

From BERT to Agents: A Systematic Literature Review of Skill Extraction and Taxonomy Alignment for Higher Education

Samson Kwaku Nkrumah , **Sara Morrissey**, **Fiona Boyle**, **Joseph Walsh**

IMaR, Research Centre, Munster Technological University, Department, Ireland

How to cite: Nkrumah, S. K.; Morrissey, S.; Boyle, F.; Walsh, J. (2026). From BERT to Agents: A Systematic Literature Review of Skill Extraction and Taxonomy Alignment for Higher Education. In: 12th International Conference on Higher Education Advances (HEAd'26). Valencia, 15-18 June 2026. <https://doi.org/10.4995/HEAd26.2026.21662>

Abstract

*Connecting higher education curricula with labour market demands requires understanding what skills are needed, and automated extraction tools now make this feasible at scale. This systematic literature review, following PRISMA 2020 guidelines, traces a decade of methodological evolution: from rule-based and statistical approaches, through transformer models (e.g., BERT), to Large Language Models (LLMs) and agentic AI pipelines. Analysing 25 peer-reviewed studies, we examine how methods have progressed, how skills are aligned to standard taxonomies (e.g., ESCO, O*NET), and what value agentic systems add. Key findings show substantial accuracy gains from contextual embeddings and LLM-enabled verification loops. Persistent challenges include handling emerging skills, multilingual vocabularies, and ensuring fairness in educational and hiring contexts. We outline practical implications for curriculum design, career guidance, and competency-based education, emphasising systems that are robust, adaptable, and ethically grounded.*

Keywords: Systematic Literature Review; Skill Extraction; Taxonomy Alignment; Higher Education; Competency-Based Education; Large Language Models; Artificial Intelligence

1. Introduction

The global economy is changing rapidly. Automation, artificial intelligence, and the lessons from COVID-19 have reshaped how we work and what skills employers need. This creates a growing mismatch: job market demands shift faster than universities can respond, since new courses take time to develop and programmatic reviews occur only every few years. The World Economic Forum estimates that one-third of jobs worldwide will require significant skills transformation by 2030. For higher education, understanding and closing these skill gaps is now essential for keeping curricula relevant and graduates employable. Artificial intelligence offers

a promising solution. Recent years have seen a surge of AI applications in education, from intelligent tutoring systems that adapt to each student, to platforms that recommend personalised learning paths. A related development is the rise of AI-powered tools in human resources (HR) technology. These tools read job postings, CVs, and course descriptions to automatically identify what skills are mentioned. For example, such a tool might scan thousands of job advertisements and discover that "data visualisation" appears frequently in analyst roles but rarely in current business curricula, signalling a gap that educators could address. This process is called *automated skill extraction*: using computers to find skill mentions in text, standardise them (so "MS Excel" and "Microsoft Excel" are recognised as the same thing), and link them to official skill lists used by governments and industries. Think of it as teaching a computer to read a CV the way a recruiter would, spotting relevant skills and matching them to job requirements. The underlying technology has evolved significantly over the past decade, from simple keyword matching to sophisticated language models and agentic systems (see Section 3).

This paper reviews a decade of progress in skill extraction and taxonomy alignment, from rules to transformers to agentic pipelines. Using peer-reviewed studies, we address three questions:

- How have methods progressed from pattern matching to LLM-assisted pipelines?
- Which taxonomy alignment strategies work best for educational applications?
- Where do agentic systems add value for higher education workflows?

We conclude with practical implications for curriculum design, competency-based assessment, and career services in higher education.

2. Methodology

We followed the PRISMA 2020 guidelines for systematic reviews. Our goal was to identify peer-reviewed studies on automated skill extraction and taxonomy alignment published between 2015 and 2025, a period spanning from early rule-based systems to today's large language models. In terms of search strategy, we searched four major databases: IEEE Xplore, ACM Digital Library, Scopus, and arXiv. Search terms combined skill-related keywords ("skill extraction," "competency mining," "taxonomy alignment") with method-related terms ("BERT," "LLM," "NLP") and taxonomy names ("ESCO," "O*NET"). Searches were conducted in December 2024. Regarding the selection process, Table 1 summarises how we narrowed from initial results to final included studies. To be included, a study had to: (i) present an automated method for extracting or aligning skills, (ii) provide technical implementation details, (iii) report empirical evaluation results, and (iv) be peer-reviewed or published at a recognised venue.

Table 1. Systematic review screening process

Stage	Studies
Records identified from databases	105
After removing duplicates	74
After title/abstract screening	43
After full-text assessment	25

The 25 included studies span rule-based methods, traditional machine learning, transformer models, and emerging agentic approaches; Section 3 cites representative work within each wave. The focused corpus reflects strict inclusion criteria prioritising technical implementation and empirical evaluation over descriptive breadth. The December 2024 cutoff precedes several 2025 developments (notably in agentic evaluation and multilingual skill models), which we flag where relevant but do not claim to cover comprehensively.

3. Evolution of Methods

3.1. Rule-Based and Traditional ML (2015–2017)

Early systems used dictionaries, regular expressions, and part-of-speech tags to match skill phrases. When dictionaries were comprehensive, precision was high, but recall was weak for new or unusual skills. Traditional machine-learning models, Support Vector Machines (SVMs) and Conditional Random Fields (CRFs), offered a more data-driven approach. Trained on labelled job advertisements, these models handled wording variations better than pure rules but required large annotated datasets and captured only shallow context.

3.2. Transformer Revolution: BERT (2018–2021)

The transformer architecture and BERT (Bidirectional Encoder Representations from Transformers) transformed how NLP models process text. BERT reads entire sentences bidirectionally, using surrounding words to disambiguate meaning, e.g., distinguishing “Java” (programming) from “Java” (island). Applied to skill extraction, BERT-based models delivered clear gains in precision and recall. Domain-specialised variants like JobBERT, continuously pre-trained on job postings, further improved performance on skill span extraction. A practical advantage is *transfer learning*: a pre-trained model can be fine-tuned with small, task-specific datasets, lowering annotation costs, particularly relevant for resource-constrained educational institutions.

3.3. Large Language Models (2022–2023)

LLMs such as Generative Pre-trained Transformer 3 (GPT-3) and more recent GPT variants opened new possibilities for skill extraction. Key advantages include:

- *Zero/few-shot learning*: Extract skills with brief instructions and examples, reducing annotation effort.
- *Synthetic data generation*: Create training examples to supplement limited labelled data.
- *Flexible reasoning*: Handle diverse phrasing and infer implied skills.

These properties make LLMs attractive for educational applications where domain-specific training data may be scarce.

3.4. Agentic AI Systems (2024–2025)

Agentic AI extends LLMs with planning and acting capabilities. Instead of single-pass extraction, agents break tasks into steps, use external tools (APIs, databases), check their work, and self-correct. For skill extraction, this enables:

- *Tool use*: Query job-posting databases or taxonomy services to verify unfamiliar terms.
- *Self-critique*: Evaluate outputs, flag low-confidence matches, and revise.
- *Iterative refinement*: Apply consistency checks before returning results.

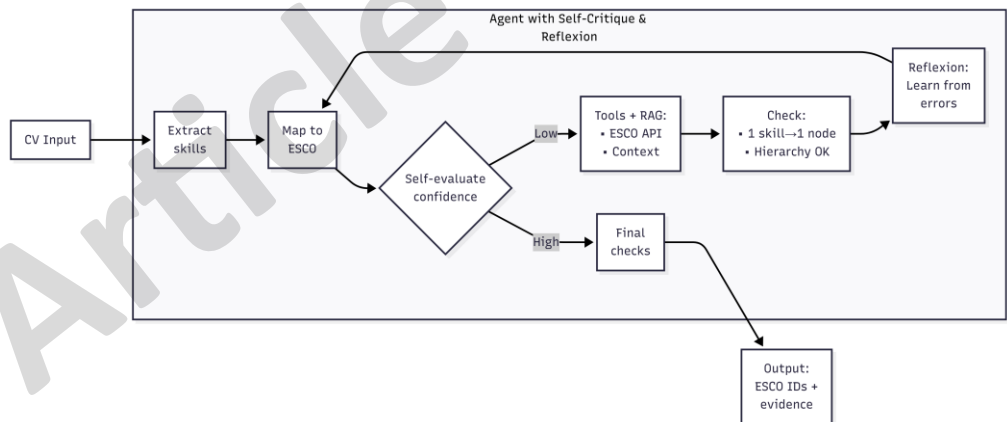


Figure 1. Agent-driven pipeline with self-assessment and refinement for taxonomy alignment.

Figure 1 shows a typical agentic pipeline: extraction, taxonomy mapping, self-evaluation, and iterative refinement when confidence is low.

4. Taxonomy Alignment

Once skills are extracted, they must be mapped to shared vocabularies for consistency and comparability across educational programmes and job markets.

4.1. Major Taxonomies

ESCO (European Skills, Competences, Qualifications and Occupations) is a multilingual EU taxonomy linking skills, occupations, and qualifications, widely used in European higher education for curriculum alignment. *O*NET* is the U.S. Department of Labor's occupation database, offering deep documentation of skills, knowledge, and abilities. For international or multi-sector educational projects, crosswalks between taxonomies are often necessary.

4.2. Alignment Methods

We identify three complementary approaches: *String similarity* compares extracted text to taxonomy labels using measures like Jaro–Winkler distance. This handles typos and minor variants (e.g., “MS Excel” ↔ “Microsoft Excel”) but struggles with synonyms. *Embedding-based methods* represent phrases as vectors (e.g., Sentence-BERT) and match by cosine similarity. This captures semantic meaning, “bookkeeping” aligns with “accounting support”, but may confuse related concepts without additional constraints. *Knowledge graph linking* treats the taxonomy as a graph, using structure (parent–child relationships, related skills) to disambiguate. For example, “R” in context of “tidyverse” links to the programming language rather than other senses. This approach offers highest precision but requires more sophisticated infrastructure. A practical pipeline runs these in sequence: string matching for exact hits, embeddings for paraphrases, and graph-based linking for difficult cases.

5. Implications for Higher Education

5.1. Curriculum Design

Automated skill extraction enables systematic mapping between course learning outcomes and labour market skill demands. Institutions can identify curriculum gaps and the skills that employers seek but academic institutions do not address and update courses regularly. The hybrid alignment approach (string + embeddings + graph) ensures accurate mapping to standard frameworks like ESCO, facilitating recognition of qualifications across European institutions.

5.2. Competency-Based Education

Skill taxonomies provide the foundation for competency-based education models. Accurate extraction and alignment allow:

- Assessment of prior learning against standardised competencies.
- Personalised learning pathways targeting specific skill gaps.
- Transparent certification that employers can interpret consistently.

5.3. Career Services and Guidance

Career centres can use these technologies to match student skills to job requirements, recommend upskilling courses, and track employment outcomes. Agentic systems add value by providing evidence-backed recommendations: rather than simple keyword matching, they can explain why certain roles fit a student's profile and what additional competencies would expand opportunities .

5.4. Emerging Challenges

Skill demands differ substantially across professional profiles, and one-size-fits-all pipelines are unlikely to succeed. Regulated professions (accounting, medicine, engineering) derive competencies from association frameworks such as IFAC, IEEE, or ABET; alignment must reconcile ESCO/O*NET entries with these mandated profiles, and professional associations act as slow-moving gatekeepers, so accredited-curriculum teams benefit from taxonomy-first approaches. Fast-moving fields like data science, by contrast, rely on emerging tooling vocabularies where agentic verification against live postings outperforms fixed taxonomies, so accountants and engineers do not face identical alignment problems. Despite advances, several challenges remain particularly relevant for educational contexts. First, skill taxonomies struggle to keep pace with rapidly evolving fields. For instance, job postings now frequently mention “prompt engineering” or “generative AI”, skills that did not exist in taxonomies created just a few years ago. Adaptive systems must detect these emerging skills and incorporate them before curricula fall behind industry needs. Second, cross-lingual alignment remains uneven. A student in Germany studying “Datenanalyse” should have their skills recognised equivalently to a student in Spain studying “análisis de datos,” yet many systems work well only in English. This gap is even wider for languages with fewer digital resources, limiting the global applicability of skill extraction tools. Third, systems trained on historical job data may unintentionally perpetuate biases. If past postings undervalued certain skills, such as caregiving or community organizing, models learn to rank them lower. Educational applications must audit these systems carefully to ensure fair treatment across disciplines and student backgrounds. Finally, transparency matters. When a system tells a student they lack “data visualisation skills,” both the student and their advisor should understand why. Showing the evidence, the exact phrases extracted, the taxonomy matches considered, and the reasoning behind the assessment, builds trust and allows meaningful feedback.

6. Discussion

Over a decade, the field has progressed from rules to transformers to agentic pipelines (Vaswani et al., 2017; Brown et al., 2020). This evolution expanded capabilities: modern systems plan workflows, use tools, verify outputs, and self-correct. For higher education, these advances enable more sophisticated curriculum analytics, personalised learning, and career guidance.

Key findings from our review: (1) Accuracy gains: Contextual embeddings raised span-level accuracy; agentic designs add verification that improves end-to-end quality. (2) Staged alignment: String matching → embeddings → graph linking balances precision, cost, and coverage. (3) Operational considerations: Bias, drift, privacy, and explainability require ongoing attention for responsible deployment in educational settings.

Future work should prioritise: (i) shared benchmarks with versioned datasets and hierarchical metrics; (ii) multilingual models with per-language evaluation; (iii) fairness audits with group-level error analysis; and (iv) adaptive mechanisms that surface emerging skills with human oversight.

7. Conclusion

We traced a decade of progress from rule-based systems to LLM-enabled agentic pipelines for skill extraction and taxonomy alignment. For higher education, these technologies offer practical benefits: systematic curriculum, labour market alignment, competency-based assessment, and evidence-based career guidance. Three patterns stand out. First, hybrid alignment strategies (string + embedding + graph) provide robust mapping across diverse skill expressions. Second, agentic verification improves reliability where errors are costly, essential for high-stakes educational decisions. Third, responsible deployment requires attention to bias, transparency, and emerging skill detection. We advocate for systems that are robust, adaptable, transparent, and ethically grounded. With coordinated effort among researchers, educators, and policymakers, these technologies can support higher education institutions in preparing graduates for a rapidly changing labour market.

References

- Atasoy, H.; Pavlou, P.A.; Banker, R.D. (2021). Information technology skills and labor market outcomes for workers. *Information Systems Research*, 32(2), 437–461. <https://doi.org/10.1287/isre.2020.0975>
- Breque, M.; De Nul, L.; Petridis, A. (2021). *Industry 5.0: Towards a sustainable, human-centric and resilient European industry*. Publications Office of the European Union. <https://doi.org/10.2777/308407>
- Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; et al. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

- Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of NAACL-HLT 2019*. Minneapolis, 2–7 June 2019, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>.
- Khaouja, I.; Kassou, I.; Ghogho, M. (2021). A survey on skill identification from online job ads. *IEEE Access*, 9, 118134–118153. <https://doi.org/10.1109/access.2021.3106120>
- Morandini, S.; Puzzo, G.; Fraboni, F.; Pietrantoni, L.; De Angelis, M.; Giusino, D. (2023). The impact of artificial intelligence on workers' skills: Upskilling and reskilling in organisations. *Informing Science*, 26, 39–68. <https://doi.org/10.28945/5078>
- OECD (2019). *OECD Skills Outlook 2019: Thriving in a Digital World*. OECD Publishing. <https://doi.org/10.1787/df80bc12-en>
- Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Pallagani, V.; Horesh, L.; Loreggia, A.; Fabiano, F.; Roy, K.; Murugesan, K.; et al. (2024). On the prospects of incorporating large language models (LLMs) in automated planning and scheduling (APS). In: *Proceedings of ICAPS 2024*. Banff, 1–6 June 2024, 432–444. <https://doi.org/10.1609/icaps.v34i1.31503>
- Senger, E.; Zhang, M.; van der Goot, R.; Plank, B. (2024). Deep learning-based computational job market analysis: A survey on skill extraction and classification from job postings. In: *Proceedings of NLP4HR Workshop at EACL 2024*. Malta, 21 March 2024. <https://doi.org/10.48550/arXiv.2402.05617>
- Shinn, N.; Cassano, F.; Gopinath, A.; Narasimhan, K.; Yao, S. (2023). Reflexion: Language agents with verbal reinforcement learning. In: *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*. New Orleans, 10–16 December 2023, 8634–8652.
- Tambe, P.; Cappelli, P.; Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, 61(4), 15–42. <https://doi.org/10.1177/0008125619867910>
- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; et al. (2017). Attention is all you need. In: *Advances in Neural Information Processing Systems 30*. Long Beach, 4–9 December 2017, 5998–6008.
- World Economic Forum (2020). *The Future of Jobs Report 2020*. World Economic Forum.
- Yao, S.; Zhao, J.; Yu, D.; Du, N.; Shafran, I.; Narasimhan, K.; Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models. In: *Proceedings of ICLR 2023*. Kigali, Rwanda, 1–5 May 2023. https://openreview.net/forum?id=WE_vluYUL-X
- Zawacki-Richter, O.; Marín, V.I.; Bond, M.; Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education*, 16, Article 39. <https://doi.org/10.1186/s41239-019-0171-0>
- Zhang, M.; Jensen, K.; Sonniks, S.; Plank, B. (2022). SkillSpan: Hard and soft skill extraction from English job postings. In: *Proceedings of NAACL 2022*. Seattle, 10–15 July 2022, 4962–4984. <https://doi.org/10.18653/v1/2022.naacl-main.366>

Zhao, M.; Javed, F.; Jacob, F.; McNair, M. (2015). SKILL: A system for skill identification and normalization. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Austin, 25–30 January 2015, 4012–4017. <https://doi.org/10.1609/aaai.v29i2.19064>

Article in Press