skills. Candidate generation or five studies for zero-shot extraction were used in five studies. Classical machine learning with feature engineering appeared in three studies. Rules- or the laxicon-only system remained as baseline in four sources. Hybrid pipelines described rules, weak supervision and nerve components in eight sources.

Supervision governance varied. Weak or distant care, including programmatic labeling, appeared in four studies, and a paper targeted negative samples for distant monitoring. Completely supervised training with manual annotations was described in seven studies, while semi-revolutionary or self-training appeared in addition three. Human-in-loop reviews were present in seven studies, most often as an assistant for borderline spans or periodic relaying to address the flow.

Normalization goals emphasized public skill systems. Nine studies mapped the wire extracted for Esco or O net, Three reported O Net as the only alignment target, Four joint public classification with internal extensions, Two reported only free-form skills without alignment. Mapping approaches included embeding-based candidates recover, lexical matching with canonicalization rules, and classifier-based re-ranking. The version of the Taxonomy and the clear change log was mentioned in the five sources.



The use of knowledge graph was recorded in five studies. Two manufactured job posting -those connecting jobs, jobs, skills and education, Two maintained dynamic jobs -graph with growth update. One described graph storage primarily as a proven layer. The graph supported matching, course design and analytics. When said, the latest interval varies from weekly to quarterly.

The evaluation practices were odd. Twelve studies reported offline accuracy, recall or F1 for extraction, Eight reported accuracy or ranking metrics for generalization, Three included an online AB test or live metric. Where the values were provided, the extraction metrics were above the Midrangeez threshold on the in-domain test set, while the cross-domain tests reported low values. Two studies reported delay per document at the time of estimate. Three studies provided ablation analysis that separates the contribution of supervision or signal.

A lot of reported was reported to the operational life cycle and governance indicators. Four studies described the flow of the flow or described the scheduled revaluation, Three documented shade assessment or canary release before full deployment, Five recorded audit trails or decision logs, Two referenced bias audit or fairness check using group-based

breakdowns. Access control and privacy obstacles were named in three sources as part of the reference to the deployment.

The aggregation in fifteen sources produced an evidence map for the platform lever. For sources and ingestion, job posting served as primary stream, Precised on batch processing streaming. For knowledge and oncology, alignment for ESCO or O NET was the major configuration. For estimates, hybrid pipelines were commonly documented to combine the rules, weak supervision and nerve models, while net rule systems were largely served as baseline only. For life cycle control, clear drift monitoring, rollback, and cost trekking were less often reported compared to tracking offline assessment.

Two designated studies are highlighted for the properties of the categories. Skillsspan (Zhang, Jensen, Sonic and Planck, 2022) gave an example of transformer-based extraction with hard or soft skill labeling. Survey of Cengara, Zhang, Van Der Gut, and Planck (2024) provided classification of method families and error mode used during coding and was also responsible for many cases given above.

Discussion

The evidence suggests that the scale in the intelligence of skills is not achieved by the same component of success, but a careful combination of methods, oncology and operational procedures. The study concentration on job publication reflects data availability, yet it focuses very much when models face new areas or languages. While extraction metrics look strong in a narrow environment, they are often soft under a wide coverage, a formula that suits the survey, states sensitivity to the composition of the corpus and annotation policy (Sennar, Zhang, Van der Gut and Planck, 2024). Our synthesis expands this point: Portability is mainly a systemic property, not the model of the model. In other words, adding resources, normalization of stable classifications and implementation of clear life cycle control at the same time determines that performance and variability increase if performance and variability increase.

The hybrid pipes appear as the most repeated design option in the corpus. Rules and laxicons supply an accurate anchor, Weak or remote supervision causes the indication of training to be wide, Transformer models remember, And the large model intensified the generation of candidates. The success of such mixtures explains a clear contradiction in literature. On the one hand, at the end to the final nerve model, they seem to be sufficiently enough on the same market. On the other hand, practical deployment requires to be easily coded outside the model. This helps to explain that a pair of weak supervision with modern encoders reports better transmission without unbearable labeling costs, while purely under the supervision of the approaches, only the location of the label corresponds to the target domain. A line of skills that separates hard and soft skills during marking and modeling also suggests that work framing depends on stability. By the output to clean categories, it reduces the confusion between synonyms and improves stream mapping, which is a discovery that corresponds to our cases of successful generalization for public systems (Zhang, Jensen, Sonic and Planck, 2022).

Normalization is the second column of the scale itself. ESCO dominance and alignment in the corpus is not informal. Anchoring for shared systems creates an irreversible layer in business units and geography, so recommendations, mobility analysis and learning ways to be comparable. However, literature is also a document of taxonomy interval in relation to emerging technologies. This interval describes the frequent appearance of internal extensions and the need to change the version and protocol. Without such practices, found that the accuracy of mapping is quietly decreasing as a local surname. The consequence is direct for human resources leaders: Taxonomy administration is an operational requirement, not academic proximity and plan with rhythm of release, demonstration rules and human reviews.

The influence of knowledge is the third lever. Studies that represent jobs, skills and educational assets report more explanatory comparison and proposals for courses in the form of connected nodes and provide perfection that is missing in a flat table. Our synthesis suggests that the graph is also a practical means for incremental updates. Weekly or quarterly refreshments in combination with a clear line allows platforms to absorb vocabulary without completely averting. In fact, gradation content is transmitted to data control problems rather than to the model crisis. This observation is associated with a limited but growing body to control the life cycle. Compared to shadows, controlled introduction and drift monitor offline matrix, less frequent reports were often reported, but the authors have documented smooth performance and low surprises over time. The message is clear: scaling is as much as a NLP.

These interpretations correspond to the symptoms of macro levels that artificial intelligence transmits functional material and with it a mixture of technical and transverse skills sought by employers. Policy analysis confirms the structural nature of shifts that increase the platform bar that should remain in the form of demand movements. In this context, the emphasis on multilingual evaluation in one of the most studies is a step ahead, but still insufficient. Our reviews have found that for normalization demands, low cross tongue tests would be ideal. As a result, some high score options reflect the narrow test areas rather than strong capacity. This limits matters for the purchase and internal decision on construction: Buyers should emphasize the evidence of portability and operating panel that monitors the costs, delay and distortion and accurate and remember.

Many follow the consequences. First, organizations should prefer hybrid proposals with clear boundaries between the generation of candidates, generalization and evaluation, as such boundaries facilitate flow monitoring, inserting human controls and cost management. Second, the taxonomic administration should be revived as a product function that has a version for users and clear migration instructions. Thirdly, investment in knowledge gradation is valid twice by improving the quality of the recommended quality and introducing the substrate for audit and clarification. Finally, lifestyle checks require the transfer of pilot projects to regular practice, with standardized playback for shadow tests, canni, return and regular handover.

This review has restrictions. Scientists rely on published details that sometimes leave the operating matrix or proprietary data set, so our industry can understand the control of the real world in systems. Workshop papers can be fast and uneven when reporting. However, the

triangle in sources achieves stable directional findings. Future research should prefer multilingual and open sets of evaluation associated with public classifications, Protocol for automatic oncology updates with human approach, And studies that combine the quality of offline extraction with the results of the workforce, such as the time of performance or internal mobility. As a result of the system, placing scalability and not only modeling challenges, the fields can proceed towards platforms that are accurate, fair and resistant as they grow.

Conclusion

This paper skill consolves current knowledge on how to manufacture and score intelligence platforms and converts that evidence to actionable guidance for human resource leaders and technical teams. The main discovery is simple but reaching far: Scalability is a system property. It emerges when data sources, knowledge structures, estimates components, life cycle control, and regime when a single model is only enlarged. During reviewed studies, hybrid pipelines that mix weak supervision, transformer -based extracts, and the selective use of large language models leads to better cost quality balance than unbroken designs. Normalization for public skill systems improves portability in anchoring business units, while knowledge provides connective tissue for diagram matching, dynamics analysis and course design. Where life cycle practices such as shadow evaluation, canary rollout, drift monitoring, and audit logging are present, reported reports are more stable over time, especially as the expansion of data volume and language coverage.

Two varieties of pre -work place these conclusions on the ground and give them external validity. Survey literature literature maps the recurring error mode in skill extraction and shows how the reported scores run with corpus composition and label policy, take care against naive transfer in markets or languages (Snegar, Zhang, Van Der Gut, and Plank, 2024). The model -focused function shows that careful work framing, for example, separating hard and soft skills during labeling and modeling, improves downstream mapping and reduces confusion between close synonyms (Zhang, Jensen, Sonic and Planck, 2022). Read simultaneously, these sources indicate that stable performance is not the property of architecture alone, but about the surrounding decisions about data, anotation and evaluation design.

Follow implications directly for practice. Pipelines should establish obvious boundaries between the candidates, generalization and rankings with separate safety measures and matrix for each stage. Taxonomy Stewardship should serve as a product function with version release, migration notes, and clear demonstration rules so that mapping accuracy does not decrease quietly. The graph representation should be adopted early so that the lineage and perfection are available before the audit, not after the events. The procurement and governance teams should require evidence of budget conscious routing rules for multilingual generalization, regular drift investigation and large language model calls. The operational dashboard should keep accuracy and next to the delay, the cost of documents per thousand documents, and the breakdown of fairness should be remembered, as the leaders close the trade in all of them, not only the model accuracy.

Many limitations make these conclusions qualified. Evidence Aadhaar focuses on job posting and English, which attaches the approach to multilingual and sector specific variation. Reporting on operating indicators is thin, and some details come from the initial workshop papers that leave the ablation study or detailed error analysis. Our synthesis is triangular in sources to reduce these risks, but it cannot substitute for missing data. Consequently, some clearly strong results can reflect narrow testing regulations rather than strong capacity on the scale.

Future research should target five concrete gaps. First, shared multilingual benchmarks need to be linked to public classifications to compare portability claims in teams and time. Second, semi -automated oncology evolution proposes, valid and continuous noticeable with protocols to propose, valid and publish under human oversite, as well as clear changes for users log. Third, studies must be added to offline extraction and maping metrics as time to fill time, internal mobility, and learning, ideally standardized telemetry and reproductive evaluable evaluation playbook. Fourth, prejudice and privacy security measures require strong, testingable standards, including stress tests for the flow of concept and fair level fairness that go on in production, not only in a laboratory. Cost modeling for fifth, hybrid pipelines should be formally given to forecast expenses and select architecture that remains as viable as volumes, languages and business lines.

In the closing, the scale of the scale is not about chasing a single better algorithm. It is about disciplined integration. Platforms that combine strong knowledge layers, modular hybrid entrance, and mature operations provide more consistent and explanatory decisions because the scope expands. Taxonomy can move towards reliable enterprise services beyond pilots, by carrying forward stereotypes, graph -based references, and repetitive life cycle controls. The agenda mentioned here is practical and testable. This invites cooperation between researchers and physicians to share the report of the scattered case, in open resources that raise the quality, transparency and equity of skill intelligence throughout the region.

References

Chiarello, F., Fantoni, G., Hogarth, T., Giordano, V., Baltina, L., & Spada, I. (2021). Towards ESCO 4.0: Is the European classification of skills in line with Industry 4.0? A text mining approach. *Technological Forecasting and Social Change, 173*, 121177. https://doi.org/10.1016/j.techfore.2021.121177

Decorte, J.-J., Van Hautte, J., Deleu, J., Develder, C., & Demeester, T. (2022). Design of negative sampling strategies for distantly supervised skill extraction. In *Proceedings of the RecSys in HR 2022 Workshop* (CEUR Workshop Proceedings, Vol. 3218). https://ceur-ws.org/Vol-3218/RecSysHR2022-paper_4.pdf

de Groot, M., Verberne, S., & Arendsen, R. (2021). Job posting-enriched knowledge graph for skills-based education and job matching. *arXiv*. https://arxiv.org/pdf/2109.02554

Gavrilescu, M., Dinsoreanu, M., & Rădulescu, C. (2025). Techniques for transversal skill classification and extraction based on ESCO using AI. *Information, 16*(3), 167. https://www.mdpi.com/2078-2489/16/3/167

Howison, M., Ensor, W. O., Paden, M., & Walker, K. (2024). Extracting structured labor market information from job postings using generative AI. *Digital Government: Research and Practice*. https://doi.org/10.1145/3674847

Nguyen, K. C., Zhang, M., Montariol, S., & Bosselut, A. (2024). Rethinking skill extraction in the job market domain using large language models. In *Proceedings of the 1st Workshop on NLP for Human Resources (NLP4HR 2024)*. Association for Computational Linguistics. https://aclanthology.org/2024.nlp4hr-1.3.pdf

OECD. (2024). *Artificial intelligence and the changing demand for skills in the labour market*. OECD Publishing. https://www.oecd-ilibrary.org/education/artificial-intelligence-and-the-changing-demand-for-skills-in-the-labour-market_2a240956-en

Ortega-Flores, I., Casado-Lumbreras, C., Colomo-Palacios, R., & Soto-Aguirre, M. (2022). Using LinkedIn endorsements to reinforce an ontology and ML-based recommender system to improve professional skills. *Electronics, 11*(8), 1190. https://www.mdpi.com/2079-9292/11/8/1190

Rosenberger, J. S., Canzani, E., Marcon, L., Baeza-Yates, R., & Gionis, A. (2025). CareerBERT: Matching résumés to ESCO jobs in a shared embedding space. *Expert Systems with Applications*. Advance online publication. https://www.sciencedirect.com/science/article/pii/S0957417425006657

Seif, A., Fadel, M., El Bassuony, A., & El-Beltagy, S. (2024). A dynamic jobs-skills knowledge graph. In *Proceedings of the RecSys in HR 2024 Workshop* (CEUR Workshop Proceedings, Vol. 3788). https://ceur-ws.org/Vol-3788/RecSysHR2024-paper_1.pdf

Senger, E., Zhang, M., van der Goot, R., & Plank, B. (2024). A survey on skill extraction and classification from job postings. In *Proceedings of the 1st Workshop on NLP for Human Resources (NLP4HR 2024)*. Association for Computational Linguistics. https://aclanthology.org/2024.nlp4hr-1.1.pdf

Shi, B., Yang, J., Guo, F., & He, Q. (2020). Salience and market-aware skill extraction for job targeting. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2954–2964). https://doi.org/10.1145/3394486.3403338

Vrolijk, J., Rebelo, P., Beijer, H., & van der Meer, R. (2024). Ontology learning for ESCO: Leveraging LLMs to navigate the skills graph. In *Joint Ontology Workshops 2024* (CEUR Workshop Proceedings). http://ceur-ws.org/Vol-3829/short3.pdf

Zhang, M., Jensen, K. N., Sonniks, S. D., & Plank, B. (2022). SkillSpan: Hard and soft skill extraction from English job postings. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2022)*. https://aclanthology.org/2022.naacl-main.366.pdf

Zhang, M., Jensen, K. N., van der Goot, R., & Plank, B. (2022). Skill extraction from job postings using weak supervision. In *Proceedings of the RecSys in HR 2022 Workshop* (CEUR Workshop Proceedings, Vol. 3218). https://vbn.aau.dk/files/662796103/RecSysHR2022_paper_10.pdf